

Dimerization specificity of all 67 B-ZIP motifs in *Arabidopsis thaliana*: a comparison to *Homo sapiens* B-ZIP motifs

Christopher D. Deppmann, Asha Acharya¹, Vikas Rishi¹, Barry Wobbes², Sjef Smeekens², Elizabeth J. Taparowsky and Charles Vinson^{1,*}

Department of Biological Sciences, Purdue University, West Lafayette, IN 47907-1392, USA, ¹Laboratory of Metabolism, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA and ²Department Molecular Plant Physiology, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

Received March 9, 2004; Revised April 16, 2004; Accepted May 26, 2004

ABSTRACT

Basic region-leucine zipper (B-ZIP) proteins are a class of dimeric sequence-specific DNA-binding proteins unique to eukaryotes. We have identified 67 B-ZIP proteins in the *Arabidopsis thaliana* genome. No *A.thaliana* B-ZIP domains are homologous with any *Homo sapiens* B-ZIP domains. Here, we predict the dimerization specificity properties of the 67 B-ZIP proteins in the *A.thaliana* genome based on three structural properties of the dimeric α -helical leucine zipper coiled coil structure: (i) length of the leucine zipper, (ii) placement of asparagine or a charged amino acid in the hydrophobic interface and (iii) presence of interhelical electrostatic interactions. Many *A.thaliana* B-ZIP leucine zippers are predicted to be eight or more heptads in length, in contrast to the four or five heptads typically found in *H.sapiens*, a prediction experimentally verified by circular dichroism analysis. Asparagine in the *a* position of the coiled coil is typically observed in the second heptad in *H.sapiens*. In *A.thaliana*, asparagine is abundant in the *a* position of both the second and fifth heptads. The particular placement of asparagine in the *a* position helps define 14 families of homodimerizing B-ZIP proteins in *A.thaliana*, in contrast to the six families found in *H.sapiens*. The repulsive interhelical electrostatic interactions that are used to specify heterodimerizing B-ZIP proteins in *H.sapiens* are not present in *A.thaliana*. Instead, we predict that plant leucine zippers rely on charged amino acids in the *a* position to drive heterodimerization. It appears that *A.thaliana* define many families of homodimerizing

B-ZIP proteins by having long leucine zippers with asparagine judiciously placed in the *a* position of different heptads.

INTRODUCTION

Basic region-leucine zipper (B-ZIP) proteins are dimeric transcription factors that bind to DNA in a sequence specific manner (1,2). This class of protein is found exclusively in eukaryotes. Previously, we predicted the genome-wide dimerization properties for all 53 *Homo sapiens* (3) and 27 *Drosophila melanogaster* (4) B-ZIP motifs. A recent experimental analysis of the leucine zipper dimerization specificity of most *H.sapiens* B-ZIP proteins (5) has verified our previous predictions. This gives us confidence that we can use the same structural rules to predict the dimerization specificity of B-ZIP proteins from other organisms. Recently, the *Arabidopsis thaliana* genome was sequenced and the B-ZIP proteins within this genome were annotated (6,7). These proteins are important for pathogen defense, light-induced signaling, seed maturation and flower development (7). Many of the features of the leucine zipper that are critical for specifying the dimerization properties of *H.sapiens* B-ZIP proteins are conserved in *A.thaliana* B-ZIP proteins. Thus, we can use our understanding of mammalian leucine zipper dimerization specificity to predict the dimerization properties of *A.thaliana* B-ZIP proteins.

When bound to DNA, each B-ZIP monomer is a long α -helix (Figure 1). The N-terminal half binds in the major groove of double-stranded DNA in a sequence-specific manner. The C-terminal half is amphipathic and can dimerize to form a parallel coiled coil, termed a leucine zipper (8). B-ZIP proteins form homodimers and/or heterodimers depending on the amino acid sequence of the leucine zipper (9). Extensive mutagenic and biophysical studies have characterized the most common interhelical amino acid interactions observed in human B-ZIP proteins by quantifying the contribution of

*To whom correspondence should be addressed. Tel: +1 301 496 8753; Fax: +1 301 496 8419; Email: vinsonc@dc37a.nci.nih.gov
Present addresses:

Christopher D. Deppmann, Department of Neuroscience, Johns Hopkins Medical School, Baltimore, MD, USA
Sjef Smeekens, Yacht Biochemistry and Life Sciences, Yacht Technology, Utrecht, The Netherlands

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

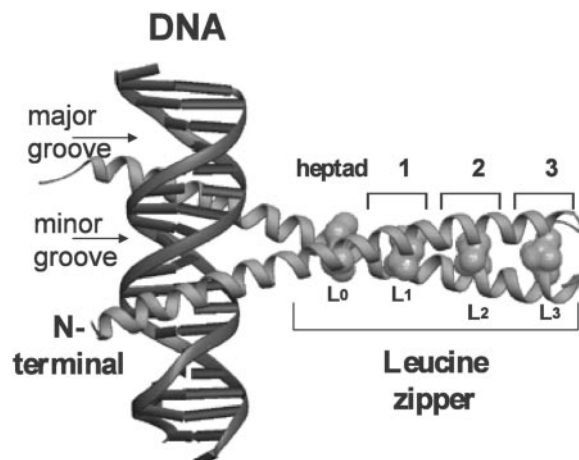


Figure 1. X-ray structure of GCN4 B-ZIP motif bound to double stranded DNA (1). The N-terminal of the protein, the basic region and leucine zipper are labeled. The first three heptads of the leucine zipper are delineated.

individual amino acids to both dimerization stability and specificity (10–13). This knowledge has allowed us to rationalize what is known about the dimerization properties of B-ZIP domains and to predict how uncharacterized B-ZIP proteins will dimerize with each other.

Each monomer in the leucine zipper has a structural repeat of two α -helical turns (seven amino acids) (supplementary Figure 1). This repeat is termed a heptad with each of the seven positions assigned the letter designation *a*, *b*, *c*, *d*, *e*, *f* and *g*. The *a*, *d*, *e* and *g* positions are near the leucine zipper interface and dictate dimerization specificity (3). Amino acids in the *a* and *d* positions lie on the same side of the α -helix and are typically hydrophobic. These hydrophobic amino acids interact interhelically with hydrophobic amino acids in the same *a* and *d* positions of the second α -helix of the leucine zipper to stabilize the structural dimer (14–16). Typically, the *d* position contains leucine that stabilizes the dimer better than other amino acids (16).

Dimerization specificity is regulated by amino acids in the *a*, *e* and *g* positions. In animal B-ZIP proteins, the second heptad *a* position typically contains asparagine which limits leucine zipper oligomerization to dimers (17,18). As depicted in supplementary Figure 1, asparagine prefers an interhelical interaction with an asparagine in the *a* position of a partner protein and will not interact efficiently with an aliphatic amino acid (12,19). Charged amino acids in the *a* position inhibit homodimer formation (12,20), while lysine residues in this position permit heterodimer formation with a wide range of amino acids (12).

The *g* and *e* positions often contain charged amino acids that form attractive electrostatic interhelical interactions (3,21–23). These interactions are denoted $g \leftrightarrow e'$ where the prime (') indicates a residue on the second α -helix of the dimeric leucine zipper. $g \leftrightarrow e'$ interactions between oppositely charged amino acids (e.g. $E \leftrightarrow R$ or $K \leftrightarrow E$) are attractive and promote dimerization specificity, while $g \leftrightarrow e'$ interactions between similarly charged amino acids (e.g. $E \leftrightarrow E$ or $R \leftrightarrow R$) are repulsive and inhibit homodimerization. For the mammalian B-ZIP protein FOS, repulsive $E \leftrightarrow E$ interactions involving glutamate residues discourage homodimer formation, while attractive

$E \leftrightarrow K$ interactions between the glutamate of FOS and the lysine of JUN promote heterodimer formation (9).

In this study, we identify 67 B-ZIP motifs in the recently sequenced *A.thaliana* genome (6). None of these proteins are homologous to human B-ZIP proteins but they have similar amino acids to regulate dimerization specificity. We have annotated these B-ZIP protein sequences and have predicted their dimerization properties. Our analysis reveals that many *A.thaliana* B-ZIP proteins have longer leucine zippers than observed in *H.sapiens*. In addition, unlike human B-ZIP proteins, dimerization of most of these zippers is not specified by attractive and repulsive $g \leftrightarrow e'$ interactions, but by the unique placement of asparagines in heptad *a* positions throughout the leucine zipper. These structural features generate 14 homodimerizing families of *A.thaliana* B-ZIP proteins, in contrast to the six homodimerizing families observed in *H.sapiens*.

METHODS

Pattern matching

The pattern matching program at <http://www.arabidopsis.org/cgi-bin/patmatch/nph-patmatch.pl> was used to identify *A.thaliana* proteins that contain putative B-ZIP motifs. Two types of regular expressions were used to query this database. The 'Basic Region' expression ([RQK][NLE][TRK]X[ASY][ASQ]XX [CSFYG][RDL]X[RK][RKL]) was modified from (24). The 'B-ZIP' expressions used were variations of ([AVK][RKASQNE][SCFYLR][RKAIFDNQ]XX XXXXXX[LIVTS]XXX[VRATSC]XX[LYVM]XXX[NKVR]XX[LIY]) previously used to identify B-ZIP proteins in *D.melanogaster* (4).

