UNI
BASEL

Rethinking Leucine Zipper: ribonuclease activity and structural dynamics of a ubiquitous oligomerization motif

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Yaroslav Nikolaev

aus Saint-Petersburg, Russia

Basel, 2011

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Stephan Grzesiek (Faculty responsible)
Prof. Dr. Konstantin Pervushin (Scientific supervisor)
Dr. Michel Steinmetz (Co-referee)

Basel, den 24.03.2009

Prof. Dr. Eberhard Parlow, Dekan

creative commons

Attribution-Noncommercial-No Derivative Works 2.5 Switzerland

---

You are free:

to Share — to copy, distribute and transmit the work

Under the following conditions:

Attribution. You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).

Noncommercial. You may not use this work for commercial purposes.

No Derivative Works. You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work. The best way to do this is with a link to this web page.
- Any of the above conditions can be waived if you get permission from the copyright holder.
- Nothing in this license impairs or restricts the author's moral rights.

Your fair dealing and other rights are in no way affected by the above.

This is a human-readable summary of the Legal Code (the full license) available in German:
http://creativecommons.org/licenses/by-nc-nd/2.5/ch/legalcode.de

Disclaimer:
The Commons Deed is not a license. It is simply a handy reference for understanding the Legal Code (the full license) — it is a human-readable expression of some of its key terms. Think of it as the user-friendly interface to the Legal Code beneath. This Deed itself has no legal value, and its contents do not appear in the actual license. Creative Commons is not a law firm and does not provide legal services. Distributing of, displaying of, or linking to this Commons Deed does not create an attorney-client relationship.

Quelle: http://creativecommons.org/licenses/by-nc-nd/2.5/ch/deed.en Datum: 3.4.2009

Acknowledgements

First and most of all I want to thank my parents, Valerij Nikolaev and Ludmila Rodina, for nurturing creativity and scientific way of thinking in our family. As well as for many other things which simply aren't expressible on paper.

I sincerely thank Prof. Bernd Gutte for giving me an opportunity to work as a student in his laboratory at University of Zurich, and encouraging to pursue the Leucine Zipper story after his retirement.

I thank Prof. Konstantin Pervushin, my primary mentor in the Doctoral studies. With his help I have revived my interest to science and developed a deep and passionate curiosity in several areas, from structural biology and molecular evolution, to astrophysics and semantic web technology. I am seriously indebted for the balanced mixture of congenial freedom and knowledgeable support he provided during my studies.

Many thanks to Prof. Stephan Grzesiek for accepting me as a member of his group in Biozentrum at University Basel, and providing with essential support during the last two years of my PhD studies.

Special thanks to Dr. Michel Steinmetz for his generous agreement in taking responsibility as a co-examiner of my Doctoral thesis.

I am very thankful to all the colleagues with whom I worked during my delocalized PhD studies. Especially I want to thank Serge Chesnov, Gongda Xue and Christine Deillon from University of Zurich for their help in adaptation to the new environment. James Masse, Simon Alioth and Beat Vögeli from ETH Zurich for their guidance through the debris of the Nuclear Magnetic Resonance. Denys Pogoryelov for a challenging collaboration and friendly support. All the members of Prof. Grzesiek group at Biozentrum for the hospitable atmosphere and refreshed perspectives. Especially Navratna Vajpaj for his positive vibes, Marco Rogowski for his assistance, patience and humor, and Martin Allan with Daniel Häussinger for their help and thought-provoking afternoon discussions.

Beyond everything else, I am full of inexpressible gratitude to my beloved one, Maria Antsiferova, for the patience, understanding and support she expressed to the disordered creature obsessed by its research activities.

And finally I want to thank all of my friends, brothers and sisters for providing me with energy and optimism that I would have never embraced in a standalone version.
Thank you all!

To my beloved family, Mashustik, and friends

Table of contents

Summary

This dissertation focuses on structural, dynamic and catalytic properties of a Leucine Zipper (LZ) motif – a family of protein oligomerization domains which belong to the structural class of coiled coil proteins. LZ possess unique stability owing to high abundance of leucine residues in the key positions of the oligomerization interface. This allows increased combinatorial flexibility for the sidechains in coiled coil positions defining oligomerization specificity, thus making LZ an ideal protein-protein interaction determinant. This potential is reflected in the omnipresence of LZ within protein signalling pathways. Summarized in the Chapter I, we review the structure, interaction specificity, folding characteristics and functional diversity of LZ motifs, revealing the molecular mechanisms underlying LZ-enabled protein signaling. This review is now under preparation for publication in ***PLoS Biology***.

Beyond the widely acknowledged role of a protein oligomerization motif, recently it was shown that LZ motifs from bZIP factors GCN4 and cJun are capable of catalyzing degradation of RNA. Moreover catalytic RNase activity is conserved within full-length bZIP factors. This discovery was made in the laboratory of Prof. Bernd Gutte (University of Zurich) and served as a basis for the structural studies of LZ presented in this thesis. The manuscript presented as the Chapter II (submitted to ***Biochemistry***) summarizes the results of the initial LZ RNase studies, performed in collaboration with Christine Deillon and Stefan Hoffman. My contribution to these studies primarily relates to investigation of inhibitor effects on LZ catalytic activity, kinetic characterization of catalysis and studies of RNase activity within full-length cJun.

Our first structural trials on LZ-GCN4 employing solution NMR led to the discovery of the x-form – a novel monomeric folding intermediate of LZ that exists in equilibrium with the classical coiled coil state. Although marginally populated at experimental in vitro conditions, x-form might represent a considerable fraction of the LZ structural ensemble in vivo, providing a transient interface for specific recombination of interaction partners within bZIP networks. Results of these studies were published in the ***Journal of the American Chemical Society*** (JACS 2007; 129:6461-6469), and are presented as Chapter III of this thesis.

Finally, our structural NMR studies of LZ–RNA interactions have shown that the substrate interacts with the coiled coil (dimeric) conformation, while the x-form is incapable of binding RNA molecules. This is supported by the fact that the catalytic site is formed at the interface of two LZ chains, and therefore is only available upon assembly of the coiled coil dimer. Experimental data show that LZ from GCN4 and cJun differ in the topology and catalytic properties of the active site, which points to the ability of LZ to provide a general scaffold for assembly of catalytic sites with different properties. These results, presented in the Chapter IV of this thesis, are currently under preparation for publication as a separate manuscript.

Abbreviations

AKAP	A-kinase anchor protein
AP-1	activator protein 1 transcription factor
bHLH-LZ	basic region helix-loop-helix leucine zipper
BRLZ	basic region leucine zipper domain
BYA	billion years ago
bZIP	basic region leucine zipper
CC	coiled coil
cJun	oncoprotein and component of transcription factor AP-1
DAPK	death associated protein kinase
DC	diffusion collision
DCD	diffusion collision desolvation
DQ	double quantum
dsRNA	double stranded RNA
ESI	electrospray ionization
Fmoc	fluorenylmethyloxycarbonyl amino protecting group
GCN4	yeast S. cerevisiae transcriptional activator
GCN4p1	33 residue peptide corresponding to LZ region of GCN4
GILZ	glucocorticoid-induced leucine zipper
HD-ZIP	homeodomain leucine zipper
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethane-sulfonic acid
HPLC	high performance liquid chromatography
HSQC	heteronuclear single quantum coherence
IKK	IkB kinase
LZ	leucine zipper
LZ35	leucine zipper of GCN4, 35 residues per chain
MALDI	matrix-assisted laser desorption/ionization
MS	mass-spectrometry
NF-κB	nuclear factor kappa-light-chain-enhancer of activated B cells
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
NR	nuclear receptor
PKG	cGMP-dependent protein kinase, aka Protein Kinase G
R42	designed 42-residue HIV-1 enhancer-binding peptide
rLZ35	35-residue retro-leucine zipper of GCN4

rLZ38	38-residue retro-leucine zipper of GCN4
rLZ67	fusion of rLZ38 and shortened R42, total 67 residues per chain
RNA18	synthetic octadecaribonucleotide
RNase	ribonuclease
RP	reverse phase
SNARE	soluble N-ethylmaleimide-sensitive factor attachment protein receptor
ssRNA	single stranded RNA
ST2-PT	single transition-to-single transition polarization transfer
TF	transcription factor
TRIS	[1,3-dihydroxy-2-(hydroxymethyl)propan-2-yl]azanium
TROSY	transverse relaxation optimized spectroscopy
VdW	Van der Waals
XYEX	XY-EXchange
ZIPK	zipper interacting kinase
ZQ	zero quantum

Chapter I

Leucine Zipper –
a universal signal transduction motif.

Abstract

In this chapter we attempt to reconsider the concept of the “Leucine Zipper” (LZ) protein oligomerization motif. Reasoning on the wealth of existing data, we suggest that despite of structural similarity with highly stable extended “Coiled Coil” motifs, on the functional level short and moderately stable “Leucine Zippers” might stand out as a distinct group. This family of oligomerization motifs apparently provides cells with basic signal transduction functionality by delivering highly specific protein-protein interaction determinants, thus going beyond the structural role of the extended “Coiled Coils”. In this perspective we summarize existing empirical knowledge on the stability and specificity of LZ and demonstrate how a simple set of rules applied in the context of a universal coiled coil scaffold is capable of producing a very robust signal transduction motif. Leucine zippers, as a common protein interaction determinant, create a universal signal transduction framework, which might couple distinct protein signalling pathways into one global cellular network. Closer to the end of the chapter we provide examples demonstrating prevalence of the LZ-mediated signal transduction and illustrate applicability of the developed “LZ code” formalism to explain existing evidences of couplings between cytoplasmic and nuclear signalling networks. Finally, going beyond protein interaction functionality, at the end of this chapter we discuss the ability of LZ motifs to act as a scaffold for establishing catalytic sites with variable properties. Investigation of this functionality in relation to the catalytic degradation of RNA is presented in Chapters II and IV of the thesis.

= 1 = Introduction

The one-dimensional code for encryption of the protein in the nucleic acid sequence has been decrypted half a century ago, providing the key component for the emergence of genetic engineering and molecular biology. Unfortunately, extreme complexity of protein 3D structures defers the key advancement required for the widespread advent of the protein engineering. Namely, the decryption of 3D protein structure from its primary sequence is not yet accessible and remains one of the fundamental frontiers in modern biology, generally referred to as the "protein folding" problem. One of the main purposes of solving this problem is the ability to understand and accurately predict interactions between proteins. This knowledge is vital for understanding a wide range of cellular processes governed by protein signal transduction pathways, for example transmittance of extracellular signals to the transcription machinery. As a rule, these interactions are defined by complex and often highly dynamic 3D protein interfaces, making ***ab initio*** prediction of these interactions an extremely daunting task, which cannot be solved at the current state of science and technology. However, a small part of this problem can be solved already today. Leucine Zippers (LZ) represent a family of highly abundant protein-protein interaction motifs. Being based on the well characterized coiled coil scaffold, Leucine Zippers allow reduction of the 3D protein structure prediction problem to a simple comparison of two linear amino acid sequences. This does not bring us much closer to solving the general "protein folding" problem, but omnipresence of Leucine Zipper-based protein interactions makes this "LZ code" formalism an extremely useful tool for evaluation of protein interactions among plethora of LZ-mediated signalling pathways. Precise understanding of LZ specificity rules shall also allow to assess and modulate catalytic activities exhibited by LZ factors within the context of cellular signalling networks.

Leucine zippers belong to the class of coiled coil structural motifs, arguably the simplest and the most ubiquitous mediators of protein-protein interactions (*1, 2*). The members of the LZ class exhibit extreme thermodynamic stability owing to the prevalence of leucine residues at the key positions of their hydrophobic interface. This allows reduction of a minimal peptide length required for oligomerization to three (*3*), sometimes even two (*4, 5*) heptad repeats. Based on this high stability per heptad the LZ motifs and fragments were proposed to serve as folding triggering sequences in the context of extended coiled coil structures (*6, 7*).

Based on the data from genome sequencing projects, coiled coils are established as the most abundant protein motif and are predicted to be found in 5-10% of all proteins (*1*). Their importance and versatility both in vivo and in vitro is underscored by the amount of literature available on the topic, with a number of valuable reviews appearing in the recent years (*2, 8-11*).

Contrary to the "elder" members of the coiled coil class of proteins, which are "obligatory oligomers" and mainly participate as structural cores in macromolecular ensembles (filaments, extracellular matrices, cytoskeletal networks, spacers, stalks, etc), LZ motifs represent "transient oligomers", predominantly found in the signalling and regulatory proteins (receptors, kinases, transcription factors), which reflects the transient nature of these interactions.

The Leucine Zipper motif was originally discovered in 1988 in the family of transcription factors named bZIP (basic region leucine zippers) (*12*). Shortly after its discovery, their presence was revealed in a much broader array of proteins (*13, 14*). During the two past decades the LZ motif has been actively employed as a model for protein folding (*15, 16*) and protein engineering studies (*17, 18*) (and references therein).

Recent discovery of the enzymatic activity associated with the LZ domains of transcription factors GCN4 and cJun (discussed in the 2nd and 4th parts of this thesis), prompted us to review existing knowledge on this abundant motif, suggesting its importance for the cellular signalling networks. Moreover, owing to the independence of LZ stability on its "surface" residues (heptad positions **b**, **c** and **f**), this motif may serve as a structural scaffold for assembly of catalytic sites with variable properties both *in vitro* and *in vivo*. This hypothesis is strongly supported by protein engineering studies performed by Reza Ghadiri and coworkers, illustrating the ability of engineered leucine zipper peptides to mediate aminoacyl transfer (peptide synthesis) reactions similar to nonribosomal peptide synthetases (*19*).

Here we review the existing data on the structure, interaction specificity and folding characteristics of LZ motifs, revealing the molecular mechanisms underlying LZ-enabled protein signalling. We discuss the omnipresence of LZ motifs and illustrate their ability to couple distinct protein signalling pathways. As well we demonstrate their potential to provide direct coupling between protein interactions and other levels of cellular signal transduction, exemplified by RNase activity of bZIP factors GCN4 and cJun.

= 2 = Structure

= 2.1 = Primary – heptad repeat

Primary structure of leucine zippers, as coiled coils class of proteins, is defined by characteristic seven residue (heptad) sequence repeat – $(\boldsymbol{a\ b\ c\ d\ e\ f\ g})_n$, where the pattern is formed by hydrophobic residues at the ***a*** and ***d*** positions, charged residues at the ***e*** and ***g*** positions, and generally polar residues elsewhere (*20*) (Figure 2.1).

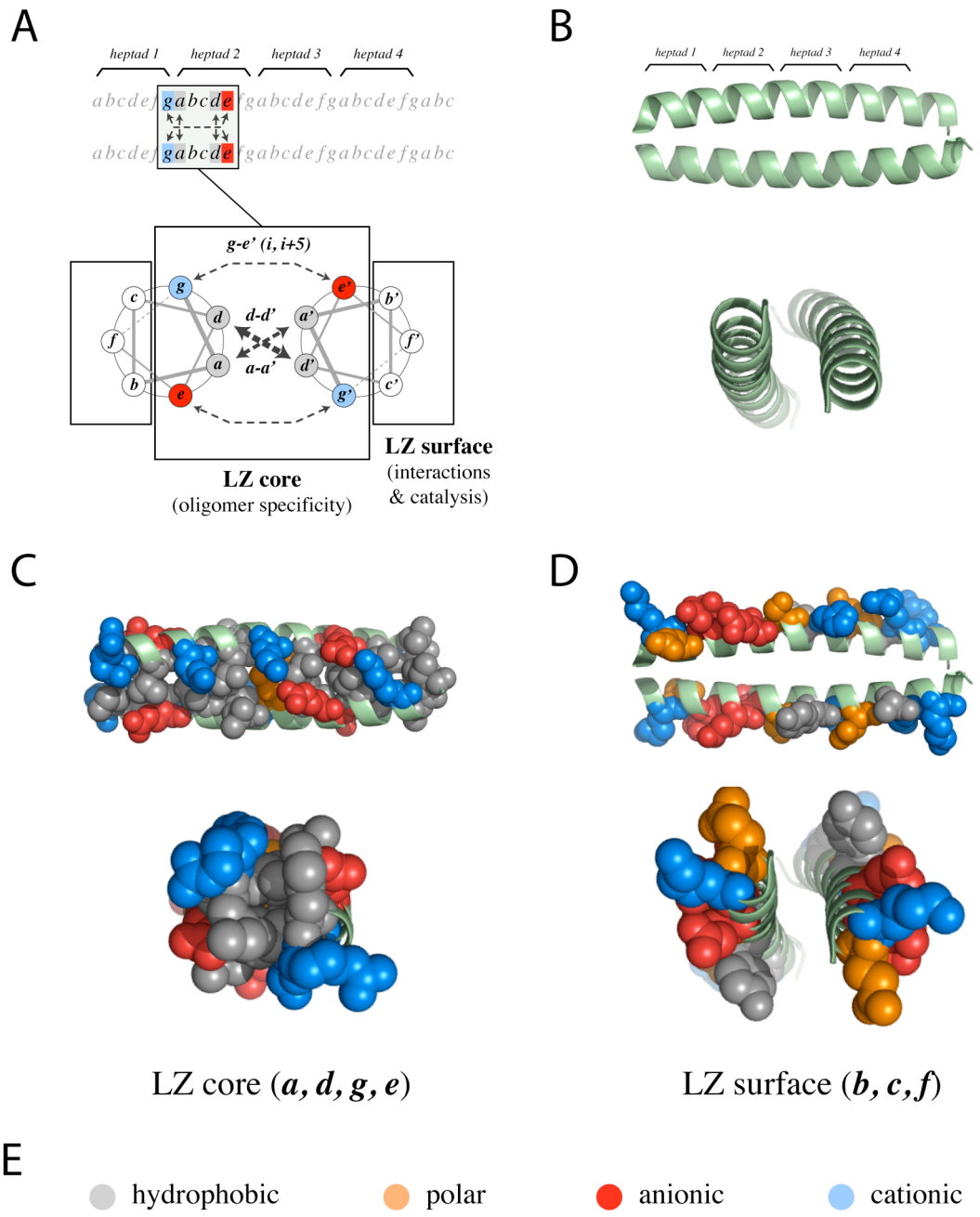


Figure 2.1. LZ structure and interactions. (**A**) Linear and wheel representation of coiled coil heptad repeat structure. (**C**) LZ core formed by hydrophobic ***d-d'***, ***a-a'*** and electrostatic ***g-e'*** interactions. (**D**) LZ surface ***b***, ***c*** and ***f*** positions generally do not affect stability and specificity of LZ structure.

= 2.2 = Secondary and tertiary – stability and stoichiometry

Regular amphiphatic primary sequence drives polypeptide assembly into a supercoiled structure, with knobs-into-holes packing of hydrophobic ***a*** and ***d*** side chains at the interacting interface (*21, 22*). Charged residues at the ***e*** and ***g*** positions pack over the hydrophobic core effectively shielding it from the solvent, stabilizing the structure by inter-chain ***g-e'*** electrostatic interactions and providing essential determinants for specificity of dimerization interface (*23, 24*) (more details below).

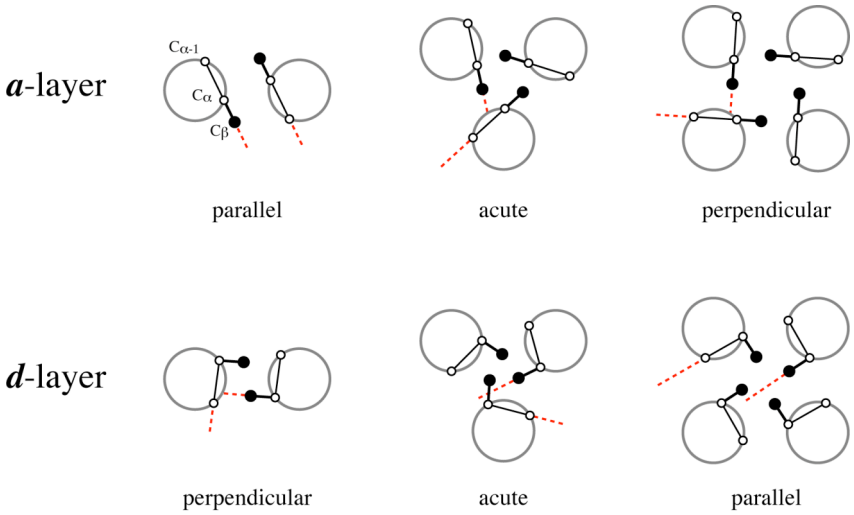


Figure 2.2. Packing interactions in the coiled coil hydrophobic core.

The key structural difference of leucine zippers from other coiled coils is almost exclusive presence of leucine residues in the ***d*** positions of the hydrophobic core (*12*), which essentially defines their dimeric nature. As shown by Pehr Harbury and colleagues (*25*) stoichiometry of a coiled coil is mainly determined by side chain packing geometry of the hydrophobic residues in the ***a*** and ***d*** positions of the interface, which varies systematically between different oligomeric states (reviewed in (*2*)). Briefly – packing topology of coiled coil hydrophobic core is distinguished by the orientation of Cα-Cβ bond of the hydrophobic residues (***a*** and ***d*** positions) relative to the peptide bond of the opposing helix (Figure 2.2). In parallel orientation the Cα-Cβ vector projects out of the dimer interior allowing more space between residues and thus favoring β-branched side chains (Ile, Val, Thr), where methyls branching from Cβ project back into the core, providing efficient Van der Waals interactions. Conversely, in perpendicular orientation Cα-Cβ vector projects directly into the core, limiting space available for the sidechains branched at Cβ, simultaneously providing excellent packing space for Cγ-branched Leucines. Folding topology of dimeric coiled coils brings residues of the ***a***-layer into parallel orientation, and ***d***-layer – into perpendicular. Thus, sequences bearing Leucines in ***d*** positions, and beta-branched residues in ***a***, are very likely to fold into dimers. The situation is reversed in the tetrameric coiled coil fold: ***a***-layer adopts

perpendicular orientation, and ***d*** – parallel. Therefore this fold is favored by the sequences containing (Ile, Val, Thr) in ***d*** positions, and Leu - in the ***a***. Topology of trimeric coiled coil fold is less restrictive - it has an intermediate ("acute") geometry in both ***a*** and ***d*** layers - thus allowing more versatile sequence patterns.

= 2.3 = Quaternary – specificity

Regular topology of interactions within the coiled coil motif together with a diverse set of available amino acid side-chains, provides LZ with a wide range of stabilities and specificities, allowing them to form both homodimeric and heterodimeric structures depending on the motif composition. Moreover, a significant fraction of natural LZ motifs exhibits a wide range of intrinsic specificity allowing them to have an extended set of heterodimeric pairs. This variability of specificities is a fundamental property that enables the transcription factors to assemble combinatorial regulatory networks based on their LZ motifs. These networks - bZIP, bHLH-LZ, HD-ZIP - are amongst the most advanced regulatory networks developed by eukaryotic species (*26*), and have evolved as key regulators in the wide variety of processes, ranging from cell metabolism to tissue differentiation (*27*). The rules governing interaction specificity within these networks have been thoroughly characterized during last two decades, and are mainly defined by electrostatics of ***g-e'*** couplings and polar interactions of the ***a-a'*** pairs, as discussed in more details below.

= 3 = Stability and specificity

Core packing at ***a*** and ***d*** positions, together with ionic interactions between ***e*** and ***g*** positions are the key factors influencing stability and specificity of the coiled coil assembly. Applying reductionist approach to the most widely studied family of LZ proteins – bZIP TFs, three main interactions can be distinguished for the analysis of thermodynamic contributions to stability and specificity of the LZ interface (Figure 2.1):

1) ***d-d'*** interactions (primarily hydrophobic/VdW > defining stability)

2) ***g-e'*** interactions (primarily electrostatic/VdW > defining specificity)

3) ***a-a'*** interactions (mixed hydrophobic/VdW/electrostatic > defining stability and specificity)

Most of currently existing data on the weights of these contributions to the stability and specificity of leucine zipper motifs was produced by Charles Vinson group through application of double-mutant thermodynamic cycle analysis (*28*) in the context of LZ motif

from bZIP factor VBP: ***d-d'*** (*29*), ***g-e'*** (*24, 30*) and ***a-a'*** (*7, 31*). Obtained results are largely corroborated by studies performed by Robert Hodges group (*32-34*), who targeted predominantly homodimeric interactions in the context of engineered coiled coils stabilized by covalent cross-linking. However, highly convoluted oligomerization equilibrium exhibited by engineered peptides in the latter cases, in absence of high-resolution structural data and double-mutant cycle free energy analysis urges to treat these data with caution when applied to canonical LZ motifs.

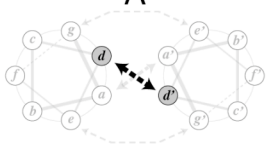
Detailed review of bZIP LZ stability and specificity, as well as specificity-based classification of bZIP transcription factors can be found elsewhere (*10*). Herein we provide a general summary on the topic, along with some contextual re-evaluation of available data.

= 3.1 = D-D' interactions (stability)

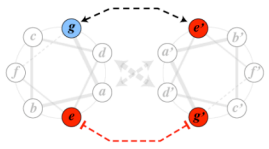
Hydrophobic ***d-d'*** interactions are the key stabilizing component and the distinctive feature of the LZ family. Efficient packing of Leucine side chains in the ***d*** positions of the knobs-into-holes topology dramatically stabilizes the dimeric coiled coil interface (*29*), to a large extent defining the stoichiometry of the complex (*25*). Importantly, stability is conferred not only by the hydrophobic effect (burial of the hydrophobic side-chain in the protein interior, shielding it from the polar solvent) but also Van der Waals interactions (efficient packing of the sidechain against neighboring residues). The latter contribution provides leucine with upto 5.2 - 5.9 kcal/mol/pair (i.e. contribution from one heptad) advantage in packing energy over similarly sized methionine and isoleucine pairs (*29*) (Table 3.1-A). 3D structure modeling suggests that the favorable rotamer conformations of beta-branched Ile and Val side-chains produce interhelical clashes between the Cγ2 methyls if placed into the ***d***-position (*29*). Thus, energy required to compensate for the thermodynamically unfavorable rotamer conformation may account for a part of the remarkable stability difference between leucine and beta-branched residues. This stability compromise does not play a significant role in the case of long structural coiled coil proteins, where a variety of hydrophobic amino acids have been shown to occupy the ***d*** position of the amphipathic helix (*35*). However, stabilizing effect of the leucine side chain appears crucial for short leucine zipper sequences involved in signal transduction, thus yielding near invariance of this residue in the ***d*** position of the interface (*29, 34*).

Analysed solely in the context of bZIP motifs, the role of ***d***-position in determining the LZ interface specificity is apparently underestimated. For example in the Myc/Max/Mxd family of bHLH-LZ transcription factors, ***d***-position histidine of Max protein forms a unique buried salt bridge with anionic sidechains in the heterodimerization partners, which defines

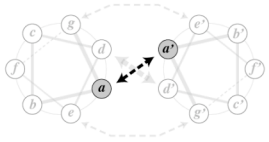
the specificity of this network (*36, 37*). Thus, it is important to recognize that empirical dimerization rules discussed here provide only a part of the "LZ code" definition.



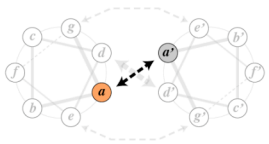
Position	Residue	$\Delta\Delta G_{A\text{-}A}$ Energy difference [kcal/mol/pair]	$\Delta\Delta\Delta G_{int}$ Coupling energy [kcal/mol/pair]	Stability	Side chain / interaction type	Oligomer specificity
d-d'	L-L	-9.2		+	aliphatic	
	M-M	-4.0				
	I-I	-3.3				
	V-V	-2.2				
	C-C	-1.9				
	A-A	0.0				
	S-S	+0.5				



Position	Residue	$\Delta\Delta G_{A\text{-}A}$ Energy difference [kcal/mol/pair]	$\Delta\Delta\Delta G_{int}$ Coupling energy [kcal/mol/pair]	Stability	Side chain / interaction type	Oligomer specificity
g-e'	R-E	-1.55	-1.07	+	oppositely charged	
	K-E	-1.42	-0.91			
	E-R	-1.30	-0.45			
	E-K	-0.95	-0.25			
	Q-Q	-1.17	-0.03			
	Q-K	-0.83	0.11			
	E-Q	-0.73	0.17			
	Q-E	-0.46	0.20	–	~ identicaly charged	(–) homodimers
	K-Q	-0.71	0.28			
	Q-R	-0.79	0.30			
	R-Q	-0.58	0.38			
	K-K	-0.34	0.45			
	K-R	-0.32	0.62			
	R-K	-0.10	0.66			
	E-E	+0.38	0.80			
	R-R	-0.10	0.81			



Position	Residue	$\Delta\Delta G_{A\text{-}A}$ Energy difference [kcal/mol/pair]	$\Delta\Delta\Delta G_{int}$ Coupling energy [kcal/mol/pair]	Stability	Side chain / interaction type	Oligomer specificity
a-a' (symmetric / homodimer)	I-I	-9.2	-0.9	+	aliphatic	(+) homodimers
	V-V	-5.4	-0.7			
	L-L	-5.2	-0.6			
	N-N	-2.4	-0.5	~	polar	
	T-T	+0.8	+0.2			
	S-S	+1.8	+0.2			
	K-K	+2.9	+0.3	–	charged	(–) homodimers
	R-R	+4.6	+1.2			
	E-E	**+6.0**	**+2.1**	––		



Position	Residue	$\Delta\Delta G_{A\text{-}A}$ Energy difference [kcal/mol/pair]	$\Delta\Delta\Delta G_{int}$ Coupling energy [kcal/mol/pair]	Stability	Side chain / interaction type
a-a' (asymmetric / heterodimer)	(N,T,S) • (K,R,E)	+1.8	+0.1	~	charged • polar
	(K,R,E) • (I,V,L)	-1.1	+0.2		charged • aliphatic
	(I,V,L) • (I,V,L)	+5.5	+0.4	–	aliphatic • aliphatic
	(K,R,E) • (K,R,E)	+3.9	+0.6		charged • charged
	(N,T,S) • (N,T,S)	+1.1	+0.9		polar • polar
	(I,V,L) • (N,T,S)	-0.6	**+2.3**	––	**polar • aliphatic**

Table 3.1. Free energy differences ($\Delta\Delta G_{A\text{-}A}$ [kcal/mol/pair] – useful to compare between LZ interaction types) and coupling energies ($\Delta\Delta\Delta G_{int}$ [kcal/mol/pair] – useful when comparing pairs within one LZ interaction type) of common LZ coupling relative to a pair of alanines. Values obtaned from LZ dimer thermal stabilities in 12 mM PO_4, 150 mM KCl, pH 7.4. Data reproduced from (**A**) ***d-d'*** (*29*), (**B**) ***g-e'*** (*24, 30*), (**C**) and (**D**) ***a-a'*** (*7*). For ***g-e'*** and ***a-a'*** interactions individual pairs are sorted according to the coupling energy strengths, and grouped in four categories: ± 0.2 kcal/mol (neutral), ≤ 0.2 kcal/mol (stabilizing), ≥ 0.2 kcal/mol (destabilizing), ≥ 2 kcal/mol (strongly destabilizing). Free and coupling energies for heterodimeric ***a-a'*** interactions (**D**) are **averaged** according to the residue type; full set of energies can be found in **Table 3.2**.

= 3.2 = G-E' interactions (specificity)

G-E' interactions primarily involve charged amino acids with long aliphatic side-chains (Arg, Lys, Glu, Gln) (*23*), which simultaneously brings electrostatic, VdW and hydrophobic effects into play.

Compared to a pair of alanines, the most common bZIP ***g-e'*** salt bridges stabilize the coiled coil dimer by 1.3-1.6 (ER-RE) and 1-1.4 (EK-KE) kcal/mol/pair (Table 3.1-B). Remarkably, even identically charged Arg-Arg and Lys-Lys ***g-e'*** pairs have stabilizing effect, contributing respectively 0.1 and 0.34 kcal/mol/pair more energy than a pair of alanines. These repulsive electrostatic interactions are considered to be largely compensated by increased hydrophobic burial and favorable VdW interactions between the methylenes of ***g***/***e*** sidechains and hydrophoboc core of the structure (*22, 30, 36, 38, 39*). The only destabilizing is a pair of glutamates as compared to alanine, which reduces the dimer stability by 0.38 kcal/mol/pair. Obviously two methylenes of a glutamate have less compensatory effect than three methylenes of an arginine and four methylenes of a lysine, with net energy differences markedly conforming ~0.5–1 kcal/mol protein stability gain commonly observed upon burial of additional methylene (*40*).

The overall contribution of interhelical salt bridges to the stability of leucine zipppers for a long time has been a matter of debate (*24, 30, 39, 41-43*). The issue has been recently resolved by Hans Bosshard and Daniel Marti, showing that the net thermodynamic contribution of a salt bridge is balanced between favorable charge-charge interaction, unfavorable desolvation energy and background interactions (such as coupling with the dipole moment of the helix) (*44, 45*). As it is evident from the Table 3.1, the effect of ionic ***g-e'*** couplings compared to hydrophobic core is rather moderate, and in the context of a canonical LZ heptade will be offset by energies of ***a-a'*** and ***d-d'*** couplings. Nevertheless, as will be shown in the next section, the ionic interactions have a potential to regulate specificity of oligomerization by modulating kinetics of early steps of LZ folding process, when ***a-a'*** and ***d-d'*** interactions have not yet stabilized the structure. In this arrangement the long-range Coulombic interactions between charged side-chains shall be able to determine the specificity of coiled coil formation. The magnitude of these interactions for particular pairs of side-cahins is most efficiently evaluated employing the concept of coupling energy, which is defined as the energy conveyed by the mutual interaction of two residues, devoid of energy contributions from isolated side-chains (*24, 28*) (Figure 3.1). For example coupling energy of E-R pair ($\Delta\Delta\Delta G_{int}$ = –0.45 kcal/mol) can be estimated as total E-R contribution to the dimer stability ($\Delta\Delta G_{A\text{-}A}$ = –1.3 kcal/mol) devoid of the stability contributions of individual E ($\Delta\Delta G_{E\text{-}A}$ = –0.15 kcal/mol) and R ($\Delta\Delta G_{A\text{-}R}$ = –0.7 kcal/mol) side-chains ($\Delta\Delta\Delta G_{int} = \Delta\Delta G_{A\text{-}A} - \Delta\Delta G_{E\text{-}A} - \Delta\Delta G_{A\text{-}R}$) (*24*).

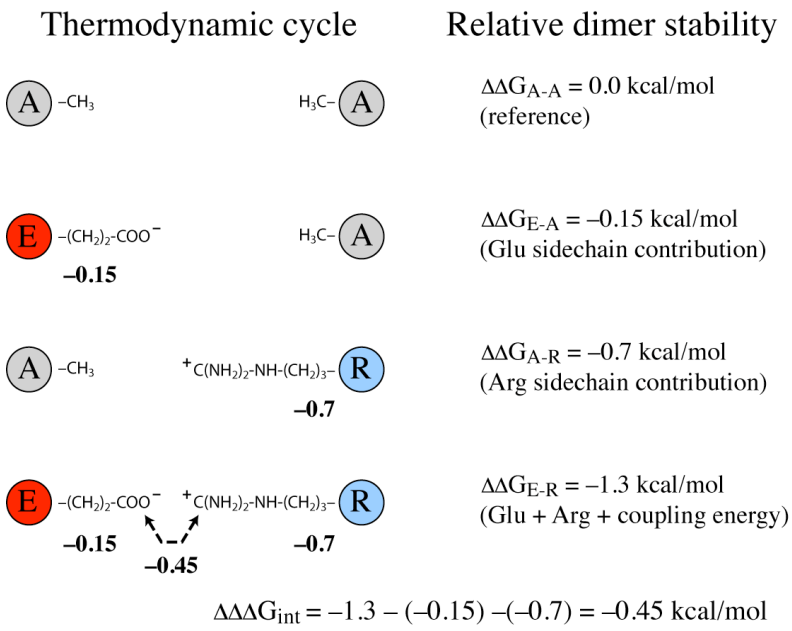


Figure 3.1. Thermodynamic double-mutant cycle for the Glu-Arg interaction. Measurement of thermal stabilities of four dimers yields three energy differences relative to a pair of alanines. Coupling energy ($\Delta\Delta\Delta G_{int}$) of Glu-Arg ionic interaction is obtained by subtracting individual contributions of Glu and Arg sidechains from overall stability of the dimer.

Employing this concept the ***g-e'*** interactions can be arranged on a more reliable thermodynamic scale, defined by pure coupling energies devoid of stabilities conferred from interactions with the core of the molecule (Table 1, column $\Delta\Delta G_{int}$). On this scale the most stabilizing interhelical coupling energies, on the order of 1 kcal/mol/pair, are shown by R-E and K-E pairs, while the most destabilizing, on the order of +0.8 kcal/mol/pair – by repulsive E-E and R-R couplings (*24, 30*). Importantly, coupling energies do not clusterize and are instead uniformly distributed over all accessible energy scale. This diverse range of attractive, neutral and repulsive couplings available within common coiled coil scaffold, multiplied by the number of variable positions (8 in an average 4-heptad LZ motif) creates a highly combinatorial key-lock mechanism for definition of interaction specificity. Distribution of specificity determinants along the whole leucine zipper sequence allows regulation of populations of different dimers in accordance with their composition (i.e. dimers with more attractive interactions and fewer repulsive interactions would be favored over dimers with fewer attractive and more repulsive interactions). This gives a potential for establishing a multiplex signalling node, capable of emitting a rich output signal instead of a simple on/off event. Moreover, as highlighted by differences in reciprocal K-E/E-K (–0.91 vs –0.25 kcal/mol) and R-E/E-R (–1.07 vs –0.45 kcal/mol) pairs (*30*), coupling energies of ***g-e'*** interactions strongly depend on the context, broadening the combinatorial nature of LZ interface even further. However this effect appears to step into place only when underlying a

positions bear polar or charged side-chains, and is negligible in the case of purely aliphatic core (*7*).

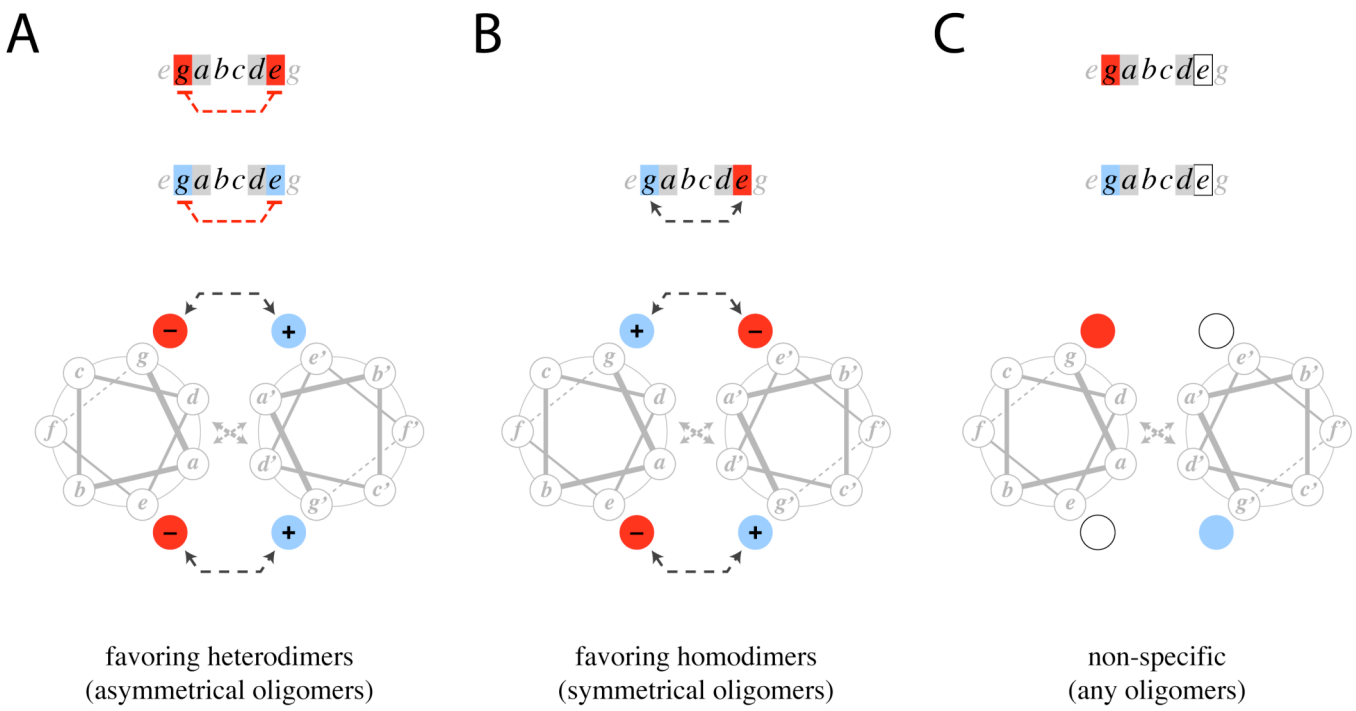


Figure 3.2. Schematic representation of interhelical ***g-e'*** interactions in defining oligomerization specificity. (A) LZ with identically charged (***i***, ***i+5***) ***g-e'*** residues – favoring heterodimerization, disfavoring homodimerization. (B) LZ with oppositely charged (***i***, ***i+5***) ***g-e'*** residues – favor homodimerization. (C) LZ with non-ionic ***g-e'*** residues are not discriminative in oligomerization.

In the simplest case of homo- versus hetero-dimer formation, a pair of ***g-e'*** residues with the same charge (acidic+acidic or basic+basic) would favor asymmetric oligomerization– favoring heterodimers and disfavoring homodimers (Figure 3.2-A). A ***g-e'*** pair with alternating charges would favor symmetric oligomers (homodimers) and disfavor asymmetrical oligomers (heterodimers with mirrored charge allocation) (Figure 3.2-B). Non-charged side-chain would give the most liberal specificity range, allowing coupling with any type of residue (Figure 3.2-C).

In vivo these selective specificity mechanisms are successfully employed to decouple LZ-TF networks that operate in different functional realms. For example, specific ***g-e'*** electrostatic interactions define a subfamily of PAR factors involved in regulation of circadian rhythms, which precludes its cross-reactivity with other bZIP families (*46*). These considerations, together with the specificity rules conveyed by residues in ***a***-positions, were successfuly employed for classification of bZIP proteins based on their dimerization properties (*47, 48*).

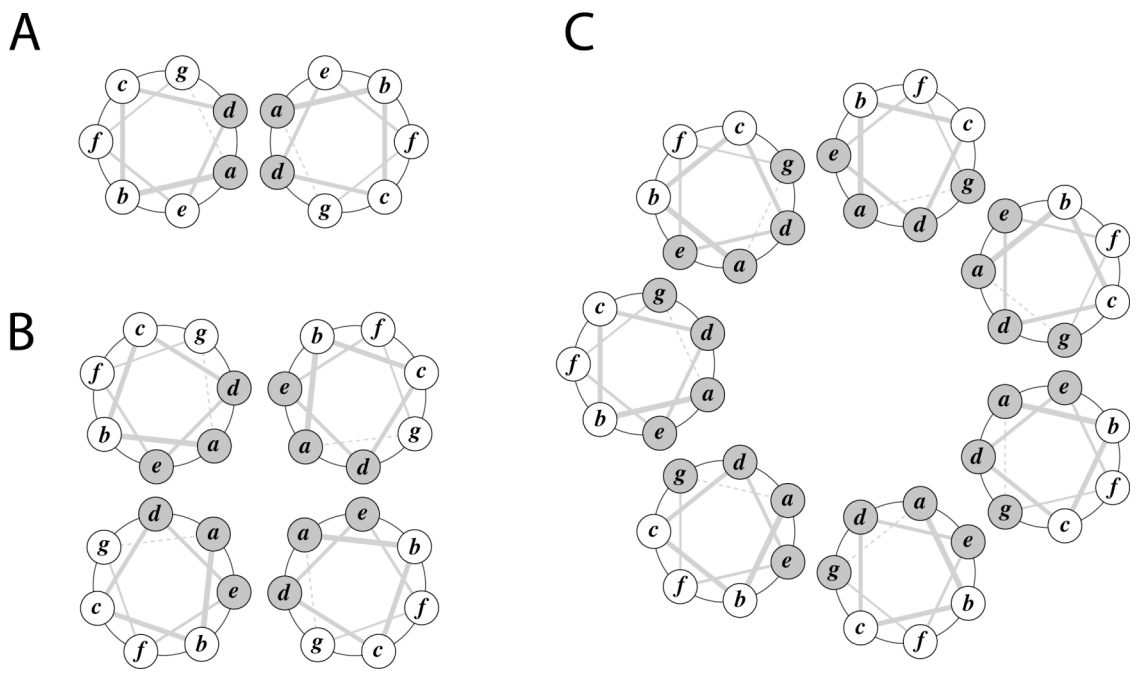


Figure 3.3. Dependence of LZ oligomer stoichiometry on the size of continuous hydrophobic core. (**A**) Canonical LZ dimer with (***a***+***d***) hydrophobic interface. (**B**) Extended hydrophobic interface (***a***+***d***+***e***) yields a tetramer (*24, 49*). (**C**) Four-residue hydrophobic interface (***a***+***d***+***e***+***g***) yields upto a heptameric ensemble (*18*).

In addition to functional specificity (selection of dimerization partners), ***g-e'*** ionic interactions contribute to the structural specificity of LZ motifs, modulating register and orientation of monomer chains in the oligomeric ensemle (*18, 50, 51*). Furthermore, though predominantly providing specificity-control in vivo, in vitro ***e*** and ***g*** positions can be employed for generation of high-order oligomers by extending the hydrophobic interface of the monomer chain. As originally shown by Harbury (*25*) (see "2.2 - secondary and tertiary structure" section above) the stoichiometry of the coiled coil oligomers is to a large extent defined by the packing geometry of the residues occupying ***a*** and ***d*** positions of the sequence. However, a simpler rule might also be of value in this respect – an estimate of continuous hydrophobic surface area carried by the coiled coil monomer. For example extension of a dimer-favoring 2-pair (***a***+***d***) hydrophobic interface (Figure 3.3-A), to a 3-pair (***a***+***d***+***e***) hydrophobic patches induces formation of tetramers (Figure 3.3-B) (*24, 49*), replacement of 14 Trp sidechains in ***a*** and ***d*** positions with bulky tryptophan residues results in pentameric bundle (*52*), and extension of a 2-pair interface (***a***+***d***) to a 4-pair (***a***+***d***+***g***+***e***) creates high-order oligomers (*53*) with a heptameric coiled coil being the most striking structurally characterized example (Figure 3.3-C)(*18*).

= 3.3 = A-A' interactions (stability and specificity)

The nature of this interaction has the most complex effect on the stability and specificity of the LZ interfaces. Similarly to the residues in ***d***-positions, packing of aliphatic side chains in ***a***-position affects the stability and stoichiometry of the complex, with prevalence of Cβ-branched amino acids (Ile, Val) (*48*) strongly favoring the dimeric structure of leucine zippers (*25*).

Similarly to Leucine in ***d***-positions, isoleucine exhibits uniquely efficient side-chain packing in ***a***-position, providing 9.2 kcal/mol/pair more energy than homotypic Ala interaction, and ~4 kcal/mol/pair over similarly sized Leu or Val sidechains (*31*). However, as opposed by the extreme conservation of leucines in ***d***-positions of the interface, isoleucine is a relatively infrequent residue in the ***a***-position, with its occurence probability twice less compared to that of either leucine, valine and even asparagine (*7*). Selection forces against most stable interactions can be explained by two evolutionary advantages. First, as will be discussed below, incorporation of destabilizing polar residues provides additional mechanism for control over transcription factor functional (defining appropriate partners) and structural (defining stoichiometry and orientation) specificities. Thus, high occurence of asparagine in the ***a***-positions of bZIP factors highlights specificity-driven rather than stability-driven evolutionary pressures acting on these motifs. Secondly, moderate stability of the interface defined by high abundance of leucine and valine sidechains in the ***a***-positions, as discussed in more detail in the "folding" section, reduces the activation energy needed for LZ dissociation, decreasing lifetime of the folded coiled coil state and elevating sensitivity of the LZ network to changes in external stimuli. This aspect underscores the notion of leucine zippers being a transient motif for signal transduction, rather than a static structural motif, as in the case of extended coiled coils.

A-A' stability scale

In addition to the "default" set of hydrophobic side chains, LZ factors often bear polar and charged residues in the ***a***-positions of the interface. This creates an additional mechanism for control of specificity allowing a dynamic range of homo- and hetero-dimerization events (*7, 47, 48*). Thermodynamic contribution of different residues to homodimeric ***a-a'*** interactions varies between stabilizing aliphatic, neutral polar and destabilizing charged sidechains (Table 3.1-C and diagonal in Table 3.2-D). This energy scale, relative to a pair of alanines, spans from –9.2 kcal/mol/pair for isoleucine to +6 kcal/mol/pair for glutamate (–0.9 kcal/mol/pair and +2.1 kcal/mol/pair in terms of coupling energies – Table 3.2-A), which signifies importance of individual ***a-a'*** couplings to the overall stability of the interface. Thus a vast 15 kcal/mol energy range is employed in regulation of LZ homodimerization

specificity. Similarly, a diverse ~11 kcal/mol range of stability contributions is available for heterodimeric ***a-a'*** interactions (Table 3.1-D), facilitating control over heterodimerization specificity. Interestingly, except interactions involving lysine sidechains, heterotypic interactions are predominantly destabilizing (Tables 3.1-D, 3.2-A&B).

