Structural insights into partner selection for MYB and bHLH transcription factor complexes

Baihui Wang, Qiang Luo, Yingping Li, Kangxi Du, Zhen Wu, Tianyang Li, Wen-Hui Shen, Chien-Hsun Huang, Jianhua Gan, Aiwu Dong

Title:

Structural insights into partner selection for MYB and bHLH transcription factor complexes

Authors:

Baihui Wang[1,†], Qiang Luo[1,†], Yingping Li[1,†], Kangxi Du[1], Zhen Wu[1], Tianyang Li[1], Wen-Hui Shen[2], Chien-Hsun Huang[3*], Jianhua Gan[4*], Aiwu Dong[1,*]

[†]B.W., Q.L. and Y.L. contributed equally to this work.

[*]Correspondence: aiwudong@fudan.edu.cn, ganjhh@fudan.edu.cn and chhuang1981@139.com

Affiliations:

[1]State Key Laboratory of Genetic Engineering, Collaborative Innovation Center for Genetics and Development, Department of Biochemistry and Biophysics, Institute of Plant Biology, School of Life Sciences, Fudan University, Shanghai 200438, P.R. China

[2]Institut de Biologie Moléculaire des Plantes, CNRS, Université de Strasbourg, 12 rue du Général Zimmer, 67084 Strasbourg Cédex, France

[3]State Key Laboratory of Genetic Engineering, Center for Evolutionary Biology, Institute of Plant Biology, School of Life Sciences, Fudan University, Shanghai 200438, P.R. China

[4]Shanghai Public Health Clinical Center, State Key Laboratory of Genetic Engineering,

Collaborative Innovation Center of Genetics and Development, Department of Biochemistry and Biophysics, School of Life Sciences, Fudan University, Shanghai 200438, P.R. China

Abstract:

MYB and bHLH transcription factors form complexes to regulate diverse metabolic and developmental processes in plants. However, the molecular mechanisms responsible for MYB–bHLH interaction and partner selection remain unclear. Here, we report the crystal structures of three MYB–bHLH complexes (WER–EGL3, CPC–EGL3, and MYB29–MYC3), uncovering two MYB–bHLH interaction modes. WER and CPC are R2R3- and R3-type MYBs, respectively, but interact with EGL3 through their N-terminal R3 domain in a similar mode. A single amino acid of CPC, Met49, is crucial for competition with WER to interact with EGL3. MYB29, a R2R3-type MYB TF, interacts with MYC3 by its C-terminal MYC-interaction motif. The WER–EGL3 and MYB29–MYC3 binding modes are widely applied among MYB–bHLH complexes in Arabidopsis and evolve independently in plants.

Introduction

By recognizing specific DNA elements within the genome, transcription factors (TFs) play central roles in gene regulatory networks in multicellular organisms. In higher plants, MYB and basic helix-loop-helix (bHLH) TFs are among the largest groups of TFs; for example, more than 300 MYB and 100 bHLH TFs are present in Arabidopsis[1,2]. MYB TFs contain a conserved DNA-binding domain (MYB domain) and extensive intrinsically disordered regions (IDRs) at the C-terminus. MYB TFs are grouped into different types on the basis of three imperfect repeats (R1, R2, and R3) within the MYB domain of c-Myb, and further divided into subgroups based on the IDRs[3]. bHLH TFs are defined by ~60 conserved amino acids, including a basic DNA-binding domain and a dimerization domain[4].

Formation of complexes between MYB and bHLH TFs is widespread and different MYB–bHLH complexes regulate diverse physiological processes, including organ development, metabolic pathways, and biotic and abiotic stress responses[1-3,5,6]. For instance, WEREWOLF (WER), a well-studied R2R3-type MYB in Arabidopsis, interacts with the bHLH TFs GL3/EGL3, directly binds to and activates *GLABRA 2* (*GL2*), the central regulator of epidermal cell fate determination, and leads to the non-hair cell fate during root hair development[7]. CAPRICE (CPC), a R3-type MYB TF in Arabidopsis, competes with WER to interact with GL3/EGL3 and inhibits *GL2* expression, leading to the hair cell fate establishment[8-10]. The Arabidopsis bHLH TFs MYC2/MYC3/MYC4 interact with MYB21/MYB24 to regulate stamen development and seed production[11]. MYC2/MYC3/MYC4 play crucial roles in jasmonate (JA)

signaling through interaction with jasmonate ZIM (zinc-finger inflorescence meristem)-domain (JAZ) repressors and participate in both development and stress responses[12,13]. MYC2/MYC3/MYC4 also regulate glucosinolate biosynthesis by interacting with glucosinolate-related MYBs, including MYB28, MYB29, MYB76, MYB34, MYB51, and MYB122, which belong to MYB subgroup 12[14]. In addition, many other MYB–bHLH complexes function in plant growth and development, stress defense, and metabolism regulation[6].

The co-crystal structure of MYC3 in complex with JAZ transcriptional repressors was resolved previously[13], and recently we resolved the complex structure of WER–DNA[15]. The structure of MYB–bHLH complexes remains uncertain, which precludes understanding the molecular mechanisms that regulate the interaction and partner selection of MYB and bHLH TFs. Here we report the crystal structures of three MYB–bHLH complexes, WER–EGL3, CPC–EGL3, and MYB29–MYC3, which reveals two distinct MYB–bHLH interacting modes. The WER–EGL3 and MYB29–MYC3 binding modes are widely applied among a set of MYB‒bHLH complexes in Arabidopsis. Our results shed light on the mechanisms by which different MYB and bHLH TFs select their partners to form complexes and indicate that the two types of MYB–bHLH interaction modes are evolutionarily conserved but evolve independently in plants.