Multiple alignment and phylogenetic tree analysis

Multiple alignments were performed using ClustalW [<http://www.ebi.ac.uk/clustalw>, (25)] using the default options. Phylogenetic analysis was performed using tree review (1.6.6) (26).

Proteins

Heterodimerizing proteins containing mutations in the *a* position were generated as described previously (12). Thermodynamic parameters were determined from denaturation curves, assuming a two-state equilibrium dissociation of α -helical dimers into unfolded monomers using ΔC_p of 2.04 kcal/mole/°C, as described previously (12). Coupling energy values ($\Delta\Delta G$) are reported at 37°C.

The Opaque B-ZIP domain amino acid sequence is as follows. The first 13 amino acids are from phi10 followed by the basic region and the leucine zipper that is divided into heptads (*gabcdef*). The asparagines in the *a* position of the second and fifth heptads are in *italic*. ASMTGGQQMGRDP-EILGFKMPTEERVVRKRKESNRESARRSRYRKAAHLKELEDQVAQLKAENSCLLRRIAALNQKYN-DANVDNRVLRA DMETLRA KVKMGEDSLKRVIE MSSVPSS. Circular dichroism (CD) studies were performed in buffer containing 12.5 mM KPO₄ (pH 7.4), 150 mM KCl and 0.25 mM DTT using a Jasco J-720 spectropolarimeter as described previously (12).

RESULTS

Identification and phylogenetic analysis of *Arabidopsis thaliana* B-ZIP domains

Previous studies by two groups utilized automated motif identification software to identify 81 and 75 (7,27) B-ZIP factors, respectively, in the *A.thaliana* genome. We used the web-based TAIR genehunter 2.0 tool at <http://www.arabidopsis.org/cgi-bin/geneform/geneform.pl> to search several different databases [TAIR, NCBI, TIGR, NASC, PIR] and assembled a list of proteins identified as B-ZIP factors. Each of these proteins was inspected visually and included in our analysis if they appeared to be bona fide B-ZIP factors based on two criteria: (i) a well-defined basic region including a near invariant asparagine critical for sequence-specific DNA binding and (ii) a canonical leucine zipper region with obvious amphipathic qualities. This analysis yielded 67 B-ZIP proteins.

We also performed a search of the *A.thaliana* genome using a web-based pattern search tool at <http://www.arabidopsis.org/cgi-bin/patmatch/nph-patmatch.pl> to search known and predicted proteins for amino acid patterns representing the entire B-ZIP domain ([RKFS]XXXNXX[ASYK][AVK][RKASQNE][SCFYLR][RKAIFDNQ]XXXXXXXXX[LIVTS]XXX[VRATSC]XX[LYVM]XXX[NKVR]XX[LIY]) or just the basic region ([RQK][NLE][TRK]X[ASY][ASQ]XX[CSFYG][RDL]X[RK][RKL]). These patterns have been used successfully to identify B-ZIP factors in the genomes of *Saccharomyces cerevisiae* (24), *D.melanogaster* (4) and *H.sapiens* (3). This strategy identified the same 67 *A.thaliana* B-ZIP proteins.

Supplementary Table 1 lists the accession numbers, synonyms, protein length and position of the B-ZIP motif in the whole protein for the 67 B-ZIP factors identified in this genome-wide search. Eight proteins (AtbZIP14, AtbZIP27, AtbZIP32, AtbZIP34, AtbZIP61, AtbZIP72, AtbZIP73 and AtbZIP75) were reported earlier as B-ZIP proteins (7) but are not represented in this study. These eight factors are leucine rich, but did not meet our B-ZIP selection criteria. For example, AtbZIP14 and AtbZIP27 each have perfect basic regions, but no leucine zipper. The sequences of AtbZIP32, AtbZIP73 and AtbZIP75 were analyzed and yielded no recognizable B-ZIP domain. Riechmann *et al.* (27) did not provide accession numbers for the 81 B-ZIP factors identified in their study. This complicated our ability to analyze the sequences from their study that are not included in our list.

The amino acid relatedness of the 67 *A.thaliana* and 53 *H.sapiens* B-ZIP motifs was determined using ClustalW and the information is presented as a cladogram (Supplementary Figure 2). No interspecies clustering is observed, indicating that there are no homologous *A.thaliana* and *H.sapiens* B-ZIP domains. This is in contrast to the near total overlap between *D.melanogaster* and *H.sapiens* B-ZIP proteins (4).

Analysis of the 67 *A.thaliana* B-ZIP leucine zippers

Alignment and amino acid content. Figure 2 presents the amino acid sequences of the 67 *A.thaliana* B-ZIP motifs using a conserved asparagine residue in the basic region of all proteins (data not shown) to align the sequences. This stringent criterion excluded from our analysis the rare B-ZIP proteins, like the mammalian GADD153 protein, that lacks this

conserved asparagine in the basic region (28). The N-terminal boundary of the leucine zipper is identical for all B-ZIP proteins because it is defined by the presence of the basic region. To identify the C-terminal boundary, three criteria were used: (i) the presence of a proline or a pair of glycines, either of which disrupts the α -helical structure of a leucine zipper, (ii) the presence of a leucine in the *d* position and (iii) the presence of charged amino acids in the *g* and *e* positions. Amino acids within the zippers that are important for dimerization specificity are color-coded using the key developed for the analysis of the *D.melanogaster* (4) and *H.sapiens* (3) B-ZIP proteins. The proteins are listed beginning with the zippers that are predicted to form only homodimeric interhelical interactions (families A–N), followed by the zippers that are predicted to form both homodimeric and heterodimeric interactions (families O–S). The TGA family of B-ZIP proteins (family T in Figure 2) contains short leucine zippers with charged amino acids in all *a* positions.

Figure 3 presents the kinds of amino acids found in the *a*, *d*, *e* and *g* positions, the coiled coil position that are known to regulate dimerization stability and specificity using the definition of the C-terminus as shown in Figure 2. Overall, *A.thaliana* B-ZIP proteins contain fewer charged amino acids in the *g* and *e* positions than *H.sapiens* zippers. The charged amino acids produce attractive and repulsive $g \leftrightarrow e'$ pairs to regulate dimerization specificity. For the *a* position, >50% of the residues in *H.sapiens* are aliphatic (I, V, L or M) and only 16% are asparagine. In contrast, in *A.thaliana* *a* positions, aliphatic amino acids are less abundant while asparagine is more abundant. In the *d* position, the uniquely stabilizing leucine is more abundant in *H.sapiens* (66%) than in *A.thaliana* (56%) while other aliphatic amino acids (I, V and M) are more frequent in *A.thaliana* (19%) than in *H.sapiens* (9%). While there are significant differences in the relative frequencies of amino acids appearing in the *a*, *d*, *e* and *g* positions of the heptads from plants and humans, the obvious conservation in amino acid usage provides assurance that the structural rules that we have applied to analyzing *H.sapiens* B-ZIP dimerization specificity (3) can be used to predict the dimerization of *A.thaliana* B-ZIP proteins.

Leucine zipper length. Leucine zipper length presented in Figure 2 is a subjective decision based on the expectation that a heptad will be α -helical. Several discrete criteria suggest the *A.thaliana* B-ZIP leucine zippers can be longer than *H.sapiens*. Supplementary Figure 3 categorizes the B-ZIP proteins from *A.thaliana* and *H.sapiens* by heptad number using the natural C-terminus or the presence of a proline or two glycines as the terminator of the α -helix. Using these criteria, we observe that over half of *H.sapiens* leucine zippers terminate after four or six heptads and none are greater than nine heptads in length. In contrast, *A.thaliana* leucine zippers are more variable. Of the *A.thaliana* B-ZIP proteins, 10% have short zippers (3 heptads) and >30% have no α -helix breakers for 10 or more heptads. This does not suggest that the zippers are that long but only that their structural limits are less well-defined than that observed in *H.sapiens* B-ZIP proteins.

The frequency of stabilizing leucine or other hydrophobic amino acids in the *d* positions of the *A.thaliana* and *H.sapiens* leucine zipper (16,29) were compared (Supplementary Figure 4). In *H.sapiens*, leucine is found in 100% of the second

		Heptad #	Consensus	0	1	2	3	4	5	6	7	8	9
				L ₀	L ₁	L ₂	L ₃	L ₄	L ₅	L ₆	L ₇	L ₈	
				g a b c d e f	g a b c d e f	g a b c d e f	g a b c d e f	g a b c d e f	g a b c d e f	g a b c d e f	g a b c d e f	g a b c d e f	g a b c d e f
A A	At5g42910	-		Q T M E V E V	E L E N L K K	D Y E E L L K	Q H V E L R K	R Q M E P G M	I S L H E R P	E R K L R R T	K S D I K ^		
B A ●	At3g56850	DPBF3	Y T H E L E I	K V S R L E E	E N E R L R K	* Q K E V E K I	L P S V P P P	D P K R Q L R	R T S S A P F ^				
	At2g41070	DPBF4	Y T H E L E I	K V S R L E E	E N E K L R R	* L K E V E K I	L P S E P P P	D P K W K L R	R T N S A S L ^				
	At5g44080	-	Y Q V E L E A	L A A K L E E	E N E L L S K	E I E D R K R	* E R Y Q K L M	E F V I P V V	E K P K Q Q P	P R F L R R I	R S L E W ^		
	At1g03970	GBF4	Y Q V E L E T	L A A K L E E	E N E Q L L K	E I E E S T K	* E R Y K K L M	E V L I P V D	E K P R P P S	R P L S R S H	S L E W ^		
C A ●	At4g34000	ABF3	Y T M E L E A	E I A Q L K E	L N E E L Q K	K Q V E I M E	* K Q K N Q L L	E P L R Q P W	G M G C K R Q	C L R R T L T	G P W ^		
	At1g49720	ABF1	Y T L E L E A	E I E S L K L	V N Q D L Q K	K Q A E I M K	* T H N S E L K	E F S K Q P P	L L A K R Q C	L R R T L T G	P W ^		
	At3g19290	ABF4	Y T L E L E A	E I E K L K K	T N Q E L Q K	K Q A E M V E	* M Q K N E L K	E T S K R P W	G S K R Q C L	R R T L T G P	W ^		
	AF093545	ABF2	Y T V E L E A	E V A K L K E	E N D E L Q R	K Q A R I M E	* M Q K N Q E T	E M R N L L Q	G G P K K K L	R R T E S G P	W ^		
	At2g36270	AB15	Y T V E L E A	E L N Q L K E	E N A Q L K H	A L A E L E R	* K R K Q Q Y F	E S L K S R A	Q P K L P K S	N G R L R T L	M R N P S C P	L	
	At3g44460	DPBF2	Y T V E L E L	E L N N L T E	E N T K L K E	I V E E N E K	* K R R Q E I I	S R S K Q V T	K E K S G D K	L R K I R R M	A S A G W ^		
D B	At3g56660	-	Y V E E L E D	K V K N M H S	T I S E L S S	K M S Y F V A	E N V T L R Q	* Q M G T R F S	S G P P M V P	I V Y P W M Q	Y P A Y M V K	P Q G S Q V A	
	At3g10800	-	Q T E E L E R	K V K S M N A	T I A E L N G	K I A Y V M A	E N V A L R Q	* Q M A V A S G	A P H L V T T ^				
	At2g40950	-	Y V E E L E E	K V R N M H S	T I T D L N G	K I S Y F M A	E N A T L R Q	* Q L G G N G M	C P P H L P P	P P M G M Y P	P M A P M P Y	P W M P C P P	
	At1g42990	-	Y V Q D L E K	K S K Y L E R	E L R L R G R	M L E C F V A	E N Q S L R Y	* C L Q K G N G	N N T T M M S	K Q E S A V L	L L E S L L L	G S L L W L L	
E H ●	At5g11260	HY5	Y L S E L E N	R V K D L E N	K N S E L E E	K R L S T L Q N	E N Q M L R H	I L K N T T G	* N K R G G G G	G S N A D A S	L ^		
	At3g17610	HYH	Y V S D L E S	N N D Q L E E	K T S T L T N	E N T M L R K	M L I N T R P	* K T D D N H					
F G ●	At4g36730	GBF1	E C E Q L Q Q	R V E S L S N	E N Q S L R D	E L Q R L S S	E C D K L K S	E N N S I Q D	E L Q R V L G	* A E A V A N L	E Q N A A G S	K D G E G T N ^	
	At4g01120	GBF2	E T E Q L S V	K V D A L V A	E N M S L R S	K L G Q L N N	E S E K L R L	E N E A I L D	Q L K A Q A T	* G K T E N L I	S R V D K N N	S V S G S K T	
	At2g46270	GBF3	E T E E L A R	K V E A L T A	E N M A L R S	E L N Q L N E	K S D K L R G	A N A T L L D	K L K C S E P	* E K R V P A N	M L S R V K N	S G A G D K N	
	At1g19490	-	M C E E L S K	K A A D L T Y	E N E N L R R	E K D W A L K	E F Q S L E T	I N K H L E E	Q V L K S V K	* P D T K E P E	E S P K P S Q	V E M S T S S	
	At1g32150	-	E C D E L A Q	R A E V L N G	E N S S L R A	E I N K L K S	Q Y E E L L A	E N S S L K N	* K F S S A P S	L E G G D L D	K N E Q E P Q	R S T R Q D V	
	At2g35530	-	E C D E L A Q	R A E V L N E	E N T N L R A	E I N K L K S	Q Y E E L T T	E N T S L K V	* K K K ^				
G C	At3g54620	AJ010860	Q M N E F D T	Q V G Q L R A	E H S T L I N	R L S D M N H	K Y D A A A V	D N R I L R A	D I E T L R T	K V K M A E E	T V K R V T G	* V N P L H W S	
	At4g02640	BZ02H1	Q T S D L E T	Q V N D L K G	E H S S L L K	Q L S N M N H	K Y D E A A V	G N R I L K A	D I E T L R A	K V K M A E E	T V K R V T G	* M N P M L L G	
	At5g24800	BZ02H2	Y L V D L E T	Q V D S L K R	D N S T L Y K	Q L I D A T Q	Q F R S A G T	N N R V L K S	D V E T L R V	K V K L A E D	L V A R G S L	* T S S L N Q L	
	At5g28770	BZ02H3	H L S E L E T	Q V S Q L R V	E N S K L M K	G L T D V T Q	T F N D A S V	E N R V L K A	N I E T L R A	K V K M A E E	T V K R L T G	* F N P M F H N	
H S	At3g62420	-	Q L G D L I N	E V T L L K N	D N A K I T E	Q V D E A S K	K Y I E M E S	K N N V L R A	Q A S E L T D	R L R S L N S	V L E M V E E	* I S G Q A L D	
	At1g75390	-	H L D D L T A	Q V T H L R K	E N A Q I V A	G I A V T T Q	H Y V T I E A	E N D I L R A	Q V L E L N H	R L Q S L N E	I V D F V E S	* S S S G F G M	
	At4g34590	ATB2	L L D D L T A	Q V N H L K K	E N T E I V T	S V S I T T Q	H Y L T E A E	E N S V L R A	Q L D E L N H	R L Q S L N D	I I E F L D S	* S N N N N N N	
	At2g18160	GBF5	H V D D L T A	Q I N Q L S N	D N R Q I L N	S L T V T S Q	L Y M K I Q A	E N S V L T A	Q M E E L S T	R L Q S L N E	I V D L V Q S	* N G A G F G V	
	At3g49760	-	H L E E M S I	Q L N Q L K I	Q N Q E L K N	Q L R Y V L Y	H Q Q R T K M	E N D R L L M	E H R I L H D	K L L N I R Q	V L M F R Q T ^		
I S	At5g60830	-	Q I E E L Q Q	Q V E Q L M M	L N H H L S E	K V I N L L E	S N H Q I L Q	E N S Q L K E	K V S S F H L	L M A D V L L	* P M R N A E S	N I N D R N V	
J S	At1g68880	-	H M E E L W S	M L V Q L I N	K N K S L V D	E L S Q A R E	C Y E K V I E	E N M K L R E	E N S K S R K	* M I G E I G L	N R F L S V E	A D Q I W T F ^	
K S	At5g15830	-	H L D E L L S	Q V A W L R S	E N H Q L L D	K L N Q V S D	N N D L V I Q	E N S S L K E	E N L E L R Q	V I T S M K K	* L G G G I H D	K Y S S P S S	
L S	At1g13600	-	H L D E L W S	Q V I R L R T	D N H C L M D	K L N R V S E	S H E L A L K	E N A K L K E	E T S D L R Q	L I S E I K S	* H N E D D N S	F L R E L E D	
	At3g30530	-	H L D E L W S	Q V M W L R I	E N H Q L L D	K L N N L S E	S H D K V L Q	E N A Q L K E	E T F E L N Q	V I S D M Q I	* Q S P F S C F	R D D I I P I	
	At5g38800	-	Q V D E L W S	Q V M W L R D	E N H Q L L R	K L N C V L E	S Q E K V I E	E N V Q L K E	E T T E L N Q	M I S D M Q L	* Q N Q S P F S	C I R D D D D	
	At2g04038	-	H L D E L W S	Q V I R L R N	E N C L I D	K L N R V S E	T Q N C V L K	E N S K L K E	E A S D L R Q	L V C E L K S	* N K N N N N S	F P R E F E D	
M I	At2g21235	-	K I R D M E L	R V E T L E N	T Q A S L F G	T M T L L E K	E N I V M M N	E N K L A K I	R L Q L L E Q	* Q A P L L T A	L T A L K Q L	D E L R R L E	
N I	At2g31370	PosF21	Y I F E L E R	K V Q T L Q T	E A T T L S A	Q L T L L Q R	D T N G L T V	E N N E L K L	R L Q T M E Q	Q V H L Q D E	L N E A L K E	E I Q H L K V *	
	At1g06070	-	Y I A E L E R	K V Q T L Q T	E A T S L S A	Q L T L L Q R	D T N G L G V	E N N E L K L	R V Q T M E Q	Q V H L Q D A	L N D A L K E	E V Q H L K V *	
	At2g40620	-	Y I L E L E R	K V Q T L Q T	E A T T L S A	Q L S L F Q R	D T T G L S S	E N T E K L R	R L Q V M E Q	Q A K L R D A	L N E Q L K K	E V E R L K F *	
	At1g43700	VIP1	Y T G E L E R	K V Q T L Q N	E A T T L S A	Q V T M L Q R	G T S E L N T	E N K H L K M	R L Q A L E Q	Q A E L R D A	L N E A L R D	E L N R L K V *	
	At1g06850	-	Y I Q E L E R	K V Q S L T Q	E A T T L S A	Q L T L Y Q R	D T N G L A N	E N T E L K L	R L Q A M E Q	Q A Q L R N A	L N E A L R K	E V E R M K M *	
	At4g38900	-	Y I V E L E H	K V Q T L Q T	E A T T L S A	Q L T L L Q R	D M M G L T N	Q N N E L K F	R L Q A M E Q	Q A R L R D A	L N E A L N G	E V Q R L K L *	
	At2g21230	-	Y M A E L E H	K V Q T L Q T	E A T T L S A	Q L T H L Q R	D S M G L T N	Q N S E L K F	R L Q A M E Q	Q A Q L R D A	L S E K L N E	E V Q R L K L *	
O I	At2g24340	-	Y I V D L E G	S V R I H V M	V N T S L Y G	M I V L L E V	R H I T L A K	K K N D V D D	* E R E Q E N E	D S S ^			
P I	At2g13150	-	Y M I N L E L	K I N F L E N	K N A S I F E	K I K L L E N	D K T M R M N	E K K E I M I	R I E S L E Q	H A E L R D A	L T E H L H V	E I E R L K A *	
	At2g12900	-	Y M I D L E H	K I K F L E N	K N A L I F E	K I K L L E K	D K T I L M N	E K K E I T I	Q I E S L E Q	Q A Q L R D A	L T E K L H V	E I E R L K V *	
Q F	At4g35040	-	K A A S L E D	E V A R L R A	V N Q Q L V K	R L Q N Q A T	L F A E V S R	L K C L L V D	L R G R I D G	* E I G S F P Y	Q K P M A A N	I P S F S H M	
	At2g16770	-	K A A S L E D	E V M R L K A	V N N Q L L K	R L Q G Q A A	L F A E V T R	L K C L L V D	I R G R I D G	* E I G A F P Y	Q K P A V T N	V P Y S Y M M	
	At3g51960	-	R T A Y L E D	E V M R L Q S	L N E Q F L R	K L Q S Q E M	V E T E L I R	L R A L L V E	M Q G K I E V	E L C S F S F	* Q K Q C N G S	G F V F K E G	
R S	At5g49450	-	L M E D T I H	E I S S L E R	R I K E N S E	R C R A V K Q	R L D S V E T	E N A G L R S	E K I W L S I	Y V S D L E N	F I A T T S L	* T L T Q S G G	
	At1g59530	-	R F E E L T E	E V N R L N I	R N Q E L K N	R L A N V V S	C G D F I S R	E N N R L K T	E S V C L E I	R L L E L Y R	F L V A M Q S	* P I S T S V N	
S S	At4g37730	-	H I D N L R E	Q V N R L D L	E N R E L G N	R L R L V L H	Q L Q R V N S	D N N R L V T	E Q E I L R L	R L S E M R R	* I L I I R Q L	Q Q Q Q W E	
	At2g22850	-	H I D N L K D	E A N R L G L	E N R E L A N	R L R I V L Y	N I A L M C T	D N N Q L L S	E Q E I L R R	R F L E M R Q	* I L I F R Q L	Q L N P S L I	
T D ●	At1g68640	PAN	Y V Q Q L E N	S R I R L A Q	L E E L K R * A	Q Q G S L	V E R G V S A	D H T H L A A	G N G V F S F	E L E Y T R W	K E E H Q R M	I N D L R S G	
	At5g06960	OBF5/TGA5	Y V Q Q L E N	S R L K L T Q	L E Q E L Q R	* A R Q Q G V F	I S S S G D Q	A H S T A G D	G A M A F D V	E Y R R W Q E	D K N R Q M K	E L S S A I D	
	At3g12250	TGA6	Y V Q Q L E N	S R L K L T Q	L E Q E L Q R	* A R Q Q G V F	I S S S G D Q	A H S T G G N	G A L A F D A	E H S R W L E	E K N R Q M N	E L R S A L N	
	At5g06950	TGA2	Y V Q Q L E N	S R L K L T Q	L E Q E L Q R	* A R Q Q G V F	I S G T G D Q	A H S T G G N	G A L A F D A	E H S R W L E	E K N K Q M N	E L R S A L N	
	At1g22070	TGA3	H V Q Q L E E	S R L K L S Q	L E Q E L V R	* A R Q Q G L C	V R N S S D T	S Y L G P A G	N M N S G I A	A F E M E Y T	H W L E E Q N	R R V S E I R	
	At5g65210	TGA1	Y V Q Q L E T	S R L K L I Q	L E Q E L D R	* A R Q Q G F Y	V N G N I D T	N S L G F S E	T M N P G I A	A F E M E Y G	H W V E E Q N	R Q I C E L R	
	At5g10030	OBF4/TGA4	Y V Q Q L E T	S R L K L I H	L E Q E L D R	* A R Q Q G F Y	V N G N V D T	N A L S F S E	N M S S G I V	A F E M E Y G	H W V E E Q N	R Q I C E L R	
	At1g77920	-	Y V Q Q L E E	S R L K L S Q	L E Q E L E K	* V Q Q G H L	G P S G S I N	T G I A S F E	M E Y S H W L	Q E Q S R R V	S E L R T A L	Q S H I S D I	
	At1g08320	-	Y V Q Q L E S	S R I K L S Q	L E Q E L Q R	* A R S Q G L F	M G G C G P P	G P N I T S G	A A I F D M E	Y G R W L E D	D N R H M S E	I R T G L Q A	
	At5g06840	-	Y V Q Q L E S	C R I K L T Q	L E Q E I Q R	* A R S Q G V F	F G G S L I G	G D Q Q Q G G	L P I G P G N	I S S E A A V	F D M E Y A R	W L E E Q Q R	

heptad *d* positions and in >80% of the *d* positions in heptads 0 to 3. In subsequent heptads, the frequency of leucine in the *d* positions drops dramatically and it is absent by heptad 7. In *A.thaliana*, the distribution of leucine in the *d* positions is biphasic. Leucine is found in ~90% of the *d* positions in heptads 0 to 2, dramatically falls to 30% in heptad 3, and rises again to >50% in heptad 5. Leucine remains frequent in heptads 6 to 8, providing additional data to support the suggestion that some *A.thaliana* leucine zippers are longer than in *H.sapiens*.

Defining features of the *a* position. Figure 4 presents the frequency of asparagine in the *a* positions of both *H.sapiens* and *A.thaliana* B-ZIP proteins. In both species, asparagine is abundant in the *a* position of heptad 2. In *H.sapiens*, asparagine is rarely observed in other heptads while in *A.thaliana*, asparagine is also common in heptad 5. Asparagine produces stable N–N interactions at the *a*–*d*' position, but does not interact favorably with other *a* position amino acids (12). Thus, the unique placement of asparagines throughout the leucine zippers of *A.thaliana* is used to define families of homodimerizing B-ZIP proteins.

The *a* positions in *A.thaliana* B-ZIP proteins also contain other polar amino acids, including serine and threonine. To experimentally evaluate the contribution of serine to dimerization specificity, we used a well-established, heterodimerizing leucine zipper system (4,12). Briefly, this system contains one monomer (EE₃₄) in which homodimer formation is inhibited by repulsive, acidic *g*↔*e*' pairs [E↔E] in heptads 3 and 4 and a second monomer, termed A-RR₃₄, in which homodimer formation is blocked by the presence of repulsive, basic *g*↔*e*' interactions [R↔R] in heptads 3 and 4. Heterodimer formation between A-RR₃₄ and EE₃₄ is favored and is stabilized further by the presence of an N-terminal, amphipathic extension in A-RR₃₄ which lengthens the dimerization interface by interacting with the N-terminal, basic region of EE₃₄. The third *a* position in both EE₃₄ and A-RR₃₄ was changed to serine and the energetic contribution of serine to homotypic and heterotypic *a*–*d*' interactions was examined. Circular dichroism spectroscopy was used to monitor thermal denaturation (Supplementary Table 2) and the coupling energy of each complex (relative to alanine) was calculated (Supplementary Table 3). Serine in the *a* position does not favor an interaction with itself, an aliphatic isoleucine, or a

polar asparagine. Serine does, however, interact with lysine and in this way can exert a positive influence on heterodimer formation.