A Coupling energies for ***a-a'*** pairs ($\Delta\Delta\Delta G_{int}$ [kcal/mol/pair])

	I	V	L	N	T	S	K	R	E
I	-0.9	0.4	0.9	4.9	2.2	2.1	-0.4	0.4	1.3
V	0.2	-0.7	0.2	3.2	0.3	1.5	-0.5	-0.2	0.7
L	0.5	0.2	-0.6	2.8	0.6	1.3	-0.1	0.1	0.4
N	4.3	3	2.6	-0.5	1.3	0.9	-0.2	-0.1	1
T	2.2	0.8	1.3	1.1	0.2	0.6	-0.6	-0.2	0.5
S	2.6	2.6	2.3	0.6	1.1	0.2	-0.1	0.1	0.2
K	-0.7	-0.5	-0.2	-0.3	-0.3	-0.4	0.3	0.8	0.5
R	-0.2	-0.3	-0.1	-0.2	0.1	-0.1	1.1	1.2	0.6
E	1.4	1.1	0.8	1.2	0.3	0.6	0.1	0.2	2.1

Total specificity ranges ($\Delta\Delta\Delta G_{int\text{-}max} - \Delta\Delta\Delta G_{int\text{-}min}$)

5.2	3.7	3.2	5.4	2.5	2.1	1.7	1.4	1.9

"specific" (favoring homodimers) | "non-specific" (favoring heterodimers)

B Coupling energy class averages w/o homodimers contributions ($\Delta\Delta\Delta\overline{G}_{int}$)

	I	V	L	N	T	S	K	R	E
aliphatic	0.4	0.3	0.6	3.6	1.0	1.6	-0.3	0.1	0.8
polar	3.0	2.1	2.1	0.9	1.2	0.8	-0.3	-0.1	0.6
charged	0.2	0.1	0.2	0.2	0.0	0.0	0.6	0.5	0.6

Class specificity ranges (e.g. $\Delta\Delta\Delta\overline{G}_{int\text{-}polar} - \Delta\Delta\Delta\overline{G}_{int\text{-}aliphatic}$)

aliphatic vs polar	3	2	2	3	0	1	0	0	0
aliphatic vs charged	0	0	0	3	1	2	1	0	0
charged vs polar	3	2	2	1	1	1	1	1	0

disfavoring polar | disfavoring aliphatic | "non-specific"

C Coupling energy double averages (w/o homodimer contributions)

aliphatic	**0.4**	**2.1**	0.19
polar	**2.4**	**0.93**	0.07
charged	0.14	0.1	**0.55**

D Free energy differences for ***a-a'*** pairs ($\Delta\Delta G_{A\text{-}A}$ [kcal/mol/pair])

	A	I	V	L	N	T	S	K	R	E
A	0.0	-4	-2.4	-2.3	-1	0.9	1.4	1.6	2	1.8
I	-4.3	-9.2	-6.3	-5.7	-0.4	-0.8	-0.8	-3.1	-1.9	-1.2
V	-2.3	-6.1	-5.4	-4.4	-0.1	-0.7	0.6	-1.2	-0.5	0.2
L	-2.3	-5.8	-4.5	-5.2	-0.5	-0.4	0.4	-0.8	-0.2	-0.1
N	-0.9	-0.6	-0.3	-0.6	-2.4	1.3	1.4	0.5	1	1.9
T	-0.3	-2.1	-1.9	-1.3	-0.2	0.8	1.7	0.7	1.5	2
S	0.2	-1.2	0.4	0.2	-0.2	2.6	1.8	1.7	2.3	2.2
K	1	-3.7	-1.9	-1.5	-0.3	2	2	2.9	3.9	3.3
R	1.4	-2.8	-1.3	-1	0.2	2.4	2.7	4.1	4.6	3.8
E	2.1	-0.5	0.8	0.6	2.3	3.3	4.1	3.8	4.3	6

Table 3.2. Specificity ranges of individual amino acids and amino acid classes in ***a***-positions of the LZ interface. (**A**) and (**D**) data adapted from (*7*). As in **Table 3.1** stabilizing coupling energies are highlighted blue, destabilizing – orange, and strongly destabilizing – red. (**A**) Coupling energies ($\Delta\Delta\Delta G_{int}$ [kcal/mol/pair]) with corresponding specificity ranges defined by the difference between highest and lowest $\Delta\Delta\Delta G_{int}$ values for particular residue. (**B**) and (**C**) Averaged coupling energies for heterodimeric couplings (i.e. devoid of homodimeric contributions) between different types of residues. (**D**) Free energy differences ($\Delta\Delta G_{A\text{-}A}$ [kcal/mol/pair]) relative to the pair of alanines.

In the case of homotypic interactions, notable outliers are asparagine and lysine. Increased stability of polar Asn is thought to be brought by its favourable self-complementing hydrogen bonding (*54*). Meanwhile repulsive electrostatic interactions of lysine sidechain, as in the case of ***g-e'*** interactions, are offset by favorable hydrophobic burial and efficient VdW packing of its aliphatic backbone (*34*).

Likewise, destabilizing effect of polar and charged sidechains placed into heterotypic aliphatic context inversely correlates with their net hydrophobicity (number of methylenes in

the sidechain) (Tables 3.1-D, 3.2-A&B): Lys [-$CH_2CH_2CH_2CH2$-] > Arg [-$CH_2CH_2CH_2$-] > Glu [-CH_2CH_2-] ≈ Thr [-CH_2CH_2-] > Ser [-CH_2-] > Asn [-CH_2-]. The highest destabilization effect is shown by asparagine, and similarly to its homotypic stabilizing effect is likely a consequence of an uncompensated hydrogen bonding (*54*). Unique properties of buried Asn sidechains for dimerization specificity control are underscored by its high abundance in naturally occuring LZ signalling networks (*48*). In addition to specificity control buried asparagines are known to be involved in control of LZ chain orientation (*55, 56*), register (*9*) and stoihiometry (*25, 57*), all factors possibly contributing to its frequent occurence within LZ motfis.

Overall the ***a-a'*** stability scale (from most stable to most unstable):

- ●● stabilizing: aliphatic (Ile > Val, Leu) and polar Asn
- ●○ neutral (hetero): charged•aliphatic, charged•polar
- ●● neutral (homo): polar (Thr, Ser), charged/aliphatic (Lys)
- ●○ moderately destabilizing (hetero): polar•polar, charged•charged, aliphatic•aliphatic
- ●● destabilizing (homo): charged Arg
- ●● strongly destabilizing: charged Glu
- ●○ strongly destabilizing (hetero): polar•aliphatic

A-A' specificity scale

As suggested by Acharya and coworkers (*7*) specificity of an individual amino acid in the ***a***-positions can be estimated via the net coupling energy range they are capable to exhibit depending on the interacting sidechain. I.e. it is net energetic difference between the most stable ($\Delta\Delta\Delta G_{int\text{-}min}$) and most unstable ($\Delta\Delta\Delta G_{int\text{-}max}$) coupling exhibited by particular sidechain (Footer of Table 3.2-A). In the case if amino acid is highly selective ("specific") it shall distinguish different pairing interactions, resulting in extended range of possible coupling energies. Conversely, non-selective ("unspecific") residue shall not distinguish between different pairing sidechains, therefore its stability contribution shall not vary much depending on the partner.

On this scale isoleucine and asparagine show the greatest difference in coupling energies, indicating that they contribute the most to dimerization specificity, while charged amino acids (K, R and E) show the least difference in coupling energies, suggesting that they contribute the least to dimerization specificity (i.e. tend to heterodimerize). Overall effect can be summarized as following - aliphatic residues (Ile, Val, and Leu) and Asn induce homotypic preferences in the LZ motif, polar Thr and Ser show neutral specificity, and charged sidechains (Lys, Arg and Glu) encourage heterodimerization (*7*).

To improve the precision of this analysis, we suggest to evaluate abilities of individual sidechains to distinguish different classes of residues (i.e. aliphatic, polar, charged). For this purpose averages of heterodimeric coupling energies (devoid of homodimeric contributions) (Table 3.2-B) for particular residue classes shall be compared (Footer of Table 3.2-B). This allows, for example, to see that "specificity" ("specificity range") of aliphatic sidechains is not uniform, and mainly relates to disfavouring polar partners, while being indifferent to aliphatic and charged sidechains. Furthermore, it becomes clear that polar serine and threonine also foster dimerization "specificity" similar to asparagine sidechain. Combined with Ser/Thr abundancies in the natural LZ motifs (*7*), this observation points to their possible role as an "intermediate" specificity restrictors, providing less stringent energy discrimination compared to the Asn sidechain.

Based on these revised "specificity ranges" the following conclusions for heterodimeric interactions can be made:

(1) aliphatic residues strongly disfavor polar partners, but do not distinguish between other sidchain types.
(2) correspondingly, polar residues strongly disvafour aliphatic partners, but are indifferent for other sidchain types.
(3) charged sidechains do not distinguish between sidechain types.

These conclusions are most strikingly revealed upon further averaging of coupling energies within particular classes (Table 3.2-C). It is apparent that among heterotypic interactions the most unfavorable are those involving aliphatic and polar sidechains, while charged residues provide most stable couplings independent of the context.

Considering default hydrophobicity of the LZ core, the ***a-a'*** position specificity scale can be reformulated as following, from favoring homodimers ("specific") to favoring heterodimers ("unspecific"):

●● ●● polar (Asn > Ser > Thr) (favoring homodimers & disfavoring heterodimers)
●● aliphatic (Ile, Val, Leu) (favoring homodimers)
●○ charged Lys (favoring heterodimers)
●○ ●○ charged (Glu, Arg) (favoring heterodimers & disfavoring homodimers)

= 3.6 = Anti-parallel leucine zippers

Along with the widespread parallel dimeric architectures, coiled coils are able to assemble complexes with an anti-parallel arrangement of helices. These structures seem to be poorly represented in nature, and therefore have not received due attention, although there seems to be an increase in interest to anti-parallel structures in the last years (*9, 58*). In principle these structures could fall under the same "leucine zipper" nomenclature, because of

the similar heptad repeat featuring conserved leucine side-chain at every seventh residue (a "***d***-position") (*9*). However, similarly to the structural roles of extended parallel coiled coils, majority of the existing examples from the anti-parallel coiled coils are involved in formation of static structural cores, rather than dynamic signalling interfaces, therefore falling beyond the scope of this review. Nevertheless, a few key characteristics of these assemblies will be shortly highlighted below.

Similarly to their parallel relatives, antiparallel assemblies feature hydrophobic core formed by apolar side-chains in the ***a*** and ***d*** positions of the heptad repeat, stabilized by the electrostatic interactions between charged residues in the ***g*** and ***e*** positions. In the case of anti-parallel structures ***a-a'*** and ***d-d'*** hydrophobic interactions are replaced by ***a-d'*** and ***d-a'*** pairs, and ***g-e'*** electrostatic couplings are replaced with ***g-g'*** and ***e-e'*** pairs. As in the case of parallel structures most of the stability is conferred via the hydrophobic core, while specificity and anti-parallel chain orientation itself is mainly defined by Coulombic interactions between side-chains in the ***g*** and ***e*** positions (*50, 51*).

In addition, the potential of buried polar residues in determining structural integrity of anti-parallel coiled coils has been demonstrated by replacement of ***a-d'*** hydrophobic residues with a pair of asparagines (*59*). However, although ***a-a'*** polar interactions are an important specificity determinant for naturally occurring leucine zippers, the equivalent ***a-d'*** polar interactions has not been reported for anti-parallel assemblies.

Summing up – anti-parallel coiled coil interfaces seem to bear all the required determinants for assembly of signalling regulatory networks similar to those based on the leucine zipper interfaces. However there is one crucial difference, which appears to be the intrinsic limitation of an anti-parallel architecture for emergence of cellular signalling cascades, let alone the formation of complex combinatorial signalling networks. This limitation stems from the packing efficiency of the hydrophobic core, which defines the structural integrity and stoichiometry of the coiled coil complex. As discussed above (see section 2.2 - secondary and tertiary structure), extreme stability and specificity of the parallel dimeric LZ interface is defined by very specific and efficient packing of hydrophobic side-chains within its core – ***a***-layer side-chains adopt so called "parallel" orientation, while ***d***-layer adopts a distinct "perpendicular" arrangement. Packing of ***d***-position side-chains delivers most of the energy required to stabilize the interface, which allows certain flexibility at ***a***-positions, thus providing a mechanism for increase of stoichiometric specificity of the complex via introduction of polar residues in ***a***-positions. In the case of anti-parallel structures the ability to differentiate between stability vs specificity contributions is eliminated, since in these structures hydrophobic layers adopt a single geometrical type of side-chain arrangement, involving mixture of ***a*** and ***d*** side-chains (*58*). This lack of intrinsic

structural specificity is neatly demonstrated by heterogeneity of structural species formed by 5-heptad coiled coil domain from hepatitis delta virus antigen (*60*) and structural instability of 10-heptad coiled coil from E.coli Seryl tRNA Synthetase (*61*).

= 3.7 = LZ network design

Reviewing the discussed above LZ specificity rules, a few general remarks can be made. In the context of an isolated heptad homodimerization specificity can by achieved by incorporation of polar residues into ***a*** positions (moderately affecting homodimerization while strongly disfavoring heterodimerization) and incorporation of residues with alternating charges into the ***g-e'*** positions (since non-alternating ***g-e'*** charges disfavor homodimerization). Increased heterodimerization specificity can be achieved by incorporation of charged residues into ***a*** positions (disfavoring homodimeric while stabilizing most of heterodimeric couplings) and also introduction of identically charged residues in ***g-e'*** positions (seriously destabilizing homodimers).

Speaking about networks of factors, in the context of prevailing aliphatic side-chains in the ***a***-positions, the combinatorial specificity of a particular network can be increased by incorporation of polar residues (especially asparagines and serines) into the unique ***a***-positions of the interface – this will create a strong destabilizing effect for all except homotypic interactions (i.e. those having polar residues in the same position). Similarly, to couple a LZ-factor to a network defined by particular allocation of a buried polar sidechain, one has to place a polar side-chain in the corresponding location in the interface. To provide coupling between two networks specified by distinct allocations of buried polar residues, one shall incorporate charged residues in corresponding ***a***-positions of the interface (thus oligomerization within either of the networks will not involve unfavorable aliphatic • polar interactions). General increase in the amount of charged side-chains in ***a***-positions decreases the specificity and increases the range of interactions available for a particular LZ motif. Thus coupling of several specialized networks via a central hub requires more "unspecific" (destabilizing) residues in ***a*** positions of the heterodimerizing zipper, putting additional pressure on the optimization of its ***d-d'*** and ***g-e'*** interactions.

To selectively decouple distinct networks one has to increase the amount of specificity determinants – introduce polar residues in non-matching ***a***-positions and repulsive interactions between ***g-e'*** sidechains. These specificity determinants will not affect oligomerization within the family, while strongly disfavoring any interactions outside of it.

= 3.8 = Conclusion

Clearly, for an adequate analysis of particular interface stability and specificity local context of described above interactions will play a very important role. For example thermodynamic contribution of aliphatic side chains in the ***d***-position varies up to 4 kcal/mol/pair (for leucine) depending on the neighboring residues in ***a***-positions (*29, 62*); contribution of buried Asn residues is also context-dependent, varying on the order of 2 kcal/mol/pair depending on the environment (*63*); similar variability is shown by electrostatic ***g-e'*** couplings (discussed in (*32*)); and not surprisingly polar and charged amino acids placed in ***a***-positions do energetically differentiate reciprocal orientations of overlying electrostatic ***g-e'*** pairs, with the greatest difference of up to 1.5 kcal/mol/dimer observed for heterodimeric Thr-Asp coupling (*7*). As well it has been suggested that not only the sums of individual energies, but also the patterns of interactions define the stability and specificity of the LZ interfaces (*48*). Therefore, the issue of context still has to be resolved in more detail to increase the accuracy of our predictions. Nevertheless, as verified by experimental data (*64*), even in the absence of more detailed contextual analysis, a simplified set of LZ specificity determinants already yields quite realistic predictions on oligomerization properties of canonical leucine zippers (*47*).

Importantly, beyond contextual dependencies, quite some gaps remain in our fundamental understanding of LZ specificity determinants. For example interactions within Myc/Max/Mxd network of oncogenic bHLH-LZ factors are specified by buried salt bridges involving ***d***-position histidine on the Max side and ***a***-position glutamate/aspartate residues on the Myc/Mxd side (*36, 37*). Another example refers to a group of plant bZIP TFs which employs a conserved proline residue in the ***f***-position of the interface to restrict formation of homodimers, thus profoundly changing the topology of the signalling network (*65*).

Therefore it seems reasonable to apply described above simplified set of determinants only in the context of "canonical" LZ motifs, and when comparing interactions with notable energetic differences, since subtle energy variances will be masked by the error imposed with these simplifications. Further advancements in our understanding of LZ interaction stability and specificity require more thorough sampling of the interaction space, and thus are expected to come from the systems biology approaches (*7, 66*).

= 4 = Folding

The stability and specificity rules derived from thermodynamic properties of LZ interfaces provide only a partial insight into the nature of LZ-mediated signal transduction, showing the network equilibrium state at an infinite time limit. In addition to this "thermodynamic control", protein signalling is highly dependent on the kinetics of particular interactions, including the presence of structured intermediates which provide specificity filters when signal transduction is coupled with folding process. These characteristics of protein folding landscape facilitate the "kinetic control" over signalling events, determining sensitivity and dynamics of response to the changes in the input signal (i.e. timescale at which the signalling event occurs). In the following section we review the existing knowledge on the folding of parallel dimeric leucine zipper motifs, to aid further understanding of LZ signalling mechanisms.

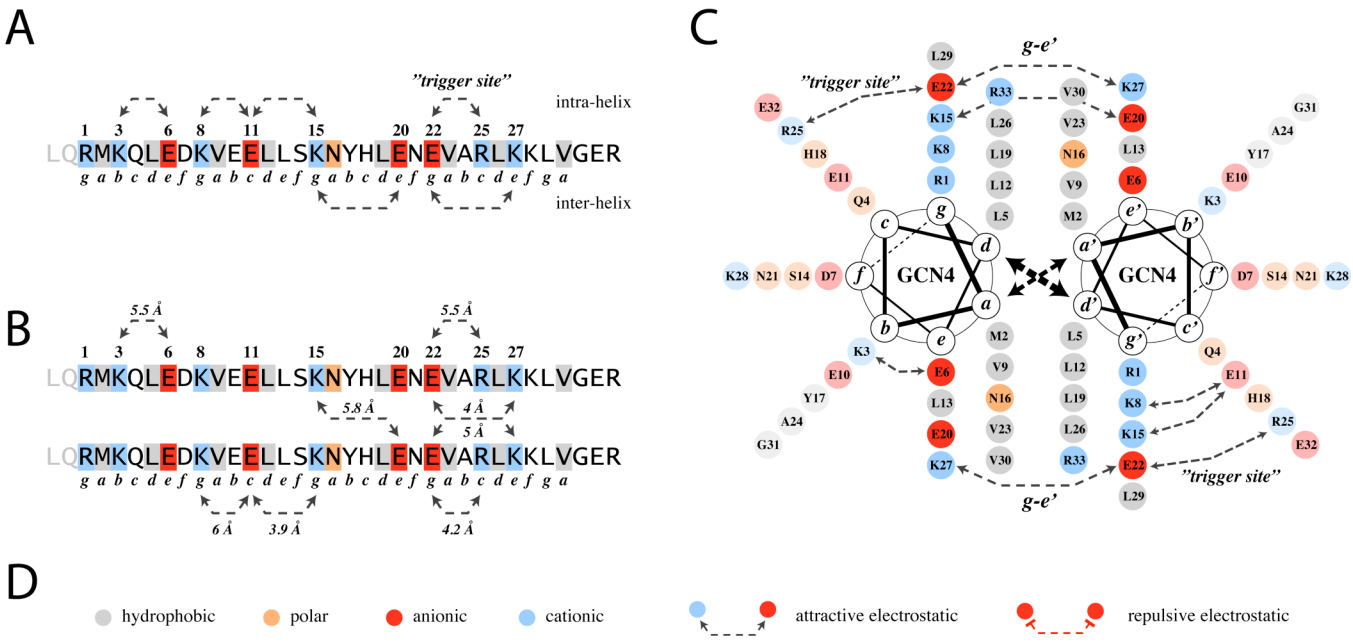


Figure 4.1. Electrostatic interactions within GCN4p1 (LZ-GCN4). Residue numbering according to GCN4p1 sequence. (**A**) One-chain linear notation, most useful for illustration of interactions within symmetrical homodimeric LZ motifs. (**B**) Two-chain linear notation, most useful for illustration of asymmetrical LZ motifs. (**C**) Helical wheel notation, useful in any situation. (**D**) Figure legend (same coloring is employed in all other figures with LZ motifs). Distances between charged atoms in (**B**) are based on the basic region + leucine zipper fragment of GCN4 bound to its consensus DNA sequence (pdb:1ysa).

For the most part our knowledge of LZ folding process is based on the studies of GCN4p1 – archetypical 33 amino acid peptide corresponding to the LZ motif of yeast transcription factor GCN4 (Figure 4.1 – LZ-GCN4 electrostatics). In addition a considerable

amount of experimental data relates to engineered LZ motifs, designed to distinguish various contributions (hydrophobics (*67, 68*), electrostatics (*69, 70*), helical propensities (*68, 71, 72*)) to LZ folding landscape. For a long time the general view on LZ folding was that monomer chains are largely unstructured at the early stages of the folding process (*68, 69, 73, 74*), and that the main energy barrier in the folding direction is highly entropic by nature (i.e. determined largely by hydrophobic and VdW interactions of ***a-a'*** and ***d-d'*** couplings) (*71, 75*). However, later it became apparent that at least one helical intermediate is populated prior to the main folding event (*6, 67, 71, 76, 77*). And finally, studies of LZ motifs from Jun and Fos transcription factors later revealed the importance of enthalpic component (electrostatic and polar ***g-e'*** interactions) within activation barrier in the folding direction (*78, 79*).

Here, reasoning on the available data, we propose that generalized folding process of short LZ motifs is best described by the Diffusion-Collision-Desolvation model (Figure 4.2). In this model, stretches of helical structure (corresponding to the "microdomain" elements of the original Diffusion-Collision model (*80*)) are primed by hydrogen bonds and stabilized by intra-helical salt bridges within LZ monomers at the early stages of the folding process (*16, 77, 81*). These intermediates collide in a diffusion-limited manner, with the probability of accessing productive transition state dependent both on the prominence of helical structures and the rate of collisions between these "microdomains". In Diffusion-Collision-Desolvation model the main activation energy barrier is also highly dependent on the long-range electrostatic interactions between the monomer chains – balanced between favorable "electrostatic guidance" (*82*) and unfavorable desolvation contributions (*45, 83*). These contributions are reflected in the enthalpic component of the free energy barrier (*78*) and provide an essential LZ specificity discrimination mechanism, based on the long-range coulombic forces.

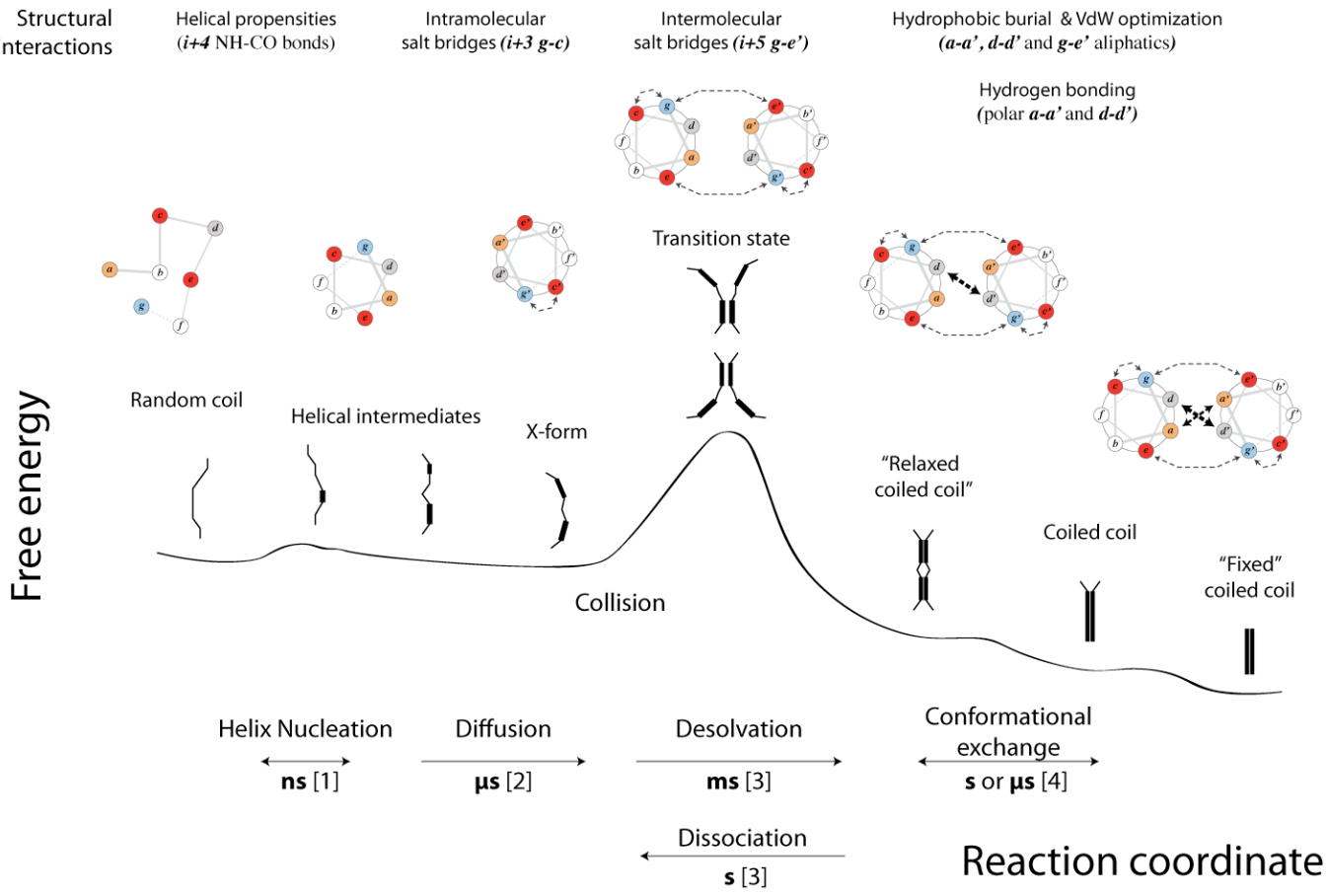


Figure 4.2. Diffusion-Collision-Desolvation model for LZ folding. References for kinetic rates:
[1] Helix nucleation, 1-17 ns (*84-87*).
[2] Theoretical diffusion-limited collision rate ~2.5 µs at 100 µM peptide (*67, 76, 88*).
[3] Forward time constant (monomer lifetime) 0.7-25 ms at 100 µM peptide; reverse time constant (dimer lifetime) 2-300 s (*16, 68, 73, 75, 76, 81*).
[4] Exchange time scale 0.4 s for GCN4-lzK analog (*89*); 0.2-1.2 s Jun-Fos analog (*79*); ~10 µs for crosslinked GCN4p1 (*15*).

Similar to the monomeric intermediates at the non-native side of the folding barrier, several groups have reported on existence of a stable dimeric intermediate at the native side of the folding barrier (*15, 89, 90*), designated here as the "relaxed coiled coil" state. The exact nature of this state has yet to be revealed, however repacking of the hydrophobic core within the central region of this structure (*90*), points to a possible rationale behind this transition. Namely, interactions involving polar buried residues in the ***a***-positions of the LZ interface (N16 in case of GCN4p1) we shown to manifest themselves only after the rate-limiting step in LZ folding process (*91*). Therefore, it seems plausible that the "relaxed coiled coil" state is defined by reorganization of VdW packing and hydrogen bonding established by buried polar sidechains. This reorganization decreases the stability of the final coiled coil state, reducing the height of unfolding activation barrier, thereby modulating the lifetime of the signalling event.

Combined DCD model explains LZ specificity discrimination mechanism based on the long-range electrostatic interactions between monomer chains, and elucidates essential "kinetic control" components on both sides of the main folding barrier. Combinatorial multiplicity of LZ interfaces discussed in the previous section, together with the flexible kinetic control of LZ folding lansdcape, provide the fundamental basis for the remarkable versatility and robustness of this motif in estabilishment of protein signalling pathways.

Summarized above aspects of LZ folding process are discussed in more detail below.

= 4.1 = Folding models

Two-state model

For a number of years folding of LZ was considered a two-state process involving predominantly unstructured monomer and a fully-fledged coiled coil dimer (*68, 69, 73, 74*). In the two-state model folding starts as a collision of two unstructured monomer chains, followed by a "downhill" hydrophobic collapse resulting in formation of a folded coiled coil dimer. This does not mean an all-or-none synchronous structuring of the whole chain, but rather refers to the situation where all molecular conformations can be organized into two general groups divided by a single high-energy barrier. In the case of leucine zippers those groups represent predominantly disordered monomers (M) and predominantly folded coiled coil dimers (CC):

$$M + M \underset{k_{-1}}{\overset{k_1}{\leftrightarrow}} CC \qquad (1a)$$

In the two-state LZ folding approximation the transition state contains little if any secondary structure, and the highest energy barrier (rate limiting step) is primarily dictated by the diffusion processes:

$$k_f \propto D \qquad (1b)$$

where D reflects the frequency of diffusion-limited collisions events.

As happened in the course of early LZ folding studies, depending on the sensitivity and time resolving capabilities of particular experimental setup, some non-two-state processes may appear as two-state because of short lifetimes and/or low stabilities of the folding intermediates.

Diffusion-collision model

Eventually experimental data started to accumulate indicating that folding of LZ is better described by a diffusion-collision model (*92*), where at least one helical intermediate is populated prior to the main folding event (*6, 67, 71, 76, 81, 93*). As opposed to the two-state

model, Diffusion-Collision theory relies on the existence of preformed structural elements, termed microdomains, which collide at diffusion-limited rates (*80*).

$$M + M \underset{k_{-1}}{\overset{k_1}{\leftrightarrow}} I^* + I^* \underset{k_{-2}}{\overset{k_2}{\leftrightarrow}} CC \qquad (2a)$$

In this model the folding rate is dependent both on diffusion-mediated processes and "coalescence probability" term:

$$k_f \propto D \times \beta \qquad (2b)$$

where beta ("coalescence probability") corresponds to the fraction of collisions which are productive (leading to the transition state), embracing both the prominence of elementary microdomains (defined by $\boldsymbol{k_1/k_{-1}}$) and barriers mediating the coalescence step (e.g. probability of productive orientation at the moment of encounter – defined by $\boldsymbol{k_2/k_{-2}}$).

Most of researchers currently support the DC concept, agreeing that simple kinetic considerations strongly favor this model. Specifically, helix nucleation (i.e. $M \rightarrow I^*$ transition) has been reported to occur on the nanosecond timescale (~1-17 ns) (*84-87*), while theoretical LZ monomer collision rate is 3 orders of magnitude slower (~2.5 microseconds at 100 μM peptide) (*67, 76, 88*)), and experimentally observed LZ folding rate is yet another 3-4 orders of magnitude slower than collision rate (0.7-24 milliseconds at 100 μM monomer concentration) (*16, 68, 73, 75, 76, 81*) (Table 4.1). Notable difference in timescales of individual folding steps (~ns helix nucleation → ~μs collision → ~ms dimer assembly), indicates that (1) at the moment of collision monomer chains contain a considerable amount of pre-formed helical structure, and (2) only a fraction of collisions leads to formation of the coiled coil dimer. Although from theoretical standpoint it was reported plausible to design a coiled coil with negligible intrinsic helicity that folds via pure two-state mechanism (*70*), natural occurrence of such monotonous sequences is extremely unlikely and thus application of the collision-first two-state model would be an oversimplification.

Helix nucleation

	k_1 (unimolecular rate constant) [M-1 s-1]	time to helix nucleation (unimolecular) $1/(k_1)$ [ms]
Hummer, 2000		0.000001
Thompson, 1997	1.E+08	0.000010
Williams, 1996	6.E+07	0.000017

Collision (theoretical)

	k_1 (bimolecular collision rate constant) [M-1 s-1]	time to collision @ 100μM $1/(k_1*100\mu M)$ [ms]
Durr, 1999	5.E+09	0.002
Zitzewitz, 2000	3.E+09	0.003
Holtzer, 2001	3.E+09	0.003

Dimerization

	peptide	NaCl [mM]	buffer [mM]		temp [ºC]	pH	k_1 (bimolecular association rate constant) [M-1 s-1]	monomer lifetime @ 100μM (association time) $1/(k_1*100\mu M)$ [ms]	k_{-1} (unimolecular dissociation rate constant) [s-1]	dimer lifetime (dissociation time) $1/k_{-1}$ [s]	K_d k_{-1}/k_1 [nM]
Zitzewitz, 1995	GCN4p1	150	50	Phosphate	5	7	4.2E+05	24	0.0033	303	8
Zitzewitz, 2000	GCN4p1	150	50	Phosphate	15	7	2.2E+06	4.5	0.0183	55	8
Moran, 1999	GCN4p1	150-200	20	Acetate	10	5.5	2.0E+06	5.0	0.0068†	146	3
Bosshard, 2001	BRLZ-GCN4 (62aa)	10	50	Phosphate	5	7.4	5.5E+06	1.8	0.01	100	2
	BRLZ-GCN4 (62aa)	10	50	Phosphate	30	7.4	1.4E+07	0.7	0.6	2	43
Ibarra-Molero, 2004	GCN4p1-M2V	150	50	Phosphate	15	7	2.2E+06	4.6	0.45	2	205
Steinmetz, 2007	GCN4p1	150	10	Phosphate	5	7.4	6.6E+05	15	0.17	6	256
					AVERAGE:		**4.E+06**	8	**0.18**	88	
Nikolaev, 2007 (X-form)	Leu-Gln-GCN4p1	50	50	Acetate	37	3.2	**8.6E+04**	116	**2.6**	0.4	30000

Dimerization (cross-linked vs dimeric)

	peptide	NaCl [mM]	buffer [mM]		temp [ºC]	pH	k_1 (association rate constant) [M-1 s-1]	association time [ms]	k_{-1} (dissociation rate constant) [s-1]	dissociation time	
Moran, 1999	**cross-linked GCN4p1**	150-200	20	Acetate	10	5.5	7.5E+03	**0.133***			
	GCN4p1	150-200	20	Acetate	10	5.5	2.0E+06	5	0.0068†	146	3
Wang, 2005	**cross-linked GCN4p1**	N/A	50	Phosphate	T-jump	7		**0.100*** **0.010***			

* - unimolecular process
† - graphical estimate from Chevron plot

Table 4.1. Summary of LZ folding rates. Corresponding references (coloumn 1): Hummer, 2000 (*85*); Thompson, 1997 (*86*); Williams, 1996 (*87*); Durr, 1999 (*67*); Zitzewitz, 2000 (*76*); Holtzer, 2001 (*88*); Zitzewitz, 1995 (*73*); Moran, 1999 (*68*); Bosshard, 2001 (*75*); Ibarra-Molero, 2004 (*81*); Steinmetz, 2007 (*16*); Nikolaev, 2007 (*77*); Wang, 2005 (*15*).

= 4.2 = Folding intermediates

The transition from the two-state to the Diffusion-Collision folding model of LZ folding, was accompanied by discoveries of stable folding intermediates at both non-native (monomeric) and native (dimeric) sides of the folding barrier. Related findings are summarized below, and importance of both intermediate types in establishing "kinetic control" over LZ signalling process is discussed.

Triggering sequence

According to the DC model, the protein folding rate increases proportionally with the endurance of preformed structural elements (helical segments in the case of coiled coils). Formation of these elements reduces conformational entropy of the peptide chain, thus increasing the population of association-competent conformational states and decreasing the activation energy required to access the transition state. The change from the two-state to DC view on LZ folding process was initially sparked by the emergence of the "triggering sequence" concept. This concept suggested that a conserved set of electrostatic interactions is present in a diverse set of coiled coil motifs, which induces formation of helical structures in the monomeric chains, increasing folding rates of short LZ sequences and providing a folding nucleation site for extended coiled coils (*6*). In archetypic GCN4p1 this "triggering sequence" is exemplified by the cluster of interactions around Glu22-Arg25 salt bridge (Figure 4.1). A diverse set of experimental studies has indeed confirmed the importance of the E22-R25 intramolecular salt bridge for helix stabilization in the early steps of the GCN4p1 folding process (*16, 71, 76, 81*).

Nonetheless, researchers later concurred that alpha-helix can be stabilized in many different ways, and that within extreme diversity of the coiled coil motifs a specific consensus "triggering sequence" is unlikely to exist. This conclusion is supported by experimental studies showing that early folding kinetics of LZ are also highly sensitive to perturbations in peptide intrinsic helicity (*68, 71*) and chain hydrophobicity (*67, 68*). Furthermore, as shown by Darin Lee and coworkers (*72*) the presence of the proposed triggering sequence ***per se*** does not guarantee successful folding of GCN4p1 analogs.

The final remark regarding the "triggering sequence" concept relates to the multiplicity of the accesible protein folding pathways, as proposed within original DC theory (*80*). Although not disallowing the existence of uniquely robust folding pathway, the DC model explicitly permits multiple folding routes to be attained (*80*). Individual folding pathways thus depend on the properties of particular microdomains, providing an error-resistant folding landscape in the case of multiple-microdomain architecture (the protein is less likely to be trapped in the local energetic minimum). As shown by Liam Moran (*68*), GCN4p1 sequence appears to bear three helix-nucleating regions – approximated at the N-term, C-term and at the center of the sequence. This correlates with the reported heterogeneity of the GCN4p1 transition state (*70, 76*) and reflects the adaptive characteristics of GCN4p1 folding landscape. When a single-site mutation is made that disrupts folding through one of these regions, folding proceeds through the other pathways with only minimal decrease in the folding rate (*68*). And only in the case of simultaneous disruption of all nucleation sites, a

large decrease in the folding rate is observed. These observations are in the perfect consent with the DC model.

Taking together the diversity of the coiled coil class of proteins, complex nature of helix-stabilizing interactions and multiplicity of folding pathways within DC model, we suggest a reformulation of definite “THE triggering sequence” into an indefinite “A triggering site”, in accordance with the original DC microdomain concept. As shown below, GCN4p1 E22-R25 salt bridge indeed appears to be not the solitary helix-stabilizing determinant within the GCN4p1 sequence.

X-form

Notwithstanding large amount of data pointing to the existence of a stable folding intermediates in the GCN4p1 folding pathway, for a long time no high-resolution experimental study has focused on their characterization. Recently, employing solution NMR methodology, we have identified and characterized the “x-form” – a novel stable conformation of GCN4p1, which exists in equilibrium with the coiled coil form. X-form is a semi-structured folding intermediate, populated at about 1% at ambient conditions, but considerably stabilized in the acidic pH. In the very first 1991 NMR structural study of GCN4p1 an additional set of resonances was observed at low protein concentration, pointing to the presence of a second conformational ensemble (*94*), which were essentially unattended by the authors. All further high resolution studies were conducted either at the very high protein concentration (above 1 mM), or in the neutral pH range, where this novel conformational state (x-form) is only marginally populated. Concentration dependence of the x-form population together with the slow exchange regime between the coiled coil and x-form conformations in the millisecond time scale, unequivocally renders the x-form as an intermediate at the monomer side of the folding barrier.

Within DC model folding proceeds via microdomains, which may be detected experimentally as structured intermediates transiently populated during the folding process. Populations of folding intermediates, as well as final folded states, are strongly dependant on the kinetic rates of transitions between these states. X-form population increases in acidic pH and at low peptide concentration due to the perturbations of the kinetic rates of the transitions linking x-form with other LZ conformations. At the first stage of LZ folding helix propagation is energetically favorable (*84*), which leads to accumulation of x-form at the main transition barrier, unless it associates to form the coiled coil dimer. When LZ peptide concentration is low, the rate of [successful] chain collisions decreases, reducing the rate of coiled coil formation and increasing the population of monomeric x-form intermediate.

Acidic pH appears to have additional double effect on the equilibrium distribution of the conformational states at the final stages of the LZ-GCN4 folding process. Firstly, protonation of the Glu side-chains eliminates the "electrostatic guidance" effect of attractive interchain ionic interactions (*82*), decreasing the apparent association rate from ~4 (average from 6 studies) to 0.08 [$10^6 * M^{-1} * s^{-1}$] (Table 4.1). Secondly, uncompensated positive charges in the final coiled coil structure destabilize the dimer, increasing the dissociation rate from 0.18 (average) to 2.6 [s^{-1}] (Table 4.1). Combination of these effects leads to accumulation of the monomeric folding intermediate – x-form, characterized in the second part of this thesis.

Structural information obtained from NMR data reveals a considerable amount of helical structure present in the x-form. Most importantly, NMR data indicates that LZ-GCN4 pre-collision intermediate bears two regions of increased helicity – allocated in vicinity of intrahelical ***i*,*i*+3** salt bridges K8-E11 and E22-R25 (GCN4p1 numbering) (Figure 4.1), showing that E22-R25 "trigger site" is not the only helix-nucleating determinant in LZ-GCN4. Although it has been reported that in acidic pH the helix structure of the "triggering site" is nearly abolished (*16*), the side-chain pK_a studies show that none of the GCN4p1 glutamates is fully protonated at the pH 3.2 (*95*). This allows a fraction of salt-bridge stabilized helical microdomains to be maintained even at acidic conditions. Further high-resolution studies are required to elucidate the exact 3D structure of the x-form and explain different endurances of the N-terminal K8-E11 and C-terminal E22-R25 triggering sites.

Importance of monomeric intermediates

Although appearing marginally populated at high protein concentrations and neutral pH conditions, x-form could represent a considerable fraction of the LZ-GCN4 structural ensemble in the ***in vivo*** conditions of low peptide concentration. Therefore, this conformational state might be biologically relevant in providing a transient interface for recombination of LZ partners in the context of cellular signalling networks. Structured alpha-helical intermediates appear to be an essential part of the generalized LZ folding model. They provide a robust scaffold with native-like positioning of electrostatic and hydrophobic residues, which enables discrimination between individual interacting partners both on primary (sequence) and tertiary (chain register and orientation) structural levels.

Intermediates at the native side of the main folding barrier

Even before first notions of monomeric helical intermediates have emerged, Bosshard group highlighted the biphasic nature of LZ folding kinetics (*96*). Soon after that Alfred Holtzer and colleagues employing equilibrium kinetic measurements by NMR have further challenged the apparent uniformity of the coiled coil ensemble at the native side of the

transition barrier (*89, 97-99*). Initially these observations had been overwhelmed by experimental data from other groups, and only several years later more evidences of stable folding intermediates at the native side of the LZ folding barrier appeared (*15, 90*). The most intuitive picture on these folding transitions can be derived from a thermodynamic study by Anatoly Dragan and Peter Privalov (*90*). According to this study, unfolding of LZ-GCN4 can be modeled by at least three step mechanism, with first two transitions being concentration independent (unimolecular) and only third one – concentration-dependent dissociation of the strands. This indicates that at least three conformational ensembles exist at the native side of LZ folding barrier. The first transition, starting from a 100% coiled coil state occurs at the temperatures around 0ºC for GCN4p1 and corresponds to the fraying of few N-terminal residues. The second transition occurs at much higher temperatures (starting at about ~20ºC for GCN4p1) and shows perturbations at both termini as well as structure repacking in the central region of LZ. Further increase in temperature (above ~45ºC for GCN4p1) eventually induces cooperative dissociation of coiled coil into monomers. Characteristics of the second structural transition, observed between 20 and 45ºC before the dissociation of the dimer, coincide with the conformational exchange between two folded LZ states reported earlier by Alfred Holtzer and colleagues (*89, 97*). We dub the pre-dissociation dimeric intermediate state as a "relaxed coiled coil" conformation.

Ting Wang and coworkers, employing T-jump relaxation experiments on the cross-linked version of GCN4p1, reported two equilibrium processes with the time relaxation constants in the order of 100 µs and 10 µs (*15*). Authors attributed the slow (100 µs) relaxation component to the biomolecular coiled coil folding reaction, and fast (10 µs) component to the conformational exchange at the native side of the folding barrier. The observed timescale of the slow transition indeed fits perfectly to the timescale of GCN4p1 association. As shown by Liam Moran (*68*) 133 µs (= $1/k_{on}$) folding timescale of the crosslinked GCN4p1 corresponds to ~5 ms for bimolecular association of 100 µM non-crosslinked GCN4p1 ($1/[100\ \mu M * k_{on}]$) (Table 4.1). Therefore 100 µs relaxation process reported by Wang fits to the millisecond timescale of GCN4p1 association reported by numerous research groups (compared at 100 µM peptide concentration)(*16, 75, 76, 81*) (Table 4.1). However the fast (10 µs) component assigned to unimolecular transitions at the native side of the folding barrier is in sharp contrast with the millisecond–second timescale reported for GCN4p1 analogs (*89, 97*), as well as analogs of Jun-Fos heterodimer (*79*). This discrepancy is also highlighted by the differences in structural characteristics of reported states. The 10 µs transition observed by Wang apparently involved a change in the secondary structure, while transitions observed by Andre d'Avignon and Jody Mason appeared to involve only repacking within the hydrophobic core while maintaining overall helical

structure, in accordance with the "relaxed coiled coil" observed by Anatoly Dragan (*90*). Further equilibrium kinetic and thermodynamic studies involving dimeric (i.e. non-cross-linked) forms of GCN4p1 and its analogs are required to clarify the nature of this conformational state.

Of all interactions defining specificity and stability of canonical leucine zippers, only buried polar residues at the ***a***-positions of the interface have not been thoroughly investigated. It has been shown that these interactions do not affect the folding reaction rates prior to the main transition barrier (*91*), and therefore are considered to manifest themselves only at the dimeric side of the folding barrier. In this perspective it is tempting to speculate that the transition between the coiled coil and "relaxed coiled coil" states involves reorganization of the VdW packing and hydrogen bonding in vicinity of these buried polar sidechains. From this standpoint, buried non-hydrophobic sidechains would decrease the stability of the coiled coil state, reducing the height of the unfolding activation barrier and thereby modulating the lifetime of the LZ signalling event or endurance of a catalytic interface in leucine zippers with enzymatic activity.

= 4.3 = Diffusion-Collision-Desolvation (DCD) model

Employing DC model LZ folding is defined by two transition barriers: minor helix nucleation barrier traversed on the nanosecond timescale, followed by major dimerization barrier traversed on the millisecond timescale (Figure 4.2). Second barrier occurs upon the collision of monomer chains and reflects the probability of pre-formed structural elements to establish a productive transition state. Notwithstanding massive efforts invested in the kinetic and thermodynamic studies of LZ folding, some controversies remain in our view on the physical nature of the main folding barrier and transition state ensemble associated with this event. One of the prevailing concepts in LZ folding is that the transition state comprises poorly structured dimer, with transiently formed helices undergoing search of complementary nucleating segments via local VdW interactions (*71, 75*), and that inter-chain electrostatic interactions do not affect the early folding and form only at the native side of the transition barrier (*81*). In this representation, the nature of the main activation barrier in the folding direction is purely entropic.

On the other hand, it seems more logical to expect that long-range Coulombic interactions would dominate over short range VdW and hydrophobic interactions in discrimination of dimerization partners and stabilization of the transition state. Indeed, some facts support this point of view. First, the low buried nonpolar surface area (~10-30%) reported for the transition state ensemble in a few cases (*75, 79*) points to insufficiency of short range VdW interactions in establishing the productive transition state. This is supported

by the notion of buried polar residues having negligible effect on LZ folding rates (*91*), an observation which is hard to envision if the transition state is stabilized by short-range interactions. As well several direct evidences point to the importance of electrostatic interactions within transition state ensemble. Hans Wendt in 1997 observed strong dependence of LZ folding rates on the ionic strength of the folding milieu, which suggested formation of an electrostatically stabilized transition intermediate during the rate-limiting step (*69*). In a more recent study, the folding rate of the Fos-JunW heterodimer decreased 6-fold upon introduction of additional charged residue by Q21R mutation, although the mutation increased local helical propensity and created a new interhelical salt bridge (*79*). Finally, group of Alfred Holtzer has shown that upon assembly of LZ-cJun dimer, nearly 65% of the free energy barrier in the folding direction is due to enthalpic contributions, seriously undermining the possibility of purely entropic nature of LZ transition state (*78*).

A similar controversy existed in discussions of importance of electrostatic contributions for the coiled coil dimer stability (*24, 30, 39, 41-43*), and has been recently resolved by Hans Bosshard and Daniel Marti, showing that the net thermodynamic contribution of a salt bridge is balanced between favorable charge-charge interaction and unfavorable desolvation energy (*44, 45*). Same logic can be applied here to explain the electrostatic contribution to the main folding barrier. Namely, that the favorable "electrostatic guidance" effect from complementary charges (*82*), while increasing the probability of productive transition ensemble formation (and thus increasing the rate of folding), is often compensated by the slow desolvation of the involved charges (which decreases the rate of folding) (*83*). In fact the impact of charge desolvation on the kinetics of LZ traversing the main folding barrier has been already discussed in a number of studies (*78, 100, 101*). In this perspective we suggest to extend the Diffusion-Collision model of LZ folding to the Diffusion-Collision-Desolvation model, where LZ folding rate equation (2b) will be explicitly complemented by appropriate terms reflecting electrostatic (Coulombic) and desolvation contributions.

Besides resolving the aforementioned controversies on entropic vs enthalpic nature of the transition barrier, these terms will rule out the argument used to question the validity of DC model in favor of the two-state mechanism. Tobin Sosnick and coworkers (*68*) argued that if the DC model was valid, 10-fold helicity drop in the A-peptide (*67*) compared to GCN4p1 shall lead to a 100-fold drop in its folding rate, while in reality the peptide folded 20 times faster than the original sequence; and, similarly, engineered GCN4-E9G4 peptide according to DC was expected to fold 100- to 10000- fold slower, while in experiment it was 4-fold faster than original GCN4p1 (*70*). The main accent in interpretation of these results was made on the differences in helical propensities of the peptides, thus undermining the DC model. Meanwhile, along with the changes in helical propensities a serious perturbation in

the Coulombic interactions was obviously introduced, enforcing unexpectedly high folding rates of uncharged GCN4p1 analogs compared to the wild-type sequence.

Finally, the DCD model resolves the argument of the kinetic (*96, 102*) versus thermodynamic (*81*) control of LZ specificity. As discussed above and in the previous section, the net contributions of charge-charge interactions to the coiled coil stability are relatively moderate (*24, 42, 44, 45*) and are generally overridden by free energies contributed through the residues at the hydrophobic core (Table 3.1). Under these conditions the control of LZ oligomerization specificity by interchain ***g-e'*** (***i, i+5***) ionic interactions (*10, 23, 48*) cannot be purely thermodynamic, since moderate thermodynamic stability conferred through ionic interactions would not be able to discriminate specific oligomerization partners in the background of highly stable coiled coil core. Hence, the only time point where electrostatic interactions can effectively modulate oligomerization specificity are the early stages of LZ folding process, when VdW interactions and hydrophobic burial have not yet accreted the coiled coil structure.

Relating to the DCD model definition of LZ specificity, two general types of side-chains in the ***g*** and ***e*** positions can be considered: nonionic and ionic. Nonionic side-chains decrease specificity of LZ, allowing the peptide to indiscriminately interact both with charged and non-charged residues without significant perturbations of the activation barrier. Meanwhile ionic sidechains in ***g***/***e*** positions foster increased specificity of oligomerization, disfavoring partners with identically charged counterions. Presence of repulsive interactions thus increases the activation barrier for folding, limiting the population of particular LZ oligomer, even though from thermodynamical standpoint this oligomer would be more favorable than two separate monomer chains. Kinetic contribution to LZ specificity control is strongly corroborated by early observations that the LZ strand exchange is predominantly governed by the dissociation rate of the coiled coil dimer (*96, 102*), especially in the presence of its consensus DNA seqeuence (*103, 104*).

Arguments against DCD

One argument against the validitiy of the DCD model relies in the fact of additivity of the activation free energy $\Delta\Delta G^{o}_{U>\ddagger}$ (unfolded to transition state) perturbations obtained by mutating cationic and anionic sidechains involved in interhelical salt bridges of GCN4p1 (*81*). This fact suggests that these interactions are not formed in the rate-limiting transition state. However, as shown by Hans Bosshard (*45*), it is plausible that favorable effects from charge-charge interactions are compensated by unfavorable desolvation energies of the charged sidechains, leading to insignificant perturbation of the overall activation energy contribution.

Another argument can be found in the early study by Bosshard group, which shows a striking 75-fold difference between folding rates of almost identical leucine zippers Flu-LZ(12A) and Flu-LZ(16A), which differ only in positioning of a Leu>Ala mutation – being either at ***d***-position (mutant 12A) or at ***a***-position (mutant 16A). The difference cannot be easily accounted for neither by perturbations in the first (helix-priming) transition barrier, nor by electrostatically stabilized transition state of the second barrier. This indicates that the second folding barrier could be formulated both by short-range entropic and long-range enthalpic contributions. However this example has to be treated with caution due to the apparent complexity of the folding model employed – peptide under study had a tendency to trimerize at slightly higher concentrations, and exhibited a three-state behaviour within the measurement timescale (*96*).

= 4.4 = Summary

Summing up, the events governing the LZ folding can be described as following (Figure 4.2). The first, minor transition barrier, is associated with the helix nucleation event and is traversed on the nanosecond timescale (*84-87*). Here the helix-coil transition is favored by helical propensities (*68, 71, 76, 105*) and opposed by the losses in conformational entropy (*105*). Further helix propagation is energetically favorable (occurs slightly down-hill), since each additional hydrogen bond entropically restrains only one residue, while during nucleation event simultaneous fixation of three residues is required (*84*). Depending on particular sequence, preformed helical intermediates can be additionally stabilized by intramolecular (***i***, ***i+3***) electrostatic interactions (*16, 79, 81*), as exemplified by E22(***g***)-R25(***c***) "triggering site" observed in GCN4p1. Overall this initial step leads to formation of partially structured helical intermediates, which create an association-competent scaffold for discrimination of specific partners.

Subsequent period is limited by diffusion processes and occurs on the microsecond timescale (Table 4.1) (*67, 76, 88*). It is followed by the second, major energetic barrier, which occurs upon the collision of monomer chains and reflects the efficiency of the pre-formed structural microdomains in establishing a productive transition state. In the forward (folding) direction, this barrier is traversed on the millisecond timescale (Table 4.1) (*16, 68, 73, 75, 76, 81*) and is predominantly modulated by establishment of specific (if any) interhelical ionic interactions. Enthalpic nature of this barrier provides a certain flexibility in definition of its activation energy, stemming from interplay between the attractive/repulsive "electrostatic guidance" (*82*) effects and energetic penalty associated with charge desolvation (*45, 83*). This flexibility provides one of the major tools for control over specificity of LZ oligomerization. Establishment of productive transition state is followed by a rapid down-hill "zipping up" of

the structure, associated with stabilization of coiled coil structure by short-range VdW interactions (*72, 91*), burial of the hydrophobic core (*67, 68, 72*) and formation of remaining ionic interactions (*75, 81, 88*).

The traverse of the main energetic barrier in the unfolding direction occurs on the scale of seconds and even munutes (Table 4.1). This transition is characterized by highly cooperative dissociation and unfolding of monomer chains, driven by favorable increase in conformational entropy (*75, 78, 88, 90*). The activation energy of this transition is dependent both on the properties of the hydrophobic core (*67, 68, 72, 91*) and interchain electrostatic interactions (*44, 75, 81, 88*).

Reflected in this summary, our knowledge regarding folding of coiled coils and leucine zippers in particular predominantly relies on the studies of LZ-GCN4 (GCN4-p1) and its analogs. Although providing a common frame of reference for different research groups, in some aspects this highly tailored approach appeared counter-productive. This was highlighted by the delayed recognition of enthalpic contributions to the main folding barrier, which became apparent only after LZ folding studies were extended to Jun and Fos LZ motifs. At the moment, stability and specificity rules described in the previous section, together with the knowledge of LZ folding landscape allow extrapolation of experimental findings to other LZ motifs and short coiled coil sequences. For example, compared to LZ-GCN4, coiled coil dimer of human LZ-cJun is considerably less stable, with over five orders magnitude difference in their equilibrium unfolding (dissociation) constants (K_d 446μM vs ~2-8 nM) (*78*). This instability can be attributed to the diminished amount of favorable ***a-a'*** interactions (50% vs 85% in LZ-GCN4) and absence of favorable electrostatic ***g-e'*** couplings (0 vs 3 interhelical ***g-e'*** salt bridges in LZ-GCN4) (*78*). In addition LZ-cJun shows a considerable decrease in the folding rate. Latter effect is likely dictated by the reduced amount of stabilizing intra-helical salt bridges (1 vs 3), absence of attractive inter-helical g-e' ionic interactions (0 vs 3) and presence of one repulsive inter-helical ***g-e'*** interaction (Lys7 – Lys12). Within the proposed Diffusion-Collision-Desolvation model these differences manifest themselves respectively by decreasing the stability of helical pre-collision intermediates, eliminating favorable electrostatic guidance of monomer chains and decreasing the amount of successful collision events.

= 4.5 = Conclusion

For a long time considered a simple two-state transition, folding of leucine zippers appears to be quite a multiplex process, very sensitive to alterations in the LZ primary

sequence and experimental conditions. Accordingly, the exact kinetics of LZ folding landscape will be highly dependant on particular LZ sequence and the folding environment.

Importantly, as demonstrated by the diversity of amino acids in the ***g***, ***e*** and ***a*** positions of the leucine zipper interfaces (*7, 10*), Leucine Zipper motifs have emerged under specificity-driven, rather then stability-driven evolutionary pressures. Combinatorial multiplicity of LZ interfaces together with the flexible kinetic control of LZ folding landscape, provide the fundamental basis for the extraordinary versatility and robustness of this motif in establishment of signalling pathways at the protein level.

= 5 = Functional diversity

In the previous sections we have reviewed the knowledge on LZ structure, specificity and folding properties, pointing to the molecular mechanisms of signal transduction by these motifs. Leucine Zipper signalling is based on shared sets of specificity determinants embedded into a common structural scaffold, and this combinatorial nature enables formation of highly sophisticated signalling networks within a cell. As will be shown in this section LZ networks cover a diverse set of regulatory pathways and the universal LZ "interaction code" indeed appears to provide a common framework for interconnecting various pathways. Currently no appropriate analytical tools are available for quantitative analysis and prediction of LZ motif interactions. Nonetheless, even qualitative "LZ code" analysis using empirical rules discussed above, already permits predictions to be made from the primary sequences of LZ motifs. Beyond widely known bZIP, bHLH-LZ and HD-ZIP protein networks, several other examples of LZ-mediated signalling pathways are presented, along with evidences of couplings between different pathways enabled via LZ motifs.

= 5.1 = Transcription factors – bZIP, bHLH-LZ, HD-ZIP

Origins

The most widely known families of LZ-containing proteins are basic region leucine zipper (bZIP, Figure 5-1A) (*12*), basic region helix-loop-helix leucine zipper (bHLH-LZ, Figure 5-1B) (*106*) and homeodomain (aka helix-turn-helix) leucine zipper factors (HD-ZIP, tentative structure shown in Figure 5-1C, since no native HD-ZIP structure available to date) (*107*). Underscoring universality of LZ interaction, several distinct LZ transcription factor networks have emerged in different lineages at different periods of evolution. For example, bZIP and bHLH-LZ families emerged around 1.6 billion years ago and are common for plant,

fungi and animal kingdoms, meanwhile HD-ZIPs emerged independently in the plant lineage around 0.7 BYA (*26*).

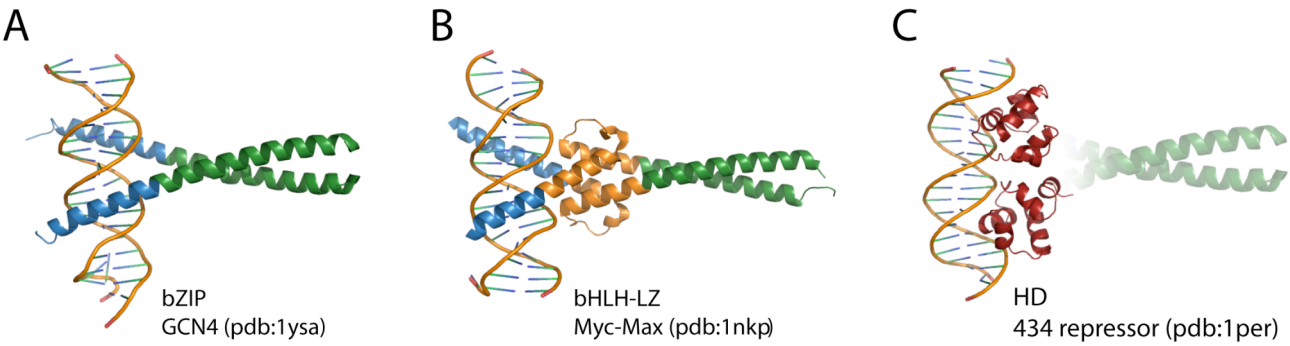


Figure 5.1. Structure of leucine zipper transcription factors. (**A**) bZIP GCN4 homodimer (pdb:1ysa). (**B**) bHLH-LZ Myc-Max heterodimer (pdb:1nkp). (**C**) Hypothetical structure of HD-ZIP factor. Homeodomain from 434 repressor protein (pdb:1per) shown along with tentative positioning of LZ motif. Leucine zipper motifs marked green, basic region – blue, helix-loop-helix (**B**) – orange, homeodomain (**C**) – dark red.

Structure

In all three LZ-TF families the leucine zipper is positioned C-terminal to the DNA-binding motif. In the case of bZIP and bHLH-LZ specific DNA-binding determinant is represented by the basic region of the molecule, and in HD-ZIP this function is mainly conferred by the third helix of the helix-turn-helix motif. While bZIP and HD-ZIP possess "two-domain" (basic region / homeodomain + leucine zipper) architecture, bHLH-LZ (and bHLH in general) possess "three-domain", sometimes even "four-domain" architecture (basic region + helix-loop-helix + leucine zipper) (*26*). This permits a more versatile network of interactions to be formed in the latter case, which is reflected in the size of bHLH family (*26*). Another adaptation for increased number of regulatory pathways is found in LZ motifs of plant bZIPs – on average those are ≥ 8 heptads long, compared to 4-6 heptads of human bZIPs, which considerably extends their combinatorial specificity (*10*).

Functions

Leucine zipper transcription factors evolved as key regulators in a wide variety of processes. Today they are truly widespread among eukarya with only human genome encoding 51 proteins with unique bZIP motifs (*108*) and at least 31 proteins with unique bHLH-ZIP motifs (*109*). In Arabidopsis similar bZIP / bHLH-LZ array is complemented with 47 unique HD-ZIP proteins (*26*).

bZIP is the second-largest family of dimerizing transcription factors in humans (*26*). They control expression of genes involved in development, environmental stresses,

metabolism, circadian clock and neuronal activity (*27*), with a number of factors being widely renowned oncogenes, such as AP-1 (*110*). Similarly, in plants bZIPs mediate diverse developmental and homeostatic processes, as well acting in various environmental stress signalling pathways (*111*).

bHLH-LZ represent a subset of bHLH proteins – the largest family of dimerizing transcription factors across eukarya (*26*). The key targets of bHLH-LZ regulation are developmental processes, differentiation and, most importantly, cell cycle (*112, 113*). Similarly to bZIP family, dysfunctions of bHLH-LZs are strongly associated with tumorigenesis, with the most prominent example being highly oncogenic transcription factor c-Myc (*114*), which regulates up to 15% of all genes in an organism (*115*).

HD-ZIP factors are unique to the plant kingdom, and, as opposed to the archetypical homeodomain proteins, do not play homeotic roles, exhibiting highly specialized functionality (*116*). These factors are split into four classes based on structural characteristics, and govern a spectrum of processes including responses to environmental conditions, hormone action, meristem regulation and organ development (*116*).

"LZ silencing" and pharmacological implications

As proposed by Grigoris Amoutzias and others, evolution of organisms starting from eukarya to a large extent depended on the expansion of TF networks, those including bHLH, bZIP, NR and MADS (*26, 109*). Emerging complexity of these networks in turn stimulated formation of additional regulatory machinery, in case of LZ proteins involving cofactor interactions (*117, 118*), compartmentalization (*119*), proteolysis (*120, 121*), phosphorylation (*122, 123*), acetylation (*124*), glycosylation (*125*) and etc (reviewed for plant bZIPs in (*126*)).

One of the most important regulatory mechanisms within the networks of dimerizing TF is suppression of signal transduction in presence of complementary dimerization partners lacking DNA-binding motifs. This modulation, targeted towards the key hubs of the TF network, provides a mechanism for robust silencing of an extensive network segments. One of the most renowned examples in this perspective are Id proteins ("inhibitor of DNA binding") of bHLH family, which employ dominant negative helix-loop-helix interactions to regulate cell cycle, tumorigenesis and cellular senescence (*127, 128*). As for Leucine Zipper transcription factors, a similar example has been recently reported in regulation of HD-ZIPs – a ZPR family represents short LZ-containing peptides which specifically heterodimerize with HD-ZIPIII transcription factors impairing their DNA-binding abilities (*129*). In the case of short LZ motifs, this mechanism becomes somewhat reminiscent to that of RNAi, supporting conceptual similarity between the linear "nucleotide code" and "LZ code" formalism. Along these lines, a number of drug-development studies are under way focusing on targeted

silencing of LZ-mediated signal transduction by small molecules (*130-132*) as well as complementary dominant-negative LZ motifs (*133-135*).

= 5.2 = Immune response signalling – NF-kappaB pathway

Leucine zipper interactions play an important role in the cytoplasmic part of antiapoptotic NFκB signalling pathway, which is conserved from flies to mammals and provides one of the key routes for the immune and stress responses (reviewed in (*136*)). Many of the extracellular signals that lead to the activation of NFκB converge on a ~900 kDa IκB kinase complex (IKK) consisting of three major subunits – catalytic IKKα and IKKβ (52% identical in primary structures), and regulatory IKKγ (also known as NEMO) (*137-139*) or reviewed in (*140*) and (*141*).

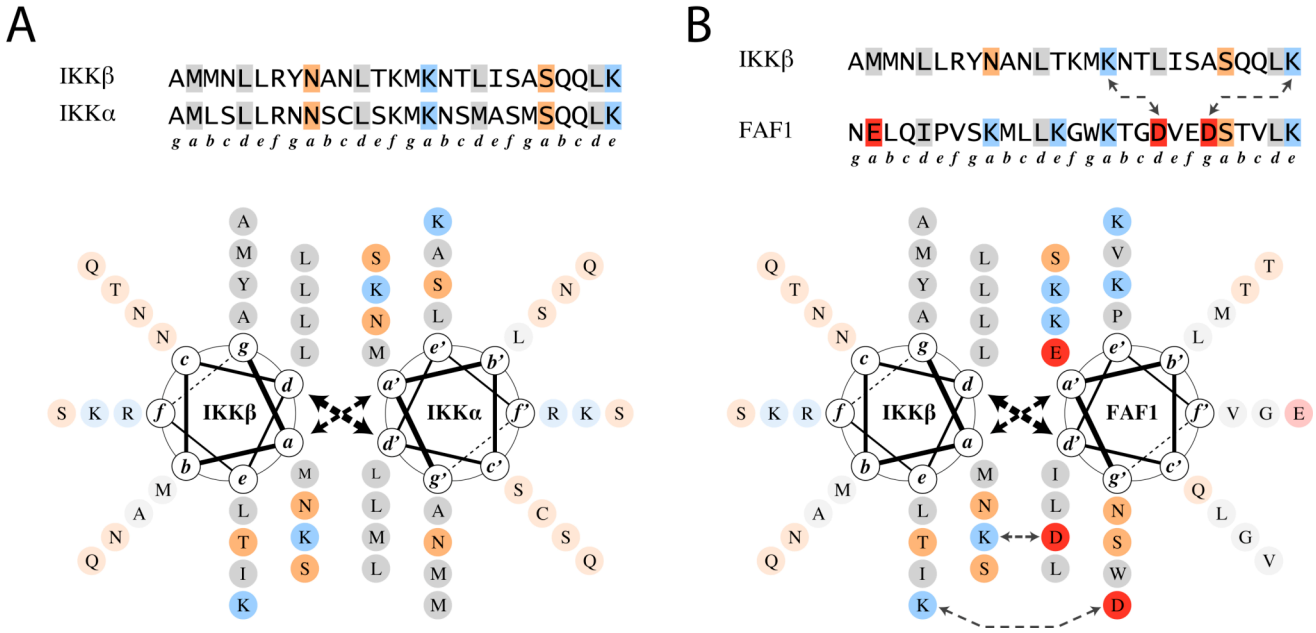


Figure 5.2. Predicted interactions of LZ motifs from factors IKKα/IKKβ (**A**) and IKKβ/FAF1 (**B**).

LZ motifs of both catalytic IKKα/IKKβ subunits exhibit identical polar residues in all four ***a***-positions of the dimer interface (Met, Asn, Lys, Ser - Figure 5.2-A), which provides a highly specific oligomerization determinant. Formation of IKKα/IKKβ heterodimer is indispensable for the activation of the whole complex and its kinase activity (*138, 142*).

The regulatory subunit IKKγ interacts with the catalytic IKKα/IKKβ heterodimer via its N-terminal coiled coil motif, while its own C-terminal LZ motif of IKKγ plays a way more intriguing role. As has been shown by Marshall Horwitz group – LZ of IKKγ enables direct interaction of the IKK complex with the components of the AP-1 complex (c-Jun and c-Fos) (*122*). Thermodynamic analysis of heterodimeric LZ interactions in the IKKγ•Jun versus IKKγ•Fos complexes employing the simplified set of LZ determinants (***a-a'***, ***d-d'*** and ***g-e'***)

shows that both complexes are comparably stable and specific. However analysis of the corresponding wheel diagrams shows that kinetically IKKγ•Jun complex is less favorable, forming one attractive and one repulsive interchain ionic interaction, opposed to 3 attractive ionic interactions in IKKγ•Fos complex (Figure 5.3). Therefore, a more precise contextual analysis of sidechain interactions within these complexes is required to make accurate predictions of AP-1 (Jun•Fos) equilibrium redistribution in the presence of IKKγ. In any case, the discovery of interactions between IKKγ and AP-1 sheds some light on the long-debated issue of molecular mechanisms underlying the balance between the proapoptotic AP-1 and antiapoptotic NF-kappaB pathways and thereby cell death versus cell survival decisions. It also provides an illustrative example of LZ-mediated coupling between cytoplasmic and nuclear protein signalling pathways.

FAF. In response to proinflammatory stimuli Fas-associated factor 1 (FAF1) interacts with the LZ motif of IKKβ, affecting its association with IKKα and IKKγ, and thus effectively suppressing NF-κB activation (*143*). The exact structural determinants responsible for this interaction from the FAF1 side are unknown, except of the knowledge that determinant is located within the N-terminal half (residues 1-201) of the polypeptide chain (*143*). Based on the characteristic set of polar and charged residues in the ***a***-positions of IKKα/IKKβ interface, a PROSITE search employing [NSKR]-x-x-x-x-x-K-x-x-x-x-x-x-S pattern, yields a unique match – FAF1 residues 139-153 (Figure 5.2-B). In terms of ***a-a'***, ***d-d'*** and ***g-e'*** interactions, this consensus appears highly similar to that of original IKKα/β leucine zipper motif. Furthermore, it bears additional intermolecular ionic bridges, making IKKα/β•FAF1 heterodimers more favorable, thus explaining suppression of the IKKα/IKKβ interaction in presence of FAF1. Being proven experimentally, this would corroborate the "LZ code" formalism and contribute a strong evidence for LZ-mediated coupling between protein signalling pathways in the cytoplasm. Notably, being optimized towards interaction specificity, FAF1 is not 100% canonical LZ, with two of four ***d***-position leucines replaced by Asp and Ile sidechains, making it not recognizable by the existing coiled coil prediction algorithms 2ZIP (*144*) and Multicoil (*145*). This again signifies the necessity of developing new tools and prediction algorithms to distinguish transient leucine zipper interfaces.

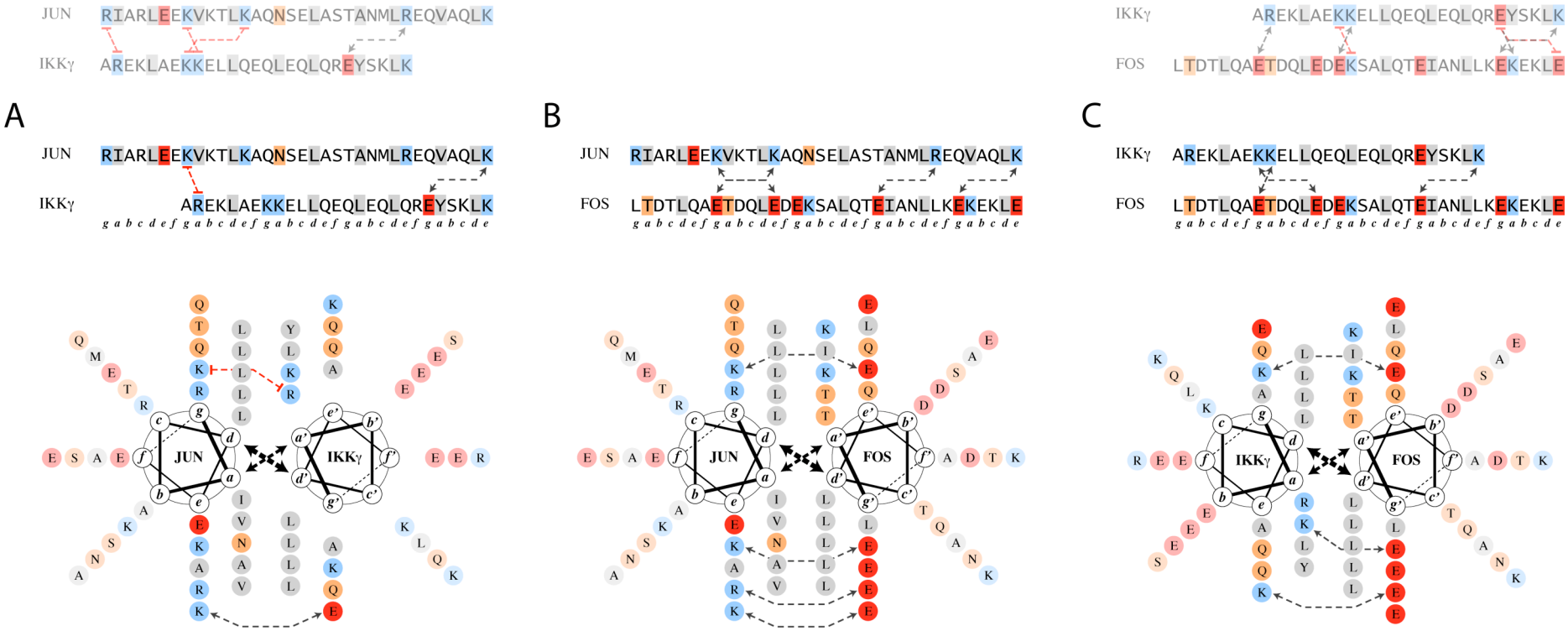


Figure 5.3. Predicted interactions between LZ motifs from factors Jun•IKKγ (A), Jun•Fos (B) and IKKγ•Fos (C). Linear representations show two alternative alignments of the 5-heptad Jun and Fos zippers with the 4-heptad IKKγ LZ motif. Thermodynamically unfavorable alignments are dimmed. Wheel representations illustrate most energetically favorable alignments of the LZ motifs.

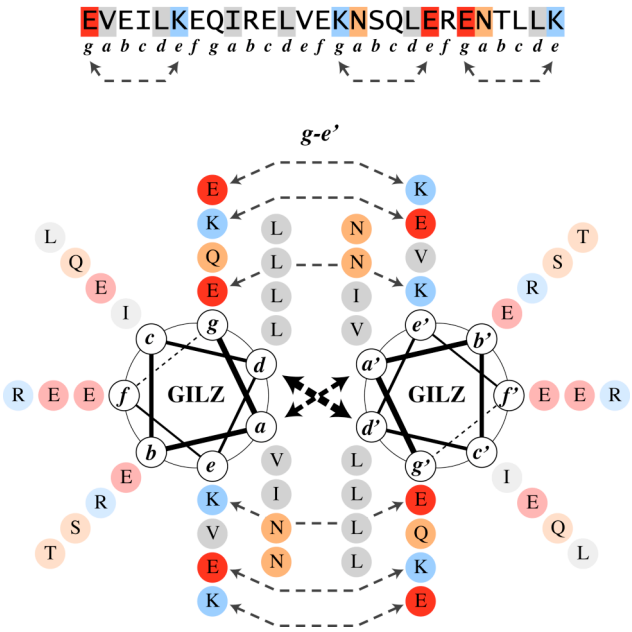


Figure 5.4. Predicted interactions of LZ motifs from factor GILZ.

GILZ. Another LZ-protein involved in NF-kappaB pathway is the Glucocorticoid-induced leucine zipper (GILZ). It is a 137 amino acid protein which is rapidly induced by treatment with glucocorticoids (*146*). Straightforward analysis of its "LZ code" shows that the protein shall be highly stable as a homodimer – containing 6 favorable intermolecular ***g-e'*** salt bridges, all-Leu in ***d***-positions and predominantly beta-branched resdidues in ***a***-positions (Figure 5.3). Presence of two polar buried asparagines increases the homodimerization specificity even further. According to experimental data LZ-mediated GILZ homodimerization is indeed essential for its function as inhibitor of NF-κB transcriptional activity (*146*).

= 5.3 = More kinases – PKG, ZIPK, DAPK

Similar to the NF-kappaB example, there are many other seemingly distinct kinase signalling pathways, coupled through the common combinatorial nature of the LZ interactions.

Regulation of vascular smooth muscle tone is mediated through the balance of myosin phosphorylation and dephosphorylation rates. In the case of nitric oxide dependant vasodilation, this involves a multiplex LZ-mediated equilibrium between homodimeric form of PKG-Iα (cGMP-dependent protein kinase Iα) and its heterotetrameric assembly with the MYPT1 (myosin-binding subunit of the myosin phosphatase) (*147, 148*).

Similarly, another enzyme – Zipper-interacting protein kinase (ZIPK) modulates the phosphorylation state of myosin light chains in the case of smooth muscle contractile in

response to Ca2+ (*149*). At the same time, ZIPK is known as the death-associated protein kinase 3 (DAPK3), enabling cell death through apoptosis (*150*). ZIPK leucine zipper motif was shown indispensable for its enzymatic activity (*150*), cellular localization and proapoptotic effects (*151*). Moreover, ZIPK LZ motif facilitates its heterodimerization with the transcription factor 4 (ATF4) from the bZIP ATF/CREB family through their LZ domains (*150*), providing direct coupling between nuclear (*27*) and cytoplasmic signalling networks.

= 5.4 = Ion channels – AKAP

Shortly after original discovery of LZ in transcription factors, the presence of this motif was revealed in the family of voltage-gated potassium channels with suggested involvement in subunit interactions, mediating voltage-dependant opening and closing of the channel (*13*). In skeletal and cardiac myocytes interaction of the LZ motifs of L-type Ca2+ channels (CaV1.1 and 1.2) and A kinase-anchoring protein (AKAP15) provides and efficient mechanism for anchoring of cAMP-dependent protein kinase (PKA) to Ca2+ channels, ensuring rapid and efficient phosphorylation of Ca2+ channels in response to local signals such as cAMP and depolarization (*152, 153*). This mechanism is very similar to the hippocampal pyramidal cells, where rapid modulation of neuronal excitability through Na(+) channels occurs by local protein phosphorylation of the channel by the protein kinase A specifically recruited via its LZ motif by the A kinase-anchoring protein AKAP15(*154*).

= 5.5 = Transport vesicles – SNARE

Many biomolecules are transferred among different cellular compartments by transport vesicles, which recognize and merge with the target compartments in a highly specific manner in addition overcoming a very high activation energy barrier during membrane fusion (nicely summarized in (*135*)). In all eukaryotic cells this task is accomplished by a family of proteins called SNARE (soluble N-ethylmaleimide-sensitive factor attachment protein receptor) (*155*). All SNARES bear a conserved cytosolic coiled-coil/leucine zipper motif of 60-70 residues, which upon assembly yields a parallel four-helix bundle – SNAREpin. Structural classification of SNAREs is based on the type of sidechain they contribute to the "zero ionic layer" – a cluster of buried ionic interactions which apparently define specificity of the interface (*156, 157*).

Essentially SNARE is a mixed "leucine zipper" (signal) / "coiled coil" (structure) motif, serving two goals – four-helix-bundle provides a high stability interface required as an energy source for membrane fusion, while "ionic zero layer" creates a unique vesicle specificity determinant, a functional analog of the buried polar residues defining specificity within bZIP signalling networks (see section 3.3). As expected, in SNARE tetramer the specificity

determinant is located in the ***d***-position of interface, adopting the same "perpendicular" geometry in tetramer as ***a***-positioned sidechains do in dimeric leucine zippers (see section 2.2). Thus SNARE represents an intermediate state between the structural coiled coil and signalling LZ motifs – extreme stability of the four-helix-bundle allows it to efficiently transmit the signal only in the "forward" direction, thus limiting thermodynamic control of the signalling event and requiring external factors for tetramer disassembly (*135*).

= 5.6 = Viral envelopes and capsides

Another discovery made shortly after the original publication of William Landschultz – notion of paramyxovirus fusion glycoproteins facilitating dimerization and tetramerization via the LZ motif (*14*). Nowadays a wealth of experimental evidence exists signifying importance of LZ in the formation of viral capsides. HIV-1 envelope glycoprotein gp41 bears two LZ motifs – in its N- and C- terminal regions. Experimental results suggest that the corresponding N-leucine zipper, along with N-terminal fusion and viral membrane-adjacent regions of HIV-1 gp41 may promote key membrane perturbations underlying the merging of the viral envelope with the cell surface (*158*). Mason-Pfizer monkey virus (M-PMV) Gag protein contains a domain p12 that is unique to this virus (simian retrovirus-3) and its close relatives. This domain incorporates a leucine zipper-like region that facilitates Gag-Gag protein oligomerization performing scaffold-like function within the viral envelope (*159*). Excapsidation of the herpes simplex virus type 1 (HSV-1) relies on the formation of 12-subunit ring structure mediated by LZ motifs of UL6 protein (*160*).

Viral LZ motifs are not always directly involved in processing of target-recognition signals, frequently conducting pure mechanistic functions like membrane fusion. Nevertheless, as will be shown below, nucleocapside LZ motifs may serve as highly specific recognition determinants enabling innate antiviral immunity responses in higher eukaryotes.

= 5.7 = Innate antiviral defense – interferon induced Mx proteins

Mx proteins are induced by alpha/beta interferons, forming one of the key components of the innate immune response against RNA viruses (*161*). Featuring highly conserved N-terminal GTPase domain, they are classified together with the dynamin-like large guanosine triphosphatases (GTPases), known to be involved in intracellular vesicle trafficking and organelle homeostasis. Beyond GTPase domain the sequence similarity fades away and at the C-terminus Mx proteins carry an unique LZ doublet (*162*), which replaces the PH-GED-PRD (Pleckstrin Homology + GTPase Effector + Proline-Rich Domain) triad characteristic for dynamins (*161*). This extended LZ motif enforces Mx proteins with antiviral activity against a wide spectrum of viruses, including members of bunyaviridae, orthomyxoviridae,

paramyxoviridae, rhabdoviridae, togaviridae, picornaviridae, and even Hepatitis B virus (*161*).

The importance of the Mx system was effectively demonstrated by Peter Stäheli and Otto Haller groups on the mouse models – disruption of a single Mx1 gene causes complete loss of innate immunity against mouse-adapted influenza, leading to an overwhelming infection and rapid death (*163, 164*). While enhanced Mx1 production efficiently protects transgenic mice against pandemic human 1918 influenza virus and highly lethal H5N1 strain (*164*).

It is proposed that Mx GTPases detect viral infection by sensing nucleocapsid-like structures, trapping viral components and making them unavailable for the generation of new viral particles (*165*). Being antivirally active as monomers, Mx proteins assemble into high-molecular-weight oligomers in the solution, possibly yielding a stable intracellular pool from which individual monomers are recruited in the presence of viral particles (*166, 167*). The detailed mechanism of Mx oligomerization is poorly understood, but is likely mediated by intermolecular domain swapping involving C-terminal LZ and coiled coil motif within central CID domain (*166, 168*). Dependence of Mx oligomerization on intramolecular backfolding of C-terminal LZ onto CID coiled coil motif also remains controversial (*168*), three-domain oligomerization (*166*).

The C-terminal leucine zipper motifs appear to be the key component required for this antiviral activity (*166, 169*). As shown by mutagenesis studies, Mx leucine zipper represents a multipurpose recognition motif, shaped for identification of a diverse array of viral species. For example a point E645R mutation (at the ***f***-position of the coiled coil) reshapes the interaction surface of human MxA, impairing recognition of the La Crosse (*170*) and Crimean-Congo hemorrhagic fever (*171*) viruses from the Bunyaviridae family. Meanwhile MxA-E645R sensitivity to the Influenza A and Thogoto viruses from Orthomyxoviridae family remains unaffected (*169, 170*). Notwithstanding all advances, the molecular mechanism of Mx antiviral activity is still poorly understood. Particularly, the exact structural determinants responsible for the recognition of virus components have yet to be reported. Nevertheless, importance of C-terminal LZ motif integrity for Mx antiviral activity, together with high abundance of coiled coil structures within viral capsids and nucleoproteins, suggests specific LZ recognition being the key interaction behind innate antiviral immunity response mediated by Mx proteins.

= 6 = Beyond protein interactions: leucine zippers with catalytic activity

Examination of amino acids in the leucine zipper region of bZIP proteins indicates that many residues beyond oligomerization-facilitating ***a***, ***d***, ***g*** and ***e*** positions are notably conserved (*10, 48*). This conservation suggests that the external surface of LZ proteins (Fig. 2.1D) might carry important functional determinants. Until today very few studies have investigated the role of conserved residues in the ***b***, ***c*** and ***f*** positions for LZ functionality. The most prominent examples include surface interactions of the AP-1 leucine zipper motif with NFAT (nuclear factor of activated T-cells) during cooperative DNA binding (*172*), and involvement of ***f***-position residues in viral recognition by Mx proteins of the innate antiviral system (*170*).

Another example signifying the importance of "LZ surface" residues is related to RNA binding by LZ motifs from transcription factors GCN4 (yeast) and cJun (human). As will be shown in the chapters II and IV of this thesis, beyond oligomerization functionality, leucine zipper motifs GCN4 and cJun are capable of catalyzing degradation of RNA. The RNA binding interfaces of these proteins involve residues located in ***b***, ***c*** and ***f*** positions on the surface of LZ structure. Notably, GCN4 and cJun show similar topological requirements for RNA binding and catalysis: (1) only dimeric form of either LZ can bind RNA; (2) in both cases RNA binding occurs in the two N-terminal heptads of LZ structure; and (3) both LZs show increased catalytic activity towards single stranded RNA. Nevertheless, the exact composition of GCN4 and cJun RNA binding interfaces is not identical. This observation shows that LZ structure can act as a scaffold for assembly of catalytical sites with variable properties. General independence of the LZ oligomer stability of the character of sidechains in the ***b***, ***c*** and ***f*** interface positions may provide considerable flexibility for binding/catalytic site formation. Moreover, as demonstrated by R. Ghadiri and co-workers additional fine-tuning of the catalytic site properties can be achieved through modulation of LZ oligomer stoichiometry and residues in the vicinity of the active site (*19*).

In the aforementioned study, researchers used coiled coil leucine zipper scaffold to engineer a set of peptides that are capable to perform aminoacyl loading and intermodular aminoacyl transfer steps of polypeptide synthesis, thus functionally mimicking nonribosomal peptide synthetases. In the reported designs catalytic site was assembled from two catalytic (cysteine and lysine) and two auxiliary X1 and X2 (histidine) residues, distributed over the two neighboring chains of LZ oligomer (Fig. 6.1A, C). Cysteine residue, located in the ***a***-position of the reference LZ assembly, performed covalent capture of substrate from solution

via trans-thiolesterification in the first step, and acted as aminoacyl donor in the second step (Fig. 6.1E). The epsilon-amine group of the Lysine residue, located in the ***c***-position of the LZ interface, acted as aminoacyl acceptor in the second reaction step. Two auxiliary Histidine residues, located in the ***e*** and ***f*** positions of engineered LZ interface, provided additional electrostatic and/or general acid-base contributions to enhance the efficiency of aminoacyl loading and transfer steps. Authors have shown that the efficiency of the performed reaction can be significantly influenced by modulation of the active site microenvironment through amino acid substitutions, as well as variation of the inter-residue distances and geometry.

First, authors demonstrated that reducing the length of Lysine acceptor and/or increasing the length of Cysteine donor sidechains decreases the aminoacyl transfer rates by up to 8-fold, suggesting that even small changes in the active site inter-residue distances can appreciably alter the efficiency of aminoacyl transfer.

Next, authors established that increasing the LZ oligomeric state from di- to tri- and tetra-meric species decreases the efficiency of the substrate loading step by 3-fold, while providing a 7-fold increase in the efficiency of the second aminoacyl transfer step. Former effect was rationalized as increased burial and reduced accessibility of the substrate-loading Cys sidechain in the context of higher-order oligomers. The latter effect was explained due to the enhanced proximity of acyl-donor and -acceptor groups in the case of tri- and tetrameric species.

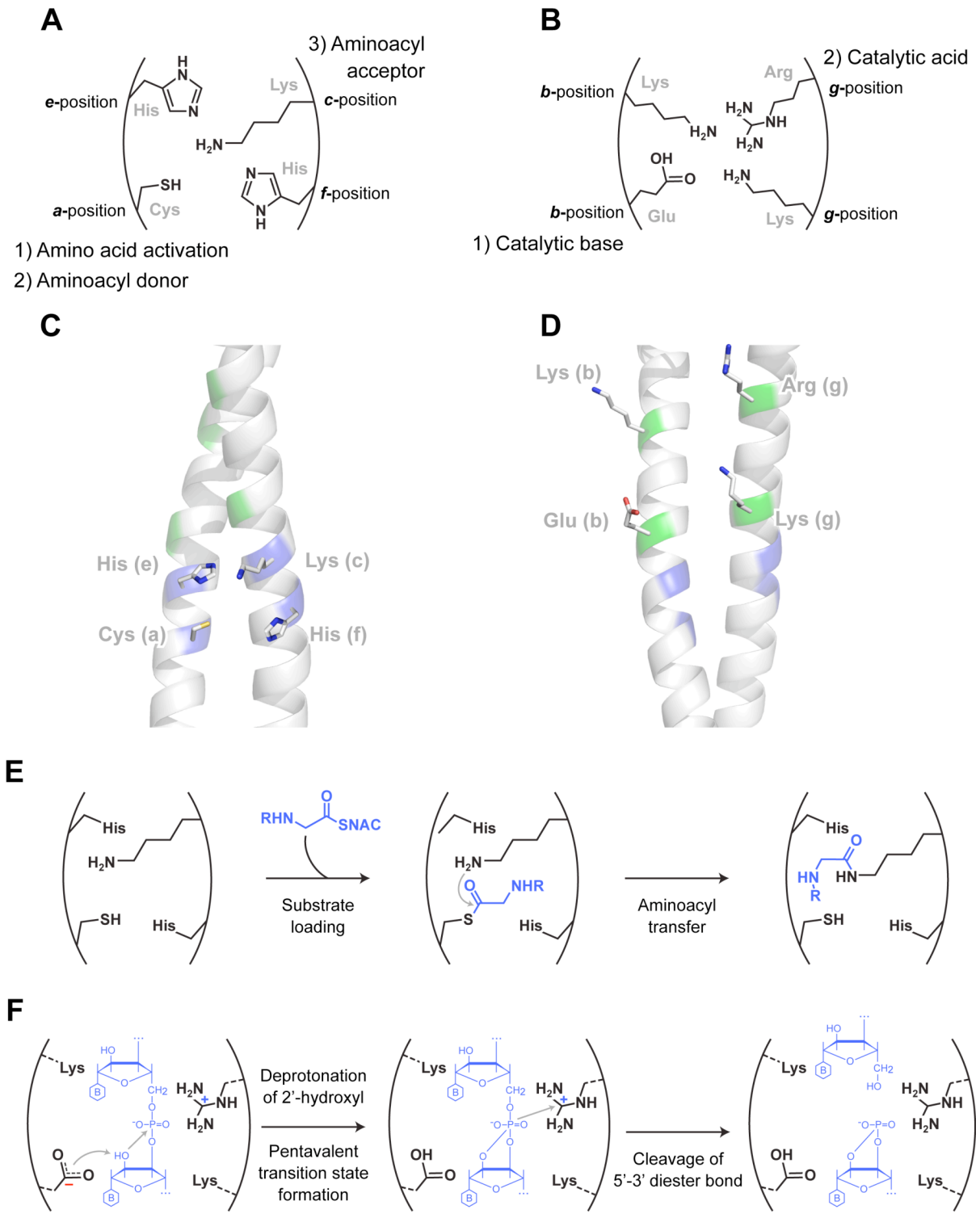


Figure 6.1. Catalytic sites and reaction mechanisms of catalytic LZ. Schematic representation (**A**) and 3D model (**C**) visualizing positioning of functional residues in the active site of aminoacyl-transferring LZ peptides designed by Leman and coworkers (adapted from (*19*) employing structure of dimeric GCN4, PDB:1ysa). Schematic (**B**) and 3D (**D**) visualization of the LZ-GCN4 RNA binding site (PDB:1ysa). (**E**) Schematic representation of substrate loading and aminoacyl transfer reaction in the LZ motifs designed by Leman and coworkers (adapted from (*19*)). SNAC – N-acetylcysteamine. (**F**) Schematic representation of possible transesterification mechanism facilitated by LZ-GCN4. Detailed discussion of the mechanism is given in Chapter IV of this thesis.

Third, the authors showed that remarkable modulation of the catalytic efficiency can be achieved by substitution of the residues in the auxiliary X1 and X2 positions. These alterations provided a ~20-fold variation of the substrate loading rates, and more than 200-fold variation in the rate of the aminoacyl transfer step. The Histidine residues in X1 and X2 positions showed cooperative behavior and simultaneous replacement of those to Alanines resulted in a 20-fold decrease in the aminoacyl transfer efficiency.

Finally, the authors showed that the residues within the secondary sphere of the catalytic site can also provide substantial contribution to the efficiency of catalysis. For example, cation-to-polar Arg14Ser mutation, neighboring the catalytic Lys8 residue, provided a 4- to 12-fold enhancement in the rate of the aminoacyl transfer step.

In summary, it was demonstrated that rational exploitation of both LZ surface and core residues allows for engineering of an artificial peptidic catalyst with desired characteristics. Importantly, appropriate positioning and rational choice of aminoacid residues in the vicinity of the catalytic site can lead to >900-fold differences in the reactivity of otherwise identical functional groups in the context of the LZ scaffold. The system provides a range of mechanisms to fine-tune the efficiency of particular catalytic reaction, and establishes LZ as a promising platform for the de-novo design of biocatalysts with desired properties. It remains to be seen whether LZ scaffold will be suitable for engineering of artificial biocatalysts to perform reactions beyond the aminoacyl transfer. Nevertheless, the research reported in chapters II and IV of this thesis, indicates that LZ scaffold may have been employed by nature for establishment of RNA degradation functionality within bZIP transcription factors. The RNA binding site and putative catalytic mechanism of LZ ribonuclease activity is shown in Figure 6.1 and discussed in detail in Chapter IV.

From biological perspective, catalytic degradation of RNA is essential to the efficient RNA turnover and decay, and is a major mechanism involved in regulation of gene expression, quality control of RNA biogenesis and antiviral defenses (*173-176*). In this context, it is plausible that ribonuclease activity associated with LZ motifs may contribute to the regulation of the RNA turnover due to the antiquity and prevalence of these motifs among transcription factors. Chapters II and IV of this thesis focus on characterization of ribonuclease activity of the Leucine Zippers GCN4 (*S. cerevisiae*) and cJun (*H. Sapiens*), as well as LZ conformational dynamics associated with this phenomenon.

= 7 = Outlook

The distinction between the structural Coiled Coil motifs and signal-transducing Leucine Zipper motifs, like in the case of SNAREs, might often remain elusive. Nevertheless, highly specific interactions enabled by short coiled coils with the hydrophobic core of predominantly leucine sidechains, makes a notable distinction for the functionality of this protein motif. We propose that on the cellular level uniformity of LZ interfaces enables one of the key components of the "protein interactome", which facilitates coupling of many functionally distinct signalling pathways into one signalling network.

Further developments in understanding of the "LZ interaction code" require more thorough sampling of natural proteins for LZ motifs with new types of specificity determinants; as well as systemic engineering approaches for acquisition of precise physico-chemical properties enabling stability and specificity variation within LZ scaffold. As exemplified by the ***a–d*** ionic bridges in the Myc/Max network (*37*) and conserved ***f***-position prolines found in some plant bZIP families (*65*), specificity rules derived from the studies of human bZIPs provide only a small part of the picture.

Another important direction of research is associated with development of advanced molecular models and prediction tools for analysis of transient LZ interfaces. Currently available contextual data is insufficient for accurate prediction of stability and specificity of particular LZ interactions even at the level of most commonly reported interaction determinants (***a-a'***, ***d-d'***, ***g-e'***). Accurate prediction algorithms together with intuitive visualization models shall facilitate a deeper understanding of functional implications of LZ signalling networks. Developments in this direction require a combination of systemic structural, biophysical and computational approaches, as well as an extended "alphabet" of available LZ specificity determinants mentioned in the previous paragraph.

Finally, the most intriguing property of LZ motifs is their role as a structural scaffold for establishment of catalytical sites with various properties. General independence of LZ oligomer stability on the residues in ***b***, ***c*** and ***f*** positions provides remarkable versatility for interaction surface and catalytic site definition. Apparent non-uniformity in LZ-GCN4 and LZ-cJun endonucleolytic site composition corroborates this flexibility. Enzymatic engineering approach is the most obvious route to extend/evaluate this hypothesis beyond transesterification of RNA phosphodiester bonds and aminoacyl transfer.

= 8 = References

1. Walshaw, J., and Woolfson, D. N. (2001) Socket: a program for identifying and analysing coiled-coil motifs within protein structures, *J Mol Biol 307*, 1427-1450.
2. Lupas, A. N., and Gruber, M. (2005) The structure of alpha-helical coiled coils, *Advances in protein chemistry 70*, 37-78.
3. Lumb, K. J., Carr, C. M., and Kim, P. S. (1994) Subdomain folding of the coiled coil leucine zipper from the bZIP transcriptional activator GCN4, *Biochemistry 33*, 7361-7367.
4. Perera, R., Owen, K. E., Tellinghuisen, T. L., Gorbalenya, A. E., and Kuhn, R. J. (2001) Alphavirus nucleocapsid protein contains a putative coiled coil alpha-helix important for core assembly, *J Virol 75*, 1-10.
5. Burkhard, P., Meier, M., and Lustig, A. (2000) Design of a minimal protein oligomerization domain by a structural approach, *Protein Sci 9*, 2294-2301.
6. Kammerer, R. A., Schulthess, T., Landwehr, R., Lustig, A., Engel, J., Aebi, U., and Steinmetz, M. O. (1998) An autonomous folding unit mediates the assembly of two-stranded coiled coils, *Proc Natl Acad Sci U S A 95*, 13419-13424.
7. Acharya, A., Rishi, V., and Vinson, C. (2006) 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 45*, 11324-11332.
8. Burkhard, P., Stetefeld, J., and Strelkov, S. V. (2001) Coiled coils: a highly versatile protein folding motif, *Trends in cell biology 11*, 82-88.
9. Woolfson, D. N. (2005) The design of coiled-coil structures and assemblies, *Advances in protein chemistry 70*, 79-112.
10. Vinson, C., Acharya, A., and Taparowsky, E. J. (2006) Deciphering B-ZIP transcription factor interactions in vitro and in vivo, *Biochimica Et Biophysica Acta-Gene Structure and Expression 1759*, 4-12.
11. Grigoryan, G., and Keating, A. E. (2008) Structural specificity in coiled-coil interactions, *Curr Opin Struct Biol.*
12. Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins, *Science 240*, 1759-1764.
13. McCormack, K., CAMPANELLI, J., RAMASWAMI, M., MATHEW, M., TANOUYE, M., IVERSON, L., and RUDY, B. (1989) Leucine-zipper motif update, *Nature 340*, 103-104.
14. Buckland, R., and Wild, F. (1989) Leucine zipper motif extends, *Nature 338*, 547.
15. Wang, T., Lau, W. L., DeGrado, W. F., and Gai, F. (2005) T-jump infrared study of the folding mechanism of coiled-coil GCN4-p1, *Biophys J 89*, 4180-4187.
16. Steinmetz, M. O., Jelesarov, I., Matousek, W. M., Honnappa, S., Jahnke, W., Missimer, J. H., Frank, S., Alexandrescu, A. T., and Kammerer, R. A. (2007) Molecular basis of coiled-coil formation, *Proc Natl Acad Sci U S A.*
17. Rieker, J. D., and Hu, J. C. (2000) Molecular applications of fusions to leucine zippers, *Meth. Enzymol 328*, 282-296.
18. Liu, J., Zheng, Q., Deng, Y., Cheng, C. S., Kallenbach, N. R., and Lu, M. (2006) A seven-helix coiled coil, *Proc Natl Acad Sci U S A 104*, 15457–15462.
19. Leman, L. J., Weinberger, D. A., Huang, Z. Z., Wilcoxen, K. M., and Ghadiri, M. R. (2007) Functional and mechanistic analyses of biomimetic aminoacyl transfer reactions in de novo designed coiled coil peptides via rational active site engineering, *J Am Chem Soc 129*, 2959-2966.