Results

Co-crystal structure of WER–EGL3 complex

Following our previous study on the crystal structure of WER in complex with its target DNA[15], we selected WER and EGL3 as a model to analyze the interaction of MYB and bHLH TFs. WER contains a R2R3-MYB domain at the N-terminus. EGL3 contains a JAZ-interacting domain (JID) and transcriptional activation domain (TAD) at the N-terminus, a bHLH domain in the central region, and a C-terminal domain (CTD) at the C-terminus (Extended Data Figs. 1a, b). Yeast two-hybrid (Y2H) and size-exclusion chromatography (SEC) experiments showed that WER interacted with EGL3 through the R3 domain of WER (WER 67–120) and the N-terminus of EGL3 (EGL3 1–205) (Extended Data Figs. 1c, d). To investigate the molecular basis underlying the interaction between WER and EGL3, we determined the crystal structure of WER 67–120 in complex with EGL3 1–205 (hereafter the WER–EGL3 complex). The complex structure was refined to 2.90 Å resolution (Table 1). The β-strands β1-β5 of EGL3 1–205 formed a flat β-sheet in the center, flanked by helices α1 and α6 on one side and by α3, α4, and α5 on the opposite side (Fig. 1a). The EGL3 JID domain consisted of helices α1 and α2, and β-strands β1, β2, and α2 located at the C-terminus of JID. Although TAD domains are generally unstructured when not bound to their targets[16-18], the EGL3 TAD was well ordered. The EGL3 TAD was composed of two helices, α3 and α4, which packed against the JID domain and β3-β5, respectively. The R3 domain of WER was composed of three helices α1-α3 linked by two short loops; the α1 and α2 helices of WER packed against the α2, α3, and α5 helices of EGL3 (Fig. 1a). Structural superposition of WER–EGL3 and WER–DNA[15] suggested that WER–EGL3 complex formation has no effect on WER–DNA interaction (Fig. 1b), which is

consistent with their inherent DNA binding and gene activation activities.

The detailed interaction between WER and EGL3 was shown in Fig. 1c. The electrostatic surface potentials of the interfaces (Extended Data Figs. 2a,b) and the electron density maps of the key residues involved in the interaction (Extended Data Fig. 3a) of WER-EGL3 were also shown. WER Gln77 at the N-terminus of α1 formed hydrogen bonds (H-bonds) with the side chain of Ser153 and the main chains of Phe156 and Leu160 of EGL3. WER Leu81 nestled in the shallow hydrophobic cavity formed by Val120, Leu160, and Leu161 of EGL3. Via the guanidine group, WER Arg84 formed two H-bonds with EGL3 Ser123. Interestingly, the guanidine group of WER Arg84 also formed an ion-π interaction with the side chain of EGL3 Phe124. Similar to Leu81, WER Leu85 and Leu88 formed hydrophobic interactions with EGL3. The side chain of Leu85 pointed toward Trp116 and Val120 of EGL3, whereas the side chain of Leu88 nestled in the pocket formed by Tyr81, Leu84, and Leu119 of EGL3. WER Arg99 at the C-terminus of α2 (residues 92–99) formed a salt bridge with Asp113 and an ion-π interaction with Trp116 of EGL3 (Fig. 1c). The crucial residues of WER involved in the interaction with EGL3 exactly overlapped with the previously identified R/B-like bHLH TF binding (RB) motif, which is responsible for interactions between MYB and R/B-like bHLH TFs [19]. *In vitro* isothermal titration calorimetry (ITC) analysis showed that mutations of all WER–EGL3 interacting residues weakened the binding affinity between WER and EGL3 (Figs. 1d, e, Supplementary Table 1). Together, these results indicated that EGL3 physically interacts with the $Q_{77}X_3L_{81}X_2R_{84}L_{85}X_2L_{88}X_{10}R_{99}$ signature motif of the WER R3 domain.

Structural basis of CPC competing with WER to bind to EGL3

Y2H assay showed that, similar to WER, CPC interacted with EGL3 1–205 through the R3 domain of CPC (Extended Data Fig. 4a). ITC analysis indicated that full-length CPC (K_d: 48.3 nM) and CPC 30–94 (CPC R3, K_d: 50 nM) displayed similar binding affinities to EGL3, thus CPC 30–94 was used in subsequent analyses (Extended Data Fig. 4b, Supplementary Table 1). We determined the crystal structure of CPC 30–94 in complex with EGL3 1–205 (hereafter the CPC–EGL3 complex) at 1.80 Å resolution (Fig. 2a, Extended Data Figs. 2c,d and 3b, Table 1). Structural superposition revealed that the overall structure of CPC–EGL3 was similar to that of WER–EGL3 (Fig. 2b), and the root mean square deviation (RMSD) between CPC–EGL3 and WER–EGL3 was only 0.70 Å, based on the superposition of 177 pairs of Cα atoms. The low RMSD value indicates that CPC and WER share the same binding mode to EGL3. The detailed CPC–EGL3 interactions are depicted in Fig. 2c. Except for Glu41 and Met49 within CPC, all residues involved in the CPC–EGL3 interaction were identical to those observed in the WER–EGL3 complex (Fig. 1c). Glu41 and Met49 of CPC corresponded to Gln77 and Leu85 of WER, respectively. Given that CPC Glu41 mainly formed weak H-bond interaction (3.0 Å) with the side chain of EGL3 Ser153, and it is variable (Asp, Glu or Gln) in other R2R3-MYB and R3-MYB TFs (Extended Data Fig. 5), thus we mainly focused on CPC Met49 and WER Leu85 for further analysis (Figs. 2d, e). The Sδ atom of CPC Met49 substitutes the Cδ2 atom of WER Leu85, forming stronger (3.4 Å vs 3.7 Å) hydrophobic interactions with the side-chain Cγ2 atom of EGL3 Val120. The Cε