Like *H.sapiens*, *A.thaliana* B-ZIP proteins also contain charged amino acids in the *a* position that are destabilizing relative to alanine (12,20). The energetics of these charged amino acids interacting with other amino acids in the *a* position has not been examined extensively. The limited data available indicate that lysine prefers to interact with amino acids other than itself (Supplementary Table 3) (12). We suspect that these charged amino acids will drive heterodimer formation in a similar manner to lysine.

Features of the *g* and *e* positions. The observed frequency of attractive and repulsive *g*↔*e*' pairs in each heptad of *H.sapiens* and *A.thaliana* leucine zippers is presented in Figure 5. In *H.sapiens*, >50% of the first four heptads contain either attractive or repulsive *g*↔*e*' pairs and are virtually absent in subsequent heptads. This is consistent with the termination of the *H.sapiens* leucine zippers (Supplementary Figure 3). The first heptad shows the highest number of repulsive *g*↔*e*' pairs while attractive *g*↔*e*' pairs predominate in the more C-terminal heptads.

Arabidopsis thaliana leucine zippers, in contrast, contain about half the number of *g*↔*e*' pairs in the first four heptads with the frequency increasing to >30% in heptad 5. This pattern is consistent with the longer, average length of the *A.thaliana* leucine zippers. In *A.thaliana*, there are few repulsive *g*↔*e*' pairs and no protein with multiple repulsive *g*↔*e*' pairs (as found in some FOS proteins) implying that this feature is not a major determinant in driving heterodimerization of plant B-ZIP proteins.

Biophysical characterization of Opaque, a plant B-ZIP protein. To test the prediction that some *A.thaliana* B-ZIP leucine zippers have long leucine zippers, we biophysically characterized Opaque, the maize homologue of *A.thaliana* G family B-ZIP proteins. Opaque is predicted to be eight heptads long and has asparagine in the *a* position of the second and fifth heptads. Figure 6 presents the circular dichroism spectra and thermal denaturation of Opaque and CREB, a human BZIP protein that is only four heptads long. Both Opaque and CREB B-ZIP domains are truncated at the same position just N-terminal of the basic region. Circular

Figure 2. Amino acid sequence of the 67 *A.thaliana* B-ZIP domains. Proteins are arranged into families, A–T shown in bold in the first column, with similar predicted dimerization properties. The second column depicts names of the families from Jakoby *et al.* (7). The solid line delineates homodimerizing proteins (A–N) from proteins that have complex dimerization pattern (P–T). The leucine zipper region is divided into heptads (*gabcdef*) to help visualize the *g*↔*e*' pairs. Amino acids predicted to regulate dimerization specificity are color-coded. If the *g* and following *e* positions contain charged amino acids we colored the heptad from *g* to the following *e*. We use four colors to represent *g*↔*e*' pairs. Green is for the attractive basic-acidic pairs (R↔E and K↔E), orange is for the attractive acidic-basic pairs (E↔R, E↔K, D↔R and D↔K), red is for repulsive acidic pairs (E↔E and E↔D), and blue is for repulsive basic pairs (K↔K and R↔K). If only one of the two amino acids in the *g*↔*e*' pair is charged, we color that residue blue for basic and red for acidic. If the *a* or *d* position is polar, it is colored black and if either is charged, it is colored purple. The prolines and glycines are colored red to indicate a potential break in the α -helical structure. The predicted C-terminal boundary of the B-ZIP leucine zipper is denoted by an asterisk that enables us to define the frequencies of amino acids in different position of the leucine zipper. The C-terminal boundary is defined by the presence hydrophobic amino acids in the *a* and *d* positions, charged amino acids in the *e* and *g* positions, the absence of proline or pairs of glycines anywhere in the structure, and the absence of charged amino acids in the *a* and *d* positions. In the majority of cases, the decision was straight forward. However, in several instances, it is more ambiguous. For example, we indicate that the DPBF3 and DPBF4 leucine zipper stops after the second heptad because the third heptad has a K and E in the *a* and *d* positions, respectively; which should prevent leucine zipper formation. For the same reason At5g44080 and GBF4 stops after the third heptad as they have a R and K in the *a* and *d* positions, respectively, in the fourth heptad. At1g06070 goes to the ninth heptad even though it has two glycines in the fourth heptad. We would normally think this would terminate a leucine zipper structure but it appears very canonical from the fifth to the ninth heptads so we propose that it continues through the fourth heptad. This type of thinking was used to define all C-terminal boundary of all the leucine zippers. We do not intend these definitions of the C-terminus to be definitive, only approximate. The natural C-terminus is denoted with circumflex accent. The protein sequence for the tenth heptad of posF21, an N family member, and At2g13150, a P family member, that are predicted to not form a coiled-coil is LTGQVAP and VLISNEK, respectively. The dot in front of a protein indicates that it has been experimentally shown to form a homodimer.

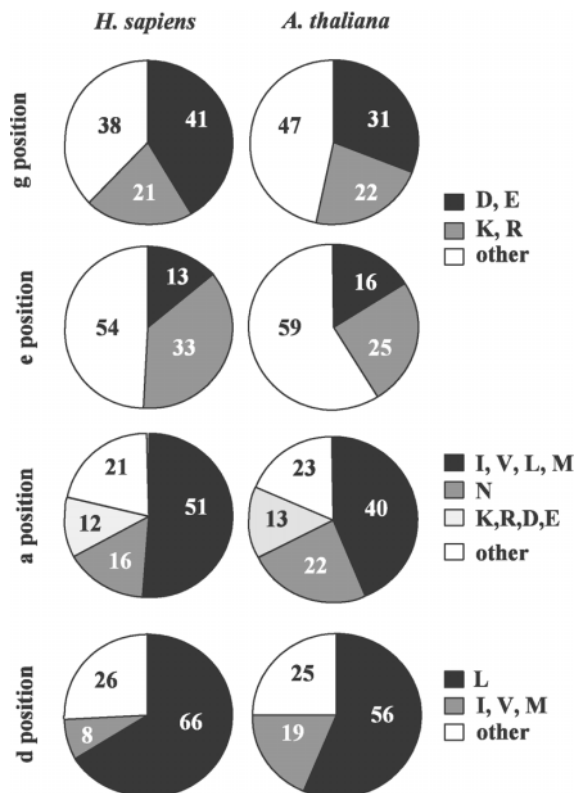


Figure 3. Pie chart presenting the frequency of amino acids in all the *g*, *e*, *a* and *d* positions of the leucine zipper for both *H.sapiens* and *A.thaliana* B-ZIP proteins.

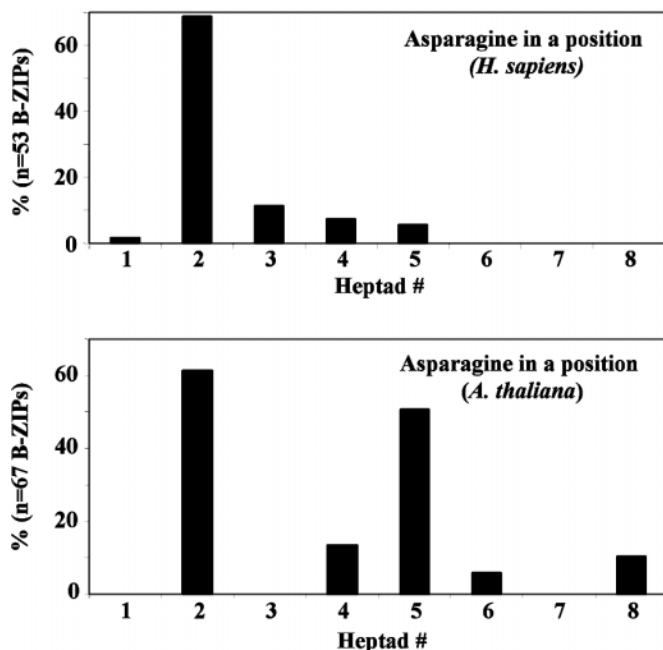


Figure 4. Histogram of the frequency of asparagine in the *a* positions of the leucine zippers for all *H.sapiens* and *A.thaliana* B-ZIP proteins.

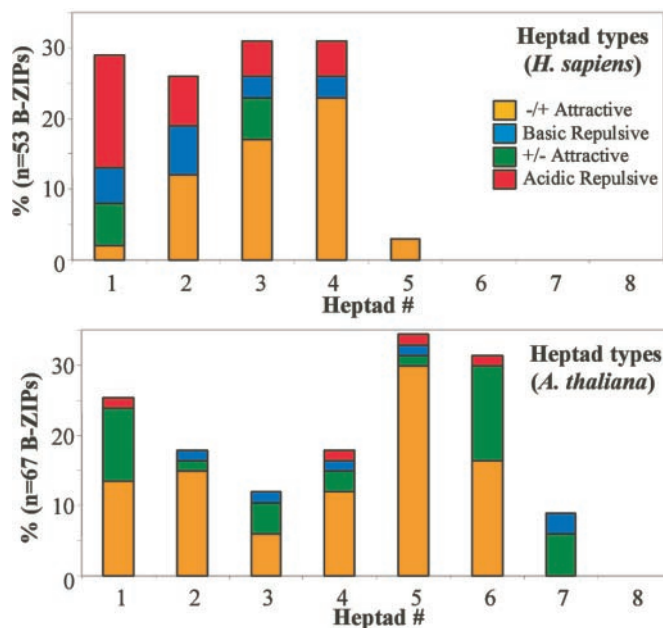


Figure 5. Histogram of the frequency of attractive or repulsive $g \leftrightarrow e'$ pairs per heptad for both *H.sapiens* and *A.thaliana* B-ZIP proteins.

dichroism spectra indicate that both Opaque and CREB have minima at 208 and 222 nM that is indicative of a protein that is primarily α -helical. However, at the same concentration, Opaque has more ellipticity than CREB, indicating that Opaque contains more α -helical structure. The truncated Opaque protein is $\sim 74.5\%$ α -helical corresponding to 12 heptads of α -helix. In contrast, CREB is only 56.5% α -helical corresponding to seven heptads of α -helix. Because the CREB protein is only four heptads long, we assume that the additional structure is the leucine zipper propagating into the basic region, as has been observed for the vertebrate VBP protein (30). The important result is that Opaque is five heptads longer than CREB indicative of a longer leucine zipper. The thermal denaturations indicate that the leucine zipper region melts cooperatively and have a simple two-state denaturation profile.

The dimerization network of *A.thaliana* B-ZIP proteins. On the basis of the structural features described above, we have organized the *A.thaliana* B-ZIP proteins into families designated A through T (Figure 2). The families have been organized further into three groups: (i) proteins whose leucine zippers contain attractive interhelical interactions and that we predict should form homodimers, (ii) proteins whose zippers contain both attractive and repulsive interhelical interactions that we predict should display more complex dimerization properties and (iii) proteins with zippers containing only repulsive interhelical interactions that are predicted to form heterodimers. Table 1 summarizes the structural rationale used in organizing these groups. Forty-seven B-ZIP proteins make up the first group and are sub-divided into 14 homodimerizing families (A–N) based on the placement of asparagine in the *a* positions of each heptad. Ten B-ZIP proteins comprise the 5 families of the second group (O–S) and possess the interhelical $g \leftrightarrow e'$ interactions favoring both homo- and heterodimer formation. We are less confident in predicting

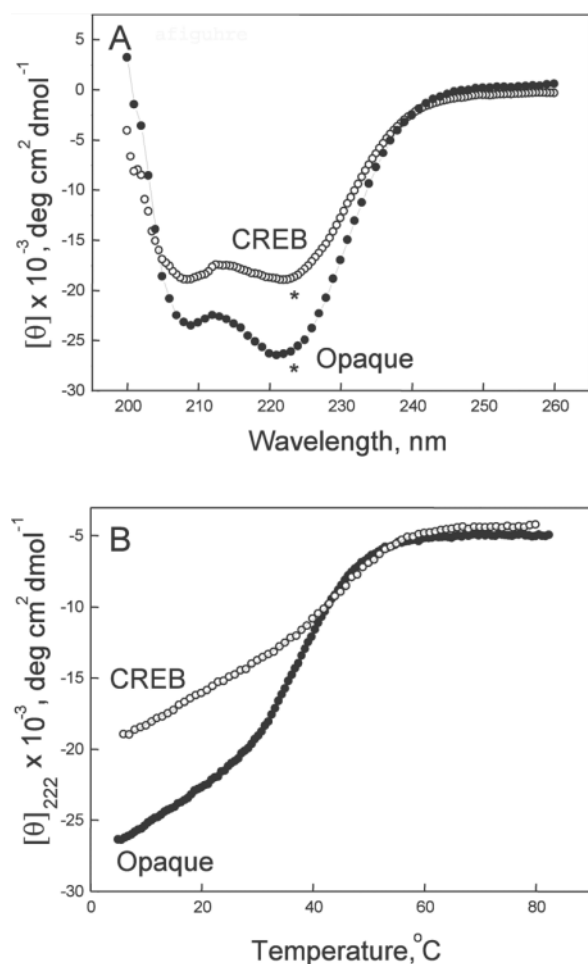


Figure 6. (A) Circular dichroism spectra from 200–260 nm of CREB and Opaque, 4 μM at 6°C. The asterisks indicate the minima at 222 nm at which the thermal denaturations of these two proteins were monitored. (B) Thermal denaturation curves of CREB and Opaque at 4 μM concentration monitored at 222 nm.

the dimerization properties of these proteins. The final group of proteins comprises a single family (*T*).

Experimental data documenting interactions between several *A.thaliana* B-ZIP proteins are available allowing us to evaluate some of our predictions. The data are typically from gel shift type experiments that indicate that two B-ZIP proteins can dimerize and bind to DNA. These results are consistent with our predictions as noted in Table 1. Cases where interactions observed experimentally were not predicted are discussed. GBF4 (family *B*) has been shown to dimerize with proteins in family *F* (31). Although the short leucine zippers of family *B* proteins could be considered a deterrent to efficient homodimer formation, there are no obvious structural details that suggest that GBF4, would interact preferentially with GBF1 or GBF2 in family *F*. Taken together, these observations indicate that GBF4, in fact, may dimerize with a large number of other B-ZIP proteins. Family *C* proteins have been shown to form intra-family dimers (32–35) and ABI5 can dimerize with proteins from family *B*, albeit at a reduced efficiency compared to intra-family dimerization (33). Once again, this

reflects the possibility that family *B* zippers are promiscuous in their interactions.

The placement of asparagines exclusively in the second and fifth *a* positions is a feature of four families of *A.thaliana* B-ZIP proteins (*F*, *G*, *H* and *K*). Segregation of these proteins into four separate families was directed by the presence of attractive or repulsive $g \leftrightarrow e'$ pairs. There is a report that family *F* proteins form dimers with family *G* proteins (36). This is surprising because an *F/G* dimer contains repulsive $g \leftrightarrow e'$ pairs in both the first and fourth heptads. Family *G* orthologs also have been shown to form dimers with family *H* proteins (37,38). In this case, there are fewer attractive $g \leftrightarrow e'$ pairs in the heterodimer compared to either homodimer, indicating that the heterodimer represents a less favored interaction. In this regard, it is important to note that the latter experiments were performed in the presence of DNA that may serve as a stabilizing influence on sub-optimal B-ZIP dimers.

DISCUSSION

We have identified 67 B-ZIP motifs in the *A.thaliana* genome and have predicted their dimerization partners. None of these B-ZIP motifs are homologous to any human B-ZIP motifs, but the kinds of amino acids observed in the *a*, *d*, *e* and *g* positions of the leucine zipper are similar to those observed in *H.sapiens* B-ZIP leucine zippers (Figure 3). We have used what is known about the contribution of these amino acids to dimerization specificity to predict *A.thaliana* B-ZIP dimerization partners. *A.thaliana* B-ZIP leucine zippers can be eight or more heptads long in contrast to the four to six heptads typically found in *H.sapiens*. Asparagine is observed throughout the *a* positions of these long leucine zippers and we used the particular placement of these asparagine residues to define 14 families of homodimerizing B-ZIP proteins (Figure 2).

In this manuscript, we have extended the biophysical examination of how different amino acids in the *a* position contribute to dimerization specificity. Quantitation of heterotypic $a-a'$ interactions have only been described for L, I, V, A, N and K (4,12). In this study, we present data to show that serine (S) in the *a* position interacts with I, N, K and S at the *a'* position. These data indicate that serine contributes less to dimerization specificity than an aliphatic amino acid, a polar asparagine, or a charged lysine residue. According to our data, serine prefers to be in a heterotypic interaction with lysine. Serine is found in the *a* position of the *A.thaliana* GBF2, GBF3 and At2g21230 proteins where it may encourage intra-group dimer formation rather than strict intra-family homodimerization. Serine also occurs in the *a* position of At1g59530. However, we predict that this protein will form heterodimers due to repulsive $g \leftrightarrow e'$ pairs.

A comparison between *A.thaliana* and *H.sapiens* B-ZIP proteins has identified clear differences. Three criteria suggest that the leucine zippers of B-ZIP domains in *A.thaliana* can be longer than in *H.sapiens*: (i) the presence of the α -helix breakers proline or a pair of glycines, (ii) the presence of leucines in the *d* position, and (iii) the presence of charged amino acids in the *g* and *e* positions. Although many *A.thaliana* leucine zippers are eight or more heptads in length, they appear to have similar stability as *H.sapiens* leucine zippers in comparing the thermal denaturation of Opaque and CREB.

The shorter, *H.sapiens* zippers have both a high frequency of stabilizing leucine residues in the *d* positions (16) and many attractive $g \leftrightarrow e'$ pairs. As a result, these zippers are optimally stable. *Arabidopsis thaliana* leucine zippers, on the other hand, have a lower frequency of these stabilizing elements and have longer leucine zippers but similar stability.

Homo sapiens and *A.thaliana* regulate the dimerization specificity of their leucine zipper proteins by different mechanisms. The six homodimerizing families of *H.sapiens* B-ZIP factors contain asparagine in the second heptad *a* position. However, only two of these six families contain additional *a* position asparagines to dictate dimerization specificity,

Table 1. *Arabidopsis thaliana* B-ZIP proteins were placed into families with similar predicted dimerization specificity

Family	Protein	Heptad with N/K* at <i>a</i> position	Length in heptads	Comments
A	At5g42910		3	Attractive $g \leftrightarrow e'$ interaction in the 1st heptad and lack of any repulsive interactions favors homodimerization
B	At5g44080 GBF4 DPBF3 DPBF4	2	3	N in <i>a</i> position of the 2nd heptad and attractive $g \leftrightarrow e'$ interactions favors homodimerization within the family (1)
C	ABF1 ABF2 ABF3 ABF4 ABI5 DPBF2	2	3	N in <i>a</i> position of the 2nd heptad and attractive $g \leftrightarrow e'$ interactions in the 1st and 2nd heptads favors intra-group homodimerization. Repulsive $g \leftrightarrow e'$ interactions in the 1st and 3rd heptads with B group proteins prevents heterodimerization (2)
D	At3g56660, At3g10800 At2g40950 At1g42990	4	4	N in <i>a</i> position of the 4th heptad and attractive $g \leftrightarrow e'$ interactions favors intra group homodimerization
E	HY5 HYH	2, 4	5	Homodimers stabilized by N at <i>a</i> position of the 2nd and 4th heptads
F	GBF1 GBF2 GBF3 At1g19490 At1g32150 At2g35530	2, 5	5	Three attractive $g \leftrightarrow e'$ interactions and the two Ns at the 2nd and 5th heptad <i>a</i> positions promote homodimerization within the family. Repulsive $g \leftrightarrow e'$ pairs in the 4th heptad of GBF3 and At1g19490 could drive preferential heterodimerization between these two proteins
G	At3g54620	2, 5	8	This long zipper is stabilized by N in <i>a</i> position of the 2nd and 5th heptads (H in the 2nd heptad for two members) and three attractive $g \leftrightarrow e'$ interactions in the 5th, 6th and 7th heptads. Repulsive $g \leftrightarrow e'$ interactions in the 1st, 2nd or 3rd heptads prevent dimerization with group F members that have the same <i>a</i> position N pattern (3)
H	At4g02640 At5g24800 At5g28770 At3g62420 At1g75390 At4g34590 At2g18160 At3g49760	2, 5	>8	Complex group (4)
I	At5g60830	2, 4, 5	7	Homodimerizes due to unique pattern of N's in <i>a</i> position
J	At1g68880	2, 5, 6	6	Unique placement of Ns at the <i>a</i> position in addition to attractive $g \leftrightarrow e'$ interactions promotes homodimerization
K	At5g15830	2, 4, 5, 6	7	Homodimerizes due to unique pattern of Ns at <i>a</i> position
L	At1g13600 At3g30530 At5g38800 At2g04038	2, 5	7	Dimerization within the family could be driven by the two Ns at the <i>a</i> position of the 2nd and 5th heptads and the two attractive $E \leftrightarrow R$ pairs in the 5th and 6th heptads
M	At2g21235	4, 5	6	Homodimerizes due to unique pattern of Ns at <i>a</i> position
N	At2g31370 At1g06070 At2g40620 At1g43700 At1g06850 At4g38900 At2g21230	5, 8	8	Unique placement of Ns at the <i>a</i> position in addition to attractive $g \leftrightarrow e'$ interactions promotes homodimerization within family
O	At2g24340	2, 5*	5	K in the 5th heptad <i>a</i> position prevents homodimerization
P	At2g13150 At2g12900	2, 4*, 5*	8	The two destabilizing Ks at the <i>a</i> position of 4th and 5th heptads could prevent homodimerization
Q	At4g35040 At2g16770 At3g51960	2, 4*, 5*, 6*	6	Charged amino acids in the <i>a</i> position of 4th, 5th and 6th heptads along with the incomplete $g \leftrightarrow e'$ pairs might lead to heterodimerization

Table 1. Continued

Group	Protein	Heptad with N/K* at a position	Length in heptads	Comments
R	At5g49450 At1g59530	2, 5, 6*	8	The repulsive $g \leftrightarrow e'$ interactions prevent homodimer formation
S	At4g37730 AT2G22850	2, 5	7	The 7th heptad repulsive $g \leftrightarrow e'$ interactions and the incomplete $g \leftrightarrow e'$ interactions promote heterodimerization
T	At1g68640 At5g06960 At3g12250 At5g06950 At1g22070 At5g65210 At5g10030 At1G77920 At1g08320 At5g06840	2*, 3*, 4*	3	Charged residues in the a position and lack of attractive $g \leftrightarrow e'$ interactions are expected to destabilize the homodimer

The first column contains the letter designation for the family. Families A to N are predicted to homodimerize, O to S have destabilizing interactions in their leucine zipper suggesting they may heterodimerize in while some may also homodimerize. The T family has an unstable leucine zipper that needs an additional protein to stabilize DNA binding. The third column notes the heptad number if there is an asparagine in the a position. If there is a charged amino acid in a position the heptad number has an asterisk. The fourth column contains a summary of the amino acids we predict help determine the dimerization specificity that defines the family of proteins. If experimental data agree with these predictions, it is referenced in the Table legend. 1) (30), 2) (31–34), 3) (35,36,46), 4) (35,36,46,47).

and the remaining four use variations of interhelical $g \leftrightarrow e'$ interactions to drive homodimer formation (3). Eleven families of homodimerizing *A.thaliana* B-ZIP proteins contain asparagine in the second heptad a position but, in contrast to *H.sapiens*, these proteins also contain asparagines in the a positions of the fourth, fifth, sixth and eighth heptads, with asparagine in the fifth heptad being nearly as abundant as in the second heptad. In the case of heterodimer formation, *H.sapiens* B-ZIP factors use repulsive electrostatic interhelical $g \leftrightarrow e'$ interactions, whereas *A.thaliana* B-ZIP factors contain few electrostatic repulsive $g \leftrightarrow e'$ interactions. Instead, *A.thaliana* B-ZIP factors contain charged amino acids in the a positions as is observed in the MAF family of *H.sapiens* B-ZIP proteins. While we suggest these proteins will heterodimerize, we are unable to predict the dimerization partners.

The use of the same amino acids to regulate dimerization specificity in both *H.sapiens* and *A.thaliana* coiled coils suggests that these amino acids are uniquely suited for this task. The coincidence between longer leucine zippers and more predicted homodimerizing B-ZIP families suggests that longer leucine zippers are needed to generate more homodimerizing families.

The leucine zipper has been a favored motif for biophysical characterization because of the assumed two-state denaturation of the structure, allowing calculation of thermodynamic parameters that describe the stability of the protein complex. All of the leucine zippers that have been biophysically characterized contain an asparagine in the second heptad a position and are approximately four to five heptads long. These proteins have been characterized by a two-state folding equilibrium. More recent work has shown that this is an oversimplification and that intermediates are observed during the folding of these proteins (40–42). Longer coiled coils such as tropomyosin display a complex denaturation profile (43). To examine if the longer leucine zippers predicted for some of the *A.thaliana* B-ZIP proteins with multiple a position asparagines, behaved as longer coiled coils, we studied the maize B-ZIP protein Opaque, which is related to the *A.thaliana* G family B-ZIP proteins. Opaque melts cooperatively and has a simple two state denaturation profile (Figure 6). At the same

concentration (4 μ M), Opaque is more elliptical than CREB, indicative of a longer coiled coil dimeric structure.

Several *A.thaliana* B-ZIP motifs would be interesting to examine in biophysical detail, including those that have short leucine zippers (families A, B and C) and those with longer bipartite zippers with a putative homodimerizing and heterodimerizing part (families Q and R). A minimum of three heptads, corresponding to six helical turns, is required for a peptide to adopt a two-stranded α -helical coiled coil conformation (39). Therefore, B-ZIP proteins in families A, B and C may need DNA binding to stabilize the structure (31). The Q and R families have several heptads of putatively homodimerizing leucine zipper structure followed by heptads with charged amino acids in the a position. This is reminiscent of the Myc leucine zipper. These proteins may form homodimers with short coiled coils and heterodimers with longer coiled coils.

Perhaps the most intriguing *A.thaliana* B-ZIP proteins are the TGA proteins (family T). These proteins possess short leucine zippers with charged amino acids in the a positions of the first, second and third heptads that are expected to destabilize the leucine zipper structure. Despite these destabilizing characteristics, however, several of these factors have been reported to form both homodimers and heterodimers (44). Interestingly, deletion analysis of the TGA proteins indicates that the DNA-binding activity of these proteins requires the B-ZIP domain and a 'dimerization stability region' that is located more than 100 amino acids C-terminal to the zipper (45). The structural understanding of this intramolecular interaction is unknown, but the observation is consistent with the suggestion that the leucine zippers of the TGA proteins are unstable. Several of the *A.thaliana* B-ZIP proteins predicted to form heterodimers (families O–Q) are similar in many ways to the TGA proteins. These proteins possess charged amino acids in the a positions of the fourth, fifth and/or sixth heptads and thus rely on the sequences of the first three heptads for stable zipper formation. Clearly, this type of heptad arrangement is non-canonical, and it would be of interest to determine if these proteins denature in a sequential manner.

A notable difference between *H.sapiens* and *A.thaliana* leucine zippers is a histidine which is present in the *d* position of the fifth heptad in one-third of all human B-ZIP factors, including all AP-1 family members. The charge of this histidine is sensitive to intracellular changes in pH and can influence the stability and specificity of B-ZIP dimers. The absence of a similarly placed histidine residue in plants suggests that such a signaling system is absent in plants.

In summary, this study examines two issues: (i) a genome-wide comparison of *A.thaliana* and *H.sapiens* B-ZIP factors and (ii) a prediction of *A.thaliana* B-ZIP protein dimerization specificity. Our analysis shows that all B-ZIP proteins use the same amino acids to regulate dimerization specificity, but that *A.thaliana* and *H.sapiens* genomes exploit the properties of these residues in different ways. *A.thaliana* B-ZIP proteins have long leucine zippers that are used to define multiple families of homodimerizing B-ZIP proteins.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

A portion of this work was supported by PHS CA-78264 awarded to E.J.T. C.D.D. was a pre-doctoral trainee of PHS T32 CA-09644-09.

REFERENCES

- Ellenberger, T.E., Brandl, C.J., Struhl, K. and Harrison, S.C. (1992) The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex. *Cell*, **71**, 1223–1237.
- Vinson, C.R., Sigler, P.B. and McKnight, S.L. (1989) Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science*, **246**, 911–916.
- 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.
- Fassler, J., Landsman, D., Acharya, A., Moll, J.R., Bonovich, M. and Vinson, C. (2002) B-ZIP proteins encoded by the Drosophila genome: evaluation of potential dimerization partners. *Genome Res.*, **12**, 1190–1200.
- Newman, J.R. and Keating, A.E. (2003) Comprehensive identification of human bZIP interactions with coiled-coil arrays. *Science*, **300**, 2097–2101.
- The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.
- Jakoby, M., Weisshaar, B., Droge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T. and Parcy, F. (2002) bZIP transcription factors in Arabidopsis. *Trends Plant Sci.*, **7**, 106–111.
- 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.
- O'Shea, E.K., Rutkowski, R. and Kim, P.S. (1992) Mechanism of specificity in the Fos–Jun oncoprotein heterodimer. *Cell*, **68**, 699–708.
- Baxevas, A.D. and Vinson, C.R. (1993) Interactions of coiled coils in transcription factors: where is the specificity? *Curr. Opin. Genet. Dev.*, **3**, 278–285.
- Zhou, J. and Goldsbrough, P.B. (1994) Functional homologs of fungal metallothionein genes from Arabidopsis. *Plant Cell*, **6**, 875–884.
- 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.
- 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.
- Thompson, K.S., Vinson, C.R. and Freire, E. (1993) Thermodynamic characterization of the structural stability of the coiled-coil region of the bZIP transcription factor GCN4. *Biochemistry*, **32**, 5491–5496.
- Landschulz, W.H., Johnson, P.F. and McKnight, S.L. (1989) The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. *Science*, **243**, 1681–1688.
- 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.
- 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.
- 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.
- Zeng, X., Herndon, A.M. and Hu, J.C. (1997) Buried asparagines determine the dimerization specificities of leucine zipper mutants. *Proc. Natl Acad. Sci. USA*, **94**, 3673–3678.
- 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.
- Alber, T. (1992) Structure of the leucine zipper. *Curr. Opin. Genet. Dev.*, **2**, 205–210.
- 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.
- Cohen, D.R. and Curran, T. (1990) Analysis of dimerization and DNA binding functions in Fos and Jun by domain-swapping: involvement of residues outside the leucine zipper/basic region. *Oncogene*, **5**, 929–939.
- Fernandes, L., Rodrigues-Pousada, C. and Struhl, K. (1997) Yap, a novel family of eight bZIP proteins in *Saccharomyces cerevisiae* with distinct biological functions. *Mol. Cell. Biol.*, **17**, 6982–6993.
- Higgins, D.G., Thompson, J.D. and Gibson, T.J. (1996) Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.*, **266**, 383–402.
- Page, R.D. (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.*, **12**, 357–358.
- Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R. et al. (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science*, **290**, 2105–2110.
- Ron, D. and Habener, J.F. (1992) CHOP, a novel developmentally regulated nuclear protein that dimerizes with transcription factors C/EBP and LAP and functions as a dominant-negative inhibitor of gene transcription. *Genes Dev.*, **6**, 439–453.
- 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.
- 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.
- Menkens, A.E. and Cashmore, A.R. (1994) Isolation and characterization of a fourth *Arabidopsis thaliana* G-box-binding factor, which has similarities to Fos oncoprotein. *Proc. Natl Acad. Sci. USA*, **91**, 2522–2526.
- Nakamura, S., Lynch, T.J. and Finkelstein, R.R. (2001) Physical interactions between ABA response loci of Arabidopsis. *Plant J.*, **26**, 627–635.
- Kim, S.Y., Ma, J., Perret, P., Li, Z. and Thomas, T.L. (2002) Arabidopsis ABI5 subfamily members have distinct DNA-binding and transcriptional activities. *Plant Physiol.*, **130**, 688–697.
- Choi, H., Hong, J., Ha, J., Kang, J. and Kim, S.Y. (2000) ABFs, a family of ABA-responsive element binding factors. *J. Biol. Chem.*, **275**, 1723–1730.
- Bensmihen, S., Ripa, S., Lambert, G., Jublot, D., Pautot, V., Granier, F., Giraudat, J. and Parcy, F. (2002) The homologous ABI5 and EEL

- transcription factors function antagonistically to fine-tune gene expression during late embryogenesis. *Plant Cell*, **14**, 1391–1403.
36. Armstrong, G.A., Weisshaar, B. and Hahlbrock, K. (1992) Homodimeric and heterodimeric leucine zipper proteins and nuclear factors from parsley recognize diverse promoter elements with ACGT cores. *Plant Cell*, **4**, 525–537.
 37. Rugner, A., Frohnmeier, H., Nake, C., Wellmer, F., Kircher, S., Schafer, E. and Harter, K. (2001) Isolation and characterization of four novel parsley proteins that interact with the transcriptional regulators CPRF1 and CPRF2. *Mol. Genet. Genomics*, **265**, 964–976.
 38. Strathmann, A., Kuhlmann, M., Heinekamp, T. and Droge-Laser, W. (2001) BZI-1 specifically heterodimerises with the tobacco bZIP transcription factors BZI-2, BZI-3/TBZF and BZI-4, and is functionally involved in flower development. *Plant J.*, **28**, 397–408.
 39. Su, J.Y., Hodges, R.S. and Kay, C.M. (1994) Effect of chain length on the formation and stability of synthetic alpha-helical coiled coils. *Biochemistry*, **33**, 15501–15510.
 40. Wendt, H., Durr, E., Thomas, R.M., Przybylski, M. and Bosshard, H.R. (1995) Characterization of leucine zipper complexes by electrospray ionization mass spectrometry. *Protein Sci.*, **4**, 1563–1570.
 41. 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.
 42. Dragon, S., Offenhauser, N. and Baumann, R. (2002) cAMP and *in vivo* hypoxia induce tob, ifr1, and fos expression in erythroid cells of the chick embryo. *Am. J. Physiol. Regul. Integr. Comp. Physiol.*, **282**, R1219–R1226.
 43. Ozeki, S., Kato, T., Holtzer, M.E. and Holtzer, A. (1991) The kinetics of chain exchange in two-chain coiled coils: alpha alpha- and beta beta-tropomyosin. *Biopolymers*, **31**, 957–966.
 44. Niggeweg, R., Thürow, C., Weigel, R., Pfitzner, U. and Gatz, C. (2000) Tobacco TGA factors differ with respect to interaction with NPR1, activation potential and DNA-binding properties. *Plant Mol. Biol.*, **42**, 775–788.
 45. Katagiri, F., Seipel, K. and Chua, N.H. (1992) Identification of a novel dimer stabilization region in a plant bZIP transcription activator. *Mol. Cell Biol.*, **12**, 4809–4816.
 46. Pysh, L.D., Aukerman, M.J. and Schmidt, R.J. (1993) OHP1: a maize basic domain/leucine zipper protein that interacts with opaque2. *Plant Cell*, **5**, 227–236.
 47. Schindler, U., Beckmann, H. and Cashmore, A.R. (1992) TGA1 and G-box binding factors: two distinct classes of Arabidopsis leucine zipper proteins compete for the G-box-like element TGACGTGG. *Plant Cell*, **4**, 1309–1319.