20. McLachlan, A. D., and Stewart, M. (1975) Tropomyosin coiled-coil interactions: evidence for an unstaggered structure, *J Mol Biol 98*, 293-304.
21. Crick, F. H. C. (1953) The Packing of Alpha-Helices - Simple Coiled-Coils, *Acta Crystallogr 6*, 689-697.
22. O'Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991) X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil, *Science 254*, 539-544.
23. Vinson, C. R., Hai, T., and Boyd, S. M. (1993) Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design, *Genes Dev 7*, 1047-1058.
24. Krylov, D., Mikhailenko, I., and Vinson, C. (1994) A thermodynamic scale for leucine zipper stability and dimerization specificity: e and g interhelical interactions, *Embo J 13*, 2849-2861.
25. Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993) A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants, *Science 262*, 1401-1407.
26. Amoutzias, G. D., Robertson, D. L., Van de Peer, Y., and Oliver, S. G. (2008) Choose your partners: dimerization in eukaryotic transcription factors, *Trends Biochem Sci 33*, 220-229.
27. Deppmann, C. D., Alvania, R. S., and Taparowsky, E. J. (2006) Cross-species annotation of basic leucine zipper factor interactions: Insight into the evolution of closed interaction networks, *Mol Biol Evol 23*, 1480-1492.
28. Serrano, L., Horovitz, A., Avron, B., Bycroft, M., and Fersht, A. R. (1990) Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles, *Biochemistry 29*, 9343-9352.
29. Moitra, J., Szilak, L., Krylov, D., and Vinson, C. (1997) Leucine is the most stabilizing aliphatic amino acid in the d position of a dimeric leucine zipper coiled coil, *Biochemistry 36*, 12567-12573.
30. Krylov, D., Barchi, J., and Vinson, C. (1998) Inter-helical interactions in the leucine zipper coiled coil dimer: pH and salt dependence of coupling energy between charged amino acids, *J Mol Biol 279*, 959-972.
31. Acharya, A., Ruvinov, S. B., Gal, J., Moll, J. R., and Vinson, C. (2002) 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 41*, 14122-14131.
32. Kohn, W. D., Kay, C. M., and Hodges, R. S. (1998) Orientation, positional, additivity, and oligomerization-state effects of interhelical ion pairs in alpha-helical coiled-coils, *J Mol Biol 283*, 993-1012.
33. Wagschal, K., Tripet, B., Lavigne, P., Mant, C., and Hodges, R. S.(1999) The role of position a in determining the stability and oligomerization state of alpha-helical coiled coils: 20 amino acid stability coefficients in the hydrophobic core of proteins, *Protein Sci 8*, 2312-2329.
34. Tripet, B., Wagschal, K., Lavigne, P., Mant, C. T., and Hodges, R. S. (2000) Effects of side-chain characteristics on stability and oligomerization state of a de novo-designed model coiled-coil: 20 amino acid substitutions in position "d", *J Mol Biol 300*, 377-402.
35. Cohen, C., and Parry, D. A. (1990) Alpha-helical coiled coils and bundles: how to design an alpha-helical protein, *Proteins 7*, 1-15.
36. Lavigne, P., Crump, M. P., Gagne, S. M., Hodges, R. S., Kay, C. M., and Sykes, B. D. (1998) Insights into the mechanism of heterodimerization from the 1H-NMR solution structure of the c-Myc-Max heterodimeric leucine zipper, *J Mol Biol 281*, 165-181.
37. Montagne, M., Naud, J. F., and Lavigne, P. (2008) Elucidation of the structural determinants responsible for the specific formation of heterodimeric Mxd1/Max b-HLH-LZ and its binding to E-box sequences, *J Mol Biol 376*, 141-152.

38. Alber, T. (1992) Structure of the leucine zipper, *Curr Opin Genet Dev 2*, 205-210.
39. Marti, D. N., Jelesarov, I., and Bosshard, H. R. (2000) Interhelical ion pairing in coiled coils: solution structure of a heterodimeric leucine zipper and determination of pKa values of Glu side chains, *Biochemistry 39*, 12804-12818.
40. Pace, C. N., Shirley, B. A., McNutt, M., and Gajiwala, K. (1996) Forces contributing to the conformational stability of proteins, *Faseb J 10*, 75-83.
41. Zhou, N. E., Kay, C. M., and Hodges, R. S. (1994) The net energetic contribution of interhelical electrostatic attractions to coiled-coil stability, *Protein Eng 7*, 1365-1372.
42. Lumb, K. J., and Kim, P. S. (1995) Measurement of interhelical electrostatic interactions in the GCN4 leucine zipper, *Science 268*, 436-439.
43. Phelan, P., Gorfe, A. A., Jelesarov, I., Marti, D. N., Warwicker, J., and Bosshard, H. R. (2002) Salt bridges destabilize a leucine zipper designed for maximized ion pairing between helices, *Biochemistry 41*, 2998-3008.
44. Marti, D. N., and Bosshard, H. R. (2003) Electrostatic interactions in leucine zippers: thermodynamic analysis of the contributions of Glu and His residues and the effect of mutating salt bridges, *J Mol Biol 330*, 621-637.
45. Bosshard, H. R., Marti, D. N., and Jelesarov, I. (2004) Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings, *J Mol Recognit 17*, 1-16.
46. Moll, J. R., Olive, M., and Vinson, C. (2000) Attractive interhelical electrostatic interactions in the proline- and acidic-rich region (PAR) leucine zipper subfamily preclude heterodimerization with other basic leucine zipper subfamilies, *J Biol Chem 275*, 34826-34832.
47. Vinson, C., Myakishev, M., Acharya, A., Mir, A. A., Moll, J. R., and Bonovich, M. (2002) Classification of human B-ZIP proteins based on dimerization properties, *Mol Cell Biol 22*, 6321-6335.
48. Deppmann, C. D., Acharya, A., Rishi, V., Wobbes, B., Smeekens, S., Taparowsky, E. J., and Vinson, C. (2004) Dimerization specificity of all 67 B-ZIP motifs in Arabidopsis thaliana: a comparison to Homo sapiens B-ZIP motifs, *Nucleic Acids Res 32*, 3435-3445.
49. Alberti, S., Oehler, S., von Wilcken-Bergmann, B., and Muller-Hill, B. (1993) Genetic analysis of the leucine heptad repeats of Lac repressor: evidence for a 4-helical bundle, *Embo J 12*, 3227-3236.
50. Monera, O. D., Kay, C. M., and Hodges, R. S. (1994) Electrostatic interactions control the parallel and antiparallel orientation of alpha-helical chains in two-stranded alpha-helical coiled-coils, *Biochemistry 33*, 3862-3871.
51. Yadav, M. K., Leman, L. J., Price, D. J., Brooks, C. L., 3rd, Stout, C. D., and Ghadiri, M. R. (2006) Coiled coils at the edge of configurational heterogeneity. Structural analyses of parallel and antiparallel homotetrameric coiled coils reveal configurational sensitivity to a single solvent-exposed amino acid substitution, *Biochemistry 45*, 4463-4473.
52. Liu, J., Yong, W., Deng, Y., Kallenbach, N. R., and Lu, M. (2004) Atomic structure of a tryptophan-zipper pentamer, *Proc Natl Acad Sci U S A 101*, 16156-16161.
53. Zeng, X., Zhu, H., Lashuel, H. A., and Hu, J. C. (1997) Oligomerization properties of GCN4 leucine zipper e and g position mutants, *Protein Sci 6*, 2218-2226.
54. Zeng, X., Herndon, A. M., and Hu, J. C. (1997) Buried asparagines determine the dimerization specificities of leucine zipper mutants, *Proc Natl Acad Sci U S A 94*, 3673-3678.
55. Holton, J., and Alber, T. (2004) Automated protein crystal structure determination using ELVES, *Proc Natl Acad Sci U S A 101*, 1537-1542.
56. Yoon, M. K., Kim, H. M., Choi, G., Lee, J. O., and Choi, B. S. (2007) Structural basis for the conformational integrity of the Arabidopsis thaliana HY5 leucine zipper homodimer, *J Biol Chem 282*, 12989-13002.

57. Gonzalez, L., Jr., Woolfson, D. N., and Alber, T. (1996) Buried polar residues and structural specificity in the GCN4 leucine zipper, *Nat Struct Biol 3*, 1011-1018.
58. Oakley, M. G., and Hollenbeck, J. J. (2001) The design of antiparallel coiled coils, *Curr Opin Struct Biol 11*, 450-457.
59. Oakley, M. G., and Kim, P. S. (1998) A buried polar interaction can direct the relative orientation of helices in a coiled coil, *Biochemistry 37*, 12603-12610.
60. Zuccola, H. J., Rozzelle, J. E., Lemon, S. M., Erickson, B. W., and Hogle, J. M. (1998) Structural basis of the oligomerization of hepatitis delta antigen, *Structure 6*, 821-830.
61. Oakley, M. G., and Kim, P. S. (1997) Protein dissection of the antiparallel coiled coil from Escherichia coli seryl tRNA synthetase, *Biochemistry 36*, 2544-2549.
62. Zhu, B. Y., Zhou, N. E., Kay, C. M., and Hodges, R. S. (1993) Packing and hydrophobicity effects on protein folding and stability: effects of beta-branched amino acids, valine and isoleucine, on the formation and stability of two-stranded alpha-helical coiled coils/leucine zippers, *Protein Sci 2*, 383-394.
63. Zhu, H., Celinski, S. A., Scholtz, J. M., and Hu, J. C. (2000) The contribution of buried polar groups to the conformational stability of the GCN4 coiled coil, *Journal of Molecular Biology 300*, 1377-1387.
64. Newman, J. R., and Keating, A. E. (2003) Comprehensive identification of human bZIP interactions with coiled-coil arrays, *Science 300*, 2097-2101.
65. Shen, H., Cao, K., and Wang, X. (2007) A conserved proline residue in the leucine zipper region of AtbZIP34 and AtbZIP61 in Arabidopsis thaliana interferes with the formation of homodimer, *Biochem Biophys Res Commun.*
66. Portwich, M., Keller, S., Strauss, H. M., Mahrenholz, C. C., Kretzschmar, I., Kramer, A., and Volkmer, R. (2007) A network of coiled-coil associations derived from synthetic GCN4 leucine-zipper arrays, *Angew Chem Int Ed Engl 46*, 1654-1657.
67. Durr, E., Jelesarov, I., and Bosshard, H. R. (1999) Extremely fast folding of a very stable leucine zipper with a strengthened hydrophobic core and lacking electrostatic interactions between helices, *Biochemistry 38*, 870-880.
68. Moran, L. B., Schneider, J. P., Kentsis, A., Reddy, G. A., and Sosnick, T. R. (1999) Transition state heterogeneity in GCN4 coiled coil folding studied by using multisite mutations and crosslinking, *Proc Natl Acad Sci U S A 96*, 10699-10704.
69. Wendt, H., Leder, L., Harma, H., Jelesarov, I., Baici, A., and Bosshard, H R. (1997) Very rapid, ionic strength-dependent association and folding of a heterodimeric leucine zipper, *Biochemistry 36*, 204-213.
70. Meisner, W. K., and Sosnick, T. R. (2004) Fast folding of a helical protein initiated by the collision of unstructured chains, *Proc Natl Acad Sci U S A 101*, 13478-13482.
71. Myers, J. K., and Oas, T. G. (1999) Reinterpretation of GCN4-p1 folding kinetics: partial helix formation precedes dimerization in coiled coil folding, *J Mol Biol 289*, 205-209.
72. Lee, D. L., Lavigne, P., and Hodges, R. S. (2001) Are trigger sequences essential in the folding of two-stranded alpha-helical coiled-coils?, *J Mol Biol 306*, 539-553.
73. Zitzewitz, J. A., Bilsel, O., Luo, J., Jones, B. E., and Matthews, C. R. (1995) Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy, *Biochemistry 34*, 12812-12819.
74. Sosnick, T. R., Jackson, S., Wilk, R. R., Englander, S. W., and DeGrado, W. F. (1996) The role of helix formation in the folding of a fully alpha-helical coiled coil, *Proteins 24*, 427-432.
75. Bosshard, H. R., Durr, E., Hitz, T., and Jelesarov, I. (2001) Energetics of coiled coil folding: the nature of the transition states, *Biochemistry 40*, 3544-3552.
76. Zitzewitz, J. A., Ibarra-Molero, B., Fishel, D. R., Terry, K. L., and Matthews, C. R. (2000) Preformed secondary structure drives the association reaction of GCN4-p1, a model coiled-coil system, *J Mol Biol 296*, 1105-1116.

77. Nikolaev, Y., and Pervushin, K. (2007) NMR Spin State Exchange Spectroscopy Reveals Equilibrium of Two Distinct Conformations of Leucine Zipper GCN4 in Solution, *J Am Chem Soc 129*, 6461-6469.
78. d'Avignon, D. A., Bretthorst, G. L., Holtzer, M. E., Schwarz, K. A., Angeletti, R. H., Mints, L., and Holtzer, A. (2006) Site-specific experiments on folding/unfolding of Jun coiled coils: thermodynamic and kinetic parameters from spin inversion transfer nuclear magnetic resonance at leucine-18, *Biopolymers 83*, 255-267.
79. Mason, J. M., Hagemann, U. B., and Arndt, K. M. (2007) Improved stability of the Jun-Fos Activator Protein-1 coiled coil motif: A stopped-flow circular dichroism kinetic analysis, *J Biol Chem 282*, 23015-23024.
80. Karplus, M., and Weaver, D. L. (1994) Protein folding dynamics: the diffusion-collision model and experimental data, *Protein Sci 3*, 650-668.
81. Ibarra-Molero, B., Zitzewitz, J. A., and Matthews, C. R. (2004) Salt-bridges can stabilize but do not accelerate the folding of the homodimeric coiled-coil peptide GCN4-p1, *J Mol Biol 336*, 989-996.
82. Getzoff, E. D., Cabelli, D. E., Fisher, C. L., Parge, H. E., Viezzoli, M. S., Banci, L., and Hallewell, R. A. (1992) Faster superoxide dismutase mutants designed by enhancing electrostatic guidance, *Nature 358*, 347-351.
83. Camacho, C. J., Kimura, S. R., DeLisi, C., and Vajda, S. (2000) Kinetics of desolvation-mediated protein-protein binding, *Biophys J 78*, 1094-1105.
84. Huang, C. Y., Getahun, Z., Zhu, Y., Klemke, J. W., DeGrado, W. F., and Gai, F. (2002) Helix formation via conformation diffusion search, *Proc Natl Acad Sci U S A 99*, 2788-2793.
85. Hummer, G., Garcia, A. E., and Garde, S. (2000) Conformational diffusion and helix formation kinetics, *Phys Rev Lett 85*, 2637-2640.
86. Thompson, P. A., Eaton, W. A., and Hofrichter, J. (1997) Laser temperature jump study of the helix<==>coil kinetics of an alanine peptide interpreted with a 'kinetic zipper' model, *Biochemistry 36*, 9200-9210.
87. Williams, S., Causgrove, T. P., Gilmanshin, R., Fang, K. S., Callender, R. H., Woodruff, W. H., and Dyer, R. B. (1996) Fast events in protein folding: helix melting and formation in a small peptide, *Biochemistry 35*, 691-697.
88. Holtzer, M. E., Bretthorst, G. L., d'Avignon, D. A., Angeletti, R. H., Mints, L., and Holtzer, A. (2001) Temperature dependence of the folding and unfolding kinetics of the GCN4 leucine zipper via 13C(alpha)-NMR, *Biophys J 80*, 939-951.
89. d'Avignon, D. A., Bretthorst, G. L., Holtzer, M. E., and Holtzer, A. (1999) Thermodynamics and kinetics of a folded-folded' transition at valine-9 of a GCN4-like leucine zipper, *Biophys J 76*, 2752-2759.
90. Dragan, A. I., and Privalov, P. L. (2002) Unfolding of a leucine zipper is not a simple two-state transition, *J Mol Biol 321*, 891-908.
91. Knappenberger, J. A., Smith, J. E., Thorpe, S. H., Zitzewitz, J. A., and Matthews, C. R. (2002) A buried polar residue in the hydrophobic interface of the coiled-coil peptide, GCN4-p1, plays a thermodynamic, not a kinetic role in folding, *J Mol Biol 321*, 1-6.
92. Karplus, M., and Weaver, D. L. (1976) Protein-folding dynamics, *Nature 260*, 404-406.
93. Kammerer, R. A., Jaravine, V. A., Frank, S., Schulthess, T., Landwehr, R., Lustig, A., Garcia-Echeverria, C., Alexandrescu, A. T., Engel, J., and Steinmetz, M. O. (2001) An intrahelical salt bridge within the trigger site stabilizes the GCN4 leucine zipper, *J Biol Chem 276*, 13685-13688.
94. Saudek, V., Pastore, A., Morelli, M. A., Frank, R., Gausepohl, H., and Gibson, T. (1991) The solution structure of a leucine-zipper motif peptide, *Protein Eng 4*, 519-529.

95. Matousek, W. M., Ciani, B., Fitch, C. A., Garcia-Moreno, B., Kammerer, R. A., and Alexandrescu, A. T. (2007) Electrostatic contributions to the stability of the GCN4 leucine zipper structure, *J Mol Biol 374*, 206-219.
96. Wendt, H., Berger, C., Baici, A., Thomas, R. M., and Bosshard, H. R. (1995) Kinetics of folding of leucine zipper domains, *Biochemistry 34*, 4097-4107.
97. Lovett, E. G., D'Avignon, D. A., Holtzer, M. E., Braswell, E. H., Zhu, D., and Holtzer, A. (1996) Observation via one-dimensional 13Calpha NMR of local conformational substates in thermal unfolding equilibria of a synthetic analog of the GCN4 leucine zipper, *Proc Natl Acad Sci U S A 93*, 1781-1785.
98. Holtzer, M. E., Lovett, E. G., d'Avignon, D. A., and Holtzer, A. (1997) Thermal unfolding in a GCN4-like leucine zipper: 13C alpha NMR chemical shifts and local unfolding curves, *Biophys J 73*, 1031-1041.
99. d'Avignon, D. A., Bretthorst, G. L., Holtzer, M. E., and Holtzer, A. (1998) Site-specific thermodynamics and kinetics of a coiled-coil transition by spin inversion transfer NMR, *Biophys J 74*, 3190-3197.
100. Kohn, W. D., Kay, C. M., and Hodges, R. S. (1997) Salt effects on protein stability: two-stranded alpha-helical coiled-coils containing inter- or intrahelical ion pairs, *J Mol Biol 267*, 1039-1052.
101. Hendsch, Z. S., and Tidor, B. (1999) Electrostatic interactions in the GCN4 leucine zipper: substantial contributions arise from intramolecular interactions enhanced on binding, *Protein Sci 8*, 1381-1392.
102. Junius, F. K., O'Donoghue, S. I., Nilges, M., Weiss, A. S., and King, G. F. (1996) High resolution NMR solution structure of the leucine zipper domain of the c-Jun homodimer, *J Biol Chem 271*, 13663-13667.
103. Patel, L. R., Curran, T., and Kerppola, T. K. (1994) Energy transfer analysis of Fos-Jun dimerization and DNA binding, *Proc Natl Acad Sci U S A 91*, 7360-7364.
104. Cranz, S., Berger, C., Baici, A., Jelesarov, I., and Bosshard, H. R. (2004) Monomeric and dimeric bZIP transcription factor GCN4 bind at the same rate to their target DNA site, *Biochemistry 43*, 718-727.
105. Pace, C. N., and Scholtz, J. M. (1998) A helix propensity scale based on experimental studies of peptides and proteins, *Biophys J 75*, 422-427.
106. Murre, C., McCaw, P. S., Vaessin, H., Caudy, M., Jan, L. Y., Jan, Y. N., Cabrera, C. V., Buskin, J. N., Hauschka, S. D., and Lassar, A. B. (1989) Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence, *Cell 58*, 537-544.
107. Ruberti, I., Sessa, G., Lucchetti, S., and Morelli, G. (1991) A novel class of plant proteins containing a homeodomain with a closely linked leucine zipper motif, *Embo J 10*, 1787-1791.
108. Tupler, R., Perini, G., and Green, M. R. (2001) Expressing the human genome, *Nature 409*, 832-833.
109. Ledent, V., Paquet, O., and Vervoort, M. (2002) Phylogenetic analysis of the human basic helix-loop-helix proteins, *Genome Biol 3*, RESEARCH0030.
110. Eferl, R., and Wagner, E. F. (2003) AP-1: a double-edged sword in tumorigenesis, *Nat Rev Cancer 3*, 859-868.
111. Correa, L. G., Riano-Pachon, D. M., Schrago, C. G., dos Santos, R. V., Mueller-Roeber, B., and Vincentz, M. (2008) The role of bZIP transcription factors in green plant evolution: adaptive features emerging from four founder genes, *PLoS ONE 3*, e2944.
112. Massari, M. E., and Murre, C. (2000) Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms, *Mol Cell Biol 20*, 429-440.
113. Luscher, B. (2001) Function and regulation of the transcription factors of the Myc/Max/Mad network, *Gene 277*, 1-14.
114. Meyer, N., and Penn, L. Z. (2008) Reflecting on 25 years with MYC, *Nat Rev Cancer 8*, 976-990.

115. Patel, J. H., Loboda, A. P., Showe, M. K., Showe, L. C., and McMahon, S. B. (2004) Analysis of genomic targets reveals complex functions of MYC, *Nat Rev Cancer 4*, 562-568.
116. Ariel, F. D., Manavella, P. A., Dezar, C. A., and Chan, R. L. (2007) The true story of the HD-Zip family, *Trends Plant Sci 12*, 419-426.
117. Baranger, A. M. (1998) Accessory factor-bZIP-DNA interactions, *Curr Opin Chem Biol 2*, 18-23.
118. Boer, U., Eglins, J., Krause, D., Schnell, S., Schofl, C., and Knepel, W. (2007) Enhancement by lithium of cAMP-induced CRE/CREB-directed gene transcription conferred by TORC on the CREB basic leucine zipper domain, *Biochem J 408*, 69-77.
119. Waldmann, I., Walde, S., and Kehlenbach, R. H. (2007) Nuclear import of c-Jun is mediated by multiple transport receptors, *J Biol Chem 282*, 27685-27692.
120. Bailey, D., and O'Hare, P. (2007) Transmembrane bZIP transcription factors in ER stress signaling and the unfolded protein response, *Antioxid Redox Signal 9*, 2305-2321.
121. Tajima, H., Iwata, Y., Iwano, M., Takayama, S., and Koizumi, N. (2008) Identification of an Arabidopsis transmembrane bZIP transcription factor involved in the endoplasmic reticulum stress response, *Biochem Biophys Res Commun 374*, 242-247.
122. Shifera, A. S., Friedman, J. M., and Horwitz, M. S. (2008) IKKgamma (NEMO) is involved in the coordination of the AP-1 and NF-kappaB pathways, *Mol Cell Biochem 310*, 181-190.
123. Takemori, H., Kajimura, J., and Okamoto, M. (2007) TORC-SIK cascade regulates CREB activity through the basic leucine zipper domain, *Febs J 274*, 3202-3209.
124. Karanam, B., Wang, L., Wang, D., Liu, X., Marmorstein, R., Cotter, R., and Cole, P. A. (2007) Multiple roles for acetylation in the interaction of p300 HAT with ATF-2, *Biochemistry 46*, 8207-8216.
125. Zhou, G. K., Kubo, M., Zhong, R., Demura, T., and Ye, Z. H. (2007) Overexpression of miR165 affects apical meristem formation, organ polarity establishment and vascular development in Arabidopsis, *Plant Cell Physiol 48*, 391-404.
126. Schutze, K., Harter, K., and Chaban, C. (2008) Post-translational regulation of plant bZIP factors, *Trends Plant Sci 13*, 247-255.
127. Norton, J. D. (2000) ID helix-loop-helix proteins in cell growth, differentiation and tumorigenesis, *Journal of cell science 113 (Pt 22)*, 3897-3905.
128. Ruzinova, M. B., and Benezra, R. (2003) Id proteins in development, cell cycle and cancer, *Trends in cell biology 13*, 410-418.
129. Wenkel, S., Emery, J., Hou, B. H., Evans, M. M., and Barton, M. K. (2007) A feedback regulatory module formed by LITTLE ZIPPER and HD-ZIPIII genes, *The Plant cell 19*, 3379-3390.
130. Kiessling, A., Sperl, B., Hollis, A., Eick, D., and Berg, T. (2006) Selective inhibition of c-Myc/Max dimerization and DNA binding by small molecules, *Chem Biol 13*, 745-751.
131. Berg, T. (2008) Inhibition of transcription factors with small organic molecules, *Curr Opin Chem Biol 12*, 464-471.
132. Aikawa, Y., Morimoto, K., Yamamoto, T., Chaki, H., Hashiramoto, A., Narita, H., Hirono, S., and Shiozawa, S. (2008) Treatment of arthritis with a selective inhibitor of c-Fos/activator protein-1, *Nat Biotechnol 26*, 817-823.
133. Zhang, J. W., Tang, Q. Q., Vinson, C., and Lane, M. D. (2004) Dominant-negative C/EBP disrupts mitotic clonal expansion and differentiation of 3T3-L1 preadipocytes, *Proc Natl Acad Sci U S A 101*, 43-47.
134. Gerdes, M. J., Myakishev, M., Frost, N. A., Rishi, V., Moitra, J., Acharya, A., Levy, M. R., Park, S. W., Glick, A., Yuspa, S. H., and Vinson, C. (2006) Activator protein-1 activity regulates epithelial tumor cell identity, *Cancer research 66*, 7578-7588.
135. Strauss, H. M., and Keller, S. (2008) Pharmacological interference with protein-protein interactions mediated by coiled-coil motifs, *Handbook of experimental pharmacology*, 461-482.
136. Li, X., and Stark, G. R. (2002) NFkappaB-dependent signaling pathways, *Exp Hematol 30*, 285-296.

137. DiDonato, J. A., Hayakawa, M., Rothwarf, D. M., Zandi, E., and Karin, M. (1997) A cytokine-responsive IkappaB kinase that activates the transcription factor NF-kappaB, *Nature 388*, 548-554.
138. Zandi, E., Rothwarf, D. M., Delhase, M., Hayakawa, M., and Karin, M. (1997) The IkappaB kinase complex (IKK) contains two kinase subunits, IKKalpha and IKKbeta, necessary for IkappaB phosphorylation and NF-kappaB activation, *Cell 91*, 243-252.
139. Rothwarf, D. M., Zandi, E., Natoli, G., and Karin, M. (1998) IKK-gamma is an essential regulatory subunit of the IkappaB kinase complex, *Nature 395*, 297-300.
140. Karin, M. (1999) How NF-kappaB is activated: the role of the IkappaB kinase (IKK) complex, *Oncogene 18*, 6867-6874.
141. Hacker, H., and Karin, M. (2006) Regulation and function of IKK and IKK-related kinases, *Sci STKE 2006*, re13.
142. Zandi, E., Chen, Y., and Karin, M. (1998) Direct phosphorylation of IkappaB by IKKalpha and IKKbeta: discrimination between free and NF-kappaB-bound substrate, *Science 281*, 1360-1363.
143. Park, M. Y., Moon, J. H., Lee, K. S., Choi, H. I., Chung, J., Hong, H. J., and Kim, E. (2007) FAF1 suppresses IkappaB kinase (IKK) activation by disrupting the IKK complex assembly, *J Biol Chem 282*, 27572-27577.
144. Bornberg-Bauer, E., Rivals, E., and Vingron, M. (1998) Computational approaches to identify leucine zippers, *Nucleic Acids Res 26*, 2740-2746.
145. Wolf, E., Kim, P. S., and Berger, B. (1997) MultiCoil: a program for predicting two- and three-stranded coiled coils, *Protein Sci 6*, 1179-1189.
146. Di Marco, B., Massetti, M., Bruscoli, S., Macchiarulo, A., Di Virgilio, R., Velardi, E., Donato, V., Migliorati, G., and Riccardi, C. (2007) Glucocorticoid-induced leucine zipper (GILZ)/NF-kappaB interaction: role of GILZ homo-dimerization and C-terminal domain, *Nucleic Acids Res 35*, 517-528.
147. Surks, H. K., Mochizuki, N., Kasai, Y., Georgescu, S. P., Tang, K. M., Ito, M., Lincoln, T. M., and Mendelsohn, M. E. (1999) Regulation of myosin phosphatase by a specific interaction with cGMP- dependent protein kinase Ialpha, *Science 286*, 1583-1587.
148. Lee, E., Hayes, D. B., Langsetmo, K., Sundberg, E. J., and Tao, T. C. (2007) Interactions between the leucine-zipper motif of cGMP-dependent protein kinase and the C-terminal region of the targeting subunit of myosin light chain phosphatase, *J Mol Biol 373*, 1198-1212.
149. Ihara, E., and MacDonald, J. A. (2007) The regulation of smooth muscle contractility by zipper-interacting protein kinase, *Can J Physiol Pharmacol 85*, 79-87.
150. Kawai, T., Matsumoto, M., Takeda, K., Sanjo, H., and Akira, S. (1998) ZIP kinase, a novel serine/threonine kinase which mediates apoptosis, *Mol Cell Biol 18*, 1642-1651.
151. Graves, P. R., Winkfield, K. M., and Haystead, T. A. (2005) Regulation of zipper-interacting protein kinase activity in vitro and in vivo by multisite phosphorylation, *J Biol Chem 280*, 9363-9374.
152. Hulme, J. T., Ahn, M., Hauschka, S. D., Scheuer, T., and Catterall, W. A. (2002) A novel leucine zipper targets AKAP15 and cyclic AMP-dependent protein kinase to the C terminus of the skeletal muscle Ca2+ channel and modulates its function, *J Biol Chem 277*, 4079-4087.
153. Hulme, J. T., Lin, T. W., Westenbroek, R. E., Scheuer, T., and Catterall, W. A. (2003) Beta-adrenergic regulation requires direct anchoring of PKA to cardiac CaV1.2 channels via a leucine zipper interaction with A kinase-anchoring protein 15, *Proc Natl Acad Sci U S A 100*, 13093-13098.
154. Few, W. P., Scheuer, T., and Catterall, W. A. (2007) Dopamine modulation of neuronal Na(+) channels requires binding of A kinase-anchoring protein 15 and PKA by a modified leucine zipper motif, *Proc Natl Acad Sci U S A 104*, 5187-5192.
155. Sollner, T., Whiteheart, S. W., Brunner, M., Erdjument-Bromage, H., Geromanos, S., Tempst, P., and Rothman, J. E. (1993) SNAP receptors implicated in vesicle targeting and fusion, *Nature 362*, 318-324.

156. Fasshauer, D., Sutton, R. B., Brunger, A. T., and Jahn, R. (1998) Conserved structural features of the synaptic fusion complex: SNARE proteins reclassified as Q- and R-SNAREs, *Proc Natl Acad Sci U S A 95*, 15781-15786.
157. Sutton, R. B., Fasshauer, D., Jahn, R., and Brunger, A. T. (1998) Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 A resolution, *Nature 395*, 347-353.
158. Mobley, P. W., Barry, J. A., Waring, A. J., Sherman, M. A., and Gordon, L. M. (2007) Membrane perturbing actions of HIV type 1 glycoprotein 41 domains are inhibited by helical C-peptides, *AIDS Res Hum Retroviruses 23*, 224-242.
159. Knejzlik, Z., Smekalova, Z., Ruml, T., and Sakalian, M. (2007) Multimerization of the p12 domain is necessary for Mason-Pfizer monkey virus Gag assembly in vitro, *Virology 365*, 260-270.
160. Nellissery, J. K., Szczepaniak, R., Lamberti, C., and Weller, S. K. (2007) A putative leucine zipper within the herpes simplex virus type 1 UL6 protein is required for portal ring formation, *J Virol 81*, 8868-8877.
161. Haller, O., Staeheli, P., and Kochs, G. (2007) Interferon-induced Mx proteins in antiviral host defense, *Biochimie 89*, 812-818.
162. Melen, K., Ronni, T., Broni, B., Krug, R. M., von Bonsdorff, C. H., and Julkunen, I. (1992) Interferon-induced Mx proteins form oligomers and contain a putative leucine zipper, *J Biol Chem 267*, 25898-25907.
163. Haller, O., Frese, M., and Kochs, G. (1998) Mx proteins: mediators of innate resistance to RNA viruses, *Revue scientifique et technique (International Office of Epizootics) 17*, 220-230.
164. Tumpey, T. M., Szretter, K. J., Van Hoeven, N., Katz, J. M., Kochs, G., Haller, O., Garcia-Sastre, A., and Staeheli, P. (2007) The Mx1 gene protects mice against the pandemic 1918 and highly lethal human H5N1 influenza viruses, *J Virol 81*, 10818-10821.
165. Haller, O., and Kochs, G. (2002) Interferon-induced mx proteins: dynamin-like GTPases with antiviral activity, *Traffic 3*, 710-717.
166. Di Paolo, C., Hefti, H. P., Meli, M., Landis, H., and Pavlovic, J. (1999) Intramolecular backfolding of the carboxyl-terminal end of MxA protein is a prerequisite for its oligomerization, *J Biol Chem 274*, 32071-32078.
167. Janzen, C., Kochs, G., and Haller, O. (2000) A monomeric GTPase-negative MxA mutant with antiviral activity, *J Virol 74*, 8202-8206.
168. Schumacher, B., and Staeheli, P. (1998) Domains mediating intramolecular folding and oligomerization of MxA GTPase, *J Biol Chem 273*, 28365-28370.
169. Zurcher, T., Pavlovic, J., and Staeheli, P. (1992) Mechanism of human MxA protein action: variants with changed antiviral properties, *Embo J 11*, 1657-1661.
170. Kochs, G., Janzen, C., Hohenberg, H., and Haller, O. (2002) Antivirally active MxA protein sequesters La Crosse virus nucleocapsid protein into perinuclear complexes, *Proc Natl Acad Sci U S A 99*, 3153-3158.
171. Andersson, I., Bladh, L., Mousavi-Jazi, M., Magnusson, K. E., Lundkvist, A., Haller, O., and Mirazimi, A. (2004) Human MxA protein inhibits the replication of Crimean-Congo hemorrhagic fever virus, *J Virol 78*, 4323-4329.
172. Chen, L., Glover, J. N., Hogan, P. G., Rao, A., and Harrison, S. C. (1998) Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA, *Nature 392*, 42-48.
173. Newbury, S. F. (2006) Control of mRNA stability in eukaryotes, *Biochem Soc Trans 34*, 30-34.
174. Mata, J., Marguerat, S., and Bahler, J. (2005) Post-transcriptional control of gene expression: a genome-wide perspective, *Trends Biochem Sci 30*, 506-514.
175. Meyer, S., Temme, C., and Wahle, E. (2004) Messenger RNA turnover in eukaryotes: pathways and enzymes, *Crit Rev Biochem Mol Biol 39*, 197-216.
176. Wilusz, C. J., and Wilusz, J. (2004) Bringing the role of mRNA decay in the control of gene expression into focus, *Trends Genet 20*, 491-497.

Chapter II

The Leucine Zippers of the Transcription Factors GCN4 and c-Jun Have Ribonuclease Activity.

Research presented in this chapter of my thesis was mainly performed in the laboratory of Prof. Bernd Gutte at the University of Zurich. The experimental studies were to a substantial extent conducted by Dr. Christine Deillon and Dr. Stephan Hoffman. My contribution to this research includes (1) purification of LZ peptides; (2) ribonuclease assays with LZ mutants, full-length cJun, and RNasin; (3) kinetic analysis of LZ activity; (4) writing and revising the manuscript for publication.

The Leucine Zippers of the Transcription Factors GCN4 and c-Jun Have Ribonuclease Activity[†]

Christine Deillon,[‡,§] Yaroslav Nikolaev,[‡,§,@,#] Stefan R.K. Hoffmann,[‡,§] Laurent Bigler,[⊥] Sebastian Friess,[@] Renato Zenobi,[@] Konstantin Pervushin,[@,#] Peter Hunziker,[¶] and Bernd Gutte[*,§]

Biochemisches Institut and Organisch-Chemisches Institut der Universität Zürich and Functional Genomics Center Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland, and Department of Chemistry and Applied Biosciences, Swiss Federal Institute of Technology Zürich, CH-8093 Zürich, Switzerland, and Biozentrum der Universität Basel, CH-4056 Basel, Switzerland

Running Title: GCN4 and c-Jun Leucine Zippers Have Ribonuclease Activity

[†] This work was supported by the Kanton of Zürich, the Mesta Foundation, and the Stiftung für wissenschaftliche Forschung an der Universität Zürich.
[*] To whom correspondence should be addressed. Tel: +41 44 6355530, E-mail: gutte@bioc.unizh.ch.
[‡] These authors contributed equally to this work.
[§] Biochemisches Institut, Universität Zürich.
[#] Present address: Biozentrum der Universität Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland.
[⊥] Organisch-Chemisches Institut, Universität Zürich.
[@] Department of Chemistry and Applied Biosciences, Swiss Federal Institute of Technology, Zürich.
[¶] Functional Genomics Center Zurich.

Abbreviations: HIV-1, human immunodeficiency virus type 1; R42, designed 42-residue HIV-1 enhancer-binding peptide; GCN4, yeast transcription activator; c-Jun, oncoprotein and component of transcription factor AP-1; Fmoc, fluorenylmethyloxycarbonyl amino protecting group; GCN4 LZ35, leucine zipper of GCN4, 35 residues per chain; *r*LZ35, single-chain 35-residue *retro*-leucine zipper of GCN4; *r*LZ38, four-chain *retro*-leucine zipper of GCN4 containing 38 residues per chain; *r*-LZ67, four-chain fusion peptide of *r*LZ38 and shortened R42 containing 67 residues per chain; c-Jun LZ36, leucine zipper of c-Jun, 36 residues per chain; RNA18, synthetic octadecaribonucleotide; U>, C>, G> and A>, 2',3'-cyclic phosphates of U, C, G and A.

ABSTRACT: The X-ray structure of the *retro*-version of the leucine zipper moiety of yeast transcriptional activator GCN4 (Mittl, P.R.E., Deillon, C., Sargent,D., Liu, N., Klauser,S., Thomas, R.M., Gutte, B., and Grütter, M.G. (2000) Proc.Natl.Acad.Sci.U.S.A. 97, 2562-2566) suggested that this *retro*-peptide may have ribonuclease activity. Here we show that not only the single-chain and four-chain *retro*-leucine zippers but also the authentic leucine zipper of GCN4 had weak but distinct ribonuclease activity. The cleavage of RNA was unspecific, included the breakage of the 3',5'-phosphodiester bond at the 3'-end of guanosine, gave 2',3'-cyclic phosphates as the final products as demonstrated by HPLC/electrospray ionization mass spectrometry, and was not inhibited by a recombinant ribonuclease A inhibitor. Several mutants of the GCN4 leucine zipper were inactive and were thus important negative controls of the activity assays. The leucine zipper moiety of c-Jun as well as the entire c-Jun protein also cleaved RNA suggesting possible biological relevance of this activity and raising the question of how many more dual-function proteins of this kind may exist. As additional result of this work, the *retro*-leucine zipper of GCN4 could be added to the list of functional *retro*-peptides.

INTRODUCTION

A designed 42-residue HIV-1 enhancer-binding peptide (R42) derived from the operator-binding domain of bacteriophage 434 repressor was a strong inhibitor of HIV-1 enhancer-controlled transcription of reporter genes in HeLa cell nuclear extracts and COS-1 cells (*1-3*). Because the two identical enhancers of the long terminal repeat of HIV-1 seemed to allow the binding of two equivalents of R42, it was conceivable that a stable dimer of this peptide could have enhanced affinity and specificity of DNA binding. To test this assumption, R42 was extended at the N-terminus with the leucine zipper domain of the yeast transcriptional activator GCN4 (*4*). Leucine zippers (*5*) are parallel alpha-helical homo- or heterodimers forming coiled coils; their basic unit is a heptad of amino acid residues labeled a-g which is repeated at least four times and contains exclusively leucine residues in the d positions. The resulting dimeric leucine zipper – R42 fusion protein (66 residues per chain) showed indeed increased enhancer binding specificity (*4*).

In a later, refined version, R42 was extended with the *retro*-leucine zipper of GCN4 which, if it also formed a coiled coil, seemed more likely to dimerize the peptide in the proposed HIV enhancer binding mode (Fig. 1 in reference 6). The *retro*-leucine zipper of GCN4 itself was a new, unnatural peptide. In order to study its physicochemical properties, it was synthesized separately in addition to the *retro*-leucine zipper – R42 fusion protein. Ultracentrifugation (*6*) and x-ray crystallography (*7*) revealed that a 38-residue *retro*-leucine zipper, covalently dimerized by an interchain disulfide bond (Figure 1), formed a parallel four-helix bundle (a dimer of dimers) even at low micromolar concentrations. In the crystal, the closest contact between two neighbouring bundles was made between exposed histidine residues, one from each tetramer (Fig. 2C in reference 7). This

juxtaposition of histidines was vaguely reminiscent of the active-site structure of ribonuclease A (RNase A)(*8*) and prompted us to test the *retro*-leucine zipper for ribonuclease activity. Surprisingly, weak ribonuclease activity distinct from that of RNase A was not only found for the GCN4 *retro*-leucine zipper but also for the authentic leucine zipper domains of GCN4 and c-Jun.

Here we have characterized mainly the ribonuclease activity of the authentic leucine zippers of GCN4 and c-Jun because this activity may have biological implications whereas that of the *retro*-leucine zippers seemed to be only of theoretical interest.

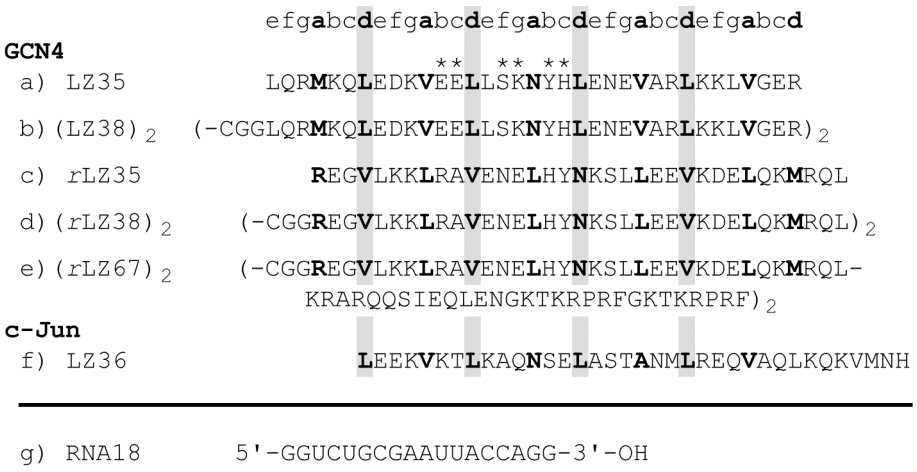


Figure 1. Relevant leucine zipper and RNA sequences. The a- and d-positions of the repeating heptads of the GCN4 and c-Jun leucine zippers are shown in bold print. In the *retro*-sequences, a- and d-positions are reversed. a) GCN4 LZ35 (residues 247-281); asterisk (*), indicates positions of point mutations; b) GCN4 LZ38, obtained through N-terminal extension of LZ35 by CysGlyGly which allowed formation of covalent disulfide dimer; c) GCN4 *r*LZ35, *retro*-sequence of LZ35; d) GCN4 *r*LZ38, *retro*-sequence of GCN4 LZ38; e) GCN4 *r*LZ67, obtained through fusion of *r*LZ38 with shortened HIV enhancer-binding peptide R42; f) c-Jun LZ36 (residues 280-315); g) RNA18.

EXPERIMENTAL PROCEDURES

Peptide Synthesis. All leucine zipper peptides were synthesized by the solid phase method (*9-10*) on an Applied Biosystems 433A Peptide Synthesizer, using Fmoc chemistry (*11*), and purified by reversed-phase HPLC. The purity of the peptides was verified by amino acid analysis and mass spectrometry (Supporting Information, Figure 1).

Ribonuclease Assays of the GCN4 and c-Jun Leucine Zippers (GCN4 LZ35 and c-Jun LZ36; Figure 1). Synthetic wild-type and mutant GCN4 leucine zippers (15 μM, based on the relative molecular mass of the coiled coils; wild-type LZ35: 8477 Da) and a commercial octadecaribonucleotide (RNA18, Figure 1; 50 μM) were incubated in 12.5 μL of 20 mM Tris x HCl and 80 mM KCl, pH 6.5 to pH 7.2, at 37° C for two to 24 h. Then 100 μL of 20 mM ammonium acetate, with 35 μM bromophenol blue as an internal reference, was added and 100 μL of the resulting mixture was fractionated on a Nucleosil C-18 300-5 HPLC column (Macherey & Nagel) using a solvent gradient (first solvent: 20 mM ammonium acetate, second solvent: 97 % acetonitrile). Uncleaved RNA18 and its cleavage products were detected at 254.4 nm. Digestions in 50 mM potassium phosphate and 50 mM KCl gave similar product patterns. *Retro*-leucine zipper 67 (*r*-LZ67), the tetrameric 67-residue fusion peptide of GCN4 *retro*-leucine zipper 38 (*r*-LZ38) and shortened HIV enhancer-binding peptide (Figure 1), was assayed in 1.5 μM and 3 μM solution, respectively (based on a relative molecular mass of 31917 Da), using 25 μM RNA18 as substrate.

The assays were repeated in the presence of two concentrations of the recombinant RNase A inhibitor RNasin (50 kDa, Promega; 5 U and 150 U per 12.5 µL assay solution). The GCN4 LZ35 and RNA18 concentrations were 25 µM and 50 µM, respectively. RNase A (9.35 pM) was used as control.

The peptide and RNA concentrations in the ribonuclease activity assays of synthetic c-Jun LZ36 were 30 µM and 85 µM, respectively, in the Tris x HCl buffer; reaction time was two to 48 h.

To detect a possible ribonuclease activity of full-length c-Jun, the recombinant protein (Promega; 8 µM) was submitted to ultrafiltration on a Millipore Ultrafree-0.5 membrane to increase its concentration to 50 µM and to change the buffer to 20 mM Tris x HCl, 85 mM KCl, pH 7.2. The final c-Jun and RNA18 concentrations in the assays were 20 µM and 90 µM, respectively, in 10 µL of the Tris x HCl buffer. The assays were performed in presence or absence of 1 U of the RNasin inhibitor. After 48 h at 37°C the reaction mixtures were analyzed by gradient elution from an Eclipse XDB-C18 RP-HPLC column (Agilent, 0.46 cm x 15 cm; solvent: 20 mM ammonium acetate, pH 7.8/acetonitrile; elution was recorded between volume ratios 98:2 and 85:15 at 254 nm; gradient volume: 16 mL).

Ribonuclease activities and kinetic data of the leucine zipper-catalyzed RNA degradation were calculated from the time-dependent decrease of the area of the RNA peak compared to the peak area at time zero.

Kinetic Measurements of the Cleavage of RNA by GCN4 LZ35. Initial velocities of RNA degradation were determined for three different concentrations (34.6 µM, 69.2 µM, and 103.8 µM, respectively) of the following octadecaribonucleotide: GGACUACGACUUACUAUU. The concentration of the LZ35 coiled coil was 27

μM. The experimental conditions and the HPLC-UV detection system used were the same as those described for the ribonuclease assays.

Mass Spectral Analysis of RNA18 – GCN4 LZ35 Complex Formation and RNA18 Cleavage by LZ35. Complex formation was shown by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry using 6-aza-2-thiothymine/citrate as the matrix. Spectra were recorded in negative ion, linear mode on a time-of-flight instrument (AXIMA CFR, Shimadzu/Kratos) equipped with a nitrogen laser (λ = 337 nm, 3 ns pulse width). The concentrations of RNA18 and LZ35 were 68 μM and 60 μM, respectively, in 20 mM ammonium acetate, 80 mM KCl, pH 6.5. The volume applied was 1 μL.

The cleavage of RNA18 by GCN4 leucine zippers and *retro*-leucine zippers was analyzed by HPLC – electrospray ionization mass spectrometry [$(M-H)^-$]. The cleavage experiments were performed in 20 mM Tris x HCl, 80 mM KCl, at pH 6.5 and 25° C. Samples (5 μL each) were applied on a Nucleosil 100-3 C_{18} HD column (Macherey & Nagel) and eluted using a stepwise gradient from 100% 20 mM ammonium acetate to 97% acetonitrile. The masses of the separated RNA18 cleavage products were determined on a Bruker ESQUIRE-LC quadrupole ion-trap instrument.

RESULTS

Assays of the Ribonuclease Activity of Wild-Type and Mutant GCN4 Leucine Zipper Peptides. The first assays were performed with GCN4 *r*LZ38, the parallel four-helix bundle (*7*) with possible ribonuclease activity, and RNA18 as substrate. Based on the positive result of these assays (Figure 2, panel A), single-chain *r*LZ35 (Figure 1) and tetrameric *r*LZ67 (Figure 1), the fusion peptide of *r*LZ38 with a shortened version of the artificial 42-residue HIV enhancer-binding peptide, were tested for ribonuclease activity and were also found to cleave RNA18 (Figure 2, panels B and D, respectively). R42 alone did not catalyze RNA18 cleavage, therefore the mixture of R42 and RNA18 served as a negative control (Figure 2, panel C).

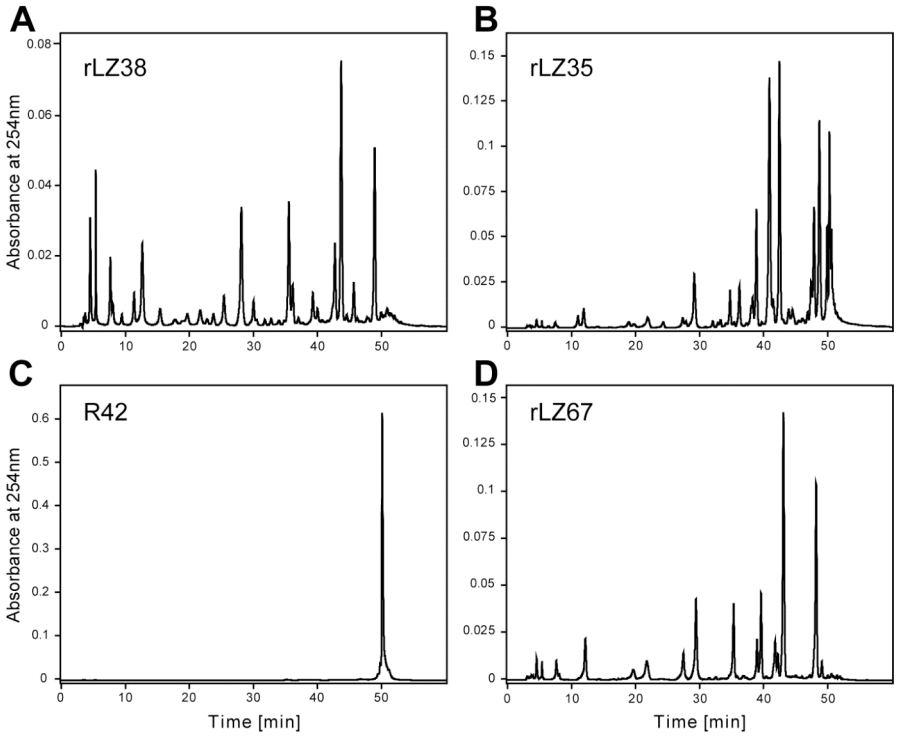


Figure 2. HPLC-electrospray ionization mass spectrometry of the products of 24-hour cleavages of RNA18 (85 μM) at 25°C by 28 μM tetrameric GCN4 *r*LZ38 (A), 56 μM monomeric GCN4 *r*LZ35 (B), and 28 μM tetrameric GCN4 *r*LZ67 (D). The mixture of R42 (14 μM) and RNA18 (48 μM) served as negative control (C). Uncleaved RNA18 substrate eluted from the HPLC column after 50 min.

Quantitative results of the cleavage of RNA18 by wild-type and mutant 35-residue leucine zipper peptides are summarized in Table 1. In the experimental conditions given, wild-type GCN4 LZ35 cleaved 47 % of RNA18 present at time 0 in 13 h. Replacement of Glu12 with alanine abolished the ribonuclease activity completely. Low activity (6 to 12 % cleavage in 13 h) was retained by the Glu12Gln, Glu13Ala, Ser16Ala, and Ser16Thr analogues whereas exchange of Lys17 and Tyr19 for alanine did not affect the enzymatic activity (47 % and 40 %, respectively). The result of the assay of the His20Ala mutant was concealed most likely by the formation of insoluble complexes between the peptide and RNA18. After 24 h, the wild-type peptide had cleaved 63 % of RNA18.

Table 1: Cleavage of RNA18 by Wild-Type (wt) and Mutant GCN4 LZ35[a]

LZ35	2h	13h
wt	17	47
E12A	0	2
E12Q	2	6
E13A	n.d.[b]	12
S16A	< 1	9
S16T	< 1	11
K17A	4	47
Y19A	13	40
H20A[c]		

[a] Percent RNA cleaved, based on the amount present at the beginning of the experiment (50 μM) and determined by HPLC at 254 nm. The concentration of the leucine zippers was 15 μM.

[b] Not determined.

[c] Result commented in DISCUSSION.

The *r*LZ67 fusion peptide (Figure 1; tetrameric, relative molecular mass: 31917 Da) was assayed in two concentrations; in 2 h, a 3 μM solution cleaved 49 %, a 1.5 μM solution 35 % of the original amount of the substrate.

Ribonuclease Assays of GCN4 LZ35 in Presence of RNase A Inhibitor. Five units of the recombinant RNase A inhibitor RNasin did not affect the ribonuclease activity of 12.5 μL of a 25 μM solution of GCN4 LZ35 whereas 150 units of the inhibitor lowered the activity by approximately 60 %. The incubation time was 13 h. In a control experiment, five units of RNasin abolished the activity of 12.5 μL of a 9 pM solution of RNase A completely.

Ribonuclease Assays of Synthetic c-Jun LZ36 and Recombinant c-Jun. In the experimental conditions applied, synthetic c-Jun LZ36 (30 μM) cleaved 32 % of RNA18 (85 μM at time zero) in 24 h. Full-size recombinant c-Jun (20 μM) produced a similar result in 48 h. Importantly, the ribonuclease activity of either c-Jun leucine zipper or full-size c-Jun protein was not affected by RNasin inhibitor. Several peaks of the RNA18 cleavage patterns produced by GCN4 LZ35 and c-Jun LZ36 had identical HPLC retention times.

Kinetic Data of the GCN4 LZ35-Catalyzed RNA Cleavage. Initial velocity values were fitted to the Michaelis-Menten equation yielding K_M=30 μM and k_{cat}=0.00044 ± 0.000029 (SE) min^{-1} (Supporting Information) assuming that the total Quasi Steady State Assumption (tQSSA) *(12)* applies within the conditions of the experiment. It must be emphasized that the numbers obtained are valid only for the initial step(s) of the RNA18 cleavage, i.e. cleavage of one or very few phosphodiester bonds that are most accessible and structurally most favorable with respect to substrate specificity leading to a decrease of the area of the RNA18 peak.

Mass Spectral Analysis of RNA18 – GCN4 LZ35 Complex Formation and RNA18 Cleavage by LZ35. MALDI mass spectrometry (Figure 3) showed signal groups starting at 11547 and 10015 Da indicating the presence of double-stranded RNA18 and a 1:1 molar complex of monomeric 35-residue peptide with single-stranded RNA18, respectively. The deviations from the masses expected (11540 and 10008 Da) were below 0.1 % and thus well within the range of accuracy of the instrument and the associated mass calibration. In both groups, the spacings of the signals with higher mass were m=38 (dominant, exchange of K for H) and m=22 (minor, exchange of Na for H), respectively. The alkali cationization of the phosphate backbone of RNA18 could not be suppressed completely despite the use of citrate in the matrix. The lower mass range of the spectrum showed the signals for the monomers of RNA18 (5771 Da) and LZ35 (4237 Da) as well as signals for major cleavage products of RNA18 at 1906 Da (AAUUAC>), 1624 Da (GGUCU), 1616 Da (GAAUU>), and 964 Da (AAU>).

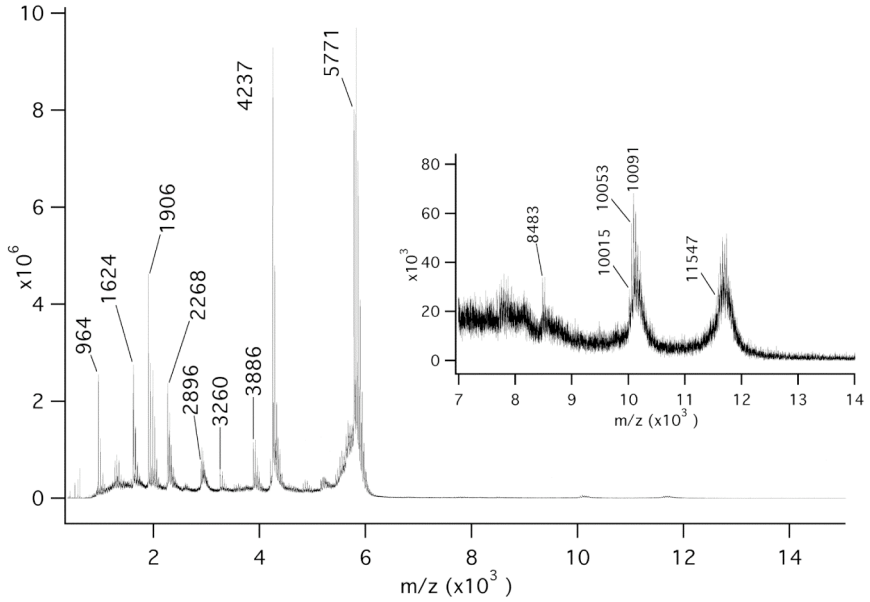


Figure 3. The 1 to 14 kDa portion of the MALDI mass spectrum of a mixture of GCN4 LZ35 (60 μM) and RNA18 (68 μM) recorded in negative ion mode using 6-aza-2-thiothymine/citrate as matrix. The nature of the signals obtained is explained in the RESULTS section.

The products of a 24-h digestion of RNA18 catalyzed by GCN4 LZ35 were separated and characterized using HPLC – electrospray ionization mass spectrometry and were compared with those of a 3-h digestion by RNase A (Figure 4, panels B and C). LZ35 cleaved the RNA18 substrate mainly at the 3'-end of U and C and to a smaller extent at the 3'-end of A and G. Almost all products were obtained as the 2',3'-cyclic phosphates (Figure 4, panel C). In contrast, RNase A could not cleave RNA at the 3'-end of G and catalyzed the hydrolysis of the intermediate 2',3'-cyclic phosphates to give the final 3'-phosphate-containing products (Figure 4, panel B). Even after 96 h no fragments of RNA18 could be detected in the absence of LZ35 or RNase A (Figure 4, panel A).

Mass spectral analysis showed also that the leucine zipper peptides emerged unaltered from the catalytic process; as true enzymes, they were not oxidized or otherwise modified.

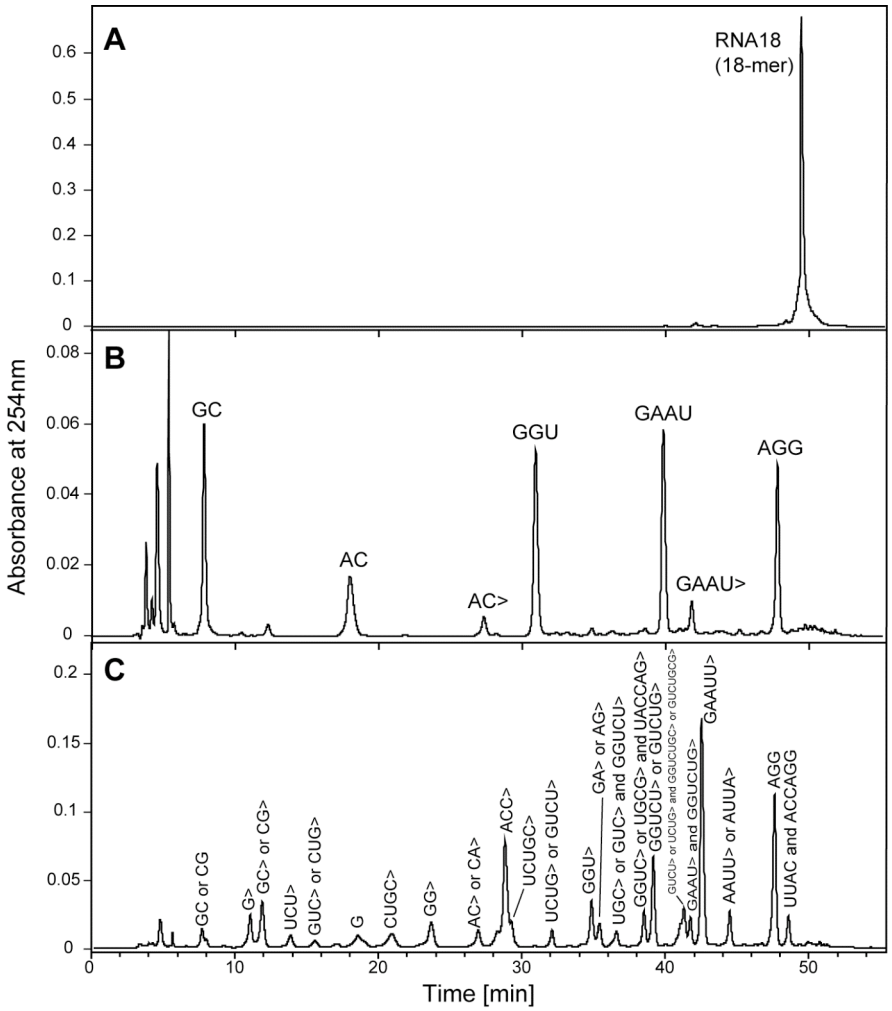


Figure 4. Comparison of the cleavage of RNA18 (80 µM) by GCN4 LZ35 and RNase A using HPLC-electrospray ionization mass spectrometry. Uncleaved RNA18 eluted from the HPLC column after 49.4 min. A, RNA18 blank (incubation time, 24 h); B, RNA18 + RNase A (90 nM; incubation time, 3 h); the major products were AGG (956 Da) lacking phosphate at the 3'-end, and the 3'-phosphate-containing GAAU (1326 Da), GGU (1013 Da), AC (651 Da), and GC (667 Da); C, RNA18 + GCN4 LZ35 (25 µM, based on the relative molecular mass of the coiled coil, i.e. 8477 Da; incubation time, 24 h); the major products were AGG (956 Da) lacking phosphate at the 3'-end, GAAUU> (1614 Da), GUCUG> or GGUCU> (1606 Da), and ACC> (938 Da); the composition of several of the minor products indicated cleavage of RNA18 at the 3'-end of guanosine phosphate or adenosine phosphate residues; for example, AAUU> or AUUA>, UACCAG>, GA> or AG>, UCUG> or GUCU>, GG>, G, UCU>, and G>. ">" – to 2',3'-cyclic phosphate.

DISCUSSION

This project started with the assumption that the synthetic 38-residue *retro*-leucine zipper of GCN4 (GCN4 *r*LZ38) could have ribonuclease activity. In a first test, *r*LZ38 was indeed found to cleave tRNA whereas bovine serum albumin was inactive (data not shown). For the following activity tests a synthetic octadecaribonucleotide (RNA18; Figure 1) was used as substrate. Figure 2 shows that *r*LZ38, *r*LZ35, and *r*LZ67 (a fusion protein of *r*LZ38 with R42, our designed HIV-1 enhancer-binding peptide) cleaved RNA18 whereas R42 alone did not have ribonuclease activity; after a 24-h incubation with R42, RNA18 eluted unchanged from the HPLC column (Figure 2, panel C). This indicated that the ribonuclease activity resided in the *retro*-leucine zipper sequences. Unexpectedly, the normal, "forward" sequences of the GCN4 leucine zipper, GCN4 LZ35, its disulfide-containing analogue, GCN4 $LZ38_{(S\text{-}S)}$, and the c-Jun leucine zipper, c-Jun LZ36, also cleaved the RNA18 substrate (sequences are given in Figure 1). The ribonuclease activity of both normal and *retro*-leucine zipper of GCN4 could be the result of their close x-ray structural similarity (*7*). As the ribonuclease activity of the GCN4 and c-Jun leucine zippers may be of cell biological relevance, all further studies were performed with the normal leucine zippers.

First it was shown that the nuclease activity of GCN4 LZ35 was RNA-specific. LZ35 cleaved tRNA, poly(C/U/A/G) (analyzed by thin-layer chromatography, data not shown) and synthetic RNA18. It did not cleave DNA and was also inactive in chymotrypsin and trypsin activity assays. Although potential substrates, the dinucleoside phosphates CpG and UpG were not split by LZ35 as shown by HPLC; most likely, their binding to the leucine zipper was too weak.

Complex formation between LZ35 and RNA18 was demonstrated by MALDI mass spectrometry (Figure 3). MALDI data of noncovalent complexes, however, have to be interpreted with caution because sample preparation and laser action are generally disruptive; in addition, nonspecific clusters of sample constituents can be produced in a dense MALDI plume. Disruption of the LZ35 x RNA18 complex by the MALDI process may have been responsible for the strong signals at 4237 and 5771 Da (monomeric 35-residue peptide and single-stranded RNA18, respectively). The intense signal groups starting at 10015 and 11547 Da (Figure 3, inset) indicated the presence of specific noncovalent complexes (monomeric 35-residue peptide x single-stranded RNA18 and double-stranded RNA18, respectively) whereas the exponentially decreasing signal intensity around 8483 Da was typical for the presence of nonspecific clusters.

To determine the nature of the ribonuclease activity of GCN4 LZ35, the cleavage products of RNA18 were separated by HPLC and then characterized by electrospray ionization mass spectrometry. Figure 4, panel C, shows that the nucleolytic activity of LZ35 seemed to be unspecific. The observations that in 24-h reactions almost all products were obtained as 3'-terminal 2',3'-cyclic phosphates and that LZ35 was completely inactive in 6 M urea, made contamination by RNase A unlikely. Furthermore, a number of products was formed through cleavage of G-A, G-C, G-G and G-U 3',5'-phosphodiester bonds (Figure 4, panel C, from right to left: AAUU>, GGUCUG>, GUCU>, UCUG>, UACCAG>, UCUGC>, GG>, the 3'-terminal G, UCU>, G>) providing additional evidence that the ribonuclease activity of LZ35 was intrinsic because contamination by a G-specific ribonuclease such as RNase T_1 could be excluded. For comparison, digestion of RNA18 by

RNase A gave the expected hydrolysis products (Figure 4, panel B) lacking completely fragments formed through cleavage after G. In the absence of LZ35 or RNase A, RNA18 was not cleaved (Figure 4, panel A). R42 (the synthetic 42-residue artificial HIV enhancer-binding peptide, net charge +14), bovine serum albumin, and glyceraldehyde-3-phosphate dehydrogenase served as negative controls.

In LZ35, potential active site residues (Glu8, Asp9, Glu12, Glu13, Ser16, Tyr19, His20) and basic residues are aligned on one side of the helix (*13*) possibly representing the binding site of the RNA18 substrate. Table 1 shows the effect of replacement of several of these residues mainly by alanine. The Glu12Ala analogue was completely inactive compared to the wild-type sequence, once more excluding contamination of the synthetic peptides by natural ribonucleases. The ribonuclease activity of the Glu12Gln, Glu13Ala, Ser16Ala, and Ser16Thr analogues was strongly reduced as compared with that of wild-type LZ35. The activity assays of the His20Ala mutant repeatedly showed a time-dependent decrease of the area of the RNA18 substrate peak in presence of the His20Ala analogue without detectable formation of cleavage products. Perhaps this mutant formed insoluble complexes with RNA18 which would have blocked a possible ribonuclease activity. If His20 was essential for activity, an RNase A-like mechanism (*8*) could be ruled out at least for *r*LZ35 which was shown to be monomeric by ultracentrifugation in the conditions of the assay *(6)* and therefore contained only one histidine residue. More likely, the cleavage of RNA by GCN4 LZ and *r*LZ followed an RNase T_1-like mechanism. In RNase T_1, Glu58/His92 were found to act as the active-site base/acid couple with His 40 participating in the electrostatic stabilization of the transition state (*14*). Later it was shown that

RNase T_1 was still active when both histidines were replaced by aspartate (*15*). It is conceivable that two of the four acidic residues which are located within approximately 5.5 Å on the same face of the GCN4 leucine zipper (Glu 8, Asp 9, Glu 12, Glu 13) formed an active site resulting in the low ribonuclease activity of LZ35 and *r*LZ35. Sequence comparison (Figure 1) suggested that there must be flexibility in the use of active site residues by the leucine zippers of GCN4 and c-Jun.

Although intrinsic and contaminating RNase A activity could be excluded, we observed an inhibitor concentration-dependent partial blocking of the ribonuclease activity of LZ35 by RNase A-specific recombinant RNasin indicating weak non-specific binding of the inhibitor by LZ35.

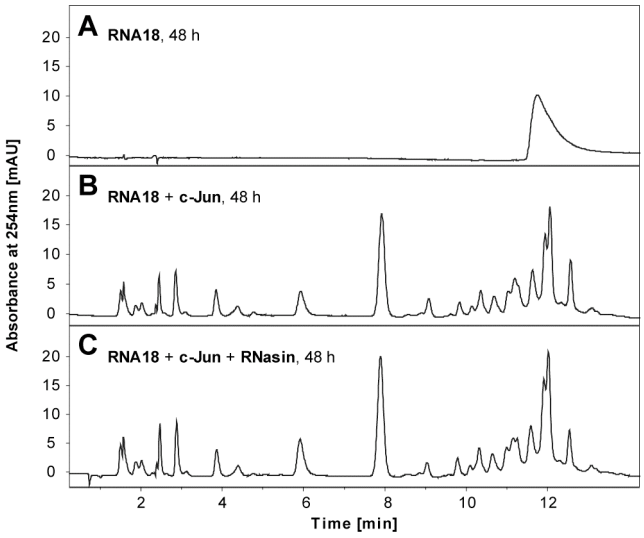


Figure 5. Cleavage of RNA18 (90 µM) by full-length 40 kDa recombinant c-Jun as shown by reversed-phase HPLC. A, RNA18 control; B, RNA18 + 20 µM c-Jun; C, RNA18 + 20 µM c-Jun in the presence of 1 U of RNasin ribonuclease inhibitor.

The k_{cat} value of the reaction of GCN4 LZ35 with RNA18 (0.00044 min^{-1}) reflected the decline of the area of the octadecaribonucleotide substrate peak by cleavage of one or very few phosphodiester bonds of equal reactivity in the experimental conditions chosen. The large number of subsequent cleavages giving rise to the complex product pattern shown in Figure 4, panel C, indicated the true ribonucleolytic potency of the GCN4 leucine zipper but did not allow determination of individual turnover numbers for these secondary cleavages. Raising the concentration of LZ35 from 15 μM to 56 μM resulted in a roughly linear increase of the catalytic activity excluding cooperativity of active sites in this concentration range. In the conditions of the assays, effects of the conformational dynamics of LZ35 (*16*) on the catalytic activity were also negligible.

Figure 2, panel D, shows that the ribonuclease activity of GCN4 *r*LZ38 was preserved in a synthetic 67-residue fusion protein of the *retro*-peptide with an artificial HIV-1 enhancer-binding peptide (sequence in Figure 1e). Even more interesting was that at nearly cellular salt concentration the ribonuclease activity of c-Jun LZ36 was retained, at least partially, in recombinant full-length 40 kDa, 331-residue c-Jun, a component of the human transcription activation complex AP-1 (Figure 5). It is conceivable that the weak unspecific RNA-degrading activity of GCN4, c-Jun and perhaps other leucine zipper-containing transcription factors contributes to controlling concentration and quality of mRNA and rRNA in the nucleus (*17*). Other natural proteins possessing very low enzymatic activities include the highly specific homing endonuclease PI-SceI from *Saccharomyces cerevisiae* (*18*), DnaK, the *Escherichia coli* Hsp70 molecular chaperone catalyzing

the isomerization of specific peptide bonds (*19*), and many antibodies converting O_2 to H_2O_2 (*20*) and catalyzing ozone formation in bacterial killing and inflammation (*21*). The turnover numbers of these reactions were comparable to that of the ribonuclease reaction of GCN4 leucine zipper. Interestingly, a 30-residue peptide corresponding to a single Cys_2His_2 zinc finger of the human male-associated ZFY transcription factor cleaved single-stranded RNA in the absence of zinc with k_{cat}=0.025 min^{-1} (*22*).

There may be many more non-enzyme proteins possessing weak enzymatic activity for the fine-tuning of their biological function or for unrelated purposes and one may speculate that mutations producing an imbalance between the major non-enzyme and the minor enzyme activity may have negative effects on the organism.

The ribonuclease activity of the *retro*-leucine zippers also raises the question to what extent *retro*-peptides and *retro*-proteins have interesting structural and functional properties (*23*).

ACKNOWLEDGMENT

We thank Bianca Bergamaschi and Vishal Agrawal for excellent experimental assistance.

SUPPORTING INFORMATION AVAILABLE

Mass spectrum of the synthetic 35-residue leucine zipper (LZ35) of GCN4. Kinetic data of the LZ35-catalyzed RNA18 cleavage. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Hehlgans, T., Stolz, M., Klauser, S., Cui, T., Salgam, P., Brenz Verca, S., Widmann,
M., Leiser, A., Städler, K., and Gutte, B. (1993) The DNA binding properties of an artificial 42-residue polypeptide derived from a natural repressor, *FEBS Lett. 315*, 51-55.
2. Städler, K., Liu, N., Trotman, L., Hiltpold, A., Caderas, G., Klauser, S., Hehlgans,
T., and Gutte, B. (1995) Design, synthesis and characterization of HIV-1 enhancer-
binding polypeptides derived from bacteriophage 434 repressor, *Int.J.Peptide Protein*
Res. 46, 333-340.
3. Caderas, G., Klauser, S., Liu, N., Bienz, A., and Gutte, B. (1999) Inhibition of HIV-1
enhancer-controlled transcription by artificial enhancer-binding peptides derived from bacteriophage 434 repressor, *Eur.J.Biochem. 266*, 599-607.
4. Liu, N., Caderas, G., Gutte, B., and Thomas, R.M. (1997) An artificial HIV enhancer-binding peptide is dimerized by the addition of a leucine zipper, *Eur.Biophys.J. 25*, 399-403.
5. Landschulz, W.H., Johnson, P.F., and McKnight, S.L. (1988) The leucine zipper: A
hypothetical structure common to a new class of DNA binding proteins, *Science 240*, 1759-1764.

6. Liu, N., Deillon, C., Klauser, S., Gutte, B., and Thomas, R.M. (1998) Synthesis,
physicochemical characterization, and crystallization of a putative retro-coiled coil,
Protein Sci. 7, 1214-1220.
7. Mittl, P.R.E., Deillon, C., Sargent, D., Liu, N., Klauser, S., Thomas, R.M., Gutte, B.,
and Grütter, M.G. (2000) The retro-GCN4 leucine zipper sequence forms a stable three-dimensional structure, *Proc.Natl.Acad.Sci.U.S.A. 97*, 2562-2566.
8. Richards, F.M., and Wyckoff, H.W. (1971) Bovine Pancreatic Ribonuclease, in *The*
Enzymes (Boyer, P.D., Ed.) 3rd ed., Vol. IV, pp 647-806, Academic Press, New York.
9. Merrifield, R.B. (1963) Solid-phase peptide synthesis. I. The synthesis of a tetrapeptide, *J.Am.Chem.Soc. 85*, 2149-2154.
10. Merrifield, R.B. (1995) Solid-Phase Peptide Synthesis, in *Peptides: Synthesis, Structures, and Applications* (Gutte, B., Ed.) pp 93-169, Academic Press, New York.
11. Carpino, L.A., and Han, G.Y. (1970) The 9-fluorenylmethoxycarbonyl function, a
new base sensitive amino-protecting group, *J.Am.Chem.Soc. 92*, 5748-5749.
12. Borghans, J. A., de Boer, R. J., and Segel, L. A. (1996) Extending the quasi-steady state approximation by changing variables, *Bull. Math. Biol. 58*, 43-63.
13. O'Shea, E.K., Klemm, J.D., Kim, P.S., and Alber, T. (1991) X-ray structure of the

GCN4 leucine zipper: A two-stranded parallel coiled coil, *Science 254*, 539-544.

14. Steyaert, J., Hallenga, K., Wyns, L., and Stanssens, P. (1990) Histidine-40 of ribonuclease T_1 acts as base catalyst when the true catalytic base, glutamic acid-58, is replaced by alanine, *Biochemistry 29,* 9064-9072.
15. Landt, O., Thölke, J., Grunert, H.-P., Saenger, W., and Hahn, U. (1997) Ribonuclease T_1 is active when both catalytic histidines are replaced by aspartate, *Biol.Chem. 378*, 553-558.
16. Nikolaev, Y., and Pervushin, K. (2007) NMR spin state exchange spectroscopy reveals equilibrium of two distinct conformations of leucine zipper GCN4 in solution, *J.Am.Chem.Soc. 129*, 6461-6469.
17. Moore, M.J. (2002) Nuclear RNA turnover, *Cell 108*, 431-434.
18. Wende, W., Grindl, W., Christ, F., Pingoud, A., and Pingoud, V. (1996) Binding, bending and cleavage of DNA substrates by the homing endonuclease PI-SceI, *Nucl.Acid Res. 24*, 4123-4132.
19. Schiene-Fischer, C., Habazettl, J., Schmid, F.X., and Fischer, G. (2002) The hsp70 chaperone DnaK is a secondary amide peptide bond cis-trans isomerase, *Nature Struct.Biol. 9*, 419-424.
20. Wentworth, A.D., Jones, L.H., Wentworth, P., Jr., Janda, K.D., and Lerner, R.A. (2000) Antibodies have the intrinsic capacity to destroy antigens, *Proc.Natl.Acad.Sci.U.S.A. 97*, 10930-10935.
21. Wentworth, P., Jr., McDunn, J.E., Wentworth, A.D., Takeuchi, C., Nieva, J., Jones, T., Bautista, C., Ruedi, J.M., Gutierrez, A., Janda, K.D., Babior, B.M., Eschenmoser, A., and Lerner R.A. (2002) Evidence for antibody-catalyzed

ozone formation in bacterial killing and inflammation, *Science 298*, 2195-2199.

22. Lima, W.F., and Crooke, S.T. (1999) Highly efficient endonucleolytic cleavage of RNA by a Cys_2His_2 zinc-finger peptide, *Proc.Natl.Acad.Sci.U.S.A. 96,* 10010-10015.

23. Juvvadi, P., Vunnam, S., and Merrifield, R.B. (1996) Synthetic melittin, its *enantio*,
retro and *retroenantio* isomers, and selected chimeric analogs: Their antibacterial, hemolytic and lipid bilayer action, *J.Am.Chem.Soc. 118*, 8989-8997.

Supporting Information for

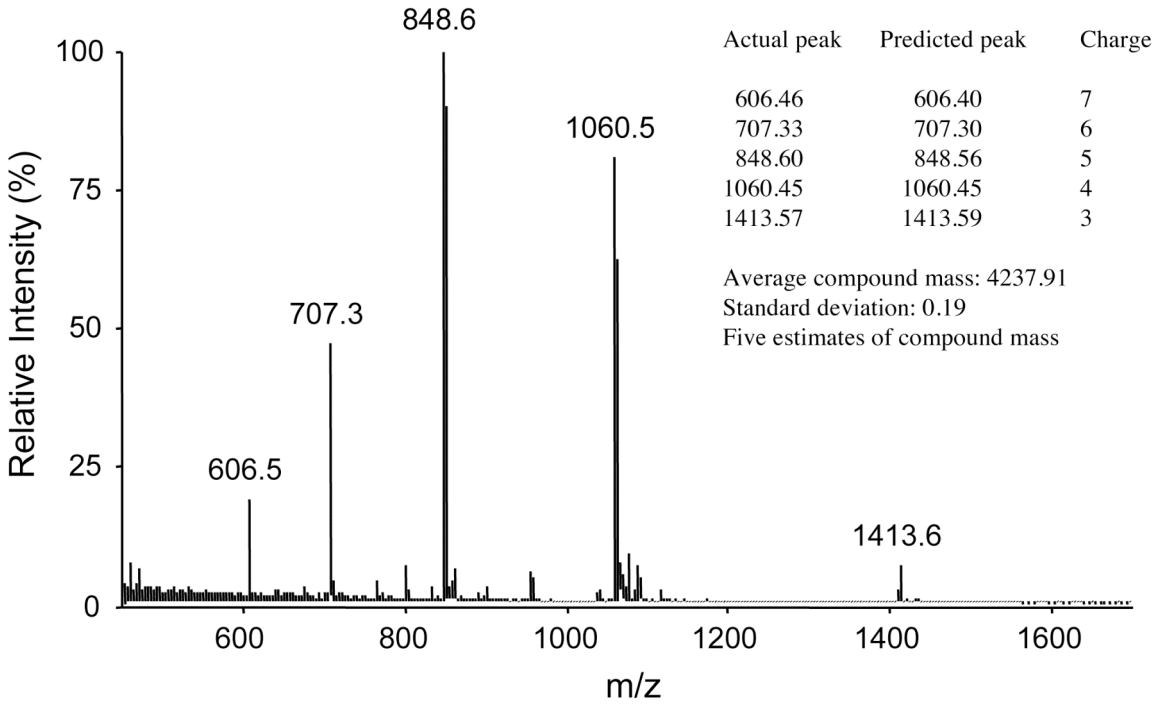
The Leucine Zippers of the Transcription Factors GCN4 and c-Jun Have Ribonuclease Activity

*Christine Deillon, Yaroslav Nikolaev, Stefan R.K. Hoffmann, Laurent Bigler, Sebastian Friess, Renato Zenobi, Konstantin Pervushin, Peter Hunziker, and Bernd Gutte**

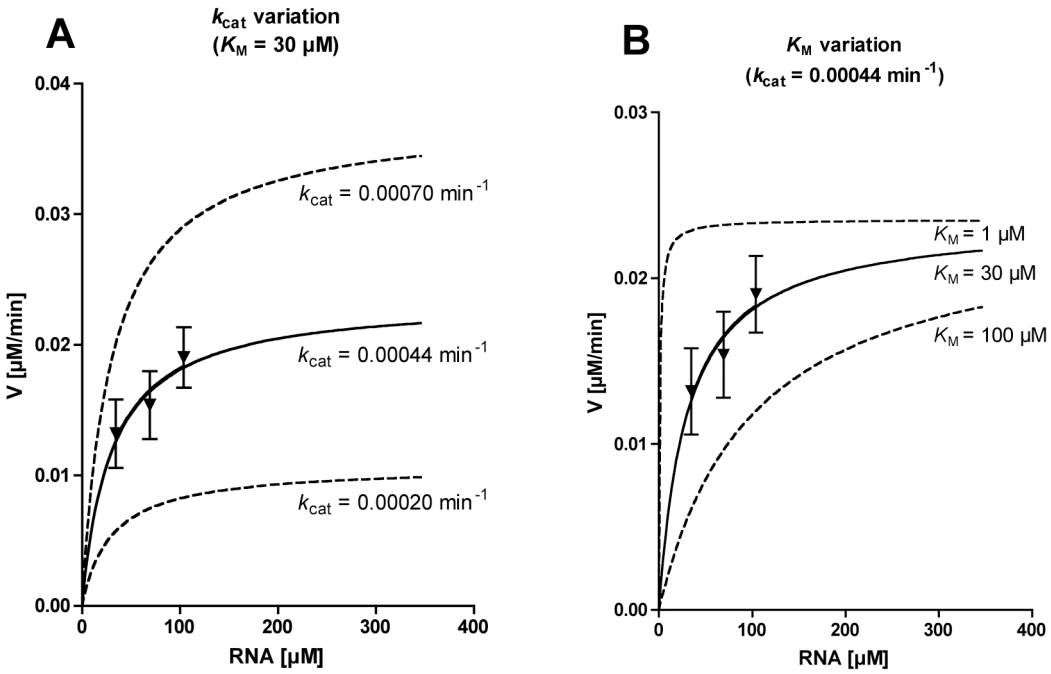
* Biochemisches Institut, Universität Zürich, Tel: +41 44 6355530, E-mail: gutte@bioc.unizh.ch

---

The quality of the synthetic leucine zipper peptides was monitored by mass spectrometry. Samples were diluted with 0.1 % formic acid/50 % acetonitrile and mass spectra were recorded on an API III+ instrument (Sciex, Toronto). Compound masses were calculated using the MacSpec software (Sciex). **Supplementary Figure 1** shows the mass spectrum of GCN4 LZ35 which is representative for the quality of the synthetic peptides used in this study.



Supplementary Figure 1. Mass spectrum of synthetic GCN4 LZ35. Experimental average compound mass and theoretical compound mass were identical (4237.91 Da).



Supplementary Figure 2. Instantaneous velocity values for RNA18 transesterification by LZ35 in relation to theoretical curves defined by the Michaelis-Menten equation (error bars correspond to SE values, n=3). Solid curves represent simulations of the Michaelis-Menten equation ($V = k_{cat} \times [LZ] \times [RNA] / (K_M + [RNA])$) at experimental k_{cat} and K_M values (0.00044 min^{-1} and 30 µM respectively). Dashed curves represent simulations of Michaelis-Menten equation using variable k_{cat} and K_M, illustrating the validity range of the obtained constants. **(A)** Best-fits obtained with constant K_M = 30 µM and varying k_{cat} values, illustrating experimental error range on the k_{cat} scale. **(B)** Best-fits obtained using constant k_{cat} = 0.00044 min^{-1} and K_M as a variable, illustrating experimental error range for the K_M parameter.

Chapter III

NMR spin state exchange spectroscopy reveals equilibrium of two distinct conformations of leucine zipper GCN4 in solution.

J|A|C|S
ARTICLES
Published on Web 05/01/2007

NMR Spin State Exchange Spectroscopy Reveals Equilibrium of Two Distinct Conformations of Leucine Zipper GCN4 in Solution

Yaroslav Nikolaev† and Konstantin Pervushin*,§,†

Contribution from the Laboratory of Physical Chemistry, Swiss Federal Institute of Technology (ETH Zurich), CH-8093 Zurich, Switzerland

Received December 6, 2006; E-mail: kpervushin@ntu.edu.sg

Abstract: The resonance assignment, secondary structure, and dynamic properties of a stable noncoiled coil conformation of the dimerization domain from yeast transcription activation factor GCN4 (Leu zipper; LZ_{GCN4}) are presented. Introduced in this paper, a new line of fully optimized spin state exchange experiments, XYEX-TROSY, applied to $^1H^N$, ^{15}N and $^1H^\alpha$,$^{13}C^\alpha$ moieties, established that in broad range of pH and buffer conditions the classical LZ_{GCN4} coiled coil dimer is in a dynamic equilibrium with another distinct conformation (denoted here as x-form) and enabled complete assignment of the resonances stemming from the x-form. The LZ_{GCN4} x-form is generally less structured in comparison with the classical GCN4-p1 coiled coil, but still retains a structured α-helical central core. The implications for folding properties and biological significance are discussed.

Introduction

Leucine zippers are a protein dimerization domain occurring in many eukaryotic enhancer-type transcription factors, referred to as basic leucine zipper proteins (bZIP)[1] and basic helix−loop−helix leucine zipper proteins (b-HLH-LZ)[2]. Leucine zipper transcription factors have evolved as regulators in many processes that are critical to the function of an organism, from cell metabolism and differentiation[3] to circadian rhythms,[4] memory,[5] and development of organs.[6] These factors are widespread among eukaryotes, with only human genome containing 56 genes encoding proteins with 53 unique bZIP motifs.[7,8]

Because of their simplicity and stability leucine zippers have been widely used as a model system for kinetic and thermodynamic studies of protein folding[9−13] and references therein. In brief, in the common neutral saline phosphate buffer solution folding of the GCN4-p1 peptide is enthalpy-driven and opposed by a loss of entropy.[11] Folding proceeds via one or more stable intermediate states where a "trigger sequence" is proposed to adopt a helical conformation in the individual monomers which might be essential for successful association[14,15] resulting in a coiled coil structure[16,17] with some local structural transitions in the folded protein.[18] GCN4-p1 exhibits a complex temperature unfolding reaction consisting of several stages[12] featuring various degrees of structural losses with increasing temperature. Intra- and interhelical salt bridges stabilize GCN4-p1 but are not critical for maintenance of its 3D fold[10,19,20] and do not accelerate the folding.[21] Intriguingly as low as 0.1 mM concentration of SDS can initiate dissociation of the homodimer while retaining the secondary structure of the individual helices,[22] apparently interfering with both electrostatic and hydrophobic stabilizing interactions.

§ Current Address: School of Biological Sciences, Nanyang Technological University, Nanyang Drive 60, Singapore 637551.
† Current address: Biozentrum of University Basel, Klinlegbergstrasse 70, CH-4056, Basel, Switzerland.

(1) Landschulz, W. H.; Johnson, P. F.; McKnight, S. L. *Science* **1988**, *240*, 1759−64.
(2) Murre, C.; McCaw, P. S.; Vaessin, H.; Caudy, M.; Jan, L. Y.; Jan, Y. N.; Cabrera, C. V.; Buskin, J. N.; Hauschka, S. D.; Lassar, A. B. *Cell* **1989**, *58*, 537−44.
(3) Darlington, G. J.; Wang, N.; Hanson, R. W. *Curr. Opin. Genet. Dev.* **1995**, *5*, 565−70.
(4) Yamaguchi, S.; Mitsui, S.; Yan, L.; Yagita, K.; Miyake, S.; Okamura, H. *Mol. Cell. Biol.* **2000**, *20*, 4773−81.
(5) Sanyal, S.; Sandstrom, D. J.; Hoeffer, C. A.; Ramaswami, M. *Nature* **2002**, *416*, 870−74.
(6) Eferl, R.; Sibilia, M.; Hilberg, F.; Fuchsbichler, A.; Kufferath, I.; Guertl, B.; Zenz, R.; Wagner, E. F.; Zatloukal, K. *J. Cell Biol.* **1999**, *145*, 1049−61.
(7) Tupler, R.; Perini, G.; Green, M. R. *Nature* **2001**, *409*, 832−33.
(8) Vinson, C.; Myakishev, M.; Acharya, A.; Mir, A. A.; Moll, J. R.; Bonovich, M. *Mol. Cell. Biol.* **2002**, *22*, 6321−35.
(9) Lovett, E. G.; D'Avignon, D. A.; Holtzer, M. E.; Braswell, E. H.; Zhu, D.; Holtzer, A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 1781−85.
(10) Kammerer, R. A.; Jaravine, V. A.; Frank, S.; Schulthess, T.; Landwehr, R.; Lustig, A.; Garcia-Echeverria, C.; Alexandrescu, A. T.; Engel, J.; Steinmetz, M. O. *J. Biol. Chem.* **2001**, *276*, 13685−88.
(11) Bosshard, H. R.; Durr, E.; Hitz, T.; Jelesarov, I. *Biochemistry* **2001**, *40*, 3544−52.
(12) Dragan, A. I.; Privalov, P. L. *J. Mol. Biol.* **2002**, *321*, 891−908.
(13) Wang, T.; Lau, W. L.; DeGrado, W. F.; Gai, F. *Biophys. J.* **2005**, *89*, 4180−7.
(14) Kammerer, R. A.; Schulthess, T.; Landwehr, R.; Lustig, A.; Engel, J.; Aebi, U.; Steinmetz, M. O. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 13419−24.
(15) Frank, S.; Lustig, A.; Schulthess, T.; Engel, J.; Kammerer, R. A. *J. Biol. Chem.* **2000**, *275*, 11672−7.
(16) O'Shea, E. K.; Klemm, J. D.; Kim, P. S.; Alber, T. *Science* **1991**, *254*, 539−44.
(17) Saudek, V.; Pastore, A.; Morelli, M. A.; Frank, R.; Gausepohl, H.; Gibson, T. *Protein Eng.* **1991**, *4*, 519−29.
(18) Holtzer, M. E.; Bretthorst, G. L.; d'Avignon, D. A.; Angeletti, R. H.; Mints, L.; Holtzer, A. *Biophys. J.* **2001**, *80*, 939−51.
(19) Spek, E. J.; Bui, A. H.; Lu, M.; Kallenbach, N. R. *Protein Sci.* **1998**, *7*, 2431−7.
(20) Zeng, X.; Herndon, A. M.; Hu, J. C. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 3673−8.
(21) Ibarra-Molero, B.; Zitzewitz, J. A.; Matthews, C. R. *J. Mol. Biol.* **2004**, *336*, 989−96.

10.1021/ja0685295 CCC: $37.00

Although GCN4-p1 has been extensively studied by NMR,[9,10,17,18,23−25] to our knowledge all these studies were conducted at high protein concentrations between 1.4 mM to 6 mM and mainly at neutral pH. However even in the first structural NMR study[17] it was observed that a set of unidentified NMR resonances stemming from GCN4-p1 appears at lower protein concentration. This urged us to conduct an NMR-based structure determination at ca. 6−8 mM of GCN4-p1 where the equilibrium is strongly shifted toward the classical coiled coil state.[16]

Here we report our findings that at concentrations below 0.5 mM in acidic and up to neutral pH conditions in TRIS ([1,3-dihydroxy-2-(hydroxymethyl)propan-2-yl]azanium), NaPi, acetate, and HEPES (2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid) buffers a conventional coiled coil state of the LZ_{GCN4} is in dynamic equilibrium with a distinctively different conformational state of the same peptide. Moreover, at pH = 3.2 and peptide concentrations below 150μM this conformation dominates the equilibrium. To characterize this dynamic equilibrium a novel pseudo-four-dimensional fully sensitivity enhanced spin-state exchange TROSY experiment was developed, together with its two-dimensional predecessors[26,27] providing evidence that the two conformation states of LZ_{GCN4} are exchanging with the rate of $16.98 \pm 0.96\ s^{-1}$, which coincides with the interconversion rate of two folded forms of GCN4-p1 found by $^{13}C^{\alpha}$ NMR[18] and is by 3 orders of magnitude slower than the main folding reaction of LZ.

Results

Sample Conditions and Assignment of NMR Resonances. To investigate conformational equilibrium in Leu-zipper dimerization domain a 36 residue uniformly ^{15}N, ^{13}C-labeled recombinant LZ_{GCN4} peptide was expressed in *E. coli* as a cleavable fusion with bacteriophageλ protein D bearing N-terminal $(His)_6$ tag. Except for the N-terminal Gly a complete assignment of $^{1}H^{N}$, $^{1}H^{\alpha}$, $^{1}H^{\beta}$, ^{15}N, $^{13}C'$, $^{13}C^{\alpha}$, and $^{13}C^{\beta}$ resonances of LZ_{GCN4} was obtained at the peptide concentration of 1 mM and pH 7.1 using standard NMR techniques confirming the assignment of ^{1}H and ^{13}C resonances reported earlier.[9,17] At lower peptide concentrations, a second distinct set of cross-peaks in [^{1}H, ^{15}N]-HSQC spectra (see Figure 1A) is evident, which we attributed to yet another, so far uncharacterized, conformation of LZ_{GCN4}. For the sake of convenience we dub this conformation of LZ_{GCN4} as "x-form" to distinguish it from the known coiled-coil conformation, designated here as "cc-form".

Analysis of the cc-form versus x-form population equilibrium based on [^{1}H, ^{15}N]-HSQC spectra revealed that the x-form dominates at low peptide concentrations in the acidic pH, reaching as high as 83% at 15 μM and pH 3.2 (Figure 2A). Intriguingly, the x-form is still detectable by NMR at neutral pH reaching 10% of the total population at the concentration of the peptide monomer of 250 μM and pH 5.5, slightly decreasing to 8% population at the concentration of 30 μM and pH 7.1 (Figure 2A). On the basis of these observations the conditions for more detailed NMR studies were chosen in such a way that both conformations are approximately equally populated (concentration of the peptide monomer of 250 μM and pH 3.2). At these conditions 3D HNCA, HNCACB, CBCA(CO)NH, HBHA(CO)HN, and ^{13}C, ^{15}N-resolved NOESY experiments were recorded. Despite the availability of assignment of the cc-form, these triple resonance experiments were proven not to be sufficient for unequivocally assigning all of the backbone resonances of the x-form. This problem has been pinpointed to the presence of conformational exchange between two forms with the characteristic exchange times coinciding with the magnetization transfer periods in triple resonance experiments resulting in an often intractable mixture of cross-peaks in the 2D [^{1}H, ^{15}N] strips taken along the ^{13}C dimension. This called for complementing the standard set of NMR experiments with an experiment selectively correlating resonances from two exchanging forms with highest sensitivity and without interference with other coherence transfer pathways.

Spin States Exchange Experiments with LZ. Figure 3 shows the pulse sequence for the pseudo-four-dimensional spin-state exchange XYEX-TROSY experiment, which is a fully sensitivity enhanced, transverse relaxation optimized NMR experiment, deployed here to unequivocally correlate $^{1}H^{N}$, ^{15}N, and $^{1}H^{\alpha}$, $^{13}C^{\alpha}$ cross-peaks stemming from two exchanging forms of LZ_{GCN4}. Mixing of the coherences in the transverse plane while preserving the individual spin-states enables implementation of the echo−anti-echo quadrature detection method[28] in all indirectly acquired spectral dimensions resulting in the fully sensitivity-enhanced 4(1)D experiment. The key feature of the proposed experiment is a convoluted evolution of the zero- and double-quantum coherencies originating from the exchanging conformations before they are mixed in the transverse plane and detected via the ST2-PT[29] or ZQ/DQ TROSY[30] schemes.

Figure 4A shows 2D [DQ/ZQ, $^{1}H^{N}$] strips taken at the corresponding ^{1}H and ^{15}N chemical shifts of the observed cross-peaks in the fingerprint [^{1}H, ^{15}N] TROSY spectra. With the use of XYEX-TROSY identification of the pairs of exchanging resonances stemming from the same amide group (or $^{1}H^{\alpha}-^{13}C^{\alpha}$ moieties, see Figure 4B) is straightforward and requires the presence of only one cross-peak usually of the most populated conformation. This is because all four chemical shifts required to link exchanging spin-systems can be inferred from the positions of a single exchange cross-peak in ZQ and DQ subspectra of the same XYEX-TROSY experiments. The opposite situation is encountered in the analysis of 3D ^{15}N or ^{13}C-resolved NOESY spectra,[31] where the presence of a pair of "forward" and "return" cross-peaks is mandatory to unequivocally link interacting spin systems.

For large proteins the DQ subspectrum of XYEX-TROSY might be significantly more sensitive comparing to the ZQ subspectrum due to the TROSY effect.[32] Even for 8.6 kDA LZ_{GCN4} dimer, a comparison between these two subspectra

(22) Meng, F. G.; Zeng, X. G.; Hong, Y. K.; Zhou, H. M. *Biochimie* **2001**, *83*, 953−56.
(23) Lumb, K. J.; Kim, P. S. *Science* **1995**, *268*, 436−9.
(24) Weiss, M. A.; Ellenberger, T.; Wobbe, C. R.; Lee, J. P.; Harrison, S. C.; Struhl, K. *Nature* **1990**, *347*, 575−8.
(25) Holtzer, M. E.; Lovett, E. G.; d'Avignon, D. A.; Holtzer, A. *Biophys. J.* **1997**, *73*, 1031−41.
(26) Wider, G.; Neri, D.; Wuthrich, K. *J. Biomol. NMR* **1991**, *1*, 93−98.
(27) Sprangers, R.; Gribun, A.; Hwang, P. M.; Houry, W. A.; Kay, L. E. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 16678−83.
(28) Kay, L. E.; Keifer, P.; Saarinen, T. *J. Am. Chem. Soc.* **1992**, *114*, 10663−5.
(29) Pervushin, K.; Wider, G.; Wuthrich, K. *J. Biomol. NMR* **1998**, *12*, 345−48.
(30) Pervushin, K.; Wider, G.; Riek, R.; Wuthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9607−12.
(31) Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G.; Skelton, N. J. *Protein NMR Spectroscopy: Principles and Practice*; Academic Press: New York, 1996.
(32) Pervushin, K.; Riek, R.; Wider, G.; Wuthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12366−71.

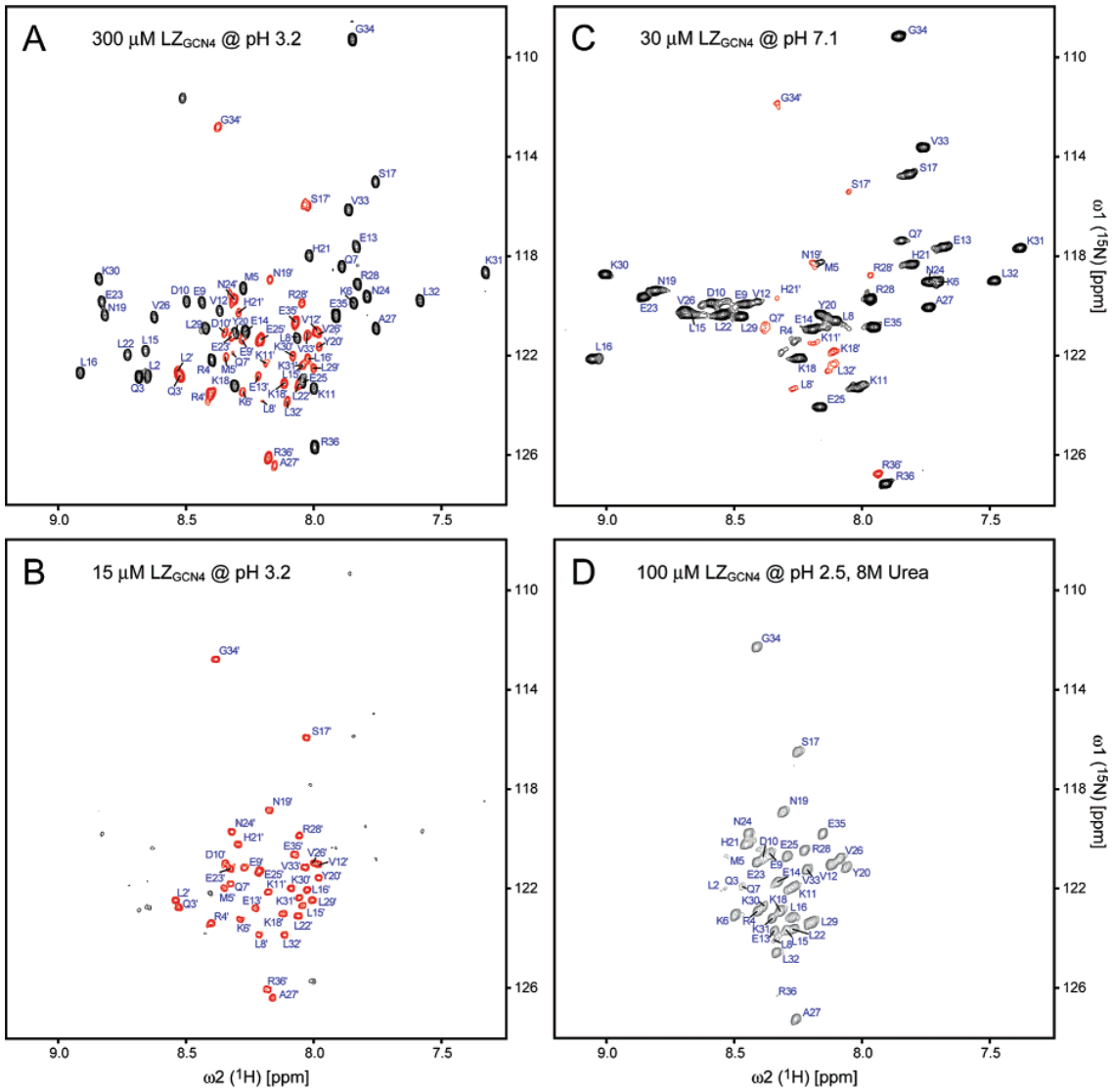


Figure 1. Equilibrium of two distinct LZ_{GCN4} conformations. [^{15}N-^{1}H]-HSQC NMR spectra of ^{15}N-^{13}C-labeled LZ_{GCN4} recorded at 37 °C and 600 MHz: (A) 300 μM and (B) 15 μM peptide (monomer concentrations) in 50 mM D_3-acetate, 40 mM KCl, 10% D_2O, pH 3.2; (C) 30 μM peptide in 20 mM D_{18}-HEPES, 80 mM KCl, 10% D_2O, pH 7.1; (D) 100 μM peptide in 8 M urea, pH 2.5. Assignment corresponds to $LZ_{cc-dimer}$ (black contours), LZ_{x-form} (red contours), $LZ_{unfold\ monomer}$ (gray contours, (D)).

(Figure 4) reveals reduced S/N in the ZQ subspectrum, with some exchange cross-peaks disappearing into the spectral noise. In this situation the only DQ subspectrum can be used to establish exchanging spin systems via search of forward and return cross-peak as it is in conventional 3D NOESY. Alternatively, ZQ evolution can be observed using the TROSY pathway in a separate experiment, which can be obtained from the one shown in Figure 3 by inverting the phase of the last ^{15}N 90° pulse.

Analysis of the [^{1}H, ^{15}N]-XYEX-TROSY strips along DQ and ZQ dimensions corresponding to the cross-peaks in [^{1}H, ^{15}N]-HSQC spectrum for the LZ_{GCN4} molecule yielded in assignment of all 35 pairs of the amino acid residues under exchange. [^{1}H, ^{15}N]-XYEX-TROSY strips correlating exchanging cross-peaks can also serve as an additional source of information to resolve resonances completely overlapping in the HSQC spectrum. We found that residues K31′/L15′, V12′/V26′, L2′/Q3′, and L22′/E25 exhibiting this kind of spectral degeneracy nonetheless can be assigned using the set of exchange peaks in the multiple-quantum dimension.

The [$^{1}H^{\alpha}$,$^{13}C^{\alpha}$]-XYEX-TROSY experiment showed slightly decreased S/N ratio in comparison with the [$^{1}H^{N}$,^{15}N]-XYEX-TROSY owing to partial dephasing of magnetization during all chemical shift evolution periods as well as the mixing period owing to the passive $^{1}J_{C\alpha C\beta}$ and $^{n}J_{H\alpha Hi}$ with $n > 2$ couplings. Analysis of the [$^{1}H^{\alpha}$,$^{13}C^{\alpha}$]-XYEX-TROSY spectra yielded resolved correlations of $^{13}C^{\alpha}$ resonances for 19 residues (with others being overlapped). As an alternative to the mixing of exchanging coherences in the transverse plane, we tested mixing in the longitudinal direction. ZEX-TROSY experiments (Supporting Information) developed to achieve this goal showed nonetheless a similar *S*/*N* ratio for the exchange cross-peaks, confirming the functionality of the proposed mixing strategy employed in XYEX-TROSY at least for LZ_{GCN4}.

Since the XYEX-TROSY series of experiments provides the maximal theoretically achievable sensitivity in the pseudo-4D

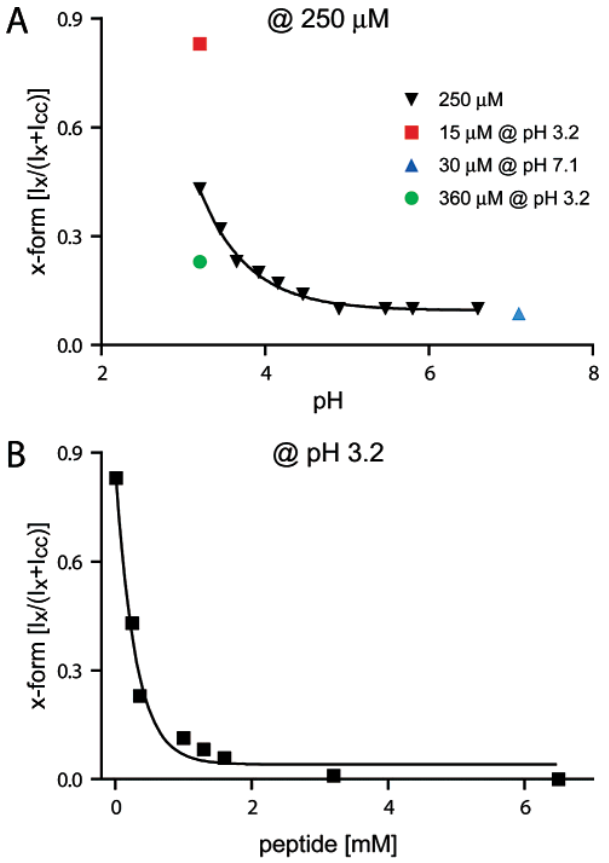


Figure 2. The pH (A) and concentration (B) dependences of LZ_{GCN4} x-form population. The x-form populations were determined from the relative volumes of [^{1}H-^{15}N]-HSQC peaks corresponding to each conformation: $I_X/(I_X + I_{CC})$ where I_X and I_{CC} refer to the intensities of peaks of the x-form and coiled coil, respectively. Peak intensities for calculations are taken as averages of 10 peaks for each conformation at each data point.

spectrum, it demonstrates practical advantages over the existing 3D versions developed earlier.[26,27,33] The maximum sensitivity requirement together with the higher resolution achieved in pseudo-4D spectra are especially important in studies of LZ_{GCN4} at low concentrations used to populate the x-form exhibiting significant spectral degeneracy. Although visible even for this small protein, the TROSY effect is of less importance for sensitivity improvement. Here the TROSY-based polarization transfer schemes were used as a convenient tool to preserve spin states during coherence transfers. Of course, for larger proteins the TROSY effect might become the most prominent source of sensitivity and resolution improvements.

To quantitatively characterize the exchange process under observation a two-dimensional version of the spin-state longitudinal exchange experiment[27] was employed (Figure 5), yielding an interconversion rate of 16.98 ± 0.96 (SE) s^{-1} for two forms of LZ_{GCN4}. This rate strongly coincides with the interconversion rate of two folded forms of GCN4-p1 found by $^{13}C^{\alpha}$ NMR[18] and, most intriguingly, is by 3 orders of magnitude slower than the main folding reaction of LZ_{GCN4}, as determined from the stopped-flow circular dichroism studies.[34]

(33) Farrow, N. A.; Zhang, O.; Forman-Kay, J. D.; Kay, L. E. *J. Biomol. NMR* **1994**, *4*, 727−34.
(34) Zitzewitz, J. A.; Bilsel, O.; Luo, J.; Jones, B. E.; Matthews, C. R. *Biochemistry* **1995**, *34*, 12812−9.

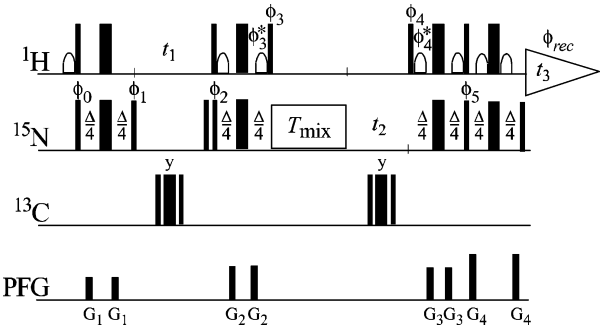


Figure 3. Scheme of 4(1)D XYEX-TROSY, pseudo-four-dimensional spin-state exchange TROSY experiment with mixing of spin coherencies in the transverse plane utilizing an entire steady-state Boltzmann thermal equilibrium magnetization and featuring sensitivity enhancement for the phase sensitive detection in all three spectral dimensions suitable for two coupled $^1/_2$ spin systems. Narrow and wide black bars indicate nonselective $\pi/2$ and π rf-pulses applied with the phase x unless indicated otherwise. Water saturation is minimized by returning the water magnetization to the $+z$-axis before data acquisition by the use of water selective 90° rf-pulses shown as open shapes on the ^{1}H channel. The time period Δ is set to $1/J_{HN}$ and is 5.78 ms for the backbone amide moieties. The ^{1}H spin-state selective ^{15}N coherences stemming from kinetically exchanging spin systems are mixed during a T_{mix} period using a planar mixing scheme IICT-4, which is a composite pulse sequence specifically designed for isotropic mixing purpose in TOCSY experiments (see Materials and Methods).[61] The IICT-4 scheme is RRRRRRRR with R = 38(0), 112.6(60.5), 205.9(68.6), 256.9(280.1), 101.9($\overline{5.2}$), 265.5(281.6), 242.6(72.6), 66.6(66.9), 44.9(0), where numbers represent rf-pulses with the corresponding flip angle and the rf-phase (in brackets), respectively, and $\underline{R}$ is phase inversion of R. In this case the durations and strengths of the pulsed magnetic field gradients (PFG) applied along the z-axis are selected as (G_1) 500 μs, 19 G/cm; (G_2) 500 μs, 15 G/cm; (G_3) 900 μs, 32 G/cm; (G_4) 1 ms, 50 G/cm. Two datasets, S_{ZQ+DQ} and S_{ZQ-DQ}, are acquired using the phases $\phi_0 = \{x, -x\}$; $\phi_1 = \{y, -y\}$; $\phi_2 = -x$; $\phi_3 = -y$; $\phi^*_3 = y$; $\phi_4 = y$; $\phi^*_4 = -y$; $\phi_5 = y$; $\phi_{rec} = \{-y, y\}$ for S_{ZQ+DQ} and $\phi_0 = \{-y, y\}$; $\phi_1 = \{x, -x\}$; $\phi_{rec} = \{x, -x\}$ for S_{ZQ-DQ}. For both S_{ZQ+DQ} and S_{ZQ-DQ} datasets the quadrature detection in t_1 and t_2 dimensions is achieved by the echo−anti-echo method.[28,64] For the dataset S_{ZQ+DQ} the anti-echo signal in t_1 dimension is obtained by inversion of the phases ϕ_2, ϕ_3, and ϕ^*_3, and the anti-echo signal in t_2 dimension is obtained by inversion of the phases ϕ_2, ϕ_3, ϕ^*_3, ϕ_4, ϕ^*_4, and ϕ_5. For the dataset S_{ZQ-DQ} the anti-echo signal in t_1 dimension is obtained by inversion of the phases ϕ_1, ϕ_2, ϕ_3, and ϕ^*_3, and the anti-echo signal in t_2 dimension is obtained by inversion of the phases ϕ_2, ϕ_3, ϕ^*_3, ϕ_4, ϕ^*_4, and ϕ_5. The processing of the datasets is described in Materials and Methods and the listings of pulse programs for Bruker Avance spectrometers are available in Supporting Information. The rf-pulses on the ^{1}H, ^{15}N, and ^{13}C nuclei are centered at 4.7, 118, and 110 ppm, respectively.

Secondary Structure of the x-Form. A complementary set of experiments (HNCA, HNCACB, CBCA(CO)NH, ^{1}H-^{15}N NOE) was carried out with LZ_{GCN4} in 8 M urea, yielding the assignment (Figure 1D) and dynamic characterization (Figure 6B) of the LZ_{GCN4} truly unfolded state. These data were used as a reference point to unequivocally exclude the possibility of the observed x-form species being unfolded LZ_{GCN4} monomer. Additionally, a set of HNCA, CBCA(CO)NH, and ^{1}H-^{15}N NOE experiments was recorded in the conditions of the x-form dominating the equilibrium (25 μM monomer concentration at pH 3.2). These spectra were used to obtain unbiased by exchange input data for characterization of secondary structure and dynamics of the x-form.

Secondary chemical shifts of the cc- and x-forms (Figure 6C-E) were calculated using the random coil chemical-shift values of the authentic LZ_{GCN4} sequence in 8 M urea. A comparison of secondary chemical shift data of the cc-form versus the x-form points out that the x-form is indeed considerably less structured. However, the appearance of the δC_{α} and δC_{β} plots

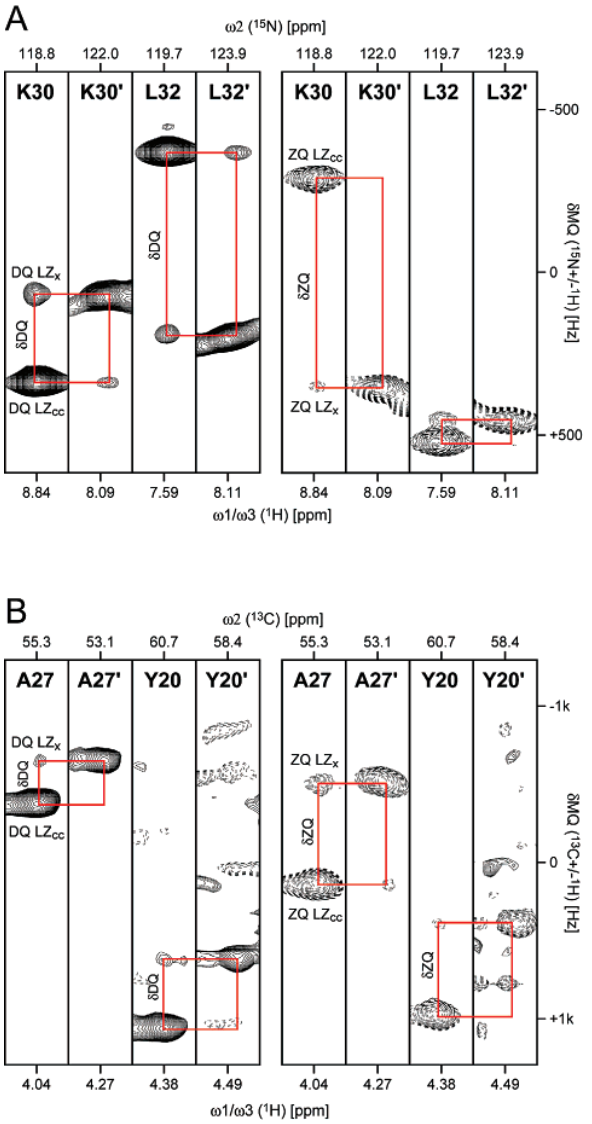


Figure 4. Correlation of exchanging conformations by XYEX-TROSY. Strip plots for K30, L32 (A) and A27, Y20 (B) residues from DQ and ZQ subspectra of the [$^1H^N$, ^{15}N]-XYEX-TROSY (A) and [$^1H^\alpha$, $^{13}C^\alpha$]-XYEX-TROSY (B) NMR experiments. Spectra recorded using 360 μM ^{15}N-^{13}C-labeled LZ_{GCN4}, at 37 °C and 600 MHz.

indicates the presence of residual α-helical structure. This residual alpha-helical propensity is mainly confined to two clusters around valine residues 12 and 26, localized in the "a" positions of leucine zipper heptad repeats[1] and contributing to the hydrophobic lock of the zipper structure. The same Val12 and Val26 residues also show the strongest helical secondary C_α shifts in the coiled coil conformation.

The observation of the steady-state heteronuclear ^1H-^{15}N NOEs (Figure 6B) indicates that the x-form is significantly less structured when compared to the LZ_{GCN4} coiled coil dimer. Nevertheless, it still retains a structural core localized to the residues Lys11 to Leu29 representing two LZ heptad repeats out of four total. At the same time the residues at the N- and C-termini exhibit strongly negative HNOE values, which is again indicative of low-spatial restriction and enhanced mobility in a nanosecond time scale. Noteworthy, x-form looks considerably more restrained in terms of the dynamics than the unfolded

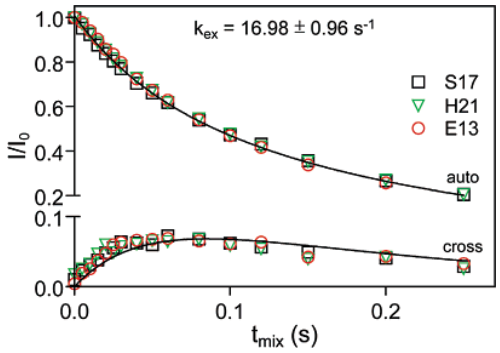


Figure 5. Characterization of exchange process. Longitudinal (H_zN_z) exchange spectroscopy recorded for LZ_{GCN4} at 37 °C. The decay of autopeaks and buildup of cross-peaks is illustrated, where I/I_0 is the intensity of a given correlation normalized to the maximum intensity of the appropriate autopeak at the zero mixing time. Experimental data points correspond to the residues Ser17 (squares), Glu13 (circles), and His21 (triangles). The solid lines correspond to the best fits of the data (all six curves fitted simultaneously) as described in Sprangers et. al.[27]

monomer. This difference shall be much more prominent if one considers the effect of 8 M urea on the viscosity of solution. Additionally the observed regular patterns of $^1H^N/^1H^N$ $d_{NN}(i,i+1)$ NOEs confined to the same central region of the x-form (e.g., residues 11−13, 15−16, 19−22, and 24−26) as well as $d_{\beta N}(i,i+1)$ NOEs (residues 9−13 and 15−29) (data not shown) are consistent with the presence of α-helical structure,[35] albeit this should be interpreted with caution since unstructured peptides may also show sequential amide protons contact patterns. The detailed structure determination of the x-form is delegated to future work.

Discussion

NMR studies of LZ_{GCN4} at low concentrations and broad range of pH and buffer conditions, revealed the presence of a second, thermodynamically stable conformation, designated as the x-form, which so far has escaped structural characterization. Introduced in this paper, a new line of fully optimized spin state exchange experiments, XYEX-TROSY, applied to $^1H^N$, ^{15}N and $^1H^\alpha$, $^{13}C^\alpha$ moieties, established that the x-form is in a dynamic equilibrium with the classical coiled coil conformation of the LZ_{GCN4} and enabled complete assignment of the resonances stemming from the x-form. This equilibrium is evident not only in ^{15}N-labeled recombinant LZ_{GCN4}, but also in a chemically synthesized LZ_{GCN4} peptide (e.g., as $^1H^N/^1H^N$ cross-peaks in [^1H, ^1H]-TOCSY spectra), as well as in the full-length recombinant bZIP motif of GCN4. This compels us to attribute this equilibrium to intrinsic properties of the LZ_{GCN4} sequence.

In fact, concentration dependent conformational equilibrium of the classical coiled coil LZ_{GCN4} conformation with some other species was noted as early as the first homonuclear NMR structural study of GCN4-p1,[17] and later in the ^{13}C studies of GCN4-p1[18] and its GCN4-lzK modification.[9] However, in all cases the x-form was considered as being either unfolded monomer or an intractable ensemble of (artificial) folding intermediates and no further investigation or characterization of the detected species was done.

The LZ_{GCN4} x-form appears to be a stable folding intermediate that is considerably populated under a wide range of conditions.

(35) Wuthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.

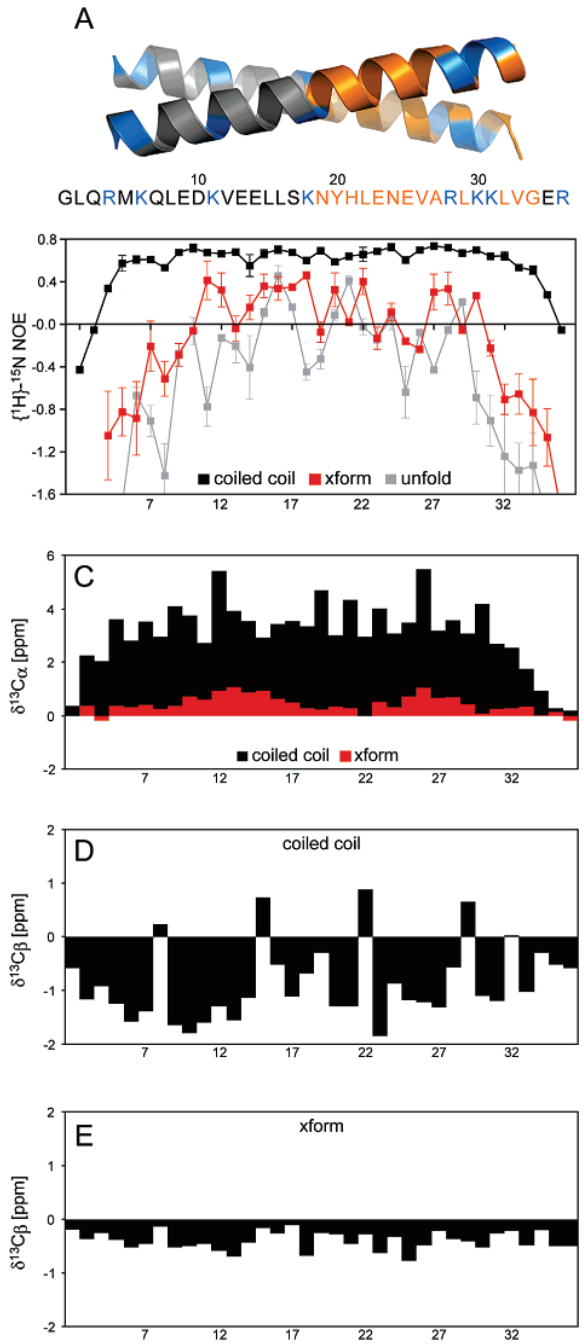


Figure 6. Preliminary structural characterization of the LZ_{GCN4} x-form. (A) 3D structure of the 31 residue LZ_{GCN4} coiled coil dimer (pdb code 2zta[16]) aligned with the sequence of the LZ_{GCN4} peptide used in the current study. The triggering sequence[14] is marked with orange; basic amino acid residues, capable of generating side-chain repulsive interactions in acidic pH, are marked with blue. (B) Heteronuclear [^{1}H-^{15}N] NOEs of cc-form (black squares), x-form (red squares), and urea-unfolded form (gray squares) measured respectively at 360 μM, 25 μM, and 100 μM of LZ_{GCN4} at pH 3.2 and 37 °C. Secondary chemical shifts for $C^{\alpha}_{cc-form}$ (black bars) and C^{α}_{x-form} (red bars) (C), $C^{\beta}_{cc-form}$ (D), and C^{β}_{x-form} (E), resonances calculated relative to the random coil chemical-shift values of the urea-unfolded form.

Notwithstanding its generally less structured appearance in comparison with the classical GCN4-p1 coiled coil, this conformation still retains a structured α-helical central core (Results section and Figure 6). Consistent with the possibility of dimeric nature of the x-form are observations that while GCN4-p11-33 (fragment of GCN4-p1 comprising residues 11 to 33) is monomeric, slightly longer GCN4-p8-33 is predominantly dimeric.[36] Data from other research groups indicate that LZ_{GCN4} folding intermediate contains a considerable amount of preformed helical structure already at the point when the collision of monomer chains occurs.[14,37−39] Experimental data of Zitzewitz and co-workers;[38] utilizing the φ-analysis[40] to assess the effect of mutations on the structure of the transition state, indicates that about one-third of the LZ_{GCN4} molecule (approximately three turns of the α-helix) may be structured in the transition state. Their data provide an evidence for the role of a preformed helix in a transition state leading to the natively folded dimer.

Interestingly, C_β secondary chemical shifts of all four Leu residues at the coiled coil d-positions forming the hydrophobic interface of the dimer, turn out to be positive in contrary to δC_β of all other residues (Figure 6D). This indicates that the location of C_βs proximally to the dimer interface makes them quite sensitive to the changes in the hydrophobic interactions. In contrast, the residues at the a-positions of the coiled coil, which are also involved in the hydrophobic interface, do not show the same effect. Elimination of denoted positive C_β secondary chemical shifts of Leu residues at the d-positions of the x-form (Figure 6E) comparing to the coiled coil structure shows that there are substantial changes in the knobs-into-holes packing characteristic to the coiled coil dimer, implying monomeric structure of the x-form. This assumption goes in line with strong concentration dependence of the x-form versus cc-form populations. However, currently we have insufficient data to make unequivocal conclusions about the monomeric versus dimeric nature of the x-form, and its exact 3D arrangement still awaits characterization.

Kammerer et al.[14] has reported the trigger sequence (autonomous folding unit) essential for oligomerization of dimeric coiled coils. In the GCN4 transcription factor this sequence is represented by third and fourth heptad repeats of the leucine zipper domain. From the current set of NMR data on the x-form including steady-state heteronuclear ^{1}H-^{15}N NOEs, $^{13}C_\alpha$ and $^{13}C_\beta$ secondary chemical shifts (Figure 6), and patterns of sequential NOE connectivities (Results section), the presence of the triggering sequence with higher α-helical content proposed by Kammerer and colleagues is not evident. It rather appears that there are two triggering sites localized proximally to the valine residues in the zipper "a" positions 12 and 26 (at the boundaries of first-second and third-fourth heptad repeats). Alternatively current NMR data would indicate that x-form represents a stable folding intermediate at the native site of the folding barrier, while the triggering sequence of more structured C-terminal heptads exists at the level of a LZ_{GCN4} random coil monomer before the chain collision event.

Additionally, it was shown that α-helical content in the peptide corresponding to this trigger sequence strongly depends on the pH, suggesting existence of repulsive interactions between

(36) Lumb, K. J.; Carr, C. M.; Kim, P. S. *Biochemistry* **1994**, *33*, 7361−7.
(37) Myers, J. K.; Oas, T. G. *J. Mol. Biol.* **1999**, *289*, 205−9.
(38) Zitzewitz, J. A.; Ibarra-Molero, B.; Fishel, D. R.; Terry, K. L.; Matthews, C. R. *J. Mol. Biol.* **2000**, *296*, 1105−16.
(39) Moran, L. B.; Schneider, J. P.; Kentsis, A.; Reddy, G. A.; Sosnick, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 10699−704.
(40) Serrano, L.; Matouschek, A.; Fersht, A. R. *J. Mol. Biol.* **1992**, *224*, 805−18.

side chains of basic amino acid residues, and indicating the importance of electrostatic interactions for the overall helix stability.[14] These data can be used to indirectly interpret the pH dependence of the x-form versus the cc-form at the equilibrium. The GCN4p16-31 peptide used by Kammerer and colleagues incorporates just third and fourth heptad repeats of GCN4, and contains only 3 basic amino acid residues total (corresponding to R28, K30, K31 in LZ_{GCN4}) within these two heptad repeats. Since all of these residues are located within the fourth repeat, the third repeat's stability must be less dependent on electrostatics. For the LZ_{GCN4} peptide used in our study the content of putatively destabilizing basic amino acid residues is 8 per 4 heptad repeats, thus making the sequence more sensitive to acidic pH. Indeed the estimated LZ_{GCN4} net charge at pH 7.1 is +1, while at pH 3.2 +8. This implies a feasibility of a coiled coil to the x-form transition being induced by these electrostatic repulsions. Interestingly most of the basic amino acid residues are located in the flanking regions of the LZ_{GCN4} sequence, with the central stretch of 16 residues containing only one, K18, in the middle (Figure 6A). Given distribution of positively charged residues coincides with the x-form ^1H-^{15}N NOEs and δC_α value patterns indicating a structured central core with two sites inclined to α helix.

Beyond the question of the x-form monomeric versus dimeric nature, is the issue of possible biological relevance of this conformation. The x-form population increases at low-peptide concentrations, which is relevant in vivo. On one hand favored by an acidic environment, on the other it is still detectable at neutral pH values. Although, the acidic buffering conditions may be considered as non-natural, one can expect a local pH decrease in the vicinity of nucleic acid polymers in the cell nucleus. Taking into account the presence of the x-form in GCN4 basic region−leucine zipper sequence (not only in isolated LZ domain), as well as the wide range of conditions where the x-form is populated, we propose that this structure is not just an off-pathway folding intermediate and not simply an unstructured monomer as in the case of LZ_{GCN4} fragments[36] or mutant sequences[38] but could be biologically relevant.

An attractive hypothesis to survey is the potential involvement of revealed LZ_{GCN4} x-conformation in the regulatory networks of bZIP and bHLH-LZ transcription factors. Common dimerization domains allow transcription factors in the LZ families to form a variety of homo- and heterodimers with different properties. Since these factors bind to promoter DNA in a sequence-specific manner, unique pairings of factors often results in unique pairings of DNA-binding preferences. Thus, by expressing different sets of subunits under different conditions, cells can generate complex regulatory circuits from a relatively small number of genes.[20] Accordingly, the larger the array of LZ transcription factors in a given genome, the greater the potential for complex transcriptional programs affecting the unique functions of individual cells, tissues, organs, and the species itself.[41] Plain monomer−dimer folding kinetics, taken together with tightness of interactions within the coiled coil interface, make it difficult to envision a mechanism allowing convenient and specific interchange of dimerization partners (and therefore transcriptional activities) within these regulatory networks. However, the existence of a stable intermediate conformation, can easily explain this phenomenon.

Besides the possible value of the x-form discovery for investigation of LZ transcription factors regulatory networks, new insights can also be gained in the direction of coiled coil folding studies. But what is more important, taking into account the multitude of cellular functions dependent on LZ transcription factors and other proteins bearing this motif, the equilibrium described in our study can be potentially employed as a model system in the development of various types of drugs, including antiviral and anticancer substances targeting the interaction interfaces within the LZ domains.[42−47]

Materials and Methods

Plasmid Constructions, Protein Expression and Purification. The LZ_{GCN4} peptide used in the current study has the following sequence: GLQRMKQLEDKVEELLSKNYHLENEVARLKKLVGER. This corresponds to the original and widely characterized GCN4-p1 peptide[16] with additional Gly-Leu-Gln N-terminal residues, two of those being GCN4 authentic residues, and Gly coming from the expression system.

The recombinant peptides were expressed in *Escherichia coli* BL21-(DE3) (Stratagene, La Jolla, CA) as fusions with bacteriophage λ-protein D[48] and released from the fusion using bovine enterokinase.[49]

The expression vector pHDe was constructed by cloning the bacteriophage λ-protein D gene (λcI857Sam7, Bohringer Mannheim, Germany) together with the downstream Gly-Ser linker and the enterokinase cleavage site (D_4K) into the pQE30 vector (Qiagen AG, Basel, Switzerland). DNA fragment coding for the leucine zipper GCN4 was amplified from pET3α-GCN4 plasmid (kindly provided by Dr. Christine Berger, University Zurich, Switzerland) and inserted into the pHDe vector via the SfoI/HindIII restriction sites downstream of the sequence encoding for the His_6-λpD-$(GS)(G_3S)_2$-(D_4K) construct. The identity of the final construct pHDe-LZgcn4 was confirmed by sequencing (Microsynth AG, Balgach, Switzerland).

For protein expression *Escherichia coli* BL21(DE3) cells (Stratagene, La Jolla, CA) were transformed with the pHDe-LZgcn4 vector following the instructions of the manufacturer. Freshly transformed cells were inoculated overnight in Spectra 9 medium (Spectra Stable Isotopes, MD) supplemented with BME vitamins solution (Sigma-Aldrich, Buchs, Switzerland). They were then diluted 1:20 with fresh medium and grown at 37 °C to a cell density $OD_{600} \approx 0.9$ before the expression was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. The cells were grown for a further 7−8 h, then harvested by centrifugation (5 min, 5000g), and frozen at −20 °C. For ^{15}N (^{15}N, ^{13}C)-labeled preparations Spectra 9 medium was supplied with $^{15}NH_4Cl$ (and ^{13}C−D-glucose) (Spectra Stable Isotopes, MD) as the sole source for nitrogen (and carbon).

Cell pellets were resuspended in Ni-NTA binding buffer $(NaCl)_{500}$-$(NaPi)_{20}$(imidazole)$_{50}$(7.4), disrupted in a French pressure cell at 11 000 psi (7.6×10^7 Pa) and centrifuged (45 min, 100000g) to separate the soluble fraction. His-tagged fusion protein was purified from the soluble fraction on a Ni-NTA metal-affinity column (Amersham-Pharmacia

(41) Deppmann, C. D.; Alvania, R. S.; Taparowsky, E. J. *Mol. Biol. Evol.* **2006**, *23*, 1480−92.
(42) Agou, F.; Courtois, G.; Chiaravalli, J.; Baleux, F.; Coic, Y. M.; Traincard, F.; Israel, A.; Veron, M. *J. Biol. Chem.* **2004**, *279*, 54248−57.
(43) Borges-Walmsley, M. I.; Beauchamp, J.; Kelly, S. M.; Jumel, K.; Candlish, D.; Harding, S. E.; Price, N. C.; Walmsley, A. R. *J. Biol. Chem.* **2003**, *278*, 12903−12.
(44) Berg, T.; Cohen, S. B.; Desharnais, J.; Sonderegger, C.; Maslyar, D. J.; Goldberg, J.; Boger, D. L.; Vogt, P. K. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 3830−5.
(45) Ahmad, A.; Yadav, S. P.; Asthana, N.; Mitra, K.; Srivastava, S. P.; Ghosh, J. K. *J. Biol. Chem.* **2006**, *281*, 22029−38.
(46) Kataoka, K.; Handa, H.; Nishizawa, M. *J. Biol. Chem.* **2001**, *276*, 34074−81.
(47) Surh, Y. J.; Kundu, J. K.; Na, H. K.; Lee, J. S. *J. Nutr.* **2005**, *135*, 2993S-3001S.
(48) Forrer, P.; Jaussi, R. *Gene* **1998**, *224*, 45−52.
(49) Liepnieks, J. J.; Light, A. *J. Biol. Chem.* **1979**, *254*, 1677−83.

Biotech, Uppsala, Sweden) according to the instruction of the manufacturer. Fractions containing the λpD-LZgcn4 fusion were dialyzed into enterokinase cleavage buffer $(\text{Tris-HCl})_{20}(\text{NaCl})_{50}(\text{CaCl}_2)_2$(7.4) and subjected to enterokinase (Novagene, Darmstadt, Germany) cleavage for 24 h at room temperature at a concentration of $(\text{fusion})_{0.5\ \text{mg}}(\text{EK})_{7\text{U}}$-$(\text{clevage buffer})_{1\ \text{mL}}$. After cleavage the 4.3 kDa LZgcn4 was separated from 15 kDa λpD by ultrafiltration on 5 kDa MWCO centrifugal filter devices (Millipore, Billerica, MA).

Final purification of the peptides was performed by reverse-phase HPLC on a C18 preparative column (Macherey-Nagel, Duren, Germany) using a water-acetonitrile gradient in the presence of 0.1% trifluoroacetic acid. Peptide identities were confirmed by MALDI-MS (FGCZ, Zurich, Switzerland). Peptide concentrations were determined from ^{1}H NMR spectra via integration of methyl-protons peak using L-Leucine (Sigma-Aldrich, Buchs, Switzerland) as a reference. Obtained values were consistent with the ones determined from absorbance at 280 nm (using extinction coefficient of the only Tyr20 tyrosine residue[50,51]).

NMR Spectroscopy. NMR experiments were performed at 37 °C on a Bruker AVANCE 600 MHz spectrometer equipped with a *z*-axis gradient triple resonance cryprobe and an AVANCE 900 MHz spectrometer. NMR data were processed with XWINNMR 3.2 (Bruker Biospin, Fallanden, Switzerland), analyzed using XEASY[52] and CARA (www.nmr.ch).[53] The ^{1}H chemical shifts were referenced to the DSS (sodium 2,2-dimethyl-2-silapentane-5-sulfonate) signal at 0 ppm, and ^{15}N and ^{13}C chemical shifts were referenced indirectly using the X/^{1}H gyromagnetic ratios.[54]

The concentration and pH dependencies of the x-form population were measured in 50 mM D_3-acetate, 40 mM KCl, pH 3.2. The pH of the samples was adjusted with potassium-D_3-acetate, and the exact values were measured at room-temperature both before and after the NMR experiments.

The assignments for the coiled coil conformation were obtained on the basis of the spectra recorded using 1 mM (monomer concentration) ^{15}N-^{13}C uniformly labeled LZ_{GCN4} peptide in 20 mM HEPES (2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid), 80 mM KCl, pH 7.1. The protein sample was further titrated down to the pH 3.2 and the changes in the chemical-shift values of assigned coiled coil residues were followed at each step of titration. The spectra for the x-form assignments were recorded using 0.25 mM LZ_{GCN4} peptide in 50 mM D_3-acetate, 40 mM KCl, pH 3.2. The same preparation was further used to obtain unfolded monomer reference sample, by diluting the peptide to 0.1 mM, adding urea to final concentration of 8 M, and adjusting the pH to 2.5 with hydrochloric acid.

Spin States in Conformationally Exchanging Molecules. For the mathematical treatment of the evolution of the density operator in the course of the reported NMR experiments, to extract quantitative information on the kinetic rates and populations from the spectra, several assumptions are made: (i) the molecule is in equilibrium between two spectroscopically distinct conformations A and A′ exchanging with a rate $k_{\text{ex}} = k_1 + k_{-1}$, which is slower than the frequency separation of the corresponding NMR resonances, for example, in slow exchange regime, a condition quantitatively defined by McConnel;[55] (ii) the chemical exchange should be a Markovian random process;[56] and (iii) the sudden jump approximation is assumed, which implies that the magnetization does not change orientation during the chemical exchange.[56] A product space between chemical configuration space and magnetization, a composite superspace, is required to account for the flow of magnetization during the chemical exchange[56,57] as it is implemented as a homogeneous version of the stochastic Liouvillian equations[58] in the NMR simulation program QSIM.[59]

To simplify product operator description of the reported NMR experiments we further assume that the spin-state operators are mixed either in the rotating frame under conditions of spin locking[60] achieved by windowless sequences of composite rf-pulses[61] or in the longitudinal direction, where only the operators representing magnetization modes[62] should be considered.[26] Under these assumptions only transverse and longitudinal subspaces of the composite superspace defined by the corresponding spin-state operators contributing to the detectable signal can be considered. The present discussion is restricted to two exchanging two-spin-$^1/_2$ systems, IS and $I'S'$, where I and I' stand for a proton and S and S' for a heteronuclear spin (e.g., ^{15}N or ^{13}C), coupled via a scalar coupling constant J_{IS}. This situation represents one of the most common spin systems encountered, e.g., for backbone HN or HC^α moieties. The conformation exchange process and longitudinal relaxation of ^{1}H (owing to the bulk of protons in protein), result into effective mapping of initial spin states entering the mixing period on the spin states subsequently detected as a signal given by

$$S_i^\alpha \rightarrow S'^\alpha_i \tag{1.1}$$

$$S_i^\alpha \rightarrow S'^\beta_i \tag{1.2}$$

In eq 1, the single transition basis operators $S_i^\alpha = {}^1/_2 S_i + S_i I_z$ and $S_i^\beta = {}^1/_2 S_i - S_i I_z$ are used, where *i* stands for "z", "+" or "−". For sufficiently short mixing periods, the buildup of the target operators is monoexponential with the corresponding buildup rates of k_{ex} and $1/(2T_{1\text{Hbulk}})$, respectively.[32] The mapping given by eq 1.1 represents the productive magnetization transfer pathway, which is selected in the NMR experiments, while the spurious transfer of eq 1.2 should be suppressed.

Spin State Exchange in the Rotating Frame. In the slow-exchange regime with the strength of the locking rf-field ω_{rf} satisfying the condition $\omega_{S,\text{rf}} \gg |\Omega_S - \Omega'_S|$, where Ω_S is the chemical shift relative to the carrier frequency of the spin S, the transfer of the ^{1}H spin-state selective ^{15}N coherence due to the exchange can be described by eq 1. For the 4(1)D XYEX-TROSY experiment of Figure 3, two equivalent (in terms of coherent evolution of the density operator) magnetization transfer pathways are given by

$$I_z^\alpha + S_z^\alpha \rightarrow I_+ S_- \exp[i(\Omega_I - \Omega_S)t_1] \rightarrow S_-^\beta \rightarrow S'^\beta_- \exp[i(-\Omega'^\beta_S)t_2] \rightarrow I_-^\alpha \exp[i(-\Omega'^\alpha_I)t_3] \tag{2.1}$$

$$I_z^\beta - S_z^\beta \rightarrow I_+ S_+ \exp[i(\Omega_I + \Omega_S)t_1] \rightarrow S_+^\alpha \rightarrow S'^\alpha_+ \exp[i(\Omega'^\alpha_S)t_2] \rightarrow I_-^\beta \exp[i(-\Omega'^\beta_I)t_3] \tag{2.2}$$

where $\Omega_i^\alpha = \Omega_i + \pi J_{IS}$, $\Omega_i^\beta = \Omega_i - \pi J_{IS}$, and $i = I$ or S. The initial steady-state I_z and S_z polarizations are the sum of the corresponding single-transition polarization operators, which can be considered separately.[29] It should be noted that the pathway 1 is optimized in terms of transverse relaxation and is enhanced by the steady-state magnetization of the spin S.[30] Three- and four- spin-$^1/_2$ systems I_2S and I_3S found in methylene and methyl groups do not result in detectable ^{1}H signal, and therefore the corresponding magnetization transfer pathways are not explicitly described.

The echo−anti-echo type Fourier transform in all spectral dimensions of the interferograms acquired with two executions of the pulse

(50) Edelhoch, H. *Biochemistry* **1967**, *6*, 1948−54.
(51) Gill, S. C.; von Hippel, P. H. *Anal. Biochem.* **1989**, *182*, 319−26.
(52) Bartels, C.; Xia, T. H.; Billeter, M.; Guntert, P.; Wuthrich, K. *J. Biomol. NMR* **1995**, *6*, 1−10.
(53) Keller, R. L. J. *The Computer Aided Resonance Assignment Tutorial*, 1st ed.; Cantina Verlag: Germany, 2004.
(54) Wishart, D. S.; Bigam, C. G.; Yao, J.; Abildgaard, F.; Dyson, H. J.; Oldfield, E.; Markley, J. L.; Sykes, B. D. *J. Biomol. NMR* **1995**, *6*, 135−40.
(55) McConnell, H. M. *J. Chem. Phys.* **1958**, *28*, 430−31.
(56) Jeener, J. *Adv. Magn. Reson.* **1982**, *10*, 1−51.
(57) Kühne, R. O.; Schaffhauser, T.; Wokaun, A.; Ernst, R. R. *J. Magn. Reson.* **1979**, *35*, 39−67.
(58) Helgstrand, M.; Hard, T.; Allard, P. *J. Biomol. NMR* **2000**, *18*, 49−63.
(59) Helgstrand, M.; Allard, P. *J. Biomol. NMR* **2004**, *30*, 71−80.
(60) Deverell, C.; Morgan, R. E.; Strange, J. H. *Mol. Phys.* **1970**, *18*, 553−59.
(61) Bai, N. S.; Ramachandran, R. *J. Magn. Reson., Ser. A* **1993**, *105*, 298−303.
(62) Canet, D. *Prog. Nucl. Magn. Reson. Spectrosc.* **1989**, *21*, 237−91.

sequence of Figure 3 results in two spectra, S_{ZQ+DQ} and S_{ZQ-DQ}, representing the sum and difference of signals described by eq 2.1 and 2.2. The extraction of the signal corresponding to the individual pathways is achieved by subsequent adding or subtracting of S_{ZQ+DQ} and S_{ZQ-DQ} spectra. The suppression of the spurious pathway of eq 1.2 can be achieved by incorporation of the relaxation compensation element in the middle of the mixing period[63] Alternatively, the spurious cross-peaks in the S_{ZQ} subspectrum resulting from the cross-talk of eq 1.2 can be, to a very good extent, eliminated by subtracting S_{DQ} spectrum shifted in ω_2 and ω_3 dimensions by J_{IS} Hz and multiplied with an empirically weighed factor typically in the range of 0.1. The latter technique we preferred in the current study, since it does not require interruption of the supercycle of the spin-locking pulse train and incorporation of an extra delay of $^1/_2J_{IS}$ in the mixing period during which no polarization transfer due to the exchange occurs.

Spin State Exchange in Longitudinal Direction. In the case of slow exchange rates, mixing of the magnetization modes[26,62] instead of spin-selective coherences can be considered owing to the longer corresponding relaxation rates and lack of need to decouple passive J couplings. With the cost of decreasing sensitivity by a factor of 2, the experimental scheme of Figure 3 can be modified in such a way, that the transfer $S_-^\beta \rightarrow S'_-{}^\beta$ is achieved in two steps implemented in two separate experiments (see ZEX-TROSY of Figure S3, Supporting Information). At first in $S_-^\beta = S_x^\beta - iS_y^\beta$, S_x^β is converted to S_z^β and subsequently mixed with $S_z'^\beta$, which is, in turn, followed by a conversion to $S_x'^\beta$. This pathway is complemented with the corresponding transfer of S_y^β. The complete pathways of eq 1 is reconstructed by addition of thus acquired two subspectra followed by the transformation procedure described earlier.

Spin State Exchange in the Doubly Rotating Frame. Two asynchronously rotating coordinate frames where the I and S spins are locked with windowless composite pulse sequences satisfying the conditions $\omega_{S,rf} \gg |\Omega_S - \Omega'_S|$ and $\omega_{I,rf} \gg |\Omega_I - \Omega'_I|$ can be used to reduce the total duration of polarization transfer periods by $1/J_{IS}$ comparing to XYEX-TROSY of Figure 3 and still obtain a fully sensitivity-enhanced spin-state exchange pseudo-4D experiment. For the 4(1)D (XY)2EX-TROSY experiment of Figure S1 (Supporting Information), two equivalent magnetization transfer pathways are given by

$$I_z^\alpha + S_z^\alpha \rightarrow I_+S_- \exp[i(\Omega_I - \Omega_S)t_1] \rightarrow I'_+S'_- \rightarrow I'_+S'_- \exp[i(-\Omega'_S)t_2] \rightarrow I_-^\alpha \exp[i(-\Omega'^{\,\alpha}_I)t_3] \quad (3.1)$$

$$I_z^\beta - S_z^\beta \rightarrow I_+S_+ \exp[i(\Omega_I + \Omega_S)t_1] \rightarrow I'_+S'_+ \rightarrow I'_+S'_+ \exp[i(\Omega'_S)t_2] \rightarrow I_-^\beta \exp[i(-\Omega'^{\,\beta}_I)t_3] \quad (3.2)$$

Note that the spurious magnetization transfer owing to proton−proton dipolar relaxation mechanisms does not occur, and thus there is no need for compensation of spurious cross-peaks in the spectra. The advantage of the ZQ/DQ version is the absence of the delays in the middle polarization transfer element, which might result in potential improvement in sensitivity. The passive coupling constants $J_{H\alpha HN}$ and $J_{H\beta HN}$ do not contribute to the signal dephasing during mixing because of the selection of relatively low rf-power to lock $^1H^N$ magnetization thus resulting in effective decoupling of $^1H^N$ spins from all other protons. However we have not achieved the maximum sensitivity with this scheme possibly because of the loss of coherences to imperfections of rf-pulses during mixing (data not shown). However, this might be improved with further optimization of mixing sequences.

Acknowledgment. This work is financially supported by grants from the Roche Research Foundation (to Y.N.) and the NTU (to K.P.).

Supporting Information Available: Figures S1, S2, S3: scheme of 4(1)D (XY)2EX-TROSY experiment; scheme of 4(1)D PFG-XYEX-TROSY experiment; scheme of 4(1)D PFG-ZEX-TROSY experiment. Listings S1 and S2: Bruker Avance pulse program for 4(1)D XYEX-TROSY experiment; Bruker Avance pulse program for 4(1)D (XY)2EX-TROSY experiment. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0685295

(63) Loria, J. P.; Rance, M.; Palmer, A. G.*J. Am. Chem. Soc.* **1999**, *121*, 2331−2.
(64) Cavanagh, J.; Palmer, A. G.; Wright, P. E.; Rance, M. *J. Magn. Reson.* **1991**, *91*, 429−36.

Supporting Information

NMR Spin State Exchange Spectroscopy Reveals Equilibrium of Two Distinct Conformations of Leucine Zipper GCN4 in Solution.

*Yaroslav Nikolaev and Konstantin Pervushin**

Laboratory of Physical Chemistry, Swiss Federal Institute of Technology (ETH Zurich), CH-8093 Zurich, Switzerland

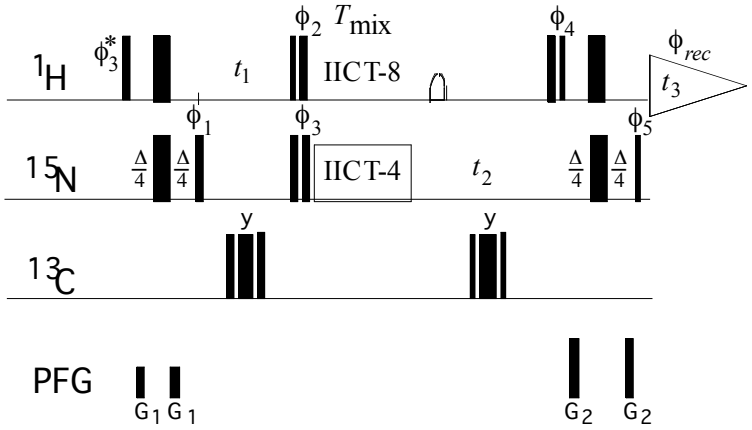


Figure S1. Scheme of 4(1)D (XY)2EX-TROSY, pseudo four-dimensional spin state exchange TROSY experiment with mixing of multiple quantum $^{1}H^{15}N$ coherencies in transverse plane utilizing entire steady-state Boltzmann thermal equilibrium magnetization and featuring sensitivity enhancement for the phase sensitive detection in all three spectral dimensions. Narrow and wide black bars indicate non-selective $\pi/2$ and π rf-pulses applied with the phase x unless indicated otherwise. Water saturation is minimized by returning the water magnetization to the $+z$ axis before data acquisition by the use of water selective 90^{o} rf-pulses shown as open shapes on the 1H channel. The time period Δ is set to $1/J_{HN}$ and is 5.78 ms for the backbone amide moieties. The multiple quantum $^{1}H^{15}N$ coherencies stemming from kinetically exchanging spin systems are mixed during T_{mix} period using two asynchronous planar mixing scheme IICT-4 and IICT-8, which are composite pulse sequences specifically designed for isotropic mixing[1]. The IICT-4 scheme is RRRRRRRR with R = 38(0) 112.6(60.5) 205.9(68.6) 256.9(280.1) 101.9(5.2) 265.5(281.6) 242.6(72.6) 66.6(66.9) 44.9(0), where numbers represent rf-pulses with the corresponding flip angle and the rf-phase (in brackets), respectively, and R is phase inversion of R. The IICT-8 scheme is RRRRRRRR RRRRRRRRR. In this case the durations and strengths of the pulsed magnetic field gradients (PFG) applied along the z-axis are selected as G_1: 800 μs, 30 G/cm; G_2: 1 ms, 80 G/ cm. Two datasets,

S_{ZQ+DQ} and S_{ZQ-DQ}, are acquired using the phases indicated in the **Listings S2 (A)** and **(B)**, respectively. For both S_{ZQ+DQ} and S_{ZQ-DQ} datasets the quadrature detection in t_1 and t_2 dimensions is achieved by the echo-anti-echo method[2,3]. The phase incrimination logics is complex and can be derived from the **Listings S2**. The processing of the datasets is described in METHODS and is identical to that of XYEX-TROSY. The rf-pulses on the ^{1}H, ^{15}N and ^{13}C nuclei are centered at 4.7 ppm, 118 ppm and 110 ppm, respectively.

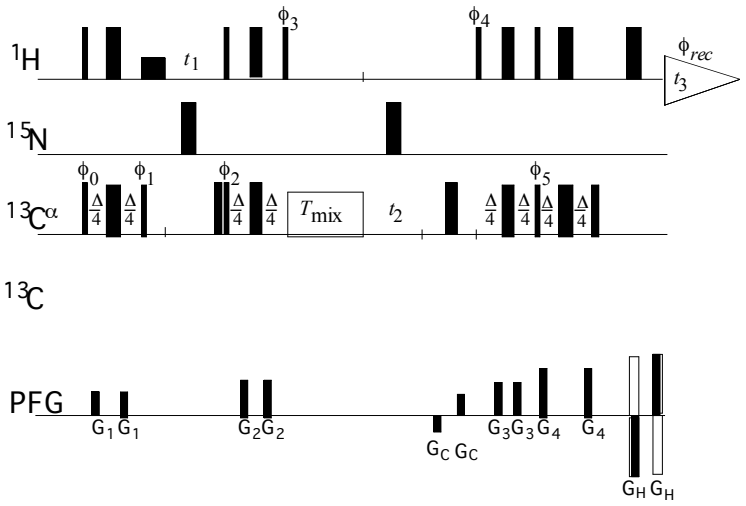


Figure S2. Scheme of 4(1)D PFG-XYEX-TROSY, pseudo four-dimensional spin state exchange TROSY experiment with mixing of spin coherencies in transverse plane utilizing entire steady-state Boltzmann thermal equilibrium magnetization and featuring sensitivity enhancement for the phase sensitive detection in all three spectral dimensions using PFG for coherence transfer pathway selection and water suppression. This pulse sequence is applied for ^{1}H-^{13}C$^\alpha$ spins systems. Narrow and wide black bars indicate non-selective $\pi/2$ and π rf-pulses applied with the phase x unless indicated otherwise. Water saturation is minimized by returning the water magnetization to the $+z$ axis before data acquisition by the use of water selective 90^o rf-pulses shown as open shapes on the 1H channel. The time period Δ is set to $1/J_{HC}$ and is 3.56 ms for the backbone amide moieties. The ^{1}H spin-state selective ^{13}C coherences stemming from kinetically exchanging spin systems are mixed during T_{mix} period using a planar mixing scheme IICT-4, which is a composite pulse sequence specifically designed for isotropic mixing purpose in TOCSY experiments (see METHODS)[1]. The IICT-4 scheme is RRRRRRRR with R = 38(0) 112.6(60.5) 205.9(68.6) 256.9(280.1) 101.9(5.2) 265.5(281.6) 242.6(72.6) 66.6(66.9) 44.9(0), where numbers represent rf-pulses with the

corresponding flip angle and the rf-phase (in brackets), respectively, and R is phase inversion of R. In this case the durations and strengths of PFG applied along the z-axis are selected as G_1: 500 μs, 19 G/cm; G_2: 500 μs, 15 G/cm; G_3: 900 μs, 32 G/ cm; G_4: 1 ms, 50 G/ cm; G_C: 1.75 ms, 80 G/ cm; G_H: 440 μs, 80 G/ cm. Two spectra, S_{ZQ} and S_{DQ}, comprising zero-quantum and double quantum pathways are collected using the following phases: for S_{ZQ} $\phi_0 = \{x, -x, -y, y\}$; $\phi_1 = \{y, -y, x, -x\}$; $\phi_2 = -x$; $\phi_3 = -y$; $\phi*_3 = y$; $\phi_4 = y$; $\phi*_4 = -y$; $\phi_5 = y$; $\phi_{rec} = \{-y, y, x, -x\}$ for S_{DQ} and $\phi_{rec} = \{y, -y, x, -x\}$ and inverted G_H (other phases and PFGs are the same). For both S_{ZQ} and S_{ZQ} datasets the quadrature detection in t_1 and t_2 dimensions is achieved by the echo-anti-echo method[2,3]. The anti-echo signal in t_1 dimension is obtained by inversion of the phases ϕ_2, ϕ_3 and $\phi*_3$, and the anti-echo signal in t_2 dimension is obtained by inversion of the phases ϕ_2, ϕ_3, $\phi*_3$, ϕ_4, $\phi*_4$ and ϕ_5 with simultaneous inversion of the sign of the G_H pulse. The processing of the datasets is described in METHODS. The rf-pulses on the ^{1}H, ^{15}N, $^{13}C^{\alpha}$ and ^{13}C' nuclei are centred at 4.7 ppm, 118 ppm, 54 ppm and 170 ppm, respectively.

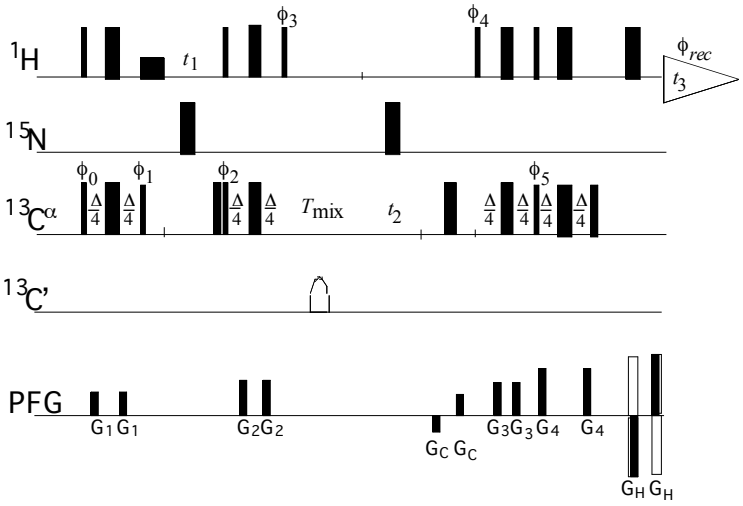


Figure S3. Scheme of 4(1)D PFG-ZEX-TROSY, pseudo four-dimensional spin state exchange TROSY experiment with mixing of spin orders in longitudinal direction utilizing entire steady-state Boltzmann thermal equilibrium magnetization and featuring sensitivity enhancement for the phase sensitive detection in all three spectral dimensions using PFG for coherence transfer pathway selection and water suppression. This pulse sequence is applied for ^{1}H-$^{13}C^{\alpha}$ spins systems. The parameters are identical for those described in **Figure S2**, except that the mixing element consists of

two ^{13}C 90° pulses ψ_1 and ψ_2. For each zero-quantum and double quantum-based dataset two subsets, S_{ZQ_x} and S_{ZQ_y}, and S_{DQ_x} and S_{DQ_y}, are measured using the pairs of phases $\psi_1 = x$ and $\psi_2 = -x$, and $\psi_1 = y$ and $\psi_2 = -y$, respectively. Before data processing the datasets S_{ZQ} and S_{DQ} are reconstructed using $S_{ZQ} = S_{ZQ\_x} + S_{ZQ\_y}$, and $S_{DQ} = S_{DQ\_x} - S_{DQ\_y}$.

Listing S1. (A) Bruker Avance pulse program for 4(1)D XYEX-TROSY experiment, dataset S_{ZQ+DQ}.

```
;pl1   : power for 1H
;pl2   : power for 13C
;pl3   : power for 15N
;sp1   : water flipback power
;spnam1: gauss128_5
;p1    : 90 degree hard pulse 1H
;p3    : 90 degree hard pulse 13C
;p4    : 180 degree hard pulse 13C
;p5    : 90 degree hard pulse 15N
;p11   : water flipback pulse (1.5m)
;p16    : IICT-4 pulse R = 38(0) 112.6(60.5) 205.9(68.6) 256.9(280.1) 101.9(5.2)
265.5(281.6) 242.6(72.6) 66.6(66.9) 44.9(0)
;p21   : 500u (Gradient in first INEPT)
;p22   : 500u (Gradient in second INEPT)
;p23   : 900u (Gradient in watergate)
;p24   : 1000u (Gradient in mixing time)
;gpz1  : 19%
;gpz2  : 15%
;gpz3  : 32%
;gpz4  : 50%
;d1    : relaxation delay
"d2=2.74m" ;d2    : INEPT delay (~2.7ms)

define delay INEPT1
define delay INEPT2
define delay INEPT3

#include <Avance.incl>

#define GRADIENT1   10u p21:gp1 200u
#define GRADIENT2   10u p22:gp2 200u
#define GRADIENT3   10u p23:gp3 200u
#define GRADIENT4   10u p24:gp4 200u

"p2=2*p1"
"p6=2*p5"
"p4=2*p3"

"d0=in0/2-(p3*2 + p5*2/3.14159 + 1.5u)"
"d10=in10/2-(p4 + 1.5u)"

"INEPT1=d2-(p21+210u)-10u"
"INEPT2=d2-(p22+p11+210u)-10u"
"INEPT3=d2-(p23+p11+210u)-10u"

"l2 = 0"
"l7 = 3"

aqseq 312

1  10u ze
2  10u
```

```
10u pl1:f1 pl2:f2 pl3:f3
10u LOCKH_OFF
d1
10u LOCKH_ON
 (p11:sp1 ph20:r):f1
 (p1 ph20):f1
GRADIENT1
INEPT1 pl1:f1
(center(p2 ph20):f1 (p6 ph20):f3)
GRADIENT1
INEPT1
if "l2 %2 == 1" goto 28
      (p5 ph29):f3
      goto 29
28    (p5 ph30):f3
29      2u
;--------------------------t1 evolution
   d0
   (p3 ph23 1.5u p4 ph20 1.5u p3 ph23):f2
   d0
   (p5 ph20 2u p5 ph4):f3

   (p1 ph20):f1
   (p11:sp1 ph22:r):f1
   GRADIENT2
   INEPT2 pl1:f1
   (center(p2 ph20):f1 (p6 ph20):f3)
   GRADIENT2
   INEPT2
   (p11:sp1 ph18:r):f1
   (p1 ph18):f1
;--------------------------mix
        10u ; IICT4 93JMRA298-105
74     (p16:sp16 ph20 10u):f3
       (p16:sp16 ph20 10u):f3
       (p16:sp16 ph22 10u):f3
       (p16:sp16 ph22 10u):f3

       (p16:sp16 ph20 10u):f3
       (p16:sp16 ph22 10u):f3
       (p16:sp16 ph22 10u):f3
       (p16:sp16 ph20 10u):f3
         lo to 74 times l7
;--------------------------t2 evolution
      d10
      (p3 ph23 1.5u p4 ph20 1.5u p3 ph23):f2
      d10
   (p1 ph11):f1
   (p11:sp1 ph11:r):f1
   GRADIENT2
   INEPT2 pl1:f1
   (center(p2 ph20):f1 (p6 ph20):f3)
   GRADIENT2
   INEPT2
   (p11:sp1 ph26:r):f1
   (p1 ph28):f1 (p5 ph15):f3
   GRADIENT3
   INEPT3
   (p11:sp1 ph20:r):f1
   (center(p2 ph20):f1 (p6 ph20):f3)
   (p11:sp1 ph20:r):f1
   GRADIENT3
   INEPT3 LOCKH_OFF
   (p5 ph27):f3
 go=2 ph31
 1m mc #0 to 2
```

```
 F1EA(iu2 & ip4*2 & ip18*2,id0)
 F2EA(rd0 & ip4*2 & ip18*2 & ip11*2 & ip15*2,id10)
10u do:f1
10u do:f2
10u do:f3
10u LOCKH_OFF
exit

ph4=0
ph11=1
ph15=1
ph18=3
ph26=2
ph27=0
ph28=0
ph29=1 3
ph30=3 1
ph31=3 1

ph20=0
ph21=1
ph22=2
ph23=3
```

Listing S1. (B) Bruker Avance pulse program for 4(1)D XYEX-TROSY experiment, dataset $S_{ZQ\text{-}DQ}$.

The pulse program is identical to that listed in **Listing S1(A)**, except for the phases ph29, ph30 and ph31:

```
ph29=0 2
ph30=0 2
ph31=0 2
```

Listing S2. (A) Bruker Avance pulse program for 4(1)D (XY)2EX-TROSY experiment, dataset S_{ZQ+DQ}.

```
;pl1   : power for 1H
;pl2   : power for 13C
;pl3   : power for 15N
;sp1   : water flipback power
;spnam1: gauss128_5
;p1    : 90 degree hard pulse 1H
;p3    : 90 degree hard pulse 13C
;p4    : 180 degree hard pulse 13C
;p5    : 90 degree hard pulse 15N
;p11   : water flipback pulse (1.5m)
;p16   : IICT-4 (15N) ,
;p14     :IICT-8  (1H)  pulse  R  =  38(0)  112.6(60.5)  205.9(68.6)  256.9(280.1)
101.9(5.2) 265.5(281.6) 242.6(72.6) 66.6(66.9) 44.9(0)
;p21   : 500u (Gradient in first INEPT)
;p22   : 500u (Gradient in second INEPT)
;p23   : 900u (Gradient in watergate)
;p24   : 1000u (Gradient in mixing time)
;gpz1  : 19%
;gpz2  : 50%
;d1    : relaxation delay
"d2=2.74m" ;d2    : INEPT delay (~2.7ms)
```

```
define delay INEPT1
define delay INEPT2
define delay INEPT3

#include <Avance.incl>

#define GRADIENT1   10u p21:gp1 200u
#define GRADIENT2   10u p22:gp2 200u

"p2=2*p1"
"p6=2*p5"
"p4=2*p3"

"d0=in0/2-(p3*2 + p5*2/3.14159 + 1.5u)"
"d10=in10/2-(p4 + 1.5u)"

"INEPT1=d2-(p21+210u)-10u"
"INEPT2=d2-(p22+p11+210u)-10u"
"INEPT3=d2-(p23+p11+210u)-10u"

"l2 = 0"
"l3 = 0"
"l7 = 3"

aqseq 312

1  10u ze
2  10u
10u pl1:f1 pl2:f2 pl3:f3
10u LOCKH_OFF
d1
10u LOCKH_ON
5  10u

      if "l2 %2 == 0" goto 6
      (p11:sp1 ph13):f1
      goto 7

6     2u
       (p11:sp1 ph14):f1
7 10u

   (p1 ph20):f1 (p5 ph15):f3
   10u
 ;  (p11:sp1 ph18:r):f1
   GRADIENT1
   INEPT1 pl1:f1
   (center(p2 ph20):f1 (p6 ph20):f3)
   GRADIENT1
   INEPT1
  ; (p11:sp1 ph19:r):f1
   (p5 ph29):f3
 ;---------------------------------------t1
   d0
   (p3 ph23 1.5u p4 ph20 1.5u p3 ph23):f2
   d0
      if "(l2+l3) %2 == 0" goto 91
      (p5 ph20 2u p5 ph4):f3
      goto fin
91    (p5 ph20 2u p5 ph3):f3
        goto fin
fin,    2u

      if "l3 %2 == 1" goto 121
      if "l2 %2 == 1" goto 121
      (p1 ph20 2u p1 ph2):f1
```

```
      goto 131

121   (p1 ph20 2u p1 ph1):f1
131   2u
;--------------------------mix
      10u ; IICT4 93JMRA298-105
74    (p14:sp14 ph20 5u p14:sp14 ph20 5u):f1 (p16:sp16 ph20 10u):f3
      (p14:sp14 ph22 5u p14:sp14 ph22 5u):f1 (p16:sp16 ph20 10u):f3
      (p14:sp14 ph20 5u p14:sp14 ph22 5u):f1 (p16:sp16 ph22 10u):f3
      (p14:sp14 ph22 5u p14:sp14 ph20 5u):f1 (p16:sp16 ph22 10u):f3

      (p14:sp14 ph22 5u p14:sp14 ph22 5u):f1 (p16:sp16 ph20 10u):f3
      (p14:sp14 ph20 5u p14:sp14 ph20 5u):f1 (p16:sp16 ph22 10u):f3
      (p14:sp14 ph22 5u p14:sp14 ph20 5u):f1 (p16:sp16 ph22 10u):f3
      (p14:sp14 ph20 5u p14:sp14 ph22 5u):f1 (p16:sp16 ph20 10u):f3
         lo to 74 times l7
;--------------------------t2 evolution
      d10
      (p2 ph20):f1 (p3 ph23 1.5u p4 ph20 1.5u p3 ph23):f2
      d10
      (p5 ph25):f3
      (p1 ph20):f1
        2u
      if "l3 %2 == 0" goto 12
      if "l2 %2 == 1" goto 12
      (p1 ph11):f1
      goto 13
12    (p1 ph10):f1
13    2u
   GRADIENT2
   INEPT3
   (p11:sp1 ph26:r):f1

   (center(p2 ph20):f1 (p6 ph20):f3)
   10u
   (p11:sp1 ph26:r):f1
   GRADIENT2
   INEPT3 LOCKH_OFF

      if "l3 %2 == 1" goto 28
      (p5 ph27):f3
      goto 29

28    (p5 ph28):f3
29      2u
if "l2 %2 == 1" goto 31
   go=2 ph30
   goto 32

31  go=2 ph31
32  5u
    1m mc #0 to 2
     F1EA(iu2,id0)
     F2EA(rd0 & iu3,id10)

10u do:f1
10u do:f2
10u do:f3

10u LOCKH_OFF
exit

ph1=2
ph2=0

ph3=2
```

```
ph4=0

ph10=0
ph11=2

ph27=1
ph28=3

ph30= 0 2
ph31= 0 2

ph13=0
ph14=2

ph25=0

ph15= 2 0
ph29= 1 3

ph20=0
ph21=1
ph22=2
ph23=3

ph26=2
```

Listing S2. (B) Bruker Avance pulse program for 4(1)D (XY)2EX-TROSY experiment, dataset S_{ZQ-DQ}. The pulse program is identical to that listed in **Listing S2(A)**, except for the phases ph15, ph30 and ph31:

```
ph15=3 1
ph30=1 3
ph31=3 1
```

REFERENCES

(1) Bai, N. S.; Ramachandran, R. *Journal of Magnetic Resonance Series A* **1993**, *105*, 298-303.

(2) Kay, L. E.; Keifer, P.; Saarinen, T. *Journal of the American Chemical Society* **1992**, *114*, 10663-10665.

(3) Cavanagh, J.; Palmer, A. G.; Wright, P. E.; Rance, M. *Journal of Magnetic Resonance* **1991**, *91*, 429-436.

Chapter IV

Catalytic properties of bZIP RNase activity.

Catalytic properties of bZIP RNase activity.

Yaroslav Nikolaev[1,3,4], Laurent Bigler[2], Bernd Gutte[1] and Konstantin Pervushin[3,4]

[1]Department of Biochemistry and [2]Institute of Organic Chemistry, University of Zurich, Winterthurerstr. 190, CH-8057 Zurich, Switzerland

[3]Laboratory of Physical Chemistry, ETH Zurich, CH-8093, Zurich, Switzerland

[4]Biozentrum of University Basel, Klinlegbergstrasse 70, CH-4056, Basel, Switzerland

Recently we have shown that leucine zipper motifs of the transcription factors GCN4 and c-Jun are capable of catalysing degradation of RNA. Owing to the antiquity and prevalence of the LZ motifs, this observation is extremely intriguing when envisaged in the perspective of *in vivo* environment. To identify possible biological role of bZIP RNase activity we examined substrate specificity and catalytic properties of leucine zippers GCN4 and cJun. We show that both peptides are more catalytically potent towards the single-stranded regions of RNA and that in both cases RNA interacts with the coiled coil (dimeric) conformation of both leucine zippers. In addition NMR data provide evidence that the reaction, at least for LZ-GCN4, might follow the classical acid-base mechanism of catalysis. We propose that *in vivo* catalytic activity is associated with the activated form of the bZIP factor and may serve as a negative feedback loop for the transcription activation. Further studies are on the way to investigate the observed catalytic function in a broader array of LZ transcription factors.

INTRODUCTION

Leucine zipper (LZ) protein dimerization domain is one of the most abundant structural motifs present in eukaryotic transcription factors (TFs). The origin of LZ transcription factors dates over a billion years ago, when those were maintaining regulatory networks in first multicellular organisms (*1*). Two main classes of LZ-TFs, basic leucine zipper (bZIP) (*2*) and basic helix-loop-helix leucine zipper (bHLH-LZ) (*3*) proteins, have evolved as key regulators in the wide variety of processes, ranging from cell metabolism to tissue differentiation (*4*). Today LZ transcription factors are truly widespread among eukarya, with only human genome encoding 53 proteins with unique bZIP motifs (*5*) and at least 31 – with unique bHLH-LZ motifs (*6*).

Leucine zipper motif has been in the focus of scientific research for almost two decades. Its conserved presence among eukaryotes, together with critically important roles in many of cellular functions, encourages a lot of effort being placed in investigation of LZ dimerization properties and architecture of their regulatory networks (*4, 7*). Due to simplicity and stability leucine zippers are extensively studied as models for protein folding (*8, 9*) and are also widely employed as dimerization motifs in protein engineering studies (*10, 11*).

Despite very thorough and multidisciplinary characterization of the LZ motif, some findings appear yet to be made. Recently we have reported on the discovery of the x-form, a stable intermediate of LZ-GCN4 folding pathway that escaped the sight of scientists for more than a decade (*12*). Another report from the group of Bernd Gutte reveals an even more intriguing observation – leucine zippers of transcription factors GCN4 and c-Jun possess intrinsic ribonuclease activity (*13*). This reported activity is apparently conserved within the full-length bZIP protein, implying a biological role for this function.

Herein we further characterize catalytic properties of the leucine zippers GCN4 and cJun. We show that both peptides are more catalytically potent towards the single-stranded regions of RNA molecules and that their RNA-binding interface is associated with the folded

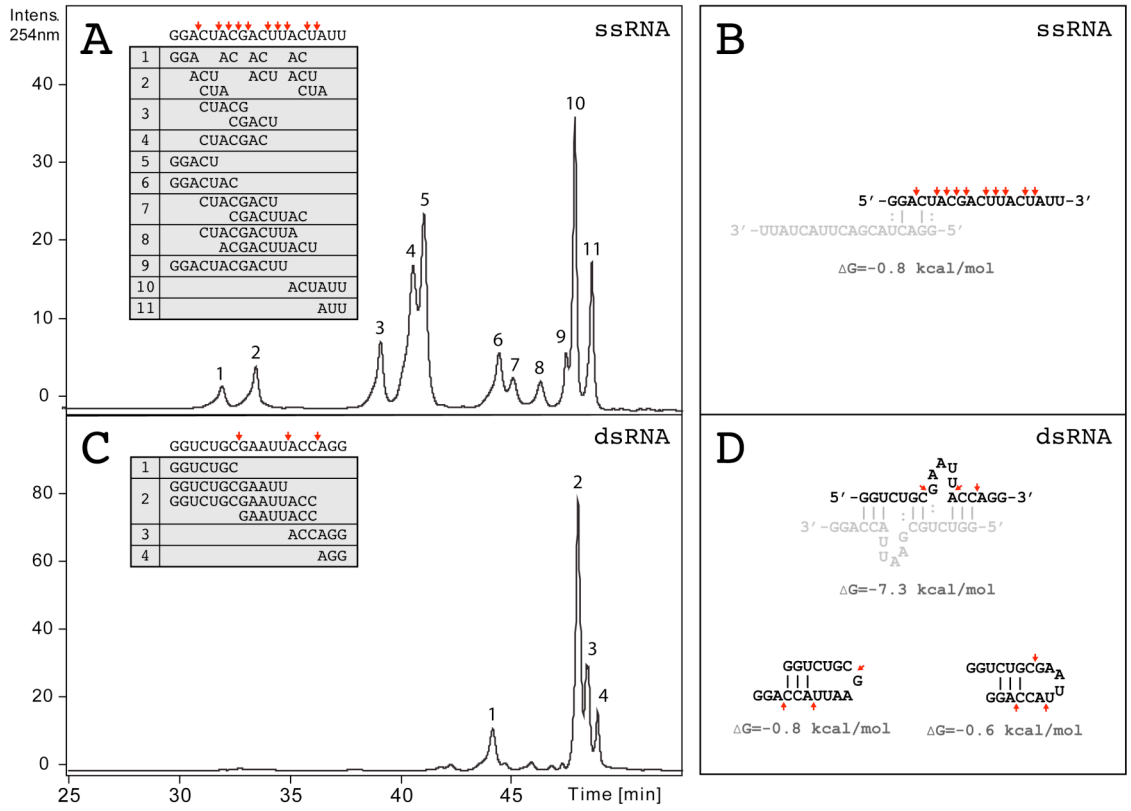
coiled coil conformation of the factor. Non-uniformity of the RNA-binding sites of GCN4 and cJun suggests topological flexibility in the active site definition within the LZ scaffold. Composition of the LZ-GCN4 active site points to the feasibility of the classical acid-base mechanism of the transesterification.

RESULTS

Substrate specificity

Synthetic peptide corresponding to 35-amino acid leucine zipper GCN4 (LQRMKQLEDK[10] VEELLSKNYH[20] LENEVARLKK[30] LVGER[35]) was subjected to RNA structure-dependent activity analysis. For this purpose original 18-mer RNA substrate (GGUCUGCGAAUUACCAGG = dsRNA) (*13*), capable of forming double-stranded structures, was redesigned at six positions to create a substrate with decreased ability for double-strand formation (GG**A**CU**A**CGA**C**UUAC**U**A**UU** = ssRNA). RNase activity was assayed by incubating 20 uM RNA with 50 uM LZ-GCN4 peptide (monomer concentrations) in 20 mM Tris-HCl, 80 mM KCl, pH 7.2 for 15 hours at 37°C. Subsequent HPLC and HPLC-MS fractionation and cleavage products analysis has shown increased cleavage efficiency of the ssRNA substrate, and revealed LZ peptide being more catalytically active towards the single stranded regions of RNA (Figure 1A-D). Analogous assay employing 670 µM dsRNA, 180 µM recombinant 42-residue leucine zipper cJun (GLERIARLEE[10] KVKTLKAQNS[20] ELASTANMLR[30] EQVAQLKQKV[40] MN) in 20 mM HEPES, 85 mM KCl, pH 7.1 at 37°C, have shown that LZ-cJun exhibits highly similar substrate affinity, targeting single-stranded regions of RNA in the first place (Figure 1E). Interestingly the observed catalytic rates in the case of LZ-cJun are significantly slower than those of the LZ-GCN4, with some intact 18-mer dsRNA substrate still detected after a week of incubation at 37°C.

LZ-GCN4



LZ-cJun

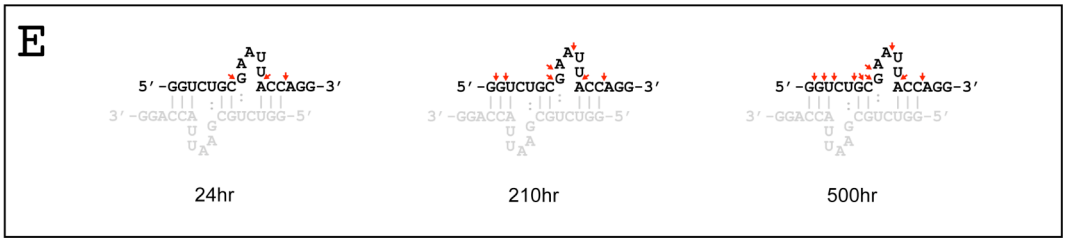


Figure 1. Products of RNA degradation by LZ-GCN4 and LZ-cJun. (**A**) and (**C**) HPLC-ESI-MS of the products of 15-hour incubation of 20 μM ssRNA (**A**) and 20 μM dsRNA (**C**) with 50 μM (monomer concentration) LZ-GCN4 in 20 mM Tris, 80 mM KCl, pH 7.2 at 37°C. (**B**) ssRNA and (**D**) dsRNA cleavage patterns mapped on the secondary ssRNA and dsRNA structures. (**E**) dsRNA cleavage patterns produced by LZ-cJun after 24, 210 and 500 hour incubation of 670 μM dsRNA in presence of ~500 μM (monomer concentration) LZ-cJun in 20 mM HEPES, 80 mM KCl, pH 7.1 at 37ºC. ssRNA and dsRNA structures calculated by MFOLD (*14*).

Active conformation

An ^{1}H-^{15}N-HSQC NMR analysis of dsRNA interactions with recombinant ^{15}N-labelled LZ-GCN4 at conditions when both coiled coil and x-form of LZ are comparably populated (15 uM LZ_{GCN4}, pH 4.0) shows that the substrate preferably interacts with the coiled coil form (Figure 2). In the conditions of this assay two LZ forms are in the slow conformational exchange in the NMR timescale (Figure 2A inset) resulting in two separate sets of resonances observed in the ^{1}H-^{15}N-HSQC spectrum (Figure 2A). Addition of dsRNA to the sample at acidic pH drives the LZ coiled coil conformation to a fast (in the NMR timescale) exchange with presumably high molecular weight peptide-RNA aggregates (which in turn irreversibly precipitate when dsRNA concentration exceeds ~50 μM). Resonances of the LZ peptide within those aggregates are broadened beyond detection (Figure 2B inset) due to the decreased tumbling rate of the aggregates. The slow exchange rate between the two LZ conformations limits the transfer of relaxation-broadening effects to the resonances of the x-form, and thus its resonances remain detectable (Figure 2B). At substrate concentrations below the aggregation/broadening boundary splitting of coiled coil Arg3 and Leu7 amide proton resonances is observed, indicating their proximity to the RNA interaction interface (data not shown). At neutral pH backbone amide of Arg3 is deprotonated (see below), limiting our ability to monitor its participation in the catalytic process.

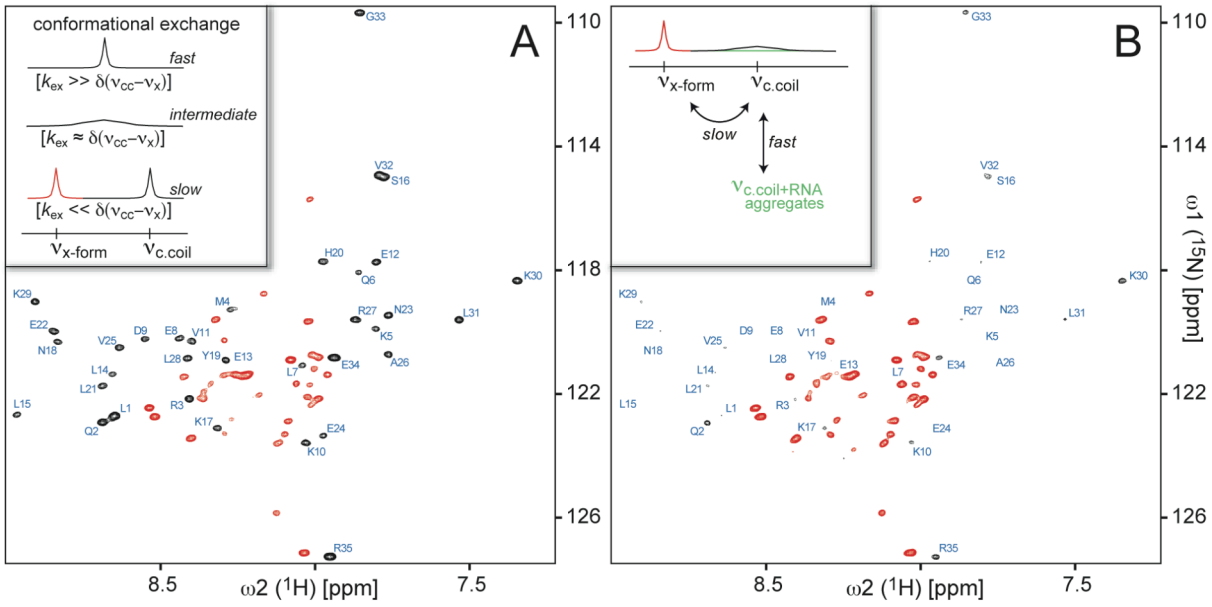


Figure 2. Catalytic conformation of the LZ-GCN4. ^{1}H-^{15}N-HSQC spectra of 15 µM LZ-GCN4 (**A**) and 15 µM LZ-GCN4 in presence of 5 µM dsRNA, in 50 mM D3-Acetate, 40 mM KCl, pH 4.0 at 37ºC. Black contours – resonances of the coiled coil conformation, red contours – resonances of the x-form. Amino acid assignment corresponds to resonances of the coiled coil conformation. Inset (**A**) illustrates the LZ resonance lineshape behaviour depending on the timescale of the exchange process. Inset (**B**) illustrates the kinetic behaviour dictating LZ resonance line broadening in presence of dsRNA substrate.

^{1}H-^{15}N-HSQC spectra of the ^{15}N-labelled 42-residue LZ-cJun at conditions approaching those of LZ-GCN4 RNase activity assays (27 uM peptide, pH 6.0) shows 40 distinct resonances (Supporting information, Figure S1), corresponding to the poorly structured monomer species as can be judged from the dispersion of resonances and chemical shift values reported for the LZ-cJun coiled coil conformation (*15*). This correlates with high unfolding equilibrium constant of LZ-c-Jun peptide (*9*). As in the case of LZ-GCN4 x-form, observed monomeric LZ-cJun species does not directly interact with dsRNA (no shifts are observed even at high substrate concentration). However decrease in LZ-cJun monomer population with increasing RNA concentration, manifested as coherent drop in amplitudes of all resonances (Supporting information, Figure S1), suggests that observed species is in a slow equilibrium with a low populated coiled coil conformation interacting with RNA. ^{1}H-^{15}N-HSQC experiments at

conditions stabilizing LZ-cJun folded conformation (~250 µM peptide, pH 7), show that the coiled coil dimer indeed interacts with the dsRNA substrate in a specific manner (Figure 3).

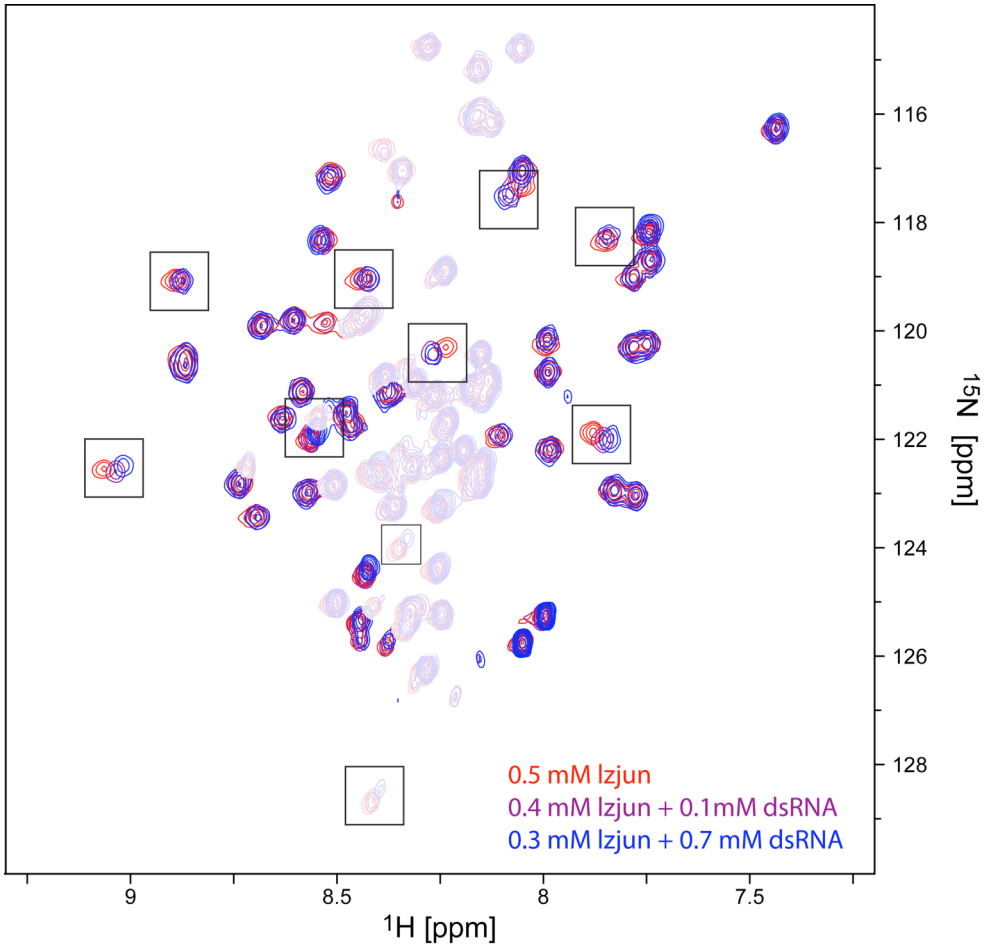


Figure 3. Catalytic conformation of the LZ-cJun. ^{1}H-^{15}N-HSQC spectra of ^{15}N-labelled LZ-cJun peptide at RNA:LZ ratios of 500:0 µM (red), 100:400 µM (magenta) and 700:300 µM (blue). Resonances of the LZ-cJun monomeric conformation (traced by increasing peptide concentration from 25 to 500 µM, see Supporting Information, Figure S1).

Active site

To characterize these interactions at physiological environment and map the amino acid residues involved in catalysis, the LZ-GCN4 binding to dsRNA was monitored at conditions replicating those employed in the RNase activity assays (20 mM HEPES-D18, 80 mM KCl, pH 7.1). ^{1}H and ^{15}N chemical shift perturbations show that LZ-GCN4 indeed interacts with the dsRNA in a specific manner, corroborating proximity of Leu7 to the binding site and

extending the interaction interface to residues Lys5, Lys10, and Glu12 (Figures 4 and 5). Due to the fast exchange of the Arg3 backbone amide proton, its resonances are not visible in neutral pH. As expected, the RNA substrate is hydrolyzed over time, resulting in return of the perturbed resonances to their original frequencies (Figure 4B).

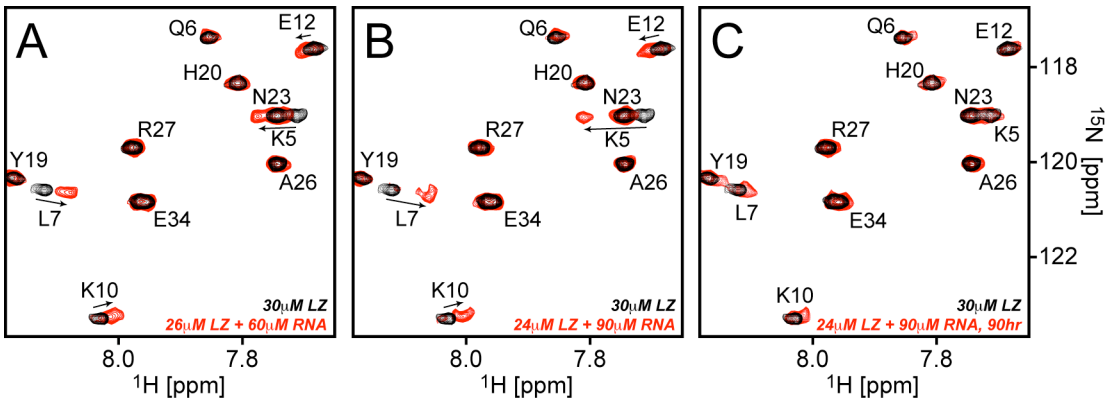


Figure 4. Mapping catalytic site of LZ-GCN4. ^{1}H-^{15}N-HSQC spectra of ~25 μM ^{15}N-labelled LZ-GCN4 in presence of 60 μM dsRNA (**A**), 90 μM dsRNA (**B**) and 90 μM dsRNA after ~90 hour incubation (**C**). Reaction performed in 20 mM HEPES-D18, 80 mM KCl, pH 7.1 at 37ºC. Black contours illustrate 30 μM LZ-GCN4 reference spectrum in absence of the RNA substrate.

Similar behavior is observed in experiments studying LZ-cJun interactions with dsRNA. However, in this case chemical shift perturbation data show that the RNA binding interface is apparently extended, involving about ten LZ-cJun residues (Figure 3), in contrast to five in the case of LZ-GCN4. Unfortunately observed amide resonances of the LZ-cJun coiled coil conformation do not unambiguously correlate with those reported in the literature (*15*), most likely because of artificial N-terminal cross-links used for coiled coil stabilization in the published studies. Currently expression and analysis of ^{15}N-^{13}C doubly labelled peptide is under way to facilitate direct assignment of the LZ-cJun active site.

End-products of catalysis

LC-MS analysis of the RNA degradation products at the point when no more peptide-RNA interaction was observed by NMR have shown that dsRNA is in fact not hydrolyzed completely. Even after seven weeks of incubation some tri-, tetra-, penta- and even hexa-nucleotides are still observed in the reaction mixture (data not shown), corresponding mainly to the regions of double-strand complementarity as derived from thermodynamic calculations with MFOLD (*14*).

LZ-GCN4 kinetic model

Within employed concentration range (~25 uM LZ & 0-90 uM dsRNA), based on fast-exchange two-state ($LZ + RNA \rightarrow [LZ \bullet RNA]$) interaction model, chemical shift perturbation data allows extraction of equilibrium dissociation constant only for residues D9 and K10 (Figure 5), yielding values of 161±232 and 289±105 (SE) μM (LZ monomer concentrations). Other affected residues (K5, L7 and E12) portray almost linear dependence of the chemical shift on the substrate concentration, implying a much higher dissociation constant. Additional experiments are under way aiming to extend the effective range of LZ:RNA concentration ratio, thus improving the accuracy of obtained dissociation constants.

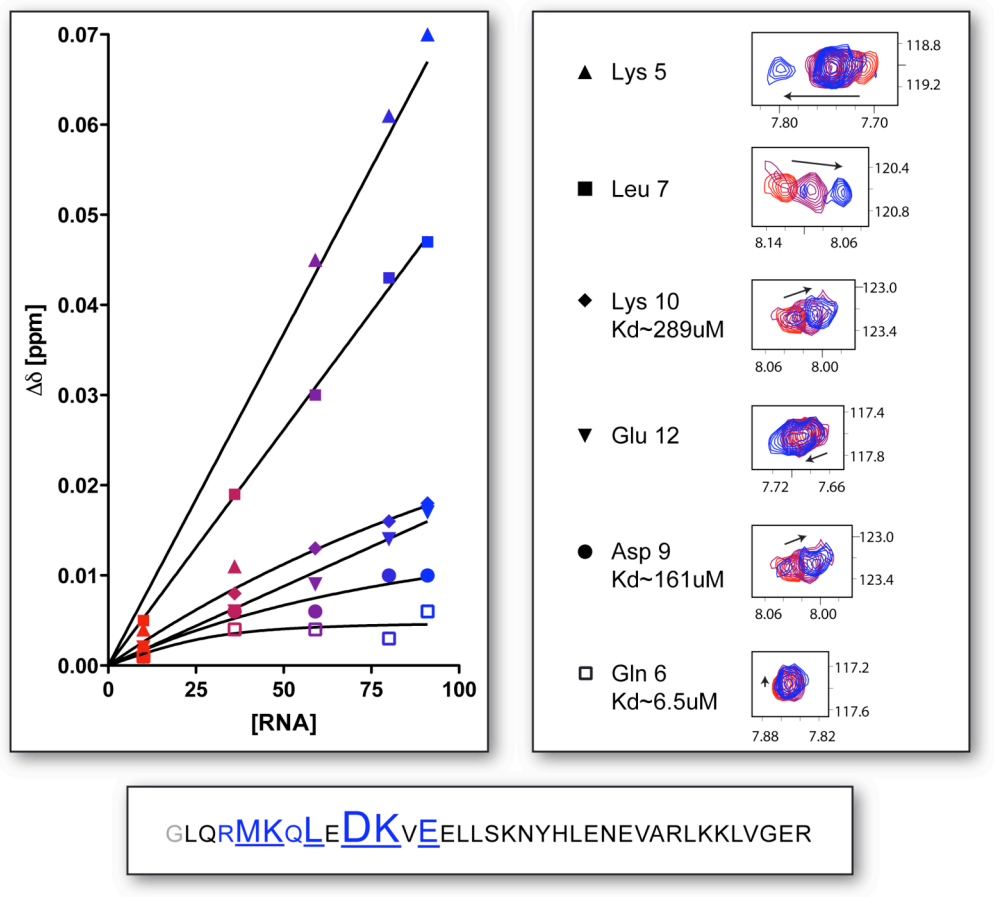


Figure 5. Characterization of $[LZ\text{-}GCN4]_2$•RNA ensemble thermodynamic properties. Data on the chemical shift perturbation of particular amino acid resonances upon RNA titration (**B**) fitted to the kinetic rate equation ($LZ + RNA \underset{k-1}{\overset{k1}{\leftrightarrow}} LZ \bullet RNA$) to obtain equilibrium dissociation constant ($K_d = \frac{k_{-1}}{k_1}$).

DISCUSSION

Catalytic degradation of RNA is an essential part of RNA turnover and decay, one of the major mechanisms for regulation of gene expression, quality control of RNA biogenesis and antiviral defenses (*16-19*). Therefore, ribonuclease activity exhibited by bZIP transcription factors may have wide biological implications because of their antiquity and prevalence among protein interaction motifs. To identify possible role of LZs RNase activity we have examined substrate specificity and catalytic properties of leucine zippers GCN4 and cJun.

Substrate specificity

Our previous results suggested that LZ activity is independent of RNA substrate sequences (*13*). Leucine zippers are simply too small to provide sufficient nucleotide-discriminating determinants, which is well supported by the notion that even the smallest of naturally occurring non-specific endoribonucleases are considerably longer than bZIP LZ motifs (*20*)(*21*). To elucidate a possibility of catalytic activity being RNA structure-dependent, we have compared the activity of LZ-GCN4 on two 18-mer RNA molecules (dsRNA and ssRNA) distinguished by their ability to form intramolecular double-stranded structures (Figure 1A-D). HPLC-MS results show that LZ-GCN4 is more catalytically active towards the single-stranded regions of RNA. Time-resolved dsRNA degradation by LZ-cJun peptide also supports this conclusion (Figure 1E).

Lower hydrolytic susceptibility of double-stranded RNA structures in the active sites of single-strand preferring endoribonucleases usually comes from the fact that nucleotide base stacking restricts the conformational space of phosohodiester backbone. This hinders formation of the transition intermediate required for the in-line phosphodiester transfer mechanism employed by these enzymes (*22*). In fact ssRNA-preferring enzymes are also able to cleave double stranded RNAs, but with considerably slower rates, limited by probability of RNase molecule binding to transient single-stranded substrate conformations induced by spontaneous thermal fluctuations (*23*).

In this perspective it has been consistently suggested that efficient dsRNA catalysis by single-strand preferring protein endoribonucleases requires additional binding determinants to disrupt the double-stranded structure. For example, a specific set of auxilary noncatalytic basic residues facilitates fastening and unwinding of ds structures in the case of pancreatic-type RNases ((*24*) and reviewed in (*25*)). Going further to double-strand specialization, enzyme classes like RNase III and RNase H bear distinct RNA-binding domains (*26*)(*27*) and catalyze depolymerization of RNA via divalent-metal ion dependent mechanisms (*28*)(*29*). Isolated leucine zipper motifs apparently lack any of the listed above determinants, and

therefore their inefficiency in transesterification of double stranded RNA structures is of a little surprise. However, a comparison of the LZ equilibrium dissociation rates for dsRNA and ssRNA is required to order unequivocally dissect transesterification kinetics from binding thermodynamics as the rate-limiting step.

Active conformation

Trial crystallization attempts aimed at structure elucidation of the catalytic complex weren't successful due to the limited solubility of 3-5 nt-long deoxyribonucleic acid oligos, used as model non-degradable substrates. In addition low catalytic rate constant and poor complex visibility in MALDI-MS underscored fragility of the LZ-GCN4–RNA interactions (*13*). This encouraged us to employ solution NMR as the only method allowing thermodynamically weak interactions to be probed at atomic resolution.

First NMR trials led to the discovery of the LZ-GCN4 x-form (*12*) – an alternative conformation of the LZ-GCN4 that exists in equilibrium with the classical coiled coil state and might be substantially populated *in vivo*. Therefore, our first aim was to elucidate which of the two observed conformations is catalytically active. At this stage application of NMR provided precise attribution of catalytic determinants even in the presence of aggregation and conformational exchange. Our data show that the substrate interacts only with the dimeric coiled coil conformation, leaving monomeric forms intact (Figures 2 and 3). In the simplest form this coupling of catalysis to particular oligomeric state may create an efficient mechanism for sensing protein concentration (*30*). However, in the case of DNA-binding transcription factors it also enables a whole array of regulatory stimuli modulating TF activity – including compartmentalization (*31*), cofactor interactions (*32*)(*33*), covalent modifications (*34*)(*35*)(*36*) and etc. Most importantly, strict association of catalysis with particulr LZ oligomeric state allows coupling of this enzymatic activity with specific state of LZ signalling pathway, thus enabling direct interconnection of protein signal transduction with the other molecular levels of information flow within the cell.

Active site

Based on the activity assays of single alanine mutants, we have previously reported the LZ-GCN4 active site localization within the 2nd and 3rd heptad repeats (residues E12, E13, S16 and, possibly, H20). Our NMR data suggest that the RNA interaction interface is rather located within the first two heptad repeats of LZ-GCN4, perturbing residues R3, K5, L7, D9, K10 and E12 (Figures 4, 5 and 6). With the exception of E12 this contradicts the previously reported data. Most likely false-positive effects of the single alanine mutations shall be attributed to the correlation of peptide activity with the local topology of the active site, as well as alterations in equilibrium between the catalytically active (coiled coil) and inactive (x-form and higher-order oligomers) forms of the LZ-GCN4. Effects of amino-acid substitutions on thermodynamic stability of the LZ-GCN4 and other coiled coils has been studied extensively (*10, 37-39*), signifying importance of both electrostatic and hydrophobic interactions for coiled coil stability and therefore catalytic activity as well.

LZ-GCN4 catalytic mechanism

Based on the chemical shift perturbation data we propose that LZ-GCN4 catalytic site is comprised of residues R3, K5, K10 and E12 and that transesterification follows the classical acid-base mechanism (*40*). In our model Glu12 acts as a catalytical base to deprotonate 2'OH of the ribose phosphate, the resulting 2'O^- nucleophile attacks the adjacent phosphodiester bond producing a pentavalent transition state, followed by Arg3 (catalytical acid) protonating the 5' leaving group which results in formation of a 2',3'-cyclic phosphate. In this model Lys5 and Lys10 stabilize the oligonucleotide in the active site via ionic interactions at ***i-1*** and ***i+1*** phosphate backbone moieties (Figure 6). This conclusion is corroborated by spatial matching of the active site dimensions with that of oligo-RNA phosphate backbone, as showed by modeling of the 3D structure of the LZ-GCN4–RNA complex (Supporting information, Figure S2). This model lacks one determinant commonly attributed to the in-line transesterification mechanism, namely – polar sidechain (Lys41 in the case of RNase A) that stabilizes the developing negative charge of the pentavalent transition state (*41*). Within the

proposed topology of LZ-GCN4 catalytic site this function could be attributed to Gln6. Albeit chemical shift of the neighbouring Leu7 is strongly affected by the interaction, Gln6 shows only minor chemical shift perturbation, not allowing unequivocal conclusions to be made. The proposed active site topology explains our earlier observations that LZ-GCN4 is unable to hydrolyze RNA dinucleotides – in this arrangement productive binding requires at least three phosphodiester bonds in the substrate molecule (Figure 6). Factual proof of the proposed catalytic mechanism based on a combined study involving mutagenesis and NMR is under way.

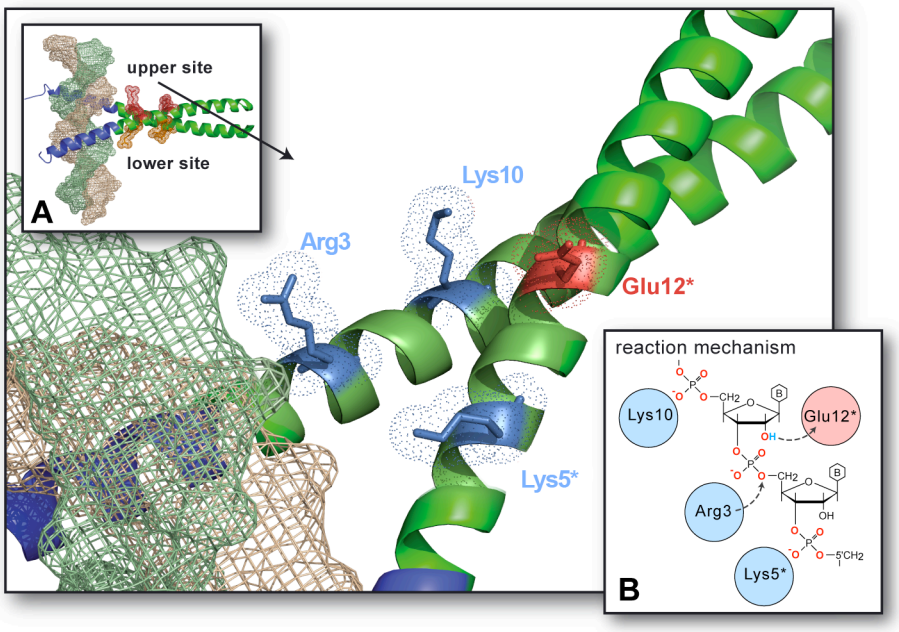


Figure 6. LZ-GCN4 active site and proposed RNase reaction mechanism. Active site residues (marked blue and red) of LZ-GCN4 mapped on the structure of GCN4 bound to DNA (pdb: 1ysa). (**A**) Side view on GCN4:DNA complex (pdb:1ysa) with two symmetrical catalytic sites marked in red and orange colour. (**B**) Proposed mechanism of transesterification reaction: Glu12 acts as a catalytic base, and Arg3 - as a catalytic acid.

Alternative metal ion-dependent phosphoryl transfer mechanism of RNA hydrolysis (*42*) can be ruled out relying on the proofs of independence of LZ-GCN4 catalytic activity on divalent ions (*13*).

Products of hydrolysis

The widely accepted acid-base mechanism of enzymatic RNA hydrolysis, beyond the transesterification step reaction yielding a 2', 3'-cyclic phosphate (cP), involves hydrolysis of this cyclic phosphate to form a 3'-phosphate in a second, independent step. However in the case of LZ-GCN4, even at extensive incubation times (7 weeks) 3'-linear phosphates comprise negligible (<1%) fractions of the corresponding 2', 3'-cP oligonucleotides (data not shown). This implies inability of LZ-GCN4 to stabilize this cyclic intermediate in the active site, and proceed with the second step of hydrolysis. Nevertheless, this fact does not counterforce the possible biological relevance of the LZ RNase activity, as it has been noted that ribonucleases as enzymes have evolved primarily to catalyze RNA transphosphorylation (transesterification) step, but not RNA hydrolysis (*43*).

K_d and kinetic model

Non-uniform equilibrium dissociation rates (fast at K5 and E12, and intermediate at D9 and K10) unveil possible order of events during binding of 18-mer RNA – with the first high-affinity interaction occurring at K10, followed by extension to low-affinity determinants involving residues R3, K5 and E12. Observed residue resonance splitting (Figure 5), on the top of the overall fast exchange between the free and bound LZ conformaitons, reflects the presence of two symmetrical binding sites within the peptide moiety (Figure 6, inset A). As inferred from the signal intensities, these sites are unequally populated, suggesting possible allosteric modulation between the two sites.

Diversity of the substrate and peptide conformational ensembles, together with allosteric effects between catalytic sites, is hindering extraction of the exact kinetic parameters within employed experimental setup. This urges for simplification of the substrate component and

derivation of a more sophisticated kinetic model. In addition the current interpretation is only valid for an isolated LZ domain, while ***in vivo*** sequence of events might be influenced by proximity of the DNA-binding interface. Studies of the full-length DNA-binding BRLZ fragment of cJun are envisaged to elucidate the role of DNA consensus-mediated TF dimerization for the bZIP catalytic activity.

Active site preservation among bZIPs

Sequence alignment of leucine zippers GCN4 and cJun shows that the active site exhibited by GCN4 is not positionally conserved in cJun (Figure 7A), suggesting considerable flexibility in topological definition of particular catalytic function. Expansion of the RNA-binding interface in LZ-cJun homodimer, apparent from HSQC titration experiments, suggests an increased control of the LZ catalytic specificity in higher eukaryote species. As evident from the alignment of all 281 factors from global bZIP family (PROSITE: ID BZIP; AC PS50217) (alignment not shown), LZ-GCN4 active site is not conserved across the whole family. However three other bZIP factors (YAP2 and HAC1 from Fungi; and NF2L1 conserved within higher eukaryotes) bear [**RK**]-x-[**RK**]-x(4)-[**RK**]-x-[**ED**] residue pattern characteristic for LZ-GCN4 active site (Figure 7A), suggesting these might possess similar catalytic properties. Importantly, this linear analysis does not address the potential of catalytic site assembly within asymmetric bZIP complexes. As exemplified by human bZIP heterodimer CEBPB/ATF4 (Figure 7B), which is not revealed by linear [**RK**]-x-[**RK**]-x(4)-[**RK**]-x-[**ED**] pattern search, analysis of sidechain topology in the context of 3D structure is essential for accurate assessment of potential catalytic properties of LZ motifs.

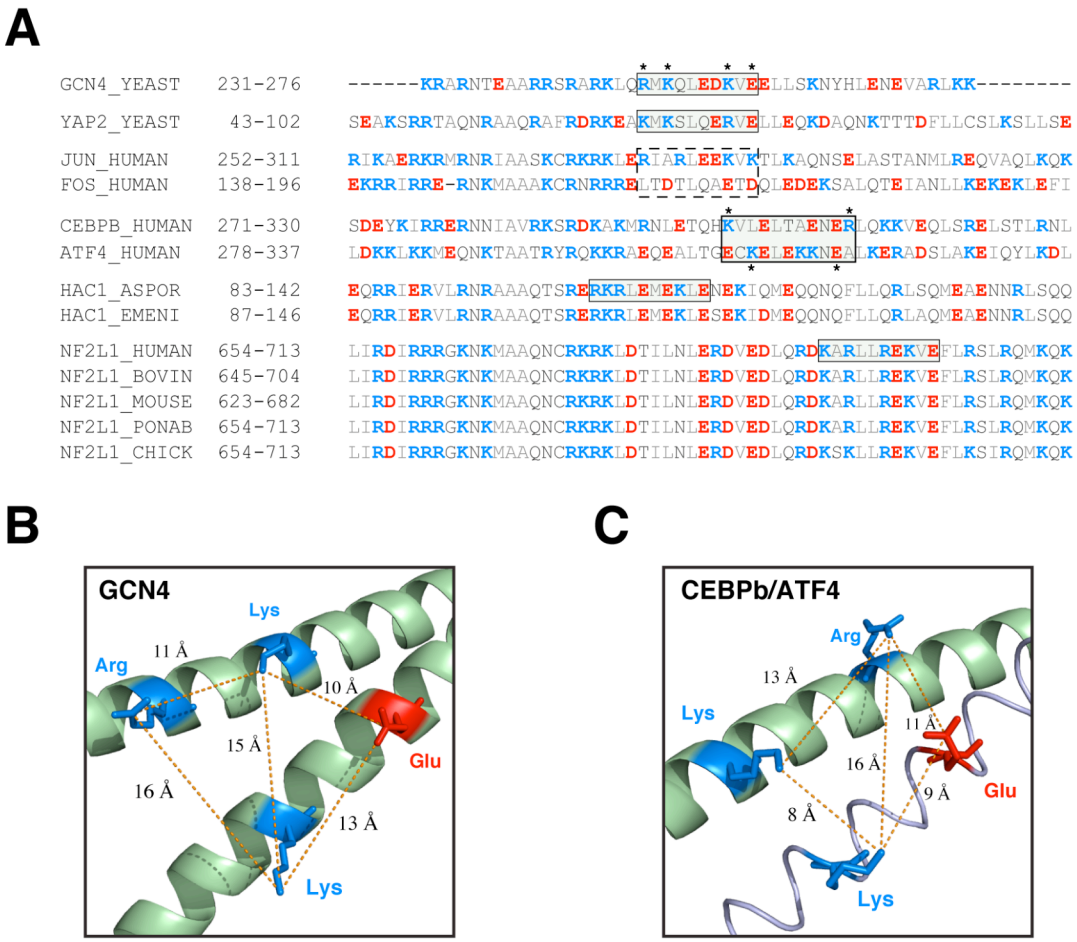


Figure 7. (**A**) Alignments of bZIP factors bearing [RK]-x-[RK]-x(4)-[RK]-x-[ED] consensus (marked with rectangle) within basic region leucine zipper region. Additionally BR-LZ sequences of human factors Jun, Fos, Atf4 and CEBP-beta are shown. (**B**) Topology of GCN4 homodimer catalytic site (based on pdb:1ysa). (**C**) Proposed topology of possible CEBPB/ATF (colored palegreen/lightblue respectively) heterodimer catalytic site (based on pdb:1ci6).

Biological role

A set of experimental facts suggests a possible rationale behind bZIP catalytic activity. First, the active site is formed on top of the coiled coil interaction interface involving residues from both monomer chains (Figure 6). This restricts LZ-GCN4 catalytic activity to the dimeric form of the motif, which is supported by observations of the substrate binding to the coiled coil, but not the (monomeric) x-form conformation. Although it has been shown (*44*) that in isolated bZIP domains the "dimer" activation pathway (dimerization occurring prior to DNA

binding) can be kinetically equivalent to the “monomer” pathway (dimerization occurring after sequential binding of monomers to the target DNA), there is a considerable experimental proof for the “monomer” pathway being more relevant *in vivo* (*45, 46*). Meaning that coiled coil assembly is limited by transcription factor binding to its DNA consensus. Low catalytic rate of LZ RNase activity suggests that ***in vivo*** this activity will increasingly manifest itself upon induced spatial colocalization with substrate molecules (i.e. local increase in substrate concentration), while in the case of pure diffusion-limited substrate encounter rate, the efficiency of catalysis will tend towards zero. Together with results on substrate specificity, these observations suggest that RNase activity of bZIPs may serve as a negative feedback loop for the transcription activation process, limiting the transcript levels of activated target genes. However alternative functions, like unspecific modulation of nuclear mRNA levels or degradation of transcripts with insufficient secondary structure are comparably plausible. A comprehensive survey of bZIP catalytic activity effects on the cellular level is required to pin-point the exact ***in vivo*** role for this novel TF functionality.

Implications for future research

Leucine zipper is arguably the simplest protein folding motif bearing a very diverse set of both intra- and inter-molecular interaction determinants. The simplicity and stability of this system attracted many research groups to study kinetics and thermodynamics governing the formation of quaternary structures encoded in this minimal polypeptide sequence. In addition long interest in LZ from functional and engineering points of view has aided to the amount of research groups involved, creating a very diverse expertise on the topic. Now that LZ role has been reconditioned from a folding model to one of the smallest natural catalysts, these research groups with multidisciplinary expertise provide a very rich environment for studies of LZ structure-function relationships and investigation of the LZ catalytic activity phenomenon in the context of different research fields.

MATERIALS AND METHODS

RNA substrates were purchased from IBA GmbH (Goettingen, Germany). RNA concentrations were determined from absorbance at 260 nm, using molar extinction coefficients of 181 and 179 mM^{-1} cm^{-1} for ss and dsRNA respectively. Concentration of dsRNA samples employed in HSQC titration experiments was additionally verified from ^{1}H-NMR spectra via integration of the aromatic (H2, H6, H8) and amino protons signals of the base using methyl peak integral of L-Leucine (Sigma-Aldrich, Buchs, Switzerland) as a reference. Obtained values were consistent with the ones determined from UV absorbance. Extended magnetization recovery delay (>5s) is essential for reliable NMR-based concentration measurements to compensate for difference in relaxation times between sample (here ~6 kDa dsRNA) and reference substance (here 0.13 kDa L-Leucine).

Synthetic LZ-GCN4 was obtained from LIPAL Biochemicals (University of Zurich) and purified using standard protocols on C18 RP-HPLC. Protein purity was verified with ESI-MS and concentration determined by amino-acid analysis (FGCZ, University Zurich).

LZ-GCN4 RNase activity was assayed as following: 10-100 uM of RNA substrate was incubated with 25-50 uM synthetic LZ-GCN4 peptide (monomer concentrations) in 20 mM Tris-HCl, 80 mM KCl, pH 7.2 overnight at 37°C in presence of 0.4 U/ul of recombinant RNAsin inhibitor (Promega). All samples were prepared and analyzed in triplicates and all assays included peptide-free RNA control. Samples were analyzed on Agilent 1100 HPLC system with 5-250 C18 column (Macherey-Nagel). Fractionation was carried out with 45 min gradient elution from 20 mM ammonium acetate pH 6.5 to 97% acetonitrile.

LC-MS analysis was performed as reported earlier (*13*).

Recombinant peptides. For ^{15}N- and ^{15}N-^{13}C-labelling LZ peptides was cloned and expressed in the pHDE (pQE30-based) expression system as described previously(*12*). In the case of LZ-cJun construct, the enterokinase cleavage step was performed on ice to limit unspecific peptide degradation at secondary cleavage sites. Ultrafiltration step in the

purification procedure was replaced with Ni-NTA metal affinity separation of the His-tagged fusion partner from the LZ fraction. After HPLC purification peptides were lyophilized 3 times from acetonitrile-H_2O mixture to remove residual TFA. Peptide purity was verified by ESI-MS on Bruker MicroTOF instrument following standard protocols of the manufacturer. Peptide concentrations were determined by 1D ^1H -NMR spectra via integration of the methyl-protons peak using L-Leucine as a reference, verified by amino acid analysis (FGCZ, University of Zurich) and Tyr19 absorbance at 280nm for LZ-GCN4.

NMR Spectroscopy. Measurements were performed at 37°C on a Bruker AVANCE 600 MHz (LPC, ETH Zurich) and 800 MHz (Biozentrum, Uni Basel) spectrometers equipped with a *z*-axis gradient triple resonance cryprobes. Data were processed with XWINNMR 3.5 (Bruker Biospin, Fallanden, Switzerland) and analyzed using CARA (www.nmr.ch)(*47*). The ^1H chemical shifts were referenced to the DSS (sodium 2,2-dimethyl-2-silapentane-5-sulfonate) signal at 0 ppm, and $^{15}N/^{13}C$ chemical shifts were referenced indirectly using the $X/^1H$ gyromagnetic ratios(*48*).

Chemical shift assignments of LZ-GCN4 coiled coil and x-form conformations was reported earlier (*12*).

RNA titration. NMR titration experiments were performed in 20 mM HEPES-D18 (Spectra Stable Isotopes), 80 mM KCl, 5% D2O, pH 7.0-7.1. NMR measurements were conducted at 37°C (LZ-GCN4) and 22°C (LZ-cJun), with decreased temperature in the latter case required to drive conformational equilibrium towards stabilization of the coiled coil. A 30 µM solution of ^{15}N-labeled LZ-GCN4 was titrated with 430 µM solution of dsRNA in total of 6 increments (0, 10, 36, 59, 80 and 91uM final dsRNA concentrations). In case of LZ-cJun, RNA titration was performed starting with ~500 µM peptide concentration in a total of four increments with 3.43 mM dsRNA solution and final RNA:LZ ratios of 100:460, 300:420, 470:350 and 670:330 µM. Monomer LZ-cJun titration (Supporting info, Figure S1) was performed in the mixture of 20 mM D18-HEPES and 50 mM D3-Acetate buffers at ~25 µM LZ and dsRNA concentration varied between 0 and 120 µM.

LZ-GCN4 Kd calculation. Equilibrium dissociation constants were determined by monitoring the chemical shift changes of LZ-GCN4 individual cross-peaks in the ^{1}H-^{15}N-HSQC spectra as a function of the dsRNA concentration. The absolute values of the chemical shift changes (calculated as $\Delta\delta_{obs} = \sqrt{\frac{(\Delta\delta_H)^2 + (\frac{\Delta\delta_N}{5})^2}{2}}$) were plotted against the total dsRNA concentration. Obtained correlations were fit into the two-state interaction model:

$$\Delta\delta_{obs} = \Delta\delta_{\max} \times \frac{(K_d + RNA_{tot} + LZ_{tot}) - \sqrt{(K_d + RNA_{tot} + LZ_{tot})^2 - 4 \times RNA_{tot} \times LZ_{tot}}}{2 \times RNA_{tot}}$$

taking into account the change in total peptide concentration (LZ_{tot}) and performing nonlinear regression analysis using both LZ_{tot} and RNA_{tot} as codependent variables. Taking into account polymeric nature of the substrate, and ~10 hr difference between the first and the last HSQC measurements – the apparent Kd values might be higher than the ones obtained from regression analysis.

REFERENCES

1. Amoutzias, G. D., Robertson, D. L., Van de Peer, Y., and Oliver, S. G. (2008) Choose your partners: dimerization in eukaryotic transcription factors, *Trends Biochem Sci 33*, 220-229.
2. Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins, *Science 240*, 1759-1764.
3. Murre, C., McCaw, P. S., and Baltimore, D. (1989) A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins, *Cell 56*, 777-783.
4. Deppmann, C. D., Alvania, R. S., and Taparowsky, E. J. (2006) Cross-species annotation of basic leucine zipper factor interactions: Insight into the evolution of closed interaction networks, *Mol Biol Evol 23*, 1480-1492.
5. Tupler, R., Perini, G., and Green, M. R. (2001) Expressing the human genome, *Nature 409*, 832-833.
6. Ledent, V., Paquet, O., and Vervoort, M. (2002) Phylogenetic analysis of the human basic helix-loop-helix proteins, *Genome Biol 3*, RESEARCH0030.
7. Vinson, C., Acharya, A., and Taparowsky, E. J. (2006) Deciphering B-ZIP transcription factor interactions in vitro and in vivo, *Biochimica Et Biophysica Acta-Gene Structure and Expression 1759*, 4-12.
8. Steinmetz, M. O., Jelesarov, I., Matousek, W. M., Honnappa, S., Jahnke, W., Missimer, J. H., Frank, S., Alexandrescu, A. T., and Kammerer, R. A. (2007) Molecular basis of coiled-coil formation, *Proc Natl Acad Sci U S A.*
9. d'Avignon, D. A., Bretthorst, G. L., Holtzer, M. E., Schwarz, K. A., Angeletti, R. H., Mints, L., and Holtzer, A. (2006) Site-specific experiments on folding/unfolding of Jun coiled coils:

thermodynamic and kinetic parameters from spin inversion transfer nuclear magnetic resonance at leucine-18, *Biopolymers 83*, 255-267.

10. Mason, J. M., Schmitz, M. A., Muller, K. M., and Arndt, K. M. (2006) 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 103*, 8989-8994.
11. Rieker, J. D., and Hu, J. C. (2000) Molecular applications of fusions to leucine zippers, *Methods Enzymol 328*, 282-296.
12. Nikolaev, Y., and Pervushin, K. (2007) NMR Spin State Exchange Spectroscopy Reveals Equilibrium of Two Distinct Conformations of Leucine Zipper GCN4 in Solution, *J Am Chem Soc 129*, 6461-6469.
13. Deillon, C., Nikolaev, Y., Hoffman, S., Bigler, L., Friess, S. D., Zenobi, R., Pervushin, K., Hunziker, P., and Gutte, B. (2009) The Leucine Zippers of the Transcription Factors GCN4 and c-Jun Have Ribonuclease

Activity, in *(in press)*.

14. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction, *Nucleic Acids Res 31*, 3406-3415.
15. Riley, L. G., Junius, F. K., Swanton, M. K., Vesper, N. A., Williams, N. K., King, G. F., and Weiss, A. S. (1994) Cloning, expression, and spectroscopic studies of the Jun leucine zipper domain, *Eur J Biochem 219*, 877-886.
16. Newbury, S. F. (2006) Control of mRNA stability in eukaryotes, *Biochem Soc Trans 34*, 30-34.
17. Mata, J., Marguerat, S., and Bahler, J. (2005) Post-transcriptional control of gene expression: a genome-wide perspective, *Trends Biochem Sci 30*, 506-514.
18. Meyer, S., Temme, C., and Wahle, E. (2004) Messenger RNA turnover in eukaryotes: pathways and enzymes, *Crit Rev Biochem Mol Biol 39*, 197-216.
19. Wilusz, C. J., and Wilusz, J. (2004) Bringing the role of mRNA decay in the control of gene expression into focus, *Trends Genet 20*, 491-497.
20. Irie, M. (1999) Structure-function relationships of acid ribonucleases: lysosomal, vacuolar, and periplasmic enzymes, *Pharmacology & therapeutics 81*, 77-89.
21. Deshpande, R. A., and Shankar, V. (2002) Ribonucleases from T2 family, *Critical reviews in microbiology 28*, 79-122.
22. Kuznetsova, I. L., Zenkova, M. A., Gross, H. J., and Vlassov, V. V. (2005) Enhanced RNA cleavage within bulge-loops by an artificial ribonuclease, *Nucleic Acids Res 33*, 1201-1212.
23. Sorrentino, S., and Libonati, M. (1997) Structure-function relationships in human ribonucleases: main distinctive features of the major RNase types, *FEBS Lett 404*, 1-5.
24. Sorrentino, S., Naddeo, M., Russo, A., and D'Alessio, G. (2003) Degradation of double-stranded RNA by human pancreatic ribonuclease: crucial role of noncatalytic basic amino acid residues, *Biochemistry 42*, 10182-10190.
25. Libonati, M., and Sorrentino, S. (2001) Degradation of double-stranded RNA by mammalian pancreatic-type ribonucleases, *Methods Enzymol 341*, 234-248.
26. Bycroft, M., Grunert, S., Murzin, A. G., Proctor, M., and St Johnston, D. (1995) NMR solution structure of a dsRNA binding domain from Drosophila staufen protein reveals homology to the N-terminal domain of ribosomal protein S5, *Embo J 14*, 3563-3571.
27. Evans, S. P., and Bycroft, M. (1999) NMR structure of the N-terminal domain of Saccharomyces cerevisiae RNase HI reveals a fold with a strong resemblance to the N-terminal domain of ribosomal protein L9, *J Mol Biol 291*, 661-669.
28. Lima, W. F., Wu, H., and Crooke, S. T. (2001) Human RNases H, *Methods Enzymol 341*, 430-440.

29. Sun, W., Pertzev, A., and Nicholson, A. W. (2005) Catalytic mechanism of Escherichia coli ribonuclease III: kinetic and inhibitor evidence for the involvement of two magnesium ions in RNA phosphodiester hydrolysis, *Nucleic Acids Res 33*, 807-815.

30. Renatus, M., Stennicke, H. R., Scott, F. L., Liddington, R. C., and Salvesen, G. S. (2001) Dimer formation drives the activation of the cell death protease caspase 9, *Proc Natl Acad Sci U S A 98*, 14250-14255.

31. Waldmann, I., Walde, S., and Kehlenbach, R. H. (2007) Nuclear import of c-Jun is mediated by multiple transport receptors, *J Biol Chem 282*, 27685-27692.

32. Baranger, A. M. (1998) Accessory factor-bZIP-DNA interactions, *Curr Opin Chem Biol 2*, 18-23.

33. Boer, U., Eglins, J., Krause, D., Schnell, S., Schofl, C., and Knepel, W. (2007) Enhancement by lithium of cAMP-induced CRE/CREB-directed gene transcription conferred by TORC on the CREB basic leucine zipper domain, *Biochem J 408*, 69-77.

34. Takemori, H., Kajimura, J., and Okamoto, M. (2007) TORC-SIK cascade regulates CREB activity through the basic leucine zipper domain, *Febs J 274*, 3202-3209.

35. Shifera, A. S., Friedman, J. M., and Horwitz, M. S. (2008) IKKgamma (NEMO) is involved in the coordination of the AP-1 and NF-kappaB pathways, *Mol Cell Biochem 310*, 181-190.

36. Karanam, B., Wang, L., Wang, D., Liu, X., Marmorstein, R., Cotter, R., and Cole, P. A. (2007) Multiple roles for acetylation in the interaction of p300 HAT with ATF-2, *Biochemistry 46*, 8207-8216.

37. Marti, D. N., and Bosshard, H. R. (2003) Electrostatic interactions in leucine zippers: thermodynamic analysis of the contributions of Glu and His residues and the effect of mutating salt bridges, *J Mol Biol 330*, 621-637.

38. Zeng, X., Zhu, H., Lashuel, H. A., and Hu, J. C. (1997) Oligomerization properties of GCN4 leucine zipper e and g position mutants, *Protein Sci 6*, 2218-2226.

39. Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993) A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants, *Science 262*, 1401-1407.

40. Findlay, D., Herries, D. G., Mathias, A. P., Rabin, B. R., and Ross, C. A. (1961) The active site and mechanism of action of bovine pancreatic ribonuclease, *Nature 190*, 781-784.

41. Raines, R. T. (1998) Ribonuclease A, *Chem Rev 98*, 1045-1066.

42. Scott, W. G. (1999) RNA structure, metal ions, and catalysis, *Curr Opin Chem Biol 3*, 705-709.

43. Sorrentino, S. (1998) Human extracellular ribonucleases: multiplicity, molecular diversity and catalytic properties of the major RNase types, *Cell Mol Life Sci 54*, 785-794.

44. Cranz, S., Berger, C., Baici, A., Jelesarov, I., and Bosshard, H. R. (2004) Monomeric and dimeric bZIP transcription factor GCN4 bind at the same rate to their target DNA site, *Biochemistry 43*, 718-727.

45. Kohler, J. J., and Schepartz, A. (2001) Kinetic studies of Fos.Jun.DNA complex formation: DNA binding prior to dimerization, *Biochemistry 40*, 130-142.

46. Kohler, J. J., Metallo, S. J., Schneider, T. L., and Schepartz, A. (1999) DNA specificity enhanced by sequential binding of protein monomers, *Proc Natl Acad Sci U S A 96*, 11735-11739.

47. Keller, R. L. J. (2004) *The Computer Aided Resonance Assignment Tutorial*, 1st ed., CANTINA Verlag, Germany.

48. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) H-1, C-13 and N-15 Chemical-Shift Referencing in Biomolecular Nmr, *Journal of Biomolecular Nmr 6*, 135-140.

SUPPORTING INFORMATION

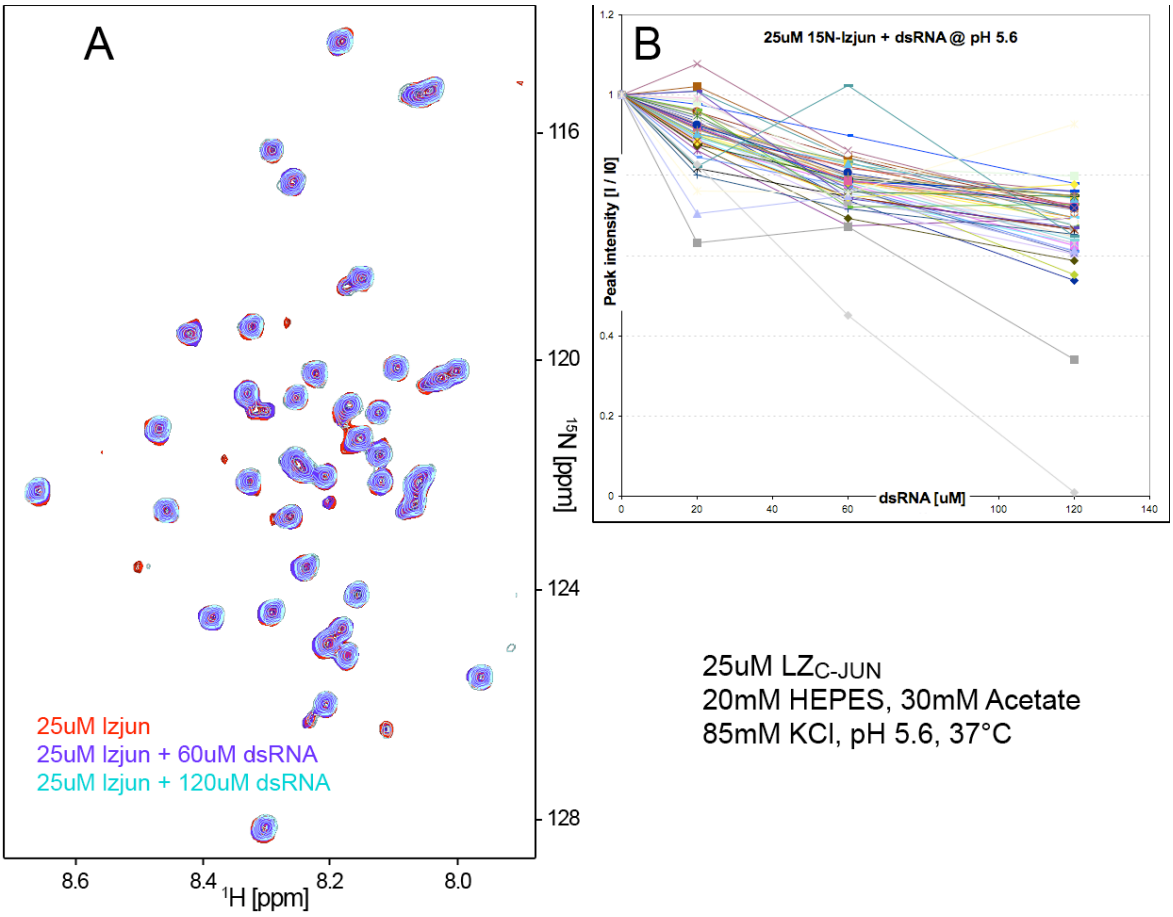


Figure S1. (**A**) ^{1}H-^{15}N-HSQC-monitored titration of 25 µM ^{15}N-labelled LZ-cJun peptide by 0-120 µM dsRNA in 20 mM HEPES, 30 mM Acetate, 85 mM KCl, pH 5.6 and 37ºC. (**B**) Dependence of LZ-cJun resonance amplitudes on the total concentration of dsRNA substrate.

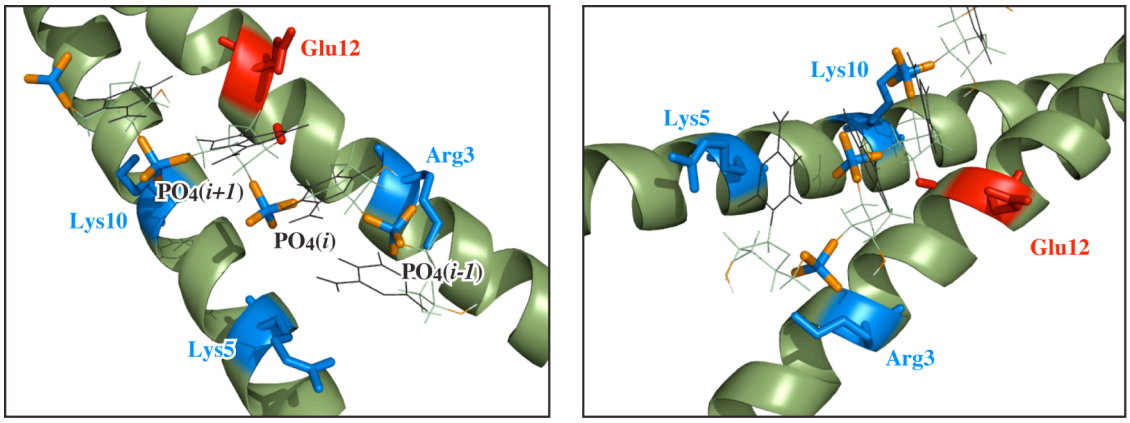


Figure S2. Complementary views of LZ-GCN4–RNA complex 3D model, revealing spatial matching of the RNA phosphate backbone to the LZ active site dimensions. Lysine and arginine sidechains of the LZ-GCN4 active site, as well as phosphate atoms of RNA backbone are coloured blue. Glutamate 12 sidechain (catalytic base) and activated 2'-hydroxy group of RNA molecule are marked red. Oxygen atoms of RNA ester backbone coloured orange.

Conclusions and Perspectives

This dissertation focuses on the characterization of functional, structural and catalytic properties of the Leucine Zipper (LZ) protein oligomerization motif. First chapter presents bibliographical analysis of LZ structural properties and their relations to the functional roles exhibited by these motifs. Chapter III presents structural NMR characterization of a novel conformational state discovered in the folding pathway of LZ from transcription factor GCN4. Chapters II and IV present biochemical and solution NMR studies of ribonuclease activity of Leucine Zippers from bZIP transcription factors GCN4 (yeast) and cJun (human).

Leucine Zippers as a signal transduction motif (Chapter I)

Bibliographical analysis presented in Chapter I reviews the structure, interaction specificity, folding characteristics and functional diversity of LZ motifs. Reasoning on the wealth of existing data, we suggest that despite of structural identity with highly stable extended "Coiled Coil" motifs, on the functional level "Leucine Zippers" stand out as a distinct group. These motifs act as specific protein-protein interaction determinants, thereby providing cells with basic signal transduction functionality. In the last part of this chapter we provide examples demonstrating prevalence of the LZ-mediated signal transduction and illustrate applicability of the "LZ interaction code" formalism to explain existing evidence of couplings between cytoplasmic and nuclear signalling networks.

Currently available LZ interaction rules are primarily derived from the studies of yeast and human bZIP proteins, and therefore reflect only a subset of possible LZ interaction space. This notion is supported by existence of "non-conventional" specificity determinants: ***a–d*** ionic bridges in the LZ motifs within Myc/Max network (1) and ***f***-position prolines in plant bZIP families (2). Neither of these interaction determinants is found in yeast or human bZIPs. Further developments in understanding of the "LZ interaction code" require more thorough sampling of natural LZ-proteins for novel specificity determinants; as well as systematic physico-chemical characterization of these determinants in relation to LZ stability and specificity.

In addition to the protein interaction functionality at the end of Chapter I we discuss the ability of LZ motifs to act as biological catalysts. Our studies of LZ RNase activity (Chapters II and IV) and research on LZ-mediated aminoacyl transfer reactions (performed in the laboratory of R. Ghadiri (3)) show that leucine zippers can act as a scaffold for establishing catalytic sites with variable properties. General independence of LZ oligomer stability of the surface residues in the ***b***, ***c*** and ***f*** positions provides substantial versatility in definition of

substrate binding and catalytic sites. Examples of LZ-mediated transesterification and aminoacyl transfer reactions lead to an intriguing question on what other (bio)catalytical functionalities may be grafted on top of the LZ scaffold.

Ribonuclease activity of Leucine Zippers (Chapters II and IV)

Studies presented in the Chapters II and IV show that Leucine Zipper motifs of bZIP factors GCN4 and cJun possess intrinsic ribonuclease activity. The activity is exhibited not only by isolated LZ motifs, but is preserved within full-length recombinant cJun protein, suggesting it may be also manifested *in vivo*. Importantly, both yeast GCN4 and human cJun leucine zippers show similar topological requirements for RNA binding and catalysis: (1) only dimeric form of either LZ can bind RNA; (2) in both cases RNA binding occurs in the two N-terminal heptads of the LZ structure; and (3) both LZs show increased catalytic activity towards single stranded RNA. This suggests that LZ ribonuclease activity originated or have been preserved under similar evolutionary pressures both in yeast and human cells.

Most valuable question for the future studies is investigation of a possible biological role of LZ ribonuclease activity. To address this question *in vivo*, one has to separate molecular determinants essential for LZ oligomerization function from those required for ribonuclease activity. Therefore, characterization of the exact mechanism of the transesterification reaction is of primary importance to permit rational investigation of LZ catalytic activity *in vivo*.

Another interesting question is possible preservation of ribonuclease activity within the general family of bZIP transcription factors, which accounts for more than 50 members in *H. sapiens* alone. Our analysis shows that RNA binding sites similar to that of LZ-GCN4 are found at least in several other bZIP proteins, and even more could be established in the context of heterodimeric LZ structures.

In case if catalytic activity is exhibited by a limited set of bZIP dimers, an intriguing question is whether RNase functionality could be associated with a particular state of transcription activation process – i.e. be dictated by relative concentration of several different bZIP factors, rather than absolute concentration of individual bZIP protein(s). If this is true, catalytic activity will become a property the bZIP network as a whole.

Leucine Zipper folding intermediates (Chapter III)

Our NMR studies of LZ-GCN4 ribonuclease activity led to the discovery of the x-form – a hitherto undetected monomeric folding intermediate that exists in equilibrium with the classical coiled coil state. X-form is considerably less structured than dimeric coiled coil conformation, but still retains semi-structured alpha-helical core. X-form is populated at about 1% at ambient pH and μM protein concentrations, but its population strongly increases

at lower concentrations and/or acidic pH. Structural characteristics of x-form <–> coiled coil equilibrium provides a valuable insight for explanation of LZ folding mechanism, supporting the hypothesis that main activation barrier in the folding direction is to a substantial extent defined by enthalpic contributions (discussed in Chapter I).

The equilibrium between monomeric and dimeric LZ conformations presented in Chapter III, establishes an attractive model for several directions of research involving LZ protein motifs.

An interesting question from biological perspective is to what extent kinetics of monomer-dimer LZ equilibrium modulate the timescales of cell signalling events mediated by LZ interactions. Reported x-form <-> coiled-coil system provides an attractive mean to develop LZ mutants exhibiting identical thermodynamic but different kinetic behaviors, which could then be employed to characterize the effect of LZ oligomerization kinetics *in vivo*.

Apart from their biological roles, LZ motifs are extensively used as interaction determinants in a variety of protein engineering studies, including basic research as well as technological and medicinal applications (4, 5). Wide range of engineering applications signifies the need for detailed understanding of LZ folding process for designing novel LZ-based protein interaction modules. High-resolution *in vitro* folding model thus provides an attractive tool for design of LZ motifs with desired properties.

Finally, taking into account the multitude of cellular functions dependent on oligomerization of LZ motifs, the reported monomer <–> dimer equilibrium between x-form and coiled coil can be employed as *in vitro* model for development of small molecule (6), as well as peptide-based (7) drugs targeting the LZ interaction interfaces.

1. Montagne, M., Naud, J. F., and Lavigne, P. (2008) *J Mol Biol* **376**(1), 141-152
2. Shen, H., Cao, K., and Wang, X. (2007) *Biochem Biophys Res Commun*
3. Huang, Z. Z., Leman, L. J., and Ghadiri, M. R. (2008) *Angew Chem Int Ed Engl* **47**(9), 1758-1761
4. Rieker, J. D., and Hu, J. C. (2000) *Meth. Enzymol* **328**, 282-296
5. Woolfson, D. N. (2005) *Advances in protein chemistry* **70**, 79-112
6. Berg, T., Cohen, S. B., Desharnais, J., Sonderegger, C., Maslyar, D. J., Goldberg, J., Boger, D. L., and Vogt, P. K. (2002) *Proc Natl Acad Sci U S A* **99**(6), 3830-3835
7. Agou, F., Courtois, G., Chiaravalli, J., Baleux, F., Coic, Y. M., Traincard, F., Israel, A., and Veron, M. (2004) *J Biol Chem* **279**(52), 54248-54257

Curriculum Vitae

Name: Yaroslav Nikolaev
Date of birth: October 10th, 1980
Place of birth: Saint-Petersburg, Russia
Nationality: Russian
Address: Bächlerstrasse 5, 8046, Zürich
Mobile: +41 76 5669787
E-mail: yar.nikolaev@gmail.com

HIGHER EDUCATION

2005 - 2009 **Ph.D. in Structural Biology**, Biozentrum, University of Basel, Switzerland.
(*2005 - 2006* performed at ETH Zurich, Switzerland).

- Supervisor Prof. K. Pervushin.
- Thesis "Rethinking Leucine Zipper: ribonuclease activity and structural dynamics of a ubiquitous oligomerization motif".

2001 - 2003 **M.Sc. in Biochemistry and Molecular Biology,** St. Petersburg State University, Russia.

- Supervisor Prof. B. Gutte, Co-supervisor Prof. E. Skvortsevich.
- Thesis "Study of endoribonuclease activity of GCN4 leucine zipper domain".

2002 - 2004 **M.Sc. qualification in Computer Design and Web Development,**
St. Petersburg State Polytechnical University, Russia.

- Graduation postponed until thesis completion.

1997 - 2001 **B.Sc. in Biology,** St. Petersburg State University, Russia.

- Supervisor Dr. M. Padkina.
- Thesis "Usage of yeast *Pichia pastoris* for recombinant bovine gamma-interferon production".

PROFESSIONAL EXPERIENCE

2007 - 2010 M.E. Müller-Institut, Biozentrum, **University of Basel** (Switzerland).
2005 - 2006 Laboratory of Physical Chemistry, **ETH Zurich** (Switzerland).
PhD student. Structural and folding NMR studies of SARS CoV protein E and *I. tartaricus* F1F0 ATP synthase. NMR and biochemical studies of mechanism and evolutionary conservation of bZIP factors RNase activity. Characterization of leucine zippers as protein signalling motifs. Conceptual and functional design of NMR experiment generator (www.nmrplus.com).

2002, 2004 Department of Biochemistry, **University Zurich** (Switzerland).
Research Assistant. Biochemical studies of GCN4 leucine zipper RNase activity.

1998 - 2003 Laboratory of Biochemical Genetics, **St. Petersburg State University** (Russia).
Bachelor / Master Student. Construction of *Pichia pastoris* strains for recombinant protein production.

1996 - 2000 Institute of Organic Chemistry, **St. Petersburg State University** (Russia).
Laboratory Assistant. Synthesis and characterization of stereoisomeric acylketenes.

SCHOOLS and COURSES

06.2010	**„Patenting in Life Sciences – Challenges and Questions“**. 2 day intensive course.
09-12.2009	**„Venture Challenge (setting up and managing a startup)“**. Semester course.
07.2008	**„Web programming using PHP, XML, SQL and AJAX**“. 2 week intensive course.
07.2007	EMBO Practical Course **„Structure determination of biomacromolecules by solution NMR“**. Basel, Switzerland.
08.2006	EMBO Theoretical course **„Multidimensional NMR in Structural Biology“.** Il Ciocco, Italy.
09.2000	Fifth John Humphrey Advanced Summer Program and Lecture Series in Immunology „**Immunology in Health and Disease**“. Pushchino, Russia.
1996 - 1997	**Academic School of St. Petersburg State University.** St. Petersburg, Russia. ▪ Major: Chemistry and Biology. ▪ Qualification: Chemical Laboratory Research Assistant.

SKILLS

Biomolecular NMR:

- Operating Bruker NMR spectrometer, spectra processing and analysis (TopSpin, XWinNMR, NMRPipe). Automated experiment setup (www.nmrplus.com).
- Protein assignment using 1H homonuclear and 1H-15N-13C heteronuclear data. Automated protein assignment using AutoLink/CARA.
- Protein interactions and dynamics, conformational exchange characterization.
- Structure calculation (CYANA).

Biochemistry and molecular biology:

- Molecular cloning and recombinant DNA technology.
- Protein expression and purification from bacteria and yeast strains.
- Liquid chromatography (HPLC, FPLC: Agilent 1100, AKTA purifier).
- LC-Mass-Spectrometry (Bruker MicroTOF).
- Protein characterization (PAGE, Circular Dichroism, ELISA, Immunoblotting, enzymatic catalysis).

IT skills:

- Proficient use of MS Windows and Mac OS X including most of the common software. Basic knowledge of *nix operating systems.
- Graphic design and publishing: Adobe Illustrator, Photoshop, ImageReady, Flash, DreamWeaver, Indesign.
- Web development: HTML, CSS, JavaScript, DHTML; familiar with XML, SQL, PHP, AJAX.
- Familiar with 3D modeling and animation in Discreet 3D StudioMax.

Organic synthesis (basic skills).

AWARDS and FELLOWSHIPS

2007 **Roche Research Foundation Fellowship**, for the work on investigation of conformational dynamics and ribonuclease activity of leucine zipper GCN4.

1999,2000,2001 **Scholarship of International Soros Science Education Program**, for outstanding academic achievements and research work.

1999, 2001 **St. Petersburg Governor's Scholarship**, for research work on construction of *Pichia pastoris* strains for recombinant interferon production.

PUBLICATIONS

* Shared corresponding authors

1. **Nikolaev Y.*** and Pervushin K. **Review:** Rethinking Leucine Zipper – a universal signal transduction motif. *(preprint available at Nature Precedings).*

2. **Nikolaev Y.***, Bigler L. and Pervushin K*. Structural basis of ribonuclease activity in basic region leucine zippers. *(submitted to JBC).*

3. **Nikolaev Y.**, Deillon C., Hoffmann S.R.K., Bigler L., Friess S., Zenobi R., Pervushin K., Hunziker P. and Gutte B. The Leucine Zipper Domains of the Transcription Factors GCN4 and c-Jun Have Ribonuclease Activity. ***PLoS ONE***, 5(5), e10765 (2010).

4. Pogoryelov D., **Nikolaev Y.**, Schlattner U., Pervushin K., Dimroth P. and Meier T. Probing the rotor subunit interface of the ATP synthase from Ilyobacter tartaricus. ***Febs J.***, 275, 4850-4862 (2008).

5. **Nikolaev Y.** and Pervushin K. NMR Spin State Exchange Spectroscopy Reveals Equilibrium of Two Distinct Conformations of Leucine Zipper GCN4 in Solution. ***J. Am. Chem. Soc.***, 129, 6461-6469 (2007).

CONFERENCE ABSTRACTS

1. **Nikolaev Y.,** Bigler L., Pervushin K. Structural basis of Leucine Zippers RNase activity. EUROMAR 2010. Florence, Italy. July 4-9, 2010.
2. **Nikolaev Y.** and Pervushin K. The X-form – a hidden intermediate of Leucine Zipper GCN4 folding pathway. 48th Experimental NMR Conference. Daytona Beach, Florida, USA. April 22-27, 2007.
3. **Nikolaev Y.**, Parfenova L., Padkina M., Smirnov M. "Pichia pastoris for recombinant γ-interferon production", International School-Conference "Cytokines. Inflammation. Immunity", St.Petersburg, Russia, June 23-26, 2002, Abstracts of papers, v.1, No.2, p.28
4. **Nikolaev Y.**, Shikhova E., Rodina L., Tidwell T. "Stereoisomeric 2,4-Dioxopyranes via [4+2] Dimerisation of the Epimeric 2-oxobutadiylketenes", 13th International Conference on Organic Synthesis, Warsaw, Poland, July 1-5, 2000, Abstracts of papers, p.211

EXTRACURRICULAR ACTIVITIES

2008 - present Coordinator of **ForeScience** project (www.forescience.org)

- Area: information management for academic research
- Activities: functional system design, market analysis, business plan development, investor presentations.
- Achievements: project received 1st funding round from Venture Kick initiative in 2009, and was rated within the second quartile in the Venture 2010 competition.

2003 - 2004 Designer and project manager at **LizArt Studio** (www.lizart-studio.com)

- Development of Web, Advertisement and Publishing solutions for: St.Petersburg State University Educational Center; Japan Tobacco International – HR Department St.Petersburg; Telphin Telecommunications; III Festival of Iberoamerican Culture.