150 atom of CPC Met49 stretches more closely (3.7 Å vs 4.5 Å for the Cδ2 atom of WER)
151 to the indole ring of EGL3 Trp116. In addition, the side chain of CPC Met49 attracts
152 the side chain of CPC Arg63. Compared with WER Arg99 in the WER–EGL3 complex,
153 the guanidine group of CPC Arg63 is rotated ~90° in the CPC–EGL3 complex, leading
154 to strong Van der Waals contacts with the indole ring of EGL3 Trp116 and H-bond
155 interaction with EGL3 Asp113. As calculated by PDBePISA[20], CPC Met49 occupies
156 an interface of ~67 Å^2, which was ~17 Å^2 broader than that of WER Leu85 (~50 Å^2).
157 Compared with WER Leu85, CPC Met49 stretches more closely to the indole ring of
158 EGL3 Trp116, probably forming a more stable hydrophobic interaction with EGL3 (Fig.
159 2e).

160 To validate the importance of CPC Met49, we first analyzed competition between
161 WER and CPC. *In vitro* SEC experiments (Fig. 3a) showed that adding CPC to the
162 WER–EGL3 complex resulted in a shifted elution peak from 15.75 mL (corresponding
163 to the WER–EGL3 complex) to 15.95 mL (corresponding to the CPC–EGL3 complex)
164 and an additional elution peak at 18 mL (corresponding to free WER). A SDS-PAGE
165 analysis confirmed the SEC results (Fig. 3a, Extended Data Fig. 6). But a weak band of
166 WER was still detected in the CPC–EGL3 peak (Extended Data Fig. 6a). Split-
167 luciferase assays showed that CPC inhibited the interaction between WER and EGL3
168 *in vivo* (Fig. 3b, upper panels). Together, the SEC and split-luciferase experiments
169 indicated that CPC competes with WER to interact with EGL3 *in vitro* and *in vivo*. To
170 test whether Met49 contributes to the competition between CPC and WER, we
171 substituted Met49 with Leu in CPC R3 (CPC M49L) because CPC M49L mimicked

WER R3. ITC analysis showed that the EGL3-binding affinity of CPC M49L (K_d: 137 nM) (Extended Data Fig. 7a, Supplementary Table 1) was distinctly weaker than that of wild-type CPC R3 (K_d: 50 nM) (Extended Data Fig. 4b), but comparable to that of wild-type WER R3 (K_d: 118.6 nM) (Fig. 1d). The mutant CPC M49A, a substitution with Ala, showed a much lower EGL3-binding affinity (K_d: 1000 nM; Extended Data Fig. 7a, Supplementary Table 1) compared with that of wild-type CPC, suggesting that Met49 is important for CPC interaction with EGL3. In addition to CPC mutants, we also constructed two WER mutants, WER L85M and WER Q77E/L85M, which mimicked CPC R3. ITC analysis showed that the EGL3-binding affinities of WER L85M (K_d: 72.5 nM) and WER Q77E/L85M (K_d: 73.0 nM) (Extended Data Fig. 7b, Supplementary Table 1) were similarly increased compared with that of wild-type WER (K_d: 118.6 nM) (Fig. 1d), further supporting that Met49 not Glu41 of CPC is crucial for competition with WER to interact with EGL3. Split-luciferase experiments *in planta* confirmed the importance of CPC Met49: wild-type CPC abolished the interaction between WER and EGL3 (Fig. 3b, upper panels) at the molecular ratio of 1:1:1, whereas CPC M49L and CPC M49A failed to block WER–EGL3 complex formation (Fig. 3b, lower panels), although similar amounts of wild-type and mutated CPC as well as WER and EGL3 proteins were expressed in the split-luciferase assays (Extended Data Fig. 8). In addition, we performed microscale thermophoresis (MST) analysis to quantify the competitive capability of CPC against WER. CPC peeled WER from the WER–EGL3 complex with an EC50 value of ~6.03×10^{-6} M, whereas the competitive capability of CPC M49L was reduced to ~5-fold to an EC50 value of ~3.01×10^{-5} M and the EC50

value of CPC M49A was undetectable (Fig. 3c). Consistently, split-luciferase experiments showed that WER failed to affect CPC-EGL3 complex formation at the molecular ratio of 1:1:1 (Extended Data Fig. 9a), and MST analysis showed that the competitive capability of WER against CPC is reduced to ~52-fold to an EC50 value of ~3.19×10^{-4} M compared with that of CPC against WER (Extended Data Fig. 9b). Since ITC results cannot well explain the strong competitive capability of CPC against WER, we performed Bio-Layer Interferometry (BLI) experiments to further analyze the binding affinities of CPC-EGL3 and WER-EGL3. BLI results (Extended Data Fig. 10) showed that the K_d value of CPC-EGL3 complex (K_d: 11.9 nM) was ~18-fold lower than that of WER-EGL3 complex (K_d: 222 nM), supporting our SEC, split-luciferase and MST results (Fig. 3). Together, these results proved that the Met49 residue within CPC plays a vital role in competition with WER to interact with EGL3.

Co-crystal structure of MYB29–MYC3 complex

Recently, split-ubiquitin assays identified a new MYC-interaction motif (MIM) within the subgroup 12 MYBs. The motif is located in the center of the non-MYB region and is responsible for interaction with MYC2/MYC3/MYC4[6], suggesting that the MYB29–MYC3 complex may represent a different interaction mode from that of WER–EGL3 and CPC–EGL3. To verify this assumption, we co-expressed the MIM motif of MYB29 (MYB29 174-222) and the N-terminal non-DNA-binding region of MYC3 (MYC3 44–238). By extensive crystallization trials, we solved the crystal structure of MYB29 174–222 in complex with MYC3 44–238 (hereafter MYB29–MYC3) at 2.5 Å resolution

(Extended Data Figs. 2e,f and 3c, Table 1). The MYB29 184–203 region is folded into an ordered α helix and a short loop, whereas other regions are disordered (Fig. 4a). MYB29 fit into the deep groove and formed extensive interactions with the TAD and JID domains of MYC3. MYB29 Thr186 formed an H-bond with TAD Glu148 of MYC3. MYB29 Leu190 formed extensive hydrophobic interactions with TAD Phe151 and Leu152 of MYC3. Via the side-chain OD1 and ND2 atoms, MYB29 Asn191 formed two H-bonds with the main chain N- and O-atoms of MYC3 JID Tyr97, respectively. The side chain of MYB29 Ala194 fit into a shallow concavity, formed by JID Trp92 as well as TAD Phe151 and Met155 of MYC3 (Fig. 4b). In the MYB29–MYC3 complex structure, Leu190 and Asn191 of MYB29 formed the core platform for MYC3 binding, which is consistent with previous findings that either Leu190 or Asn191 mutation within MYB29 abolishes the interaction between MYB29 and MYC3, and the mutated MYB29 fails to rescue *myb29* mutant phenotypes[6]. MST analysis further showed that all mutations of the interacting residues within MYC3 weakened the interaction between MYB29 and MYC3 (Supplementary Fig. 1). To our surprise, sequence similarity between MYB29 and JAZ proteins was very low (Supplementary Fig. 2), but the proteins displayed the same binding mode to MYC3 (Fig. 4c). In the structure of the MYB29–MYC3 and MYC3–JAZ1[13] complexes, the MYC3 JID helices flipped outward and attached loosely to the main body of MYC3, and the helices of MYB29 or JAZ proteins fit into the grooves formed by TAD and JID of MYC3. Together, these results showed that MYB29 interacts with MYC3 by its C-terminal MYC-interaction motif, different from the binding modes of WER–EGL3 and CPC–EGL3.

Discussion

Two MYB–bHLH binding modes are widely applied in Arabidopsis

Sequence alignment revealed that the MYB29-interacting residues within MYC3 were conserved in EGL3 (Supplementary Fig. 3). However, MYB29 does not interact with GL3/EGL3[6]. To investigate the basis for the two MYB–MYC binding modes, we carefully compared the structures of EGL3 and MYC3. The overall structures of EGL3 and MYC3 were similar but their JID helices behaved differently (Supplementary Fig. 4). The conformation of the EGL3 JID helix was stable and formed extensive interactions with TAD and the other regions of JID (Fig. 5a, Supplementary Fig. 4a-c). Presence or absence of WER had no effect on the conformation of the EGL3 JID helix (Supplementary Fig. 4d). In contrast, the conformation of the MYC3 JID helix was highly dynamic[13] and underwent substantial conformational change to expose TAD and other regions of JID when binding to MYB29 or JAZ1 (Fig. 5a, Supplementary Fig. 4e, f). In EGL3, the corresponding interface for adopting MYB29 was buried by the stable JID helix, which prevented MYB29 binding to EGL3 (Fig. 5a, Supplementary Fig. 4g). Except for certain residues within the MYC3 TAD, most residues responsible for EGL3–WER interaction were not conserved in MYC3 (Supplementary Fig. 3), preventing the MYC3 surface (α4 and α7) from binding to WER. These results suggested that the JID helices of EGL3 and MYC3 played critical roles in partner selection. In detail, the conformational change of the MYC3 JID helix provided the interface for binding to the subgroup 12 MYB TFs or JAZ repressors. The rigidity of

the EGL3 JID helix prevented an interface forming, but the α3 and α5 helices of EGL3 formed a novel interface that specifically recognized the RB motif in the R3 domain of MYB TFs. Thus, we uncovered two MYB–bHLH interaction modes: in the WER–EGL3 mode, α3 and α5 of EGL3 formed an interface and specifically interacted with the R3 domain of WER; in the MYB29–MYC3 mode, the MIM motif of MYB29 fit into the groove formed by the TAD and JID domains of MYC3 (Fig. 5b).

To determine the number of MYB and bHLH TFs that may utilize the two MYB–bHLH interaction modes, we searched the Arabidopsis genome for MYB and bHLH proteins that shared the crucial interaction motifs (Fig. 5c). Fifteen WER- and seven CPC-type MYBs and four EGL3-type MYCs were detected (Fig. 5c). All crucial residues within the RB motifs were highly conserved among WER- and CPC-type MYBs, and the crucial residues, especially those within the TAD domains, were conserved among the EGL3-type MYCs, indicating that many MYB–bHLH complexes likely share the same interaction mode as WER–EGL3. Similarly, eight MYB29-type MYBs and seven MYC-type MYCs were detected in the Arabidopsis genome. The crucial residues within the JID and TAD domains were highly conserved among MYC2, MYC3, MYC4, MYC5, bHLH13, AIB (bHLH17) and bHJH3. For MYB29-type MYBs, the crucial residues within the core MIM domains were highly conserved (Fig. 5c), whereas those in the loop region (corresponding to Ala197 and Ile203 in MYB29) were conserved in MYB28, MYB29, and MYB76 but varied in the other MYBs. However, physical interaction between the eight MYBs and MYC2/3/4 has been reported previously[6,14,21], indicating that MYB29-type MYBs contain a functional MIM domain

to apply the same interaction mode as MYB29–MYC3. To prove the utilization of the two MYB–bHLH interaction modes in Arabidopsis, we chose GL1 and MYB28, homologs of WER and MYB29, respectively, and predicted the structures of GL1-EGL3 and MYB28-MYC3 complexes by AlphaFold2. Structure superposition showed that the interaction modes of GL1-EGL3 and MYB28-MYC3 are exactly similar to those of WER-EGL3 and MYB29-MYC3, respectively (Supplementary Fig. 5), supporting that the interaction modes revealed by this study are conserved for other homologous MYB-bHLH complexes. Taken together, our results indicated that the two types of MYB–bHLH binding modes are widely applied in MYB–bHLH complexes in Arabidopsis.

Two MYB–bHLH interaction modes evolved independently

To investigate the occurrence of the two types of MYB–bHLH interaction modes during the evolution of land plants, we searched the genomes of representative species across the plant kingdom (Supplementary Tables 2-4) for the genes encoding homologs of EGL3, WER, CPC, MYC3, and MYB29 and reconstructed their phylogenies (Supplementary Figs. 6-10). The major results were summarized in Fig. 5d, where homologs of EGL3 and MYC3 of the bHLH family showed divergence as early as from mosses during evolution, whereas homologs of WER, CPC and MYB29 of the MYB family showed divergence later from gymnosperms during evolution. For the WER–EGL3 interaction mode, we found that the crucial motifs in EGL3 homologs were conserved in land plants (Supplementary Figs. 6 and 11), whereas those in WER

homologs were conserved in gymnosperms (Supplementary Figs. 7 and 12), which indicates that a functional WER–EGL3 interaction mode has co-evolved from the formation of motifs responsible for interacting with EGL3 in WER homologs in seed plants. For the MYB29–MYC3 interaction mode, the crucial motifs in MYC3 homologs were conserved in land plants (Supplementary Figs. 9 and 13), whereas the MIM motif of MYB29 was only detected in Arabidopsis among the 23 representative plant species (Supplementary Figs. 10 and 14, Supplementary Table 2). To further trace the origination of the MIM motif, we selected additional 20 species of eudicots (Supplementary Fig. 15, Supplementary Table 4), and found that the MIM motif was highly conserved in Brassicales species (Supplementary Fig. 16), indicating that the MYB29–MYC3 interaction mode is functional since the occurrence of the Brassicales.

To further dissect the evolution of the two interaction modes, we examined the amino acid residues of their motifs to assess whether the interaction may be substituted by other proteins involved in a different mode. From sequence alignment, we found that the residues corresponding to Tyr81, Ser123, Phe124, Leu160 and Leu161 in EGL3 have been replaced by amino acids with distinctive chemical characterizations in MYC3 homologs (Supplementary Fig. 11), which would hinder the potential for MYC3 interaction with WER/CPC homologs. Regarding the JID domain, in the major regions that distinguish the two MYB–bHLH interaction modes, the residues corresponding to Ser293 (neutral) and Tyr300 (hydrophobic) of EGL3 were replaced by Lys/Arg (hydrophilic) and Gln/His/Asn (hydrophilic or neutral), respectively, in MYC3 homologs (Supplementary Fig. 17). These results suggested that at an early stage of

land plant evolution (probably before the emergence of the mosses), the EGL3 and MYC3 clades have already evolved distinctive MYB-binding motifs. By examining the key residues in the alignment of all homologs of WER and MYB29, the crucial residues corresponding to Leu81 (hydrophobic) and Arg99 (hydrophilic) in the RB motif of WER were replaced by Thr (neutral) and His (neutral), respectively, in MYB29 homologs (Supplementary Fig. 12), and the crucial residues within the MIM motif of MYB29 were highly variable among WER homologs (Supplementary Fig. 14). These observations suggest that the two MYC-binding motifs had diverged prior to the separation of the WER and MYB29 clades in gymnosperms. Taken together, the two MYB–bHLH interacting modes are interpreted to have evolved independently.

Members applying the two MYB–bHLH interaction modes also evolved diverged biological functions. EGL3 and WER/CPC homologs are mostly involved in development of hairs (root hairs or trichomes) and biosynthesis of secondary metabolites associated with antioxidation (anthocyanin and flavonol) (From TAIR on www.arabidopsis.org) (Supplementary Table 5), implying that the WER–EGL3 interaction mode might be implemented in stress resistance, for example to cold, heat, drought or insects, to adapt the varied environments in ancestral seed plants. For the MYB29–MYC3 interaction mode, MYB28, MYB29 and MYB76 are functional in regulation of aliphatic glucosinolate biosynthesis[22-30], while MYB34, MYB51 and MYB122 modulate indolic glucosinolate biosynthesis[27, 30-32]. Glucosinolates are primarily found in Brassicales and can be modified into toxins against insect herbivory. It is possible that the MYB29–MYC3 interaction mode might be evolved to allow

higher plants to resist to insect attacks. Taken together, probably due to the adaptation to different environmental stress, the two MYB–bHLH interaction modes evolved independently during plant evolution.

Methods

Yeast two-hybrid (Y2H) assay

The coding regions of full-length or truncated *WER*, *CPC* and *EGL3* were PCR- or RT-PCR-amplified and cloned into pGADT7 or pGBKT7 vectors (Clontech) (Supplementary Table 6). The yeast two hybrid (Y2H) assay was performed following the manufacturer's protocol (Clontech) and protein-protein interaction was detected on media lacking leucine (Leu), tryptophan (Trp) and histidine (His).

Protein expression and purification

The DNA fragments encoding WER 12-120, EGL3 1-205, full-length CPC, CPC 30-94 and their mutants were generated by PCR and subcloned into pSUMO vector, respectively (Supplementary Table 6). All the recombined proteins were expressed in *Escherichia coli* strain BL21 (DE3) cells and induced with 0.2 mM isopropyl β-D-thiogalactoside (IPTG) when the OD_{600} reached ~0.8. The induced cultures were grown at 18°C for an additional 16-18 hours. For purification of WER and its mutants, cells were enriched and suspended in Ni-NTA (GE Healthcare) binding buffer (500 mM NaCl, 20 mM Tris-HCl and 25 mM imidazole, pH 8.0). The cells were lysed by high-pressure cell disruptor and then centrifuged at 34,000 g for 1 hour. The supernatant was loaded onto a Ni-NTA column. The target proteins were gradually eluted using elusion

buffer (500 mM NaCl, 20 mM Tris-HCl and 500 mM imidazole, pH 8.0). The proteins were treated by ULP1 protease for 3 hours. The cleaved proteins were diluted to 200 mM NaCl and 20 mM Tris-HCl (pH 8.0), loaded onto HiTrap S column (GE Healthcare) and eluted by elution buffer (1 M NaCl and 20 mM Tris-HCl, pH 8.0). The eluted proteins were concentrated and applied to a pre-treated HiLoad 16/60 Superdex 75 gel filtration column in buffer containing 300 mM NaCl and 20 mM Tris-HCl, pH 8.0.

For co-purification of WER 67-120/EGL3 1-205, CPC 30-94/EGL3 1-205 and MYB29 174-222/MYC3 44-238, the DNA fragments encoding WER 67-120, CPC 30-96 and MYB29 174-222 were generated by PCR or RT-PCR and constructed into pCDF-duet vector, respectively (Supplementary Table 6). The DNA fragment encoding MYC3 44-238 was generated by RT-PCR and subcloned into pSUMO vector (Supplementary Table 6). The plasmids expressing WER 67-120 and EGL3 1-205, CPC 30-94 and EGL3 1-205, or MYB29 174-222 and MYC3 44-238 were co-transferred into *Escherichia coli* strain BL21 (DE3) cells. Protein expression was induced by 0.2 mM IPTC when the OD_{600} reached 0.8. The cells were collected and the target complexes were purified using the same procedures as that of EGL3 1-205.

Size-exclusion chromatography (SEC) analysis

EGL3 1-205, WER 12-120, and CPC 30-94 proteins were respectively dissolved in SEC buffer (20 mM Tris-HCl, 150 mM NaCl, 5 mM DTT, pH 8.0). To prepare WER–EGL3 and CPC–EGL3 complexes, EGL3 1-205 was mixed with equivalent molar of WER 12-120 and CPC 30-94, respectively and the mixtures were incubated on ice for 30 min. For the competition assay, equivalent molar of CPC 30-94 was added into the

pre-prepared WER–EGL3 complex and incubated on ice for 30 min. The concentration of EGL3 1-205, WER 12-120, and CPC 30-94 were fixed at 50 μM in all samples, which were sequentially analyzed by Superdex^{TM} increase 200 10/300 GL column.

Crystallization and structure determination

The apo EGL3 1-205 crystals were grown at 18°C using the sitting-drop vapor diffusion method. The drop contains 0.2 μl 25 mg/ml EGL3 1-205 protein and 0.2 μl reservoir solution composed of 100 mM CAPS, 200 mM lithium sulfate, and 2 M ammonium sulfate, pH 10.5. All crystals of protein complexes were grown at 18°C by using the hanging-drop vapor diffusion method. For crystallization of WER 67-120 and EGL3 1-205 complex, the co-purified protein complex was concentrated to 20 mg/ml and the crystals were grown in well solution containing 20% PEG3350, 200 mM magnesium formate. The crystals of CPC 30-94 and EGL3 1-205 complex (20 mg/ml) were grown in well solution (0.1 M Tris-HCl, 20% PEG3350, 0.2 M ammonium sulfate, pH 6.5). For crystallization of MYB29 174-222 and MYC3 44-238 complex, the co-purified complex was concentrated to 20 mg/ml and the crystals were grown in well solution (0.1 M HEPES, 20% (w/v) PEG 6000, 0.2 M Sodium chloride, pH 7.0).

The X-ray data were collected at BL17U and BL18U beamlines at Shanghai Synchrotron Radiation Facility (SSRF). Diffractive data were indexed, integrated and scaled with HKL3000 program[33]. The apo EGL3 structure and WER–EGL3 complex were all solved by molecular replacement (MR) method using the Phaser program of CCP4i[34], using the MYC3 structure (PDB_ID：4RRU) as the search model. The CPC–EGL3 complex structure was solved by MR method using the apo EGL3 structure as

414 the search model. The MYB29–MYC3 complex structure was solved by MR method
415 using the MYC3 structure (PDB_ID：4RRU) as the search model. The model building
416 and refinement were performed with COOT[35] and PHENIX[36]. All the structural images
417 were generated by using the PyMOL program (http://www.pymol.org).

418 **Isothermal titration calorimetry (ITC) assays**

419 All ITC experiments were performed using an iTC200 MicroCalorimeter from
420 MicroCal. Interaction was performed in a buffer (20 mM Tris-HCl, 100 mM NaCl,
421 pH8.0) at 25°C. 38 μl of EGL3 1-205 (200 μM) was titrated into the cell containing
422 200 μl wild-type or mutated WER 12-120 (20 μM). A total of 25 injections (each of 1.5
423 μl) were performed. Binding curves were generated by plotting the heat change of the
424 binding reaction, and the data were fitted using one-site binding model with Origin 7.0
425 (Supplementary Table 1).

426 **Spilt luciferase assays**

427 The DNA fragment encoding C-terminus of luciferase was fused with the DNA
428 fragment encoding EGL3 1-205 to create p35S::3×Flag-EGL3 1-205-nLUC construct
429 (Supplementary Table 6). The DNA fragments encoding WER 12-120 and CPC 30-94
430 were respectively fused with the DNA fragment encoding N-terminus of luciferase to
431 create p35S::cLUC-WER 12-120-4×Myc and p35S::cLUC-CPC 30-94-4×Myc
432 (Supplementary Table 6). The DNA fragments encoding CPC, CPC M49A, CPC M49L
433 and WER were combined with DNA fragment encoding YFP and cloned into
434 pCAMBIA 1300 to create p35S::CPC 30-94-YFP, p35S::CPC 30-94 M49A-YFP,
435 p35S::CPC 30-94 M49L-YFP and p35S::WER 12-120-YFP fusion constructs,

respectively (Supplementary Table 6). Plasmids were transformed into Agrobacterium strain *Agrobacterium tumefaciens* GV3101 with different transgenic genes and were cultivated in LB medium at 28°C overnight. Different combinations were then co-infiltrated into young leaves of *N. benthamiana*. After treated in darkness for 1 day, the plants were exposed to light for 2 days. Luciferin were injected into the Agrobacterium-infiltrated positions and luciferase activity was measured. To verify the expression level of proteins, equal-size tobacco leaves were cut into small size, cells were lysed and total proteins were extracted and analyzed by Western blotting using anti-MYC (M20002L, Abmart, 1:5000 dilution), anti-GFP (M20004L, Abmart, 1:5000 dilution), anti-Actin (M20009L, Abmart, 1:5000 dilution), and anti-FLAG (F1804, Sigma, 1:2000 dilution) antibodies, respectively.

Microscale thermophoresis (MST) analysis

MST experiments were performed according to the published methods [37]. All proteins were dialyzed into MST reaction buffer (20 mM HEPES and 150 mM NaCl, pH 7.5), labelled and purified using the Protein labelling kit RED-NHS (Nanotemper, cat. no. L001) and the recommended procedures. WER protein was labeled with cy5. 60 nM EGL3 1-205 and 30 nM cy5-labelled WER were mixed in reaction buffer (20 mM HEPES, 150 mM NaCl and 0.03% tween-20, pH 7.5) and incubated on ice for 30 min. Then diluted CPC or mutants (from 0.4 mM to 1.22×10^{-5} mM) were added into the reaction. For the MYB29-MYC3 interaction, Fluorescein amidites (FAM) labeled MYB29 (184-205) was purchased from Scilight-Peptide company (http://www.scilight-peptide.com/) and were dialyzed into MST reaction buffer before reaction. 50 nM

458 FAM-MYB29 were incubated with MYC3 (from 25 μM to 0.763 nM) in reaction buffer
459 and incubated on ice for 30 min. The MST experiments were conducted with Monolith
460 NT.115 instrument (NanoTemper Technologies) and the data were collected under 100%
461 infrared laser power and 20% light-emitting diode power at 25°C. The data were
462 analyzed by MO.Affinity Analysis v2.3 software and the EC50 determined.

463 **Bio-Layer Interferometry (BLI) analysis**

464 BLI (OctetRed96) assay was performed for measuring the binding affinities and kinetic
465 parameters of WER-EGL3 and CPC-EGL3 complexes, by using 6×His-SUMO tagged
466 WER 12-120, 6×His-SUMO tagged CPC 30-94, EGL3 1-205 and 6×His-SUMO as a
467 negative control with Octet RED96 instrument (ForteBio). Experiments were
468 conducted at 30 °C with a shaking speed of 1000 rpm. Nitrilotriacetic acid (NTA)
469 capture tips were soaked into the kinetics buffer (20 mM Tris, 100 mM NaCl,
470 0.02%Tween 20, pH 8.0) for 10 min to baseline equilibration. 6×His-SUMO tagged
471 WER 12-120 or CPC 30-94 was immobilized on the biosensor tips for 300 s, then
472 washed by kinetics buffer till all the parameters were stable. Next, the association of
473 EGL3 1-205 was determined at 500, 125, 31.3, 15.6 nM for 300 s, followed by 300 s
474 of dissociation in kinetics buffer. Affinity constants were calculated by a 1:1 global fit
475 model via ForteBio 10.0 data analysis software.

476 **Structure prediction by Alphafold**

477 Alphafold 2.0.1 was used to predict the structures of GL1-EGL3 and MYB28-MYC3
478 complexes according to the published methods[38-39].

479 **Homologs searching in *Arabidopsis thaliana***

Genes encoding homologous proteins of EGL3 (AT1G63650), WER (AT5G14750), CPC (AT2G46410), MYC3 (AT5G46760) and MYB29 (AT5G07690) were searched by all-against-all BLASTP with the E-value of 1e-^{5}, and the identity >20% in *Arabidopsis thaliana*. Sequences were aligned by MAFFT[40] with accurate aligning options "- maxiterate 1000 - localpair", adjusted manually with the use of AliView[41]. Those candidate homologs were carefully examined for their interaction motifs. The homologs with opposite chemical characteristics in the key residues with our query proteins were filtered out. Due to the great variability, the Ala197 and Ile203 in MYB29s homologs were not selected during this step.

Phylogenetic reconstruction and logo comparison

We selected a total of 23 publicly-available genomes (Supplementary Table 2) representing major plant lineages including seven eudicots, two monocots, two of ANA clade, two gymnosperms, three moniliophytes/lycophytes, three mosses/liverworts/hormworts, and four algae for phylogenetic reconstruction and logo comparison of key motifs. Genomes were retrieved from the Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html), NCBI (https://www.ncbi.nlm.nih.gov), FernBase (https://www.fernbase.org), GIGADB datasets (http://gigadb.org/), Spruce Genome Project (http://congenie.org/), MarpolBase (https://marchantia.info) and from files in previous studies[42-44]. Genes encoding homologous proteins of EGL3, WER, CPC, MYC3 and MYB29 were searched and aligned as above, and then trimmed by trimAl[45] with -gt 0.1. We used RAxML to reconstruct the phylogenetic trees[46] with default settings (-m PROTGAMMAJTT). We examined the phylogenetic trees,

extracted the clades containing the query sequences (*EGL3*, *WER*, *CPC*, *MYC3* and *MYB29*) with bootstrap value > 90, and repeated the steps of alignment and phylogenetic reconstruction. By repeating these steps, the sizes of the trees were gradually reduced until each tree contains only the clade including the query sequences and its sister clade. In order to reveal the evolutionary history of *CPC* and *MYB29*, 43 and 21 species were selected for further analysis, respectively (Supplementary Table 3, 4). The logo comparison analyses were performed on WebLogo3 (http://weblogo.threeplusone.com/create.cgi). The same method was also applied for homologs in Arabidopsis for Fig. 5d.

Data availability

Structural factors and coordinates have been deposited in the Protein Data Bank (PDB) under accession codes 7FDL, 7FDM, 7FDN and 7FDO for WER–EGL3, MYB29–MYC3, EGL3, and CPC–EGL3. And the structure of WER-DNA complex is available in the PDB by accession code 6KKS.

Acknowledgements

We thank Drs. Wei Yang, Hong Ma and Xiaoya Chen for a critical reading of the manuscript. We thank Dr. Yu Ding for help in BLI and ITC experiments. We thank the staff of beamlines BL17U1, BL18U1, and BL19U1 at the Shanghai Synchrotron Radiation Facility for assistance with data collection. We thank the staff members of

the Microscale thermophoresis System at the National Facility for Protein Science in Shanghai (NFPS), Zhangjiang Lab, Shanghai Advanced Research Institute, Chinese Academy of Science, China for providing technical support and assistance in data collection and analysis. We thank Robert McKenzie, PhD, from Liwen Bianji (Edanz) (www.liwenbianji.cn/ac), for editing the English text of a draft of this manuscript. This work was supported by the National Natural Science Foundation of China (NSFC31930017) for A.D. and the National Basic Research Program of China (2012CB910500) for W-H.S..

Author Contributions

A.D. conceived and designed the research. B.W., Q.L., Y.L., K.D., Z.W., and T.L. performed the experiments. Q.L., B.W., J.G., and C.H. analyzed the data. Q.L., B.W., W-H.S., C.H., J.G., and A.D. wrote the manuscript. All authors read, revised, and approved the manuscript.

Competing Interests

The authors declare no competing interests.

Additional Information

542 Supplementary Information is available for this paper. Correspondence and requests

543 for materials should be addressed to Aiwu Dong.

544

Table 1. Data collection and refinement statistics.

	EGL3	WER-EGL3	CPC-EGL3	MYB29-MYC3
Data collection				
Space group	P32	C2221	C121	P32
Cell parameter				
a (Å)	70.4	75.9	86.0	85.2
b (Å)	70.4	193.2	50.3	85.2
c (Å)	88.5	224.6	81.6	57.0
α (°)	90.0	90.0	90.0	90.0
β (°)	90.0	90.0	97.2	90.0
γ (°)	120.0	90.0	90.0	120.0
Wavelength(Å)	0.97930	0.97930	0.97915	0.97930
Resolution (Å)	30.0-1.90	30.0-2.90	30.0-1.80	30.0-2.50
Last shell (Å)	1.97-1.90	3.0-2.90	1.89-1.80	2.59-2.50
Completeness (%)	99.8(99.2)	99.7(98.4)	97.5(98.8)	97.5(91.3)
Redundancy	9.7(7.7)	8.2(6.1)	6.1(5.7)	6.2(3.2)
I/σ(I)	37.9(2.0)	12.4(2.0)	60.8(7.5)	32.1(2.0)
Rmerge (%)	5.9(95.2)	14.8(70.8)	2.4(18.2)	4.1(42.1)
CC(1/2)	0.998(0.876)	0.994(0.114)	0.997(0.983)	0.999(0.114)
Refinement				
Resolution (Å)	30.0-1.90	30.0-2.90	30.0-1.80	30.0-2.50
R_{work} (%) / R_{free} (%)	18.5/22.7	25.4/29.9	18.8/21.5	23.0/28.4
No. of atoms				
Protein	2671	8858	1715	2840
Ligand/ion	25	0	5	0
Water	126	0	139	14
R.m.s. deviations				
Bond length (Å)	0.012	0.009	0.013	0.006
Bond angle (°)	1.689	1.061	1.557	0.935
Ramachandran plot (%)				
Most favored	97.54	95.15	95.0	97.71
Additional allowed	2.46	4.85	5.0	2.29
PDB number	7FDN	7FDL	7FDO	7FDM

*Values in parentheses are for the highest-resolution shell.

Figure Legends

Figure 1. **Co-crystal structure of the WER–EGL3 complex. a,** Overall structure of the WER–EGL3 complex. The JID and TAD domains of EGL3 are shown in cyan and green, respectively, and WER is shown in pink. The invisible regions (aa 55-66 and 87-96) of EGL3 in electron density map are indicated by dashed grey lines. **b,** Structural superposition of the WER–EGL3 complex with the WER–DNA complex (PDB:6KKS). The WER–EGL3 complex is colored as in **a**. DNA and WER in the WER-DNA complex are colored in brown and yellow, respectively. WