



Article

# Evolution of a bHLH Interaction Motif

Peter S. Millard<sup>1,2,3</sup> , Birthe B. Kragelund<sup>1,2</sup> and Meike Burow<sup>3,4,\*</sup>

- <sup>1</sup> The Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, DK-2200 Copenhagen N, Denmark; peter.millard@bio.ku.dk (P.S.M.); bbk@bio.ku.dk (B.B.K.)
  - <sup>2</sup> REPIN and Structural Biology and NMR Laboratory, Department of Biology, University of Copenhagen, DK-2200 Copenhagen N, Denmark
  - <sup>3</sup> DynaMo Center, Department of Plant and Environmental Sciences, University of Copenhagen, 1871 Frederiksberg C, Denmark
  - <sup>4</sup> Copenhagen Plant Science Centre, Department of Plant and Environmental Sciences, University of Copenhagen, 1871 Frederiksberg C, Denmark
- \* Correspondence: mbu@plen.ku.dk

**Abstract:** Intrinsically disordered proteins and regions with their associated short linear motifs play key roles in transcriptional regulation. The disordered MYC-interaction motif (MIM) mediates interactions between MYC and MYB transcription factors in *Arabidopsis thaliana* that are critical for constitutive and induced glucosinolate (GLS) biosynthesis. GLSs comprise a class of plant defense compounds that evolved in the ancestor of the Brassicales order. We used a diverse set of search strategies to discover additional occurrences of the MIM in other proteins and in other organisms and evaluate the findings by means of structural predictions, interaction assays, and biophysical experiments. Our search revealed numerous MIM instances spread throughout the angiosperm lineage. Experiments verify that several of the newly discovered MIM-containing proteins interact with MYC TFs. Only hits found within the same transcription factor family and having similar characteristics could be validated, indicating that structural predictions and sequence similarity are good indicators of whether the presence of a MIM mediates interaction. The experimentally validated MIMs are found in organisms outside the Brassicales order, showing that MIM function is broader than regulating GLS biosynthesis.



**Citation:** Millard, P.S.; Kragelund, B.B.; Burow, M. Evolution of a bHLH Interaction Motif. *Int. J. Mol. Sci.* **2021**, *22*, 447. <https://doi.org/10.3390/ijms22010447>

**Keywords:** MYB; bHLH; MYC; transcription factor; short linear motif (SLiM); intrinsically disordered protein (IDP); protein interaction; interaction motif; plant defense metabolism; jasmonate signaling

Received: 1 December 2020

Accepted: 30 December 2020

Published: 5 January 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In plants, many transcription factors (TFs) function in a combinatorial manner, where TF interactions serve to integrate internal and external cues to modulate gene expression levels [1]. Interactions between proteins from the two most expanded plant TF families, the MYB (v-myb avian myeloblastosis viral oncogene homolog) family and the bHLH (basic helix-loop-helix) family, are common, and together they regulate a large variety of biological processes [2,3]. Recently, we determined the binding site responsible for interactions between eight *Arabidopsis thaliana* MYB TFs and their bHLH partners, MYC2, MYC3, and MYC4 [4]. The binding site constitutes a short linear motif (SLiM) and coincides with a sequence motif defining a subgroup of six MYB TFs (MYB28, MYB29, MYB76, MYB34, MYB51, and MYB122) that require MYC-interaction for activity [4,5]. All members of this subgroup are involved in regulating glucosinolate (GLS) biosynthesis [6,7], and have homologs throughout the Brassicaceae [8]. Two additional MYB TFs in *A. thaliana* contain the MYC-interaction motif (MIM), MYB47 and MYB95. Their biological functions are less studied, although MYB47 was recently implicated in the regulation of seed longevity [9]. The MIM consists of the six residues [L/F]LN[K/R][V/L]A, with the core motif (residue positions that are absolutely required) being xLNxxA. Mutations in these core residues abolish the interaction [4].

Plant MYB TFs contain a DNA-binding domain (DBD) composed of one (1R MYB), two (R2R3 MYB), three (3R MYB) or four (4R MYB) MYB repeats, each comprising ~50–55 residues that fold separately into a three-helix structure [10–12]. Most common are the R2R3 MYB TFs, termed like this as the repeats are most similar to the second and third repeats in vertebrate MYB TFs, which have three repeats [13]. Apart from their DBD, TFs (including R2R3 MYB TFs) generally contain extensive intrinsically disordered regions (IDRs) [14–18] that provide specific molecular functions modulating TF activity. For example, SLiMs within IDRs facilitate interactions with SLiM-binding pockets in globular domains [19]. The MIM is a SLiM located outside the DBD [4], and the short length combined with its presence in a long disordered context increases the chance of *ex nihilo* evolution [20], i.e., the likelihood that proteins from other families could evolve an analogous interaction motif.

The specialized metabolites GLSs evolved in the ancestor of Brassicales ~90 million years ago [21,22]. They accomplish several specific functions, including defense against microbial pathogens and animal herbivory [23–25], but also regulate root growth and development [26–29]. Although they are constitutive defense compounds, GLSs can be induced upon stimulation, especially through jasmonate (JA) signaling upon herbivore and pathogen attack [6,30,31]. MYC2, MYC3, and MYC4 regulate JA-responsive genes, as JA signaling relieves repression of MYC2, MYC3, and MYC4 [32–34]. The molecular function of the MIM is to mediate interaction with MYC TFs and thereby link activity to JA signaling.

JA signaling has an established role in regulating plant defense and immunity [35]. In contrast to the evolutionarily young GLSs, the molecular components required for JA biosynthesis, signaling, and perception, including homologs of MYC2, are much more widespread in the plant lineage, being omnipresent in land plants [36,37]. Although the eight known MIM-containing TFs from *A. thaliana* are all R2R3 MYB TFs, and the MYC-interaction is essential for activating GLS biosynthesis [4], the direct role of the MIM is to link TF activity to JA signaling through MYC-interaction. We therefore hypothesized that MIMs could be present in proteins from other families, and in TFs that are not involved in GLS regulation but instead regulate other JA-responsive genes.

Two central questions arise as to whether MIMs are present in non-GLS producing organisms, and whether MIMs are present in proteins outside the R2R3 MYB TF family. Answers to these questions will contribute to a far better understanding of how SLiM interactions in TFs evolve and how they develop to link activity to pre-existing signaling networks. Using bioinformatics approaches, we identified and validated MIM-containing R2R3 MYB TFs throughout the angiosperms, including in non-GLS-containing plants. We further identified putative MIMs in trihelix TFs, located in a protein region compositionally dissimilar from the MIM-containing region of R2R3 MYB TFs. The trihelix TFs did not interact with MYC4 in our assay, which we rationalize through sequence analysis combined with biophysical experiments.

Together, our results show that the MIM is found in non-GLS producing organisms spread throughout the angiosperms, implying that the MIM is involved in regulating other biological processes besides GLS biosynthesis.

## 2. Results

### 2.1. Discovery of New MIMs

Initially, we wanted to investigate whether MIM-containing proteins exist outside the Brassicales, as this would imply that the MIM is involved in other biological processes besides regulation of GLS biosynthesis. We applied different search strategies and databases and invested several types of sequence information as a query. One approach was to use the full-length sequences of the eight AtMYB TFs with confirmed MIMs, and PHI-BLAST patterns such as [LVFI]LN[KR][IFLV]A. Using different PHI-BLAST patterns allowed fine-tuning the tolerance for variation at the different positions in the motif. As the full-length protein sequence including the DBD was used as a query, this approach yielded significant hits only in other proteins from the same family (the R2R3 MYB TF family). Yet, as a binding site that provides a regulatory link to JA signaling, the MIM might be relevant in other protein families.

Therefore, to circumvent this limitation, an additional approach was used, where an alignment of the motif region only (from the eight MIM-containing AtMYB TFs) comprised the query (Figure 1a). This way we did not limit the search to protein types similar to those we already knew, but instead focused on the motif itself. The resulting query sequence is short and has some degeneracy at non-core positions (only three residues in the xLNxxA core motif were strictly required for interaction [4]). Further, not to disregard information from the flanking regions [38], we performed a number of those searches, where the alignment used as query included between 0 and 20 flanking residues on each side [39]. The use of iterative search tools, such as jackhmmmer, that repeatedly adjusts the query according to significant hits, further facilitated the discovery of more diverged candidates.

We initially performed taxonomy-restricted searches, excluding all sequences from Brassicales, as the already known MIM-containing proteins from *A. thaliana* and their homologs from other GLS-producing plants otherwise dominated the search results and the readjustment of the hmm query in the iterative searches.

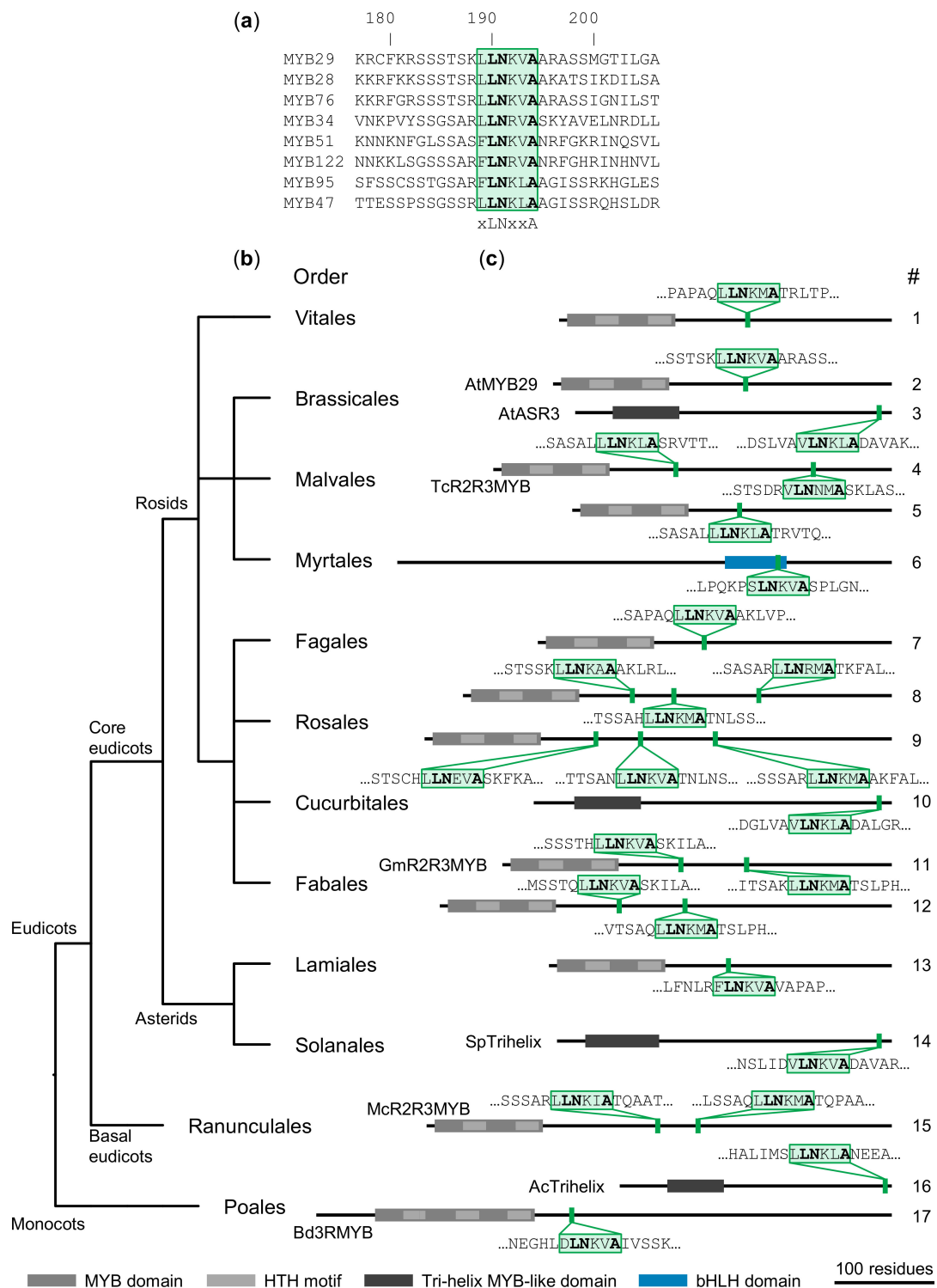
Outside angiosperms (flowering plants), we identified only a few significant hits. They were sporadic and all contained repetitive units with similarities to the motif. Further, these motif hits were present in unrelated proteins, in species where we could not find homologs of MYC TFs. Two examples of this are (i) (UniProt IDs) A0A0L7LIE3 from *Operophtera brumata* (winter moth), a putative  $\alpha/\beta$  hydrolase, which had 16 hit-instances, and (ii) I4N4U3 from the bacterium *Pseudomonas sp. M47T1*, an uncharacterized protein with 6 hit-instances, showing 49% sequence identity to A0A1B4WVT0, a putative transferase. Most likely, these were significant hits because they contained a higher number of instances that aligned to the query but were disregarded because of the absence of MYC TFs in these species.

Within angiosperms (still excluding Brassicales), we found numerous significant hits to investigate further. Representative examples of our findings are shown in Figure 1 (see Table 1 for more information on each hit). Most of the putative MIM-containing proteins we identified were R2R3 MYB TFs. Within the Rosids, we could find MIM-containing R2R3 MYB TFs in most, but not all, orders (e.g., not found in Myrtales). We also found few R2R3 MYB TFs with a putative MIM in the Asterids, whereas other species with published annotated genomes in this clade (e.g., *Nicotiana benthamiana* [40]), gave no hits. The failure to identify putative MIM-containing proteins in certain organisms included in the databases does not necessarily imply their absence in that organism, as genomic information may be missing, genomes may be poorly annotated, or the instance simply fell below the significance threshold. However, our findings suggest that putative MIM-containing R2R3 MYB TFs are widespread in the eudicots.

**Table 1.** Putative MIM-containing proteins outside the Brassicales, and AtMYB29 and AtASR3 (from Figure 1). Bold: selected for further analysis.

#	Taxonomic Order	Species	Protein Type	UniProt/UniParc ID
1	Vitales	<i>Vitis vinifera</i>	R2R3 MYB	D7SI94
2 <sup>1</sup>	Brassicales	<i>Arabidopsis thaliana</i>	R2R3 MYB	Q9FLR1
3 <sup>1</sup>	<b>Brassicales</b>	<b><i>Arabidopsis thaliana</i></b>	<b>Tri-helix MYB-like</b>	<b>Q8VZ20</b>
4	<b>Malvales</b>	<b><i>Theobroma cacao</i></b>	<b>R2R3 MYB</b>	<b>A0A061FYF1</b>
5	Malvales	<i>Gossypium raimondii</i>	R2R3 MYB	A0A0D2MLR5
6	Myrtales	<i>Eucalyptus grandis</i>	bHLH	A0A059C313
7	Fagales	<i>Quercus suber</i>	R2R3 MYB	UPI000CE19059 <sup>1</sup>
8	Rosales	<i>Morus notabilis</i>	R2R3 MYB	W9QTW6
9	Rosales	<i>Trema orientalis</i>	R2R3 MYB	A0A2P5F728
10	Cucurbitales	<i>Cucumis sativus</i>	Tri-helix MYB-like	A0A0A0KDF5
11	<b>Fabales</b>	<b><i>Glycine max</i></b>	<b>R2R3 MYB</b>	<b>I1JWTO</b>
12	Fabales	<i>Phaseolus vulgaris</i>	R2R3 MYB	V7AMU5
13	Lamiales	<i>Handroanthus impetiginosus</i>	R2R3 MYB	A0A2G9H705
14	<b>Solanales</b>	<b><i>Solanum pennellii</i></b>	<b>Tri-helix MYB-like</b>	<b>UPI0007346319<sup>2</sup></b>
15	<b>Ranunculales</b>	<b><i>Macleaya cordata</i></b>	<b>R2R3 MYB</b>	<b>A0A200QES0</b>
16	Poales	<i>Ananas comosus</i>	Tri-helix MYB-like	UPI0009817C91 <sup>2</sup>
17	Poales	<i>Brachypodium distachyon</i>	3R MYB	I1HT97

<sup>1</sup> Proteins from GLS-producing plant species; <sup>2</sup> UniParc ID.



**Figure 1.** Examples of putative MYC-interaction motifs (MIMs) identified throughout the angiosperms. (a) Alignment of MIM-containing MYB transcription factors (TFs) from *A. thaliana*, with numbering according to AtMYB29, and with the MIM highlighted by a green box. (b) Phylogenetic tree showing the evolutionary relationship of orders with putative MIM-containing proteins. The tree was built with the NCBI Taxonomy Common Tree tool (<https://www.ncbi.nlm.nih.gov/taxonomy>) and is not to scale. (c) Domain organization and localization of each individual protein with a putative MIM. Domain information was obtained from PROSITE [41]. The location of the putative motif is shown in green, and the core motif residues (xLNxxA) in bold. Numbers (#) refer to each putative MIM-containing protein in Table 1.

Basal eudicots are separated from the core eudicots by more than 100 million years of evolution [42]. The Ranunculales order is the most basal clade in the basal eudicots, hence the sister group to the remaining eudicots [43]. Within this order, we found a single significant hit, in an R2R3 MYB TF of *M. cordata* (McR2R3MYB). Besides the R2R3 MYB TFs, we identified hits in proteins belonging to a few other protein families. These were all TFs, although we did not impose that limit by our search criteria.

First, in several species, we identified trihelix TFs with a putative MIM near their C-terminal. Trihelix TFs have low sequence similarity to MYB TFs, and are usually not included in the MYB family although their DBD bears structural resemblance [44]. While the helices are longer in trihelix TF DBDs, they form a similar triple helix structure that interacts with DNA in a sequence-specific manner [44]. As the searches resulting in putative MIM hits amongst trihelix TFs had been restricted to exclude Brassicales, we used these sequences as queries to search for homologs in Brassicales. This resulted in the identification of the *A. thaliana* trihelix TF ASR3 (UniProt ID Q8VZ20), with a putative MIM near the C-terminus, like the trihelix TFs identified in other organisms.

Second, in Myrtales, an order that is evolutionarily rather close to the Brassicales, we did not find putative MIM-containing R2R3 MYB TFs. However, we found a significant hit within the bHLH DBD of a bHLH TF. In this case, we were not able to find a putative MIM-containing homolog in *A. thaliana* by using the bHLH sequence to search back against Brassicales. Members of the bHLH family have a DBD structurally dissimilar to the MYB DBD, and bind DNA in a different manner [45]. As the confirmed MIMs are SLiMs [4] that are accessible because of their location within IDRs, we did not further consider this putative MIM as it was found within a conserved, highly structured domain.

Third, apart from trihelix TFs with putative MIMs, the only significant hit outside the eudicots that we considered further was a 3R MYB TF in the Poales order. Like the MIMs we found in R2R3 MYB TFs, the putative MIM in this 3R MYB TF is located outside the DBD. Similar to the bHLH hit, we were not able to find Brassicales homologs of this protein with a putative MIM.

Thus, although GLS biosynthesis is the main biological process known to be regulated by the characterized MIM-containing proteins found in *A. thaliana*, we were able to locate putative MIMs in proteins of many species that do not produce GLSs (Table 1). Further, our approach enabled us to find a putative MIM in an additional *A. thaliana* protein, belonging to a different protein family, namely the trihelix TF family.

Recently, tools dedicated to SLiM discovery have been developed, for instance, SLiM-Search [46] and PSSMSearch [47]. They improve SLiM discovery by including attributes such as sequence conservation of core positions versus the surroundings, structure propensity and accessibility, sequence context, ontology, and known interaction data, directly when filtering and ranking putative hits. They are so far limited to fewer target organisms, which for Viridiplantae (phylogenetic group that contains all land plants and green algae) means only *A. thaliana* and *Physcomitrella patens*. Our searches with these tools in *A. thaliana* revealed no new hits with SLiMSearch, but numerous putative hits using PSSMSearch (Supplemental Table S1). However, ASR3 was not detected, and neither were seven out of the eight R2R3 MYB TFs used as a query, which are known to contain the motif and to interact with MYC TFs. Of these, only MYB122 came above the disorder threshold. These tools are, however, currently not suitable for addressing the main objective of this study, namely whether MIM-containing proteins are present in non-GLS producing organisms, and thus involved in other biological functions.

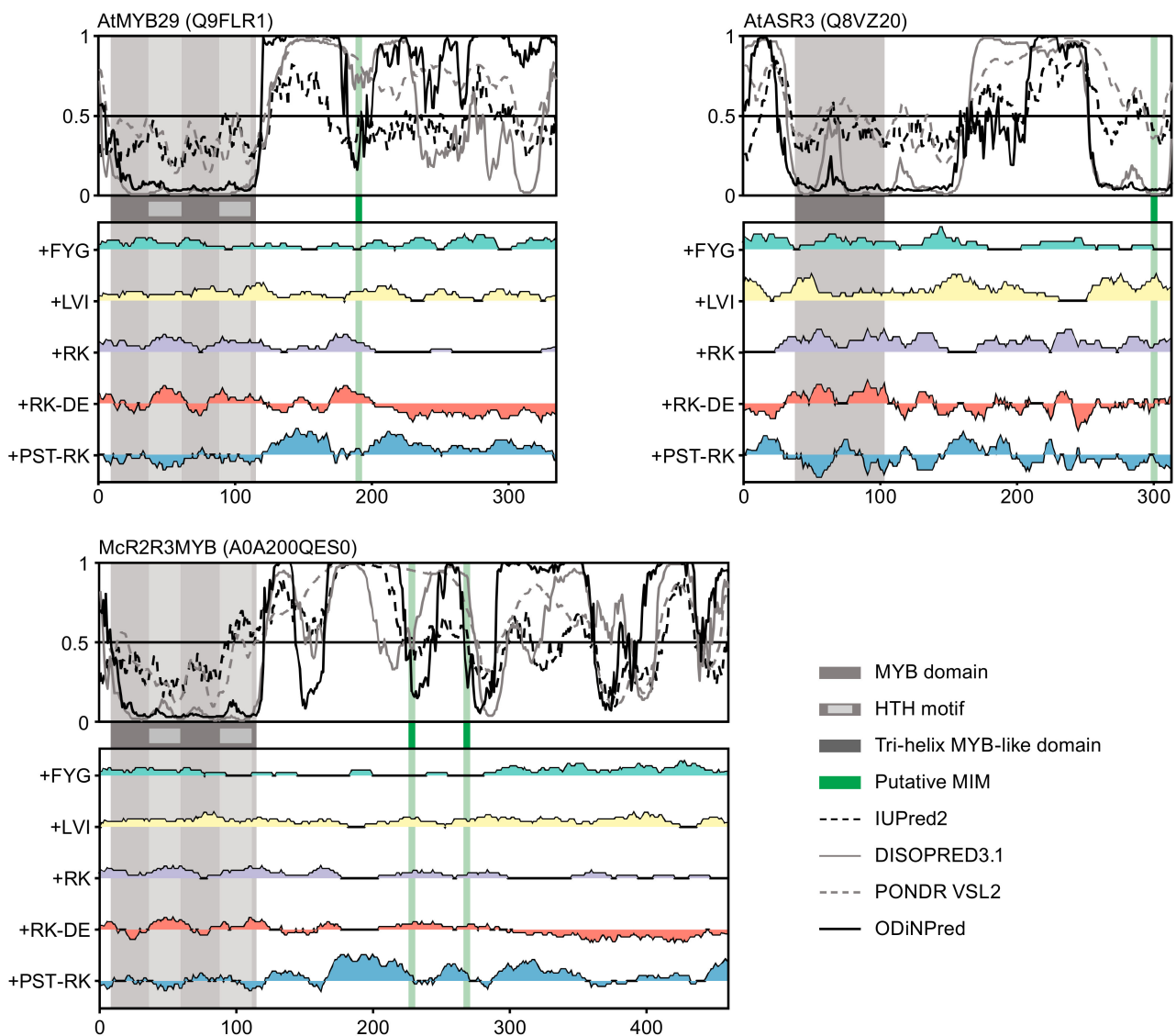
For the following analysis, we focused on the putative MIM-containing proteins found outside Brassicales, and ASR3, as it has homologs with putative MIMs throughout the angiosperms.

## 2.2. Putative MIMs in Different Protein Families Are Present in Dissimilar Sequence Contexts

The MIM is a SLiM [4], i.e., a short sequence constituting a binding motif within a larger IDR. Thus, we may expect the newly discovered MIMs to be located within similar

disordered regions. Although the core positions contribute most of the affinity and specificity of SLiM-mediated interactions [19], the flanking regions and general sequence context might allow/disallow, or at least modulate, interaction [39]. A previous attempt to introduce a functional MIM by changing a few residues in the IDR of AtMYB75 [4] indicated that there are decisive attributes besides just the xLNxxA core motif. Therefore, we could further hypothesize other MIM-containing regions to have similar chemical properties. On the other hand, previous results have shown that non-core residues and certain general properties of the motif flanking region (such as high positive charge), although conserved amongst the proteins from *A. thaliana*, were not required for interaction [4]. We selected six putative MIM-containing proteins for further analysis (named in Figure 1, bold in Table 1), to simultaneously capture the different protein families and cover a broad evolutionary distance. To assess chain properties, the proteins were first analyzed using several disorder predictors, and to visualize local compositional biases to detect disordered domains, we used IDDomainSpotter [48]. This tool calculates the relative fraction of residues within a sliding window, some residues yielding a positive contribution (+) and some a negative contribution (−) and hence allows comparison of structural and chemical properties between proteins.

Three features distinguish the MIM region in the confirmed MIM proteins [4,48] as shown here for MYB29 (Figure 2). First, the MIM coincides with local order (seen as a dip in disorder profile) within a large IDR (the region N-terminal to the MIM has been experimentally verified to be highly disordered and dynamic [48]). Second, the MIM region has an overall positive net charge (+RK−DE) compared to surrounding regions. Third, the MIM is flanked by regions with a compositional bias towards certain small, hydrophilic, and disorder-promoting residues and few positively charged residues (+PST−RK). These features were shared amongst all MIM regions in the newly discovered R2R3 MYB TFs (Figure 2 and Supplemental Figure S1).



**Figure 2.** Disorder prediction and IDDomainSpotter profiles for putative MIM-containing proteins. Top panels show sequence-specific disorder predictions using IUPred2 [49], DISOPRED3.1 [50], PONDV VSL2 [51,52], and ODINPred [53]. Bottom panels show IDDomainSpotter profiles: Each curve shows the fraction of a different set of residues within sliding windows of 15 residues [48]. Domain information was obtained from PROSITE [41]. The putative motif is shown by a green bar.

For the 3R MYB TF from *Brachypodium distachyon*, the MIM was also located within a large IDR, although not associated with a local dip in disorder (Supplemental Figure S1). For both 3R MYB and trihelix TFs, the MIM region was not characterized by the same compositional biases (as per the IDDomainSpotter profiles) as the MIM in R2R3 MYB TFs.

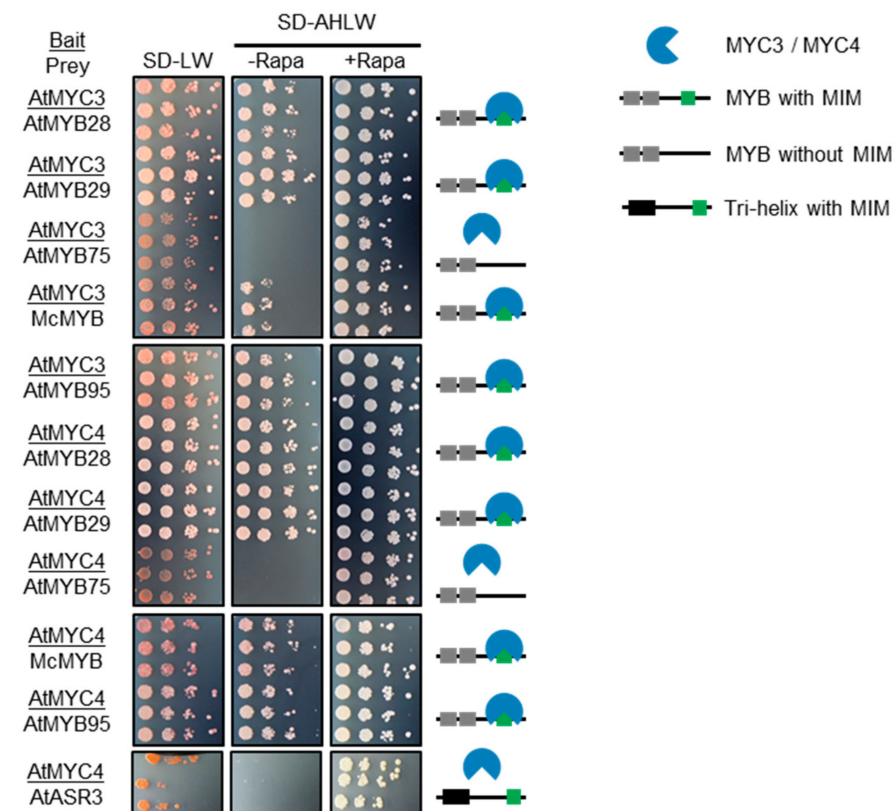
Two observations on the trihelix TFs draw special attention. First, the annotation of the trihelix domain by PROSITE does not cover the entire region suggested to be structured by the disorder predictors. Different domain databases are not in agreement for this class of trihelix TFs: for PROSITE [41] used in Figure 2, the DBD covers residues R38 to A104, while Pfam [54] indicates the residues L36 to N129 as belonging to the DBD. However, only a single three-dimensional structure of this class of DBDs has yet been determined [44], but its helices only cover parts of the alignment, leaving the conundrum unresolved. Second, the putative MIMs in the trihelix TFs were predicted to be within a more ordered region, although the different disorder predictors disagree, especially for the trihelix TFs from *Ananas comosus* and *Solanum pennellii* (Figure 2 and Supplemental Figure S1). The confirmed MIMs were predicted to be located within IDRs, and it is

therefore possible that, similar to most SLiMs [19], accessibility and flexibility is required for interaction, although this has yet to be confirmed by experiment.

To sum up, the discovered MIM-containing R2R3 MYB TFs share context characteristics with the confirmed MIM proteins, whereas the other protein families differ. Therefore, it is essential to validate our search by testing whether the newly identified putative MIMs can mediate interactions with MYC TFs.

### 2.3. Only MIMs in R2R3 MYB TFs Mediate MYC-Interaction

We next performed interaction assays with each of the R2R3 MYB and trihelix TFs shown in Figure 2 and Supplemental Figure S1, against MYC4 from *A. thaliana* (Supplemental Figure S2) or MYC3 and MYC4 (Figure 3, Supplemental Figure S2). Interaction was tested using a chemically-inducible split-ubiquitin system [55], and as a negative control we used AtMYB75, which does not have a MIM. All of the R2R3 MYB TFs tested interacted with MYC4 (Figure 3 and Supplemental Figure S2). This shows that MIM-containing R2R3 MYB TFs are present in organisms as diverse as *T. cacao*, *G. max* and *M. cordata*.



**Figure 3.** Testing interaction between *A. thaliana* (At) MYC3 and MYC4 against AtMYB28, AtMYB29, AtMYB75, AtMYB95, the MIM-containing trihelix MYB-like TF from *A. thaliana* (AtASR3), and MIM-containing R2R3 MYB TF from *Macleaya cordata* (McMYB). Growth on SD-LW medium confirms successful co-transformation with bait and prey construct and growth on SD-AHLW + Rapa medium indicates that bait and prey proteins are expressed and present in the same cellular compartment. The SD-AHLW-Rapa medium is selective for physical interaction of bait and prey protein. Rapa, rapamycin.

For physiological relevance, MIM-containing proteins must encounter interaction partners in vivo. As the molecular constituents of JA signaling, including MYC2 homologs, are conserved in land plants [36,37], the species where we identified putative MIMs should also contain MYC TFs. Therefore, to extend our analysis to a physiologically relevant MYC TF, we performed a BLAST search using AtMYC2, AtMYC3 or AtMYC4 protein sequence as query against *M. cordata* protein sequences. All three queries result in the same top candidate: (UniProt ID) A0A200PXR5 (>50% sequence identity to each AtMYC TF),

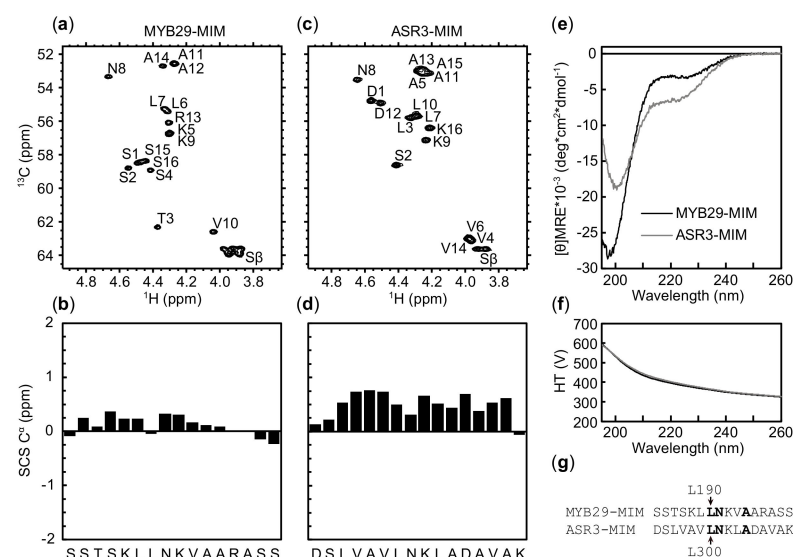


here termed McMYC. We tested interaction between McMYC and McR2R3MYB and four different MYB TFs from *A. thaliana*. In this split-ubiquitin system MYC2 is auto-activating as a bait, enabling growth on selection media even when the prey plasmid is empty. McMYC, like AtMYC2, resulted in auto-activation (Supplemental Figure S3), rendering us unable to draw any conclusions on its possible interactions.

The trihelix TFs we found, although containing the core motif residues, did not interact in our split ubiquitin assays (Figure 3, Supplemental Figure S2), possibly because of different structural context. The growth on SD-AHLW + Rapa indicates that the fusion proteins are capable of interacting, i.e., that they are expressed, correctly folded, and located in the same subcellular compartment. Thus, our next question pertained to the structural propensities of the motif region.

#### 2.4. Similar Pattern but Different Amplitude of Helical Propensity in MYB29-MIM and ASR3-MIM Peptides

As the trihelix TFs we discovered through our MIM search do not interact with MYC4 although the core positions are identical to those of the R2R3 MYB TFs, we asked whether a difference in secondary structure between the MIMs in MYB29 and ASR3 could explain this discrepancy. To answer this, synthetic peptides covering the MIM with flanking residues (16 residues in total) for MYB29 and ASR3 (Figure 4g) were subjected to structural analysis. Exploiting the natural abundance of  $^{13}\text{C}$  and  $^{15}\text{N}$ , we assigned the  $\text{C}^\alpha$  chemical shifts of MYB29-MIM and ASR3-MIM from two-dimensional nuclear magnetic resonance (NMR) spectra (Figure 4a,c and Supplemental Table S2) and calculated the secondary chemical shifts (SCS) based on sequence-corrected random coil chemical shifts [56,57], reporting on structure propensities (Figure 4b,d). Both peptides formed transient helical structures ( $\text{SCS } \text{C}^\alpha > 0.1$ ), with a slight dip around residues L190-N191 for MYB29 and L300-N301 for ASR3 (the LN residues in the core motif xLNxxA). ASR3-MIM was substantially more helical ( $\text{SCS } \text{C}^\alpha$  values reaching  $>0.75$ ;  $\sim 25\%$  helicity [58,59]) than MYB29-MIM (highest  $\text{SCS } \text{C}^\alpha$  value at 0.36;  $\sim 12\%$ ). Further, for ASR3-MIM, helical propensity was observed throughout the entire length of the peptide, whereas for MYB29-MIM it was observed only for residues S185-A195.



**Figure 4.** Structure elucidation of MYB29-MIM and ASR3-MIM.  $^{13}\text{C}$ -HSQC spectra of (a) MYB29-MIM and (c) ASR3-MIM with assignments. SCS of (b) MYB29-MIM and (d) ASR3-MIM. (e) Far-UV CD spectra and (f) HT voltage of MYB29-MIM and ASR3-MIM. (g) Sequence alignment of MYB29-MIM and ASR3-MIM (core motif residues in bold).

Far-UV circular dichroism (CD) spectra revealed that both peptides exhibited a pronounced negative band at 200 nm originating from disordered structures, but also some

negative ellipticity around 220 nm, indicative of residual helicity. The magnitude of these features supports a larger helix population in ASR3-MIM compared to MYB29-MIM.

Together, our results suggest that, although the core positions are conserved, and the peptides share a transient helical nature, the difference in the magnitude of their helicity and/or other dissimilar properties of their sequence context result in different interaction profiles. To further investigate the differences between this region in trihelix and R2R3 MYB TFs, we next turned to investigate their broader evolutionary conservation.

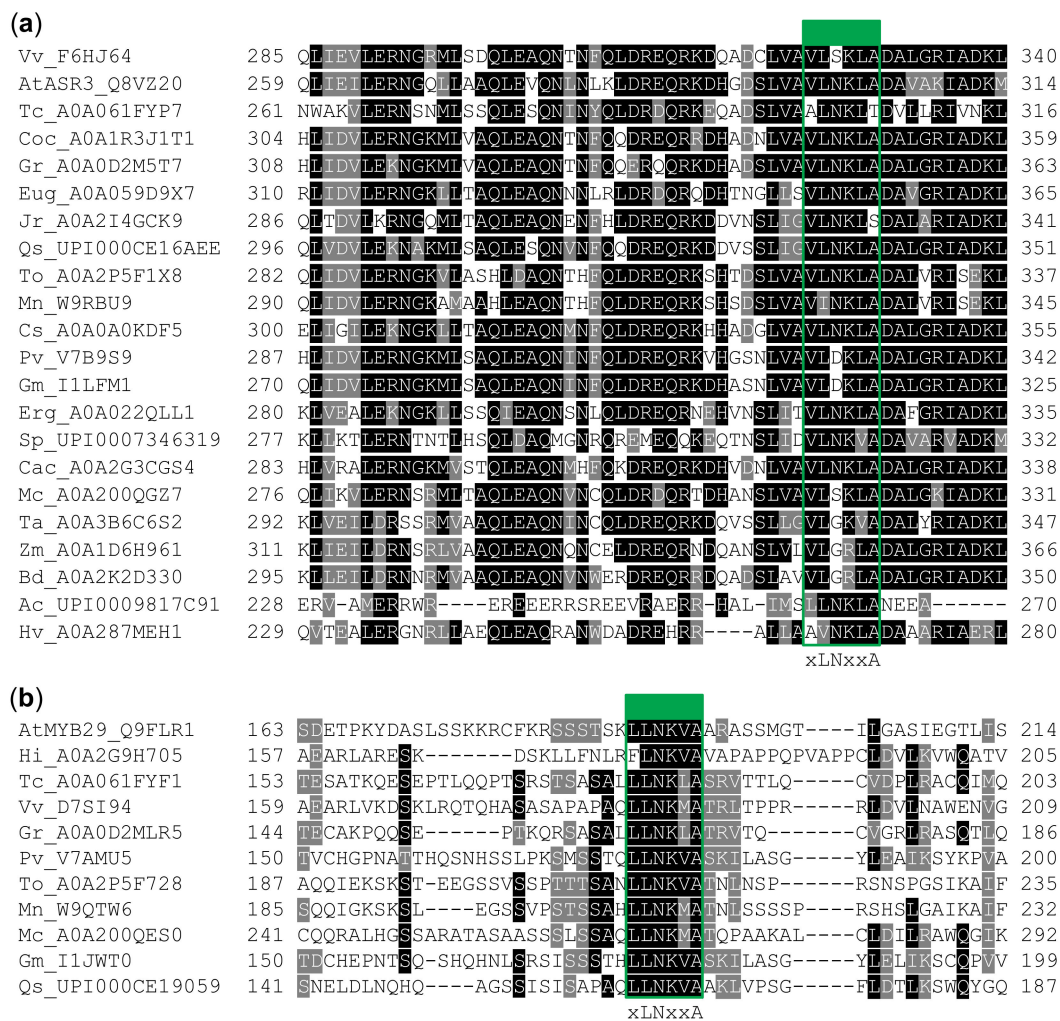
#### 2.5. Conservation of MIM and Its Context in R2R3 MYB and Trihelix TFs

SLiMs are often distinguished by relatively high conservation within otherwise rapidly evolving protein regions [20]. Low conservation is considered typical of IDRs, although that is not always the case [60–62]. Observation of the evolutionary conservation of SLiMs and flanking regions thus sheds light on the probable importance of different sequence characteristics, such as e.g., compositional biases or structural propensities.

To this end, we collected the sequences of R2R3 MYB and trihelix TF homologs that were most similar to the MIM-containing proteins, originating from species spanning the different orders of the angiosperms, also shown in Figure 1, aligned them, and colored the alignment according to conservation (Figure 5).

The results reveal a pronounced difference between the motif regions of the trihelix and R2R3 MYB TFs. The flanking regions surrounding the motifs are much more conserved in the trihelix TFs than in the R2R3 MYB TFs. In fact, the putative motif in the trihelix TFs does not stand out as more conserved than the region it is contained within, as expected for a SLiM, and as it does for the R2R3 MYB TFs.

The difference in (i) secondary structure propensity of ASR3-MIM and MYB29-MIM peptides (Figure 4), (ii) sequence characteristics of the MIM-region (Figure 2), (iii) conservation of the MIM versus MIM region (Figure 5), and (iv) ability to interact with MYC3 and MYC4 (Figure 3), indicate that the trihelix TFs were false-positives in our search for MIM-containing proteins. These differences highlight both the importance of validating *in silico* findings and the power of sequence-based predictions and analyses of intrinsic disorder, compositional biases, and evolutionary conservation.



**Figure 5.** Alignments of (a) trihelix TFs and (b) R2R3 MYB TFs colored according to conservation, performed using BoxShade ([https://embnet.vital-it.ch/software/BOX\\_form.html](https://embnet.vital-it.ch/software/BOX_form.html)). For each position where a consensus (>50%) could be defined, residues identical to this consensus were colored black, while those similar were colored gray. A green border and box highlight the putative motifs. Vv: *Vitis vinifera*, At: *A. thaliana*, Tc: *Theobroma cacao*, Coc: *Corchorus capsularis*, Gr: *Gossypium raimondii*, Eug: *Eucalyptus grandis*, Jr: *Juglans regia*, Qs: *Quercus suber*, To: *Trema orientale*, Mn: *Morus notabilis*, Cs: *Cucumis sativus*, Pv: *Phaseolus vulgaris*, Gm: *Glycine max*, Erg: *Erythranthe guttata*, Sp: *Solanum pennellii*, Cac: *Capsicum chinense*, Mc: *Macleaya cordata*, Ta: *Triticum aestivum*, Zm: *Zea mays*, Bd: *Brachypodium distachyon*, Ac: *Ananas comosus*, Hv: *Hordeum vulgare*, Hi: *Handroanthus impetiginosus*.

### 3. Discussion

#### 3.1. SLiM-Hunting Challenges

SLiMs are defined as short (3–10 residue) segments in IDRs that mediate interactions with SLiM-binding pockets [19]. SLiM-binding pockets often have several or many interaction partners that constrain the shape and characteristics of the pocket, and are consequently conserved over large evolutionary distances [20]. That only three or four residues define the SLiM makes them challenging to identify [46], as sequence databases of a certain size would by chance contain segments that match and give rise to false positives. Statistical descriptors, such as the E value, help to assess the risk of such false positives and to filter out irrelevant results. However, if the assumptions leading to the query are correct, e.g., only very few positions and sequence characteristics are required for interaction, this will likewise lead to filtering out relevant hits as false negatives. Still, predictions alone cannot be used for SLiM identification.

In the present work, we addressed these challenges by performing many searches with different information in the query. Further, iterative searches concede, and to some degree, make up for the possibility that different sequence properties may have arisen along evolutionary paths leading to other MIM-containing proteins than those known in advance. Finally, we proceeded to validate our hits by experiment.

Experimental validation is especially important for hits that either differ in the properties expected to allow interaction, such as accessibility or sequence composition, or when the biological relevance is not obvious—for example, when the hit is present in a species lacking the expected binding partner candidate or where the inferred biological function is absent. In this work, we identified hits that could belong to both of those cases: trihelix TFs have their putative MIM within a different sequence context, and all of the new MIM-containing MYB TFs identified are present in species lacking GLSs.

The trihelix TFs we identified turned out to be false positives, highlighting the importance of local structure and sequence composition in the MIM region, and thus the power of sequence predictions when comparing proteins. When the sequence and structure predictions of putative MIMs matched those of confirmed MIMs, we were able to validate MYC interactions experimentally (Figure 3 and Supplemental Figure S2). That the MIM is functional in several R2R3 MYB TFs outside Brassicales indicates that the biological function of the MIM is broader than regulation of GLS biosynthesis.

Most of the hits found using PSSMSearch (Supplemental Table S1)—like the trihelix TFs—have a different sequence composition in the motif flanking region (e.g., not an overrepresentation of small hydrophilic or positive residues), and residues with different properties than the experimentally validated MIMs at non-core positions, e.g., negatively charged, bulky, structure-disruptive (Pro), or aromatic. Although the core residues (those that by themselves disrupt interaction upon mutation), and a few other hypothesis-driven sequence characteristics (e.g., positive charges) have been investigated [4], we are still unaware of residues that are *disallowed* at certain positions. This information would facilitate MIM discovery especially when using tools such as PSSMSearch. Still, these hits were filtered by disorder (the default disorder cut-off that also filtered out ASR3 and seven out of eight known MIM-containing R2R3 MYB TFs in *A. thaliana*). Whether any of the hits found using PSSMSearch (Supplemental Table S1) can interact with MYC TFs is not known.

Although not indicated as a domain by either PROSITE and Pfam, the C-terminal predicted-to-be-ordered region in ASR3 and the other identified trihelix TFs, is conserved among trihelix TFs from the SH4 clade (which ASR3 belongs to) [63]. The structure has not been determined experimentally, but it has been proposed to form a coiled-coil responsible for homodimerization. ASR3 does homodimerize, a process that was blocked when residues K255 to Q288 were deleted [64] (the putative MIM covers <sub>299</sub>VLNKLA<sub>304</sub>). That the entire region is conserved (Figure 5) and not involved in MYC interaction (Figure 3) suggests that the presence of many instances matching the core motif was not because of selection for MYC interaction, but instead because of other evolutionary constraints, possibly of structural origin.

### 3.2. Biological Function of MIM-Containing R2R3 MYB TFs Not Involved in GLS Regulation

We identified and validated several MIM-containing TFs from species that do not produce GLSs (Figure 1 and Table 1). Furthermore, two additional R2R3 MYB TFs from *A. thaliana*, MYB95 and MYB47, contain a MIM. The MIMs present in TFs of non-GLS producing plants (and MYB47 and MYB95) may be involved in regulating a more ancestral biological process. Since JA has an established role in defense [35], and the MIM offers a direct regulatory connection to JA signaling, this could, for example, be biosynthesis of other defense compounds. However, JA also plays roles in regulation of e.g., abiotic stress responses [65] and plant development, including reproductive organs [66]. MYB47 is involved in seed longevity [9]. Whether interactions between MYB47 and MYC TFs are involved in coordinating JA signaling into the regulation of seed longevity is unknown.

MYC2 has been suggested to be a hub that interacts with countless other proteins to fine-tune most JA responses [67]. None of these already known interaction partners were identified by our search for new MIM instances, suggesting that they use other binding sites to interact with MYC2.

### 3.3. Evolutionary History of the MIM

The short length of SLiMs increases their chances of evolving convergently, as they may arise from only a few base substitutions within an IDR. Convergent evolutionary events challenge the establishment of evolutionary relationships. We identified putative MIMs in R2R3 MYB TFs of several species spread throughout the eudicot lineage (Figure 1). Furthermore, through experiment, we validated that out of three tested newly discovered MIM-containing R2R3 MYB TFs, all interacted with AtMYC4 (Figure 3). Confirmed MIM-containing R2R3 MYB TFs thus belong to species as diverse as *A. thaliana*, *T. cacao*, *G. max*, and *M. cordata*. If all the R2R3 MYB TFs we found (Figure 1 and Table 1) share evolutionary origin, this would suggest that the MIM evolved before the split between basal and core eudicots, which is over 100 million years ago [42]. An important question is therefore why there are many species within the eudicots where we failed to identify MIM-containing proteins, even though we explicitly searched in several model organisms (for instance, *Populus trichocarpa*, *Nicotiana benthamiana*, and *Solanum lycopersicum*). A combination of the following factors could explain this apparent discrepancy. (1) For many species, high quality annotated genomes are unavailable. Thus, they may still be present but remain to be annotated. (2) Our analysis may not have been exhaustive due to a too restrictive SLiM definition. (3) The SLiM may in some species have diverged at core SLiM positions, making them silent to our searches. (4) Some species could have lost putative MIMs or MIM-containing proteins. (5) The numerous instances of MIM-containing proteins could reflect a high possibility of convergence.

Several proteins confirmed to interact with MYC TFs have more than one putative MIM. These include AtMYB34 (<sub>148</sub>SGSARLLNVRASKYAV<sub>163</sub> and <sub>192</sub>SPTSTLLNKMAATS-VL<sub>207</sub>) and McR2R3MYB (<sub>222</sub>SSSARLLNKIATQAAT<sub>237</sub> and <sub>262</sub>LSSAQLLNKMATQPAA<sub>277</sub>). Further, the R2R3 MYB TFs identified in *Rosales* each contain three putative MIMs (Figure 1). Whether just one or more MIMs are functional remains to be experimentally evaluated, but if more are functional this could further support that MIMs may evolve convergently, and possibly in some cases appear in multiples to tune MYB-affinity [68,69].

If by chance a motif evolves ex nihilo, which enables a new, beneficial interaction, the interaction may initially have binding characteristics (such as affinity and specificity) that are far from optimal [19]. Subsequent mutations in the core motif and in surrounding residues then fine-tune the interaction. We most frequently identified MIMs in R2R3 MYB TFs, which could suggest that these TFs have properties that facilitate the evolution of a MIM, such as specific protein segments where it can readily evolve. Alternatively, these TFs regulate pathways where the connection to JA signaling is beneficial. Further work is needed to determine whether the many MIM instances in R2R3 MYB TFs are the result of convergent evolution or a shared evolutionary origin.

### 3.4. Evolution of Regulation in Novel Biological Processes

Since the MIM defines a subgroup of MYB TFs that are all involved in GLS regulation in *A. thaliana*, and they have homologs throughout Brassicaceae, it seems that once GLS biosynthesis evolved, early species in this taxon or their ancestor implemented the MIM to regulate GLS biosynthesis. As JA signaling is elevated by e.g., wounding and herbivore feeding [70], this allowed induction by relevant environmental stimuli to control the biosynthesis of this novel defense compound. JA signaling is much older than MIMs [36,37], and other MYB TFs bypass MYC-interaction through direct interactions with JAZ proteins [71]. Thus, when first evolved, the MIM was likely employed as a new way to connect to JA signaling. Further study of the evolution of protein interactions such as these will reveal how organisms employ already existing interaction motifs to regulate newly evolved biological processes.

Ex nihilo evolution of SLiMs allows adding molecular function to previously non-functional regions of proteins [20], providing a resource for evolutionary plasticity through new regulatory connections. Disorder can be a driver of innovation and divergence in protein function, and IDRs are capable of rapidly rewiring their interaction networks, compared to structured protein domains [72]. R2R3 MYB TFs have extensive IDRs [14], like TFs in general [15–18]. Thus, the ability of SLiMs to evolve convergently may have been key for the development and expansion of combinatorial control of gene expression as observed through cross-family TF interactions in plants [1]. Exuberated combinatorial control of gene expression is essential for plants to conduct coordinated physiological responses depending on complex environmental conditions.

## 4. Materials and Methods

### 4.1. Bioinformatics

Several tools were applied to search for MIM-containing proteins, including PHI-BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) and jackhmmer (<https://www.ebi.ac.uk/Tools/hmmer/>). The searches were performed against several different databases (non-redundant protein sequences (nr), Reference Proteomes, Ensembl, and Ensembl Genomes), to cover as many organisms as possible (the nr database includes all non-redundant CDS translations of GenBank along with all RefSeq, UniProtKB/Swiss-Prot, PDB and PRF proteins [73]).

Disorder prediction was performed using the following webservers: IUPred2 [49], DISOPRED3.1 [50], PONDR VSL2 [51,52], and ODINPred [53]. IDDomainSpotter [48] was used for calculating local compositional biases.

### 4.2. Cloning

Coding sequences with SfiI overhangs were ordered as synthetic gene fragments from Twist Bioscience (San Francisco, CA, USA), cloned into the pFRB (prey) or pFKBP12 (bait) vector [55] by restriction digest and ligation, and verified by Sanger sequencing (Eurofins Genomics, Ebersberg, Germany).

### 4.3. Split-Ubiquitin Assays

The NMY51 *TOR1-1 Δfpr1* yeast strain [55] was transformed according to DUAL-hunter™ instructions (DualsystemsBiotech, Schlieren, Switzerland). Overnight cultures (25 mL) in 2xYPAD were diluted to  $OD_{600} = 0.2$  in 500 mL, incubated at 30 °C for 4–6 h with shaking 200 rpm, and sedimented at 700 g for 5 min. The yeast cells were washed in 300 mL MQ water, then 100 mL 0.1 M LiOAc, resuspended in 15 mL 0.1 M LiOAc with 15% (v/v) glycerol, aliquoted and stored at –80 °C. Double transformation was performed by thawing the cells on ice (50 µL per transformation), mixing with transformation mix (240 µL 50% (v/v) PEG-3350, 35 µL 1 M LiOAc and 25 µL 2 mg/mL ssDNA per transformation) and 200 ng each of the bait and prey plasmids, and transferred to a 42 °C water bath for 45 min. The cells were then chilled on ice, sedimented at  $3000 \times g$  for 1 min, resuspended in sterile water and plated on SC-LW plates.

Three individual colonies of each bait/prey combination were inoculated in liquid SD-LW media and grown overnight at 30 °C with shaking 200 rpm. The cultures were adjusted to  $OD_{600} = 0.05$  and spotted on plates (SD-LW, SD-AHLW and SD-AHLW with 10 µg/mL Rapamycin) in 10-fold dilution series ( $OD_{600} = 0.05$ ,  $OD_{600} = 0.005$ ,  $OD_{600} = 0.0005$ , and  $OD_{600} = 0.00005$ ). The plates were examined and pictures taken after 2 or 3 days of growth at 30 °C.

### 4.4. CD and NMR Spectroscopy

Amidated and acetylated peptides MYB29-MIM (SSTSKLLNKVAARASS) and ASR3-MIM (DSLVAVLNKLADAVAK) were synthesized by TAG Copenhagen A/S (Copenhagen, Denmark).

For far-UV CD spectroscopy, the peptides were dissolved in 10 mM sodium phosphate pH 7.0 and centrifuged at  $15,000\times g$  for 20 min. The peptide concentration was estimated from absorbance at 214 nm [74] and the peptides were diluted to 60  $\mu\text{M}$  with 10 mM sodium phosphate pH 7.0, and pH corrected to 7.0. Far-UV CD spectra were recorded from 260 nm to 190 nm at 20 °C on a Jasco J-815 CD spectropolarimeter with 2 s D.I.T., 1 nm bandwidth, and 0.1 nm data pitch at 10 nm/min. 20 scans were accumulated. Buffer spectra were recorded with identical settings and subtracted before analysis.

For NMR spectroscopy, ASR3-MIM was dissolved in 50 mM sodium phosphate pH 7.0. MYB29-MIM was dissolved in a small amount of 25% (*v/v*) acetic acid, before being diluted with 50 mM sodium phosphate, pH 7.0. The 1D  $^1\text{H}$  NMR spectrum of this sample was identical to that of a similar sample of lower concentration MYB29-MIM that failed to dissolve completely in 50 mM sodium phosphate pH 7.0. All NMR samples were recorded in 600  $\mu\text{L}$  total volume, supplemented with 5% (*v/v*)  $\text{D}_2\text{O}$ , 125  $\mu\text{M}$  2,2-Dimethyl-2-silapentane-5-sulfonate (DSS) and 0.02% (*w/v*)  $\text{NaN}_3$ , and adjusted to pH 7.0. Final peptide concentrations were approximately 1.8 mM for MYB29-MIM and 0.9 mM for ASR3-MIM, as ASR3-MIM self-interacted at high concentrations.  $^1\text{H}$ -,  $^1\text{H}$ - $^1\text{H}$  TOCSY-,  $^1\text{H}$ - $^1\text{H}$  ROESY-,  $^{13}\text{C}$ - $^1\text{H}$  HSQC- and  $^{15}\text{N}$ - $^1\text{H}$  HSQC- spectra were recorded at 5 °C on a Bruker 800 MHz spectrometer. For MYB29-MIM, the  $^1\text{H}$ - $^1\text{H}$  TOCSY- and  $^{13}\text{C}$ - $^1\text{H}$  HSQC- spectra were also recorded at 25 °C to reveal the overlap between the lysine and arginine residues. Topspin (Bruker, Billerica, MA, USA) was used for recording, transforming and referencing the spectra.  $^1\text{H}$  chemical shifts were referenced directly to DSS at 0.00 ppm, and heteronuclei referenced indirectly by relative gyromagnetic ratios. CcpNmr analysis software [75] was used for manually assigning the spectra. Sequence-corrected random coil  $\text{C}^\alpha$  shifts were calculated with correction for temperature and pH [56,57].

**Supplementary Materials:** Supplementary Materials can be found at <https://www.mdpi.com/1422-0067/22/1/447/s1>.

**Author Contributions:** Conceptualization, P.S.M., B.B.K. and M.B.; investigation, P.S.M.; Writing—Original draft preparation, P.S.M.; Writing—Review and editing, P.S.M., B.B.K. and M.B.; visualization, P.S.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by The Danish National Research Foundation, DNRF grant 99 (PSM, MB), Villumfonden, project no. 13169 (PSM, MB), the Novo Nordisk Foundation Challenge grant REPIN (to B.B.K) #NNF18OC0033926 and the NMR spectra were recorded in cOpenNMR, an infrastructure supported by the Novo Nordisk Foundation #NNF18OC0032996.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study is contained within the article or supplementary material.

**Acknowledgments:** We thank Andreas Prestel for help with NMR experiments and analysis of NMR spectra and Henriette Kehlet Heldig Jepsen and Freja Bibi Knudsen for help with the Y2H assays.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Abbreviations

MIM	MYC-interaction motif
bHLH	basic helix-loop-helix
DBD	DNA-binding domain
JA	jasmonic acid
TF	transcription factor
SLiM	short linear motif
GLS	glucosinolate
CD	circular dichroism

SCS	secondary chemical shift
NMR	nuclear magnetic resonance
IDR	intrinsically disordered region

## References

- Bemer, M.; van Dijk, A.D.J.; Immink, R.G.H.; Angenent, G.C. Cross-Family Transcription Factor Interactions: An Additional Layer of Gene Regulation. *Trends Plant Sci.* **2017**, *22*, 66–80. [[CrossRef](#)] [[PubMed](#)]
- Pireyre, M.; Burow, M. Regulation of MYB and bHLH transcription factors: A glance at the protein level. *Mol. Plant* **2015**, *8*, 378–388. [[CrossRef](#)] [[PubMed](#)]
- Feller, A.; Machemer, K.; Braun, E.L.; Grotewold, E. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J.* **2011**, *66*, 94–116. [[CrossRef](#)]
- Millard, P.S.; Weber, K.; Kragelund, B.B.; Burow, M. Specificity of MYB interactions relies on motifs in ordered and disordered contexts. *Nucleic Acids Res.* **2019**, *47*, 9592–9608. [[CrossRef](#)] [[PubMed](#)]
- Frerigmann, H. Glucosinolate regulation in a complex relationship—MYC and MYB—no one can act without each other. In *Glucosinolates. Advances in Botanical Research*; Elsevier: Amsterdam, The Netherlands, 2016; Volume 80, pp. 57–97.
- Schweizer, F.; Fernández-Calvo, P.; Zander, M.; Diez-Diaz, M.; Fonseca, S.; Glauser, G.; Lewsey, M.G.; Ecker, J.R.; Solano, R.; Reymond, P. Arabidopsis basic helix-loop-helix transcription factors MYC2, MYC3, and MYC4 regulate glucosinolate biosynthesis, insect performance, and feeding behavior. *Plant Cell* **2013**, *25*, 3117–3132. [[CrossRef](#)]
- Frerigmann, H.; Berger, B.; Gigolashvili, T. bHLH05 is an interaction partner of MYB51 and a novel regulator of glucosinolate biosynthesis in Arabidopsis. *Plant Physiol.* **2014**, *166*, 349–369. [[CrossRef](#)]
- Seo, M.-S.; Kim, J.S. Understanding of MYB transcription factors involved in glucosinolate biosynthesis in Brassicaceae. *Molecules* **2017**, *22*, 1549.
- Renard, J.; Niñoles, R.; Martínez-Almonacid, I.; Gayubas, B.; Mateos-Fernández, R.; Bissoli, G.; Bueso, E.; Serrano, R.; Gadea, J. Identification of novel seed longevity genes related to oxidative stress and seed coat by genome-wide association studies and reverse genetics. *Plant Cell Environ.* **2020**, *43*, 2523–2539. [[CrossRef](#)]
- Ogata, K.; Hojo, H.; Aimoto, S.; Nakai, T.; Nakamura, H.; Sarai, A.; Ishii, S.; Nishimura, Y. Solution structure of a DNA-binding unit of Myb: A helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 6428–6432. [[CrossRef](#)]
- Dubos, C.; Stracke, R.; Grotewold, E.; Weisshaar, B.; Martin, C.; Lepiniec, L. MYB transcription factors in Arabidopsis. *Trends Plant Sci.* **2010**, *15*, 573–581. [[CrossRef](#)]
- Du, H.; Liang, Z.; Zhao, S.; Nan, M.-G.; Tran, L.-S.P.; Lu, K.; Huang, Y.-B.; Li, J.-N. The Evolutionary History of R2R3-MYB Proteins across 50 Eukaryotes: New Insights into Subfamily Classification and Expansion. *Sci. Rep.* **2015**, *5*, 11037. [[CrossRef](#)] [[PubMed](#)]
- Stracke, R.; Werber, M.; Weisshaar, B. The R2R3-MYB gene family in Arabidopsis thaliana. *Curr. Opin. Plant Biol.* **2001**, *4*, 447–456. [[CrossRef](#)]
- Millard, P.S.; Kragelund, B.B.; Burow, M. R2R3 MYB Transcription Factors—Functions outside the DNA-Binding Domain. *Trends Plant Sci.* **2019**, *24*, 934–946. [[CrossRef](#)] [[PubMed](#)]
- Liu, J.; Perumal, N.B.; Oldfield, C.J.; Su, E.W.; Uversky, V.N.; Dunker, A.K. Intrinsic disorder in transcription factors. *Biochemistry* **2006**, *45*, 6873–6888. [[CrossRef](#)]
- Shammas, S.L. Mechanistic roles of protein disorder within transcription. *Curr. Opin. Struct. Biol.* **2017**, *42*, 155–161. [[CrossRef](#)] [[PubMed](#)]
- Babu, M.M. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochem. Soc. Trans.* **2016**, *44*, 1185–1200. [[CrossRef](#)]
- Wright, P.E.; Dyson, H.J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 18–29. [[CrossRef](#)]
- Van Roey, K.; Uyar, B.; Weatheritt, R.J.; Dinkel, H.; Seiler, M.; Budd, A.; Gibson, T.J.; Davey, N.E. Short linear motifs: Ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem. Rev.* **2014**, *114*, 6733–6778. [[CrossRef](#)]
- Davey, N.E.; Cyert, M.S.; Moses, A.M. Short linear motifs—ex nihilo evolution of protein regulation. *Cell Commun. Signal.* **2015**, *13*, 43. [[CrossRef](#)]
- Agerbirk, N.; Olsen, C.E. Glucosinolate structures in evolution. *Phytochemistry* **2012**, *77*, 16–45. [[CrossRef](#)]
- Edger, P.P.; Heidele-Fischer, H.M.; Bekaert, M.; Rota, J.; Glöckner, G.; Platts, A.E.; Heckel, D.G.; Der, J.P.; Wafula, E.K.; Tang, M.; et al. The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 8362–8366. [[CrossRef](#)] [[PubMed](#)]
- Pastorczyk, M.; Bednarek, P. The function of glucosinolates and related metabolites in plant innate immunity. In *Glucosinolates. Advances in Botanical Research*; Elsevier: Amsterdam, The Netherlands, 2016; Volume 80, pp. 171–198.
- Beekwilder, J.; van Leeuwen, W.; van Dam, N.M.; Bertossi, M.; Grandi, V.; Mizzi, L.; Soloviev, M.; Szabados, L.; Molthoff, J.W.; Schipper, B.; et al. The impact of the absence of aliphatic glucosinolates on insect herbivory in Arabidopsis. *PLoS ONE* **2008**, *3*, e2068. [[CrossRef](#)] [[PubMed](#)]
- Jeschke, V.; Kearney, E.E.; Schramm, K.; Kunert, G.; Shekhov, A.; Gershenzon, J.; Vassão, D.G. How glucosinolates affect generalist lepidopteran larvae: Growth, development and glucosinolate metabolism. *Front. Plant Sci.* **2017**, *8*, 1995. [[CrossRef](#)] [[PubMed](#)]



26. Malinovsky, F.G.; Thomsen, M.-L.F.; Nintemann, S.J.; Jagd, L.M.; Bourguine, B.; Burow, M.; Kliebenstein, D.J. An evolutionarily young defense metabolite influences the root growth of plants via the ancient TOR signaling pathway. *Elife* **2017**, *6*, e29353. [[CrossRef](#)]
27. Francisco, M.; Joseph, B.; Caligagan, H.; Li, B.; Corwin, J.A.; Lin, C.; Kerwin, R.E.; Burow, M.; Kliebenstein, D.J. Genome Wide Association Mapping in *Arabidopsis thaliana* Identifies Novel Genes Involved in Linking Allyl Glucosinolate to Altered Biomass and Defense. *Front. Plant Sci.* **2016**, *7*, 1010. [[CrossRef](#)]
28. Francisco, M.; Joseph, B.; Caligagan, H.; Li, B.; Corwin, J.A.; Lin, C.; Kerwin, R.; Burow, M.; Kliebenstein, D.J. The Defense Metabolite, Allyl Glucosinolate, Modulates *Arabidopsis thaliana* Biomass Dependent upon the Endogenous Glucosinolate Pathway. *Front. Plant Sci.* **2016**, *7*, 774. [[CrossRef](#)]
29. Katz, E.; Nisani, S.; Yadav, B.S.; Woldemariam, M.G.; Shai, B.; Obolski, U.; Ehrlich, M.; Shani, E.; Jander, G.; Chamovitz, D.A. The glucosinolate breakdown product indole-3-carbinol acts as an auxin antagonist in roots of *Arabidopsis thaliana*. *Plant J.* **2015**, *82*, 547–555. [[CrossRef](#)]
30. Kliebenstein, D.J.; Figuth, A.; Mitchell-Olds, T. Genetic architecture of plastic methyl jasmonate responses in *Arabidopsis thaliana*. *Genetics* **2002**, *161*, 1685–1696.
31. Mewis, I.; Appel, H.M.; Hom, A.; Raina, R.; Schultz, J.C. Major signaling pathways modulate *Arabidopsis* glucosinolate accumulation and response to both phloem-feeding and chewing insects. *Plant Physiol.* **2005**, *138*, 1149–1162. [[CrossRef](#)]
32. Chini, A.; Fonseca, S.; Fernández, G.; Adie, B.; Chico, J.M.; Lorenzo, O.; García-Casado, G.; López-Vidriero, I.; Lozano, F.M.; Ponce, M.R.; et al. The JAZ family of repressors is the missing link in jasmonate signalling. *Nature* **2007**, *448*, 666–671. [[CrossRef](#)]
33. Cheng, Z.; Sun, L.; Qi, T.; Zhang, B.; Peng, W.; Liu, Y.; Xie, D. The bHLH transcription factor MYC3 interacts with the Jasmonate ZIM-domain proteins to mediate jasmonate response in *Arabidopsis*. *Mol. Plant* **2011**, *4*, 279–288. [[CrossRef](#)] [[PubMed](#)]
34. Fernández-Calvo, P.; Chini, A.; Fernández-Barbero, G.; Chico, J.M.; Gimenez-Ibanez, S.; Geerinck, J.; Eeckhout, D.; Schweizer, F.; Godoy, M.; Franco-Zorrilla, J.M.; et al. The *Arabidopsis* bHLH transcription factors MYC3 and MYC4 are targets of JAZ repressors and act additively with MYC2 in the activation of jasmonate responses. *Plant Cell* **2011**, *23*, 701–715. [[CrossRef](#)] [[PubMed](#)]
35. Gimenez-Ibanez, S.; Solano, R. Nuclear jasmonate and salicylate signaling and crosstalk in defense against pathogens. *Front. Plant Sci.* **2013**, *4*, 72. [[CrossRef](#)] [[PubMed](#)]
36. Han, G.-Z. Evolution of jasmonate biosynthesis and signaling mechanisms. *J. Exp. Bot.* **2017**, *68*, 1323–1331. [[CrossRef](#)]
37. Schluttenhofer, C. Origin and evolution of jasmonate signaling. *Plant Sci.* **2020**, *298*, 110542. [[CrossRef](#)]
38. Prestel, A.; Wichmann, N.; Martins, J.M.; Marabini, R.; Kassem, N.; Broendum, S.S.; Otterlei, M.; Nielsen, O.; Willemoës, M.; Ploug, M.; et al. The PCNA interaction motifs revisited: Thinking outside the PIP-box. *Cell Mol. Life Sci.* **2019**, *76*, 4923–4943. [[CrossRef](#)]
39. Bugge, K.; Brakti, I.; Fernandes, C.B.; Dreier, J.E.; Lundsgaard, J.E.; Olsen, J.G.; Skriver, K.; Kragelund, B.B. Interactions by Disorder—A Matter of Context. *Front. Mol. Biosci.* **2020**, *7*, 110. [[CrossRef](#)]
40. Bombarely, A.; Rosli, H.G.; Vrebalov, J.; Moffett, P.; Mueller, L.A.; Martin, G.B. A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research. *Mol. Plant Microbe Interact.* **2012**, *25*, 1523–1530. [[CrossRef](#)]
41. Sigrist, C.J.A.; de Castro, E.; Cerutti, L.; Cucho, B.A.; Hulo, N.; Bridge, A.; Bougueleret, L.; Xenarios, I. New and continuing developments at PROSITE. *Nucleic Acids Res.* **2013**, *41*, D344–D347. [[CrossRef](#)]
42. Chaw, S.-M.; Chang, C.-C.; Chen, H.-L.; Li, W.-H. Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. *J. Mol. Evol.* **2004**, *58*, 424–441.
43. The Angiosperm Phylogeny Group An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botan. J. Linn. Soc.* **2016**, *181*, 1–20. [[CrossRef](#)]
44. Nagata, T.; Niyada, E.; Fujimoto, N.; Nagasaki, Y.; Noto, K.; Miyanoiri, Y.; Murata, J.; Hiratsuka, K.; Katahira, M. Solution structures of the trihelix DNA-binding domains of the wild-type and a phosphomimetic mutant of *Arabidopsis* GT-1: Mechanism for an increase in DNA-binding affinity through phosphorylation. *Proteins* **2010**, *78*, 3033–3047. [[CrossRef](#)]
45. Lian, T.-F.; Xu, Y.-P.; Li, L.-F.; Su, X.-D. Crystal Structure of Tetrameric *Arabidopsis* MYC2 Reveals the Mechanism of Enhanced Interaction with DNA. *Cell Rep.* **2017**, *19*, 1334–1342. [[CrossRef](#)] [[PubMed](#)]
46. Krystkowiak, I.; Davey, N.E. SLIMSearch: A framework for proteome-wide discovery and annotation of functional modules in intrinsically disordered regions. *Nucleic Acids Res.* **2017**, *45*, W464–W469. [[CrossRef](#)] [[PubMed](#)]
47. Krystkowiak, I.; Manguy, J.; Davey, N.E. PSSMSearch: A server for modeling, visualization, proteome-wide discovery and annotation of protein motif specificity determinants. *Nucleic Acids Res.* **2018**, *46*, W235–W241. [[CrossRef](#)]
48. Millard, P.S.; Bugge, K.; Marabini, R.; Boomsma, W.; Burow, M.; Kragelund, B.B. IDDomainSpotter: Compositional bias reveals domains in long disordered protein regions—Insights from transcription factors. *Protein Sci.* **2020**, *29*, 169–183. [[CrossRef](#)]
49. Mészáros, B.; Erdos, G.; Dosztányi, Z. IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **2018**, *46*, W329–W337. [[CrossRef](#)]
50. Jones, D.T.; Cozzetto, D. DISOPRED3: Precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **2015**, *31*, 857–863. [[CrossRef](#)]
51. Obradovic, Z.; Peng, K.; Vucetic, S.; Radivojac, P.; Dunker, A.K. Exploiting heterogeneous sequence properties improves prediction of protein disorder. *Proteins* **2005**, *61* (Suppl. S7), 176–182. [[CrossRef](#)]
52. Peng, K.; Radivojac, P.; Vucetic, S.; Dunker, A.K.; Obradovic, Z. Length-dependent prediction of protein intrinsic disorder. *BMC Bioinform.* **2006**, *7*, 208. [[CrossRef](#)]

53. Dass, R.; Mulder, F.A.A.; Nielsen, J.T. ODINPred: Comprehensive prediction of protein order and disorder. *Sci. Rep.* **2020**, *10*, 14780. [[CrossRef](#)] [[PubMed](#)]
54. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A.; et al. The Pfam protein families database. *Nucleic Acids Res.* **2019**, *47*, D427–D432. [[CrossRef](#)] [[PubMed](#)]
55. Andersen, T.G.; Nintemann, S.J.; Marek, M.; Halkier, B.A.; Schulz, A.; Burow, M. Improving analytical methods for protein-protein interaction through implementation of chemically inducible dimerization. *Sci. Rep.* **2016**, *6*, 27766. [[CrossRef](#)] [[PubMed](#)]
56. Kjaergaard, M.; Brander, S.; Poulsen, F.M. Random coil chemical shift for intrinsically disordered proteins: Effects of temperature and pH. *J. Biomol. NMR* **2011**, *49*, 139–149. [[CrossRef](#)] [[PubMed](#)]
57. Kjaergaard, M.; Poulsen, F.M. Sequence correction of random coil chemical shifts: Correlation between neighbor correction factors and changes in the Ramachandran distribution. *J. Biomol. NMR* **2011**, *50*, 157–165. [[CrossRef](#)] [[PubMed](#)]
58. Wishart, D.S.; Sykes, B.D.; Richards, F.M. The chemical shift index: A fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **1992**, *31*, 1647–1651. [[CrossRef](#)] [[PubMed](#)]
59. Wishart, D.S.; Sykes, B.D. The <sup>13</sup>C chemical-shift index: A simple method for the identification of protein secondary structure using <sup>13</sup>C chemical-shift data. *J. Biomol. NMR* **1994**, *4*, 171–180. [[CrossRef](#)]
60. Paraskevopoulos, K.; Kriegenburg, F.; Tatham, M.H.; Rösner, H.I.; Medina, B.; Larsen, I.B.; Brandstrup, R.; Hardwick, K.G.; Hay, R.T.; Kragelund, B.B.; et al. Dss1 is a 26S proteasome ubiquitin receptor. *Mol. Cell* **2014**, *56*, 453–461. [[CrossRef](#)]
61. Zhou, J.; Oldfield, C.J.; Yan, W.; Shen, B.; Dunker, A.K. Intrinsically disordered domains: Sequence → disorder → function relationships. *Protein Sci.* **2019**, *28*, 1652–1663. [[CrossRef](#)]
62. Ahrens, J.; Dos Santos, H.G.; Siltberg-Liberles, J. The nuanced interplay of intrinsic disorder and other structural properties driving protein evolution. *Mol. Biol. Evol.* **2016**, *33*, 2248–2256. [[CrossRef](#)]
63. Kaplan-Levy, R.N.; Brewer, P.B.; Quon, T.; Smyth, D.R. The trihelix family of transcription factors—Light, stress and development. *Trends Plant Sci.* **2012**, *17*, 163–171. [[CrossRef](#)] [[PubMed](#)]
64. Li, B.; Jiang, S.; Yu, X.; Cheng, C.; Chen, S.; Cheng, Y.; Yuan, J.S.; Jiang, D.; He, P.; Shan, L. Phosphorylation of trihelix transcriptional repressor ASR3 by MAP KINASE4 negatively regulates Arabidopsis immunity. *Plant Cell* **2015**, *27*, 839–856. [[CrossRef](#)] [[PubMed](#)]
65. Kazan, K. Diverse roles of jasmonates and ethylene in abiotic stress tolerance. *Trends Plant Sci.* **2015**, *20*, 219–229. [[CrossRef](#)] [[PubMed](#)]
66. Yuan, Z.; Zhang, D. Roles of jasmonate signalling in plant inflorescence and flower development. *Curr. Opin. Plant Biol.* **2015**, *27*, 44–51. [[CrossRef](#)] [[PubMed](#)]
67. Kazan, K.; Manners, J.M. MYC2: The master in action. *Mol. Plant* **2013**, *6*, 686–703. [[CrossRef](#)]
68. Olsen, J.G.; Teilum, K.; Kragelund, B.B. Behaviour of intrinsically disordered proteins in protein-protein complexes with an emphasis on fuzziness. *Cell Mol. Life Sci.* **2017**, *74*, 3175–3183. [[CrossRef](#)]
69. Clark, S.; Myers, J.B.; King, A.; Fiala, R.; Novacek, J.; Pearce, G.; Heierhorst, J.; Reichow, S.L.; Barbar, E.J. Multivalency regulates activity in an intrinsically disordered transcription factor. *Elife* **2018**, *7*, e36258. [[CrossRef](#)]
70. Farmer, E.E.; Gasperini, D.; Acosta, I.F. The squeeze cell hypothesis for the activation of jasmonate synthesis in response to wounding. *New Phytol.* **2014**, *204*, 282–288. [[CrossRef](#)]
71. Huang, H.; Gao, H.; Liu, B.; Qi, T.; Tong, J.; Xiao, L.; Xie, D.; Song, S. Arabidopsis MYB24 Regulates Jasmonate-Mediated Stamen Development. *Front. Plant Sci.* **2017**, *8*, 1525. [[CrossRef](#)]
72. Ahrens, J.B.; Nunez-Castilla, J.; Siltberg-Liberles, J. Evolution of intrinsic disorder in eukaryotic proteins. *Cell Mol. Life Sci.* **2017**, *74*, 3163–3174. [[CrossRef](#)]
73. NCBI. Resource Coordinators Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2015**, *43*, D6–D17. [[CrossRef](#)] [[PubMed](#)]
74. Kuipers, B.J.H.; Gruppen, H. Prediction of molar extinction coefficients of proteins and peptides using UV absorption of the constituent amino acids at 214 nm to enable quantitative reverse phase high-performance liquid chromatography-mass spectrometry analysis. *J. Agric. Food Chem.* **2007**, *55*, 5445–5451. [[CrossRef](#)] [[PubMed](#)]
75. Vranken, W.F.; Boucher, W.; Stevens, T.J.; Fogh, R.H.; Pajon, A.; Llinas, M.; Ulrich, E.L.; Markley, J.L.; Ionides, J.; Laue, E.D. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins* **2005**, *59*, 687–696. [[CrossRef](#)] [[PubMed](#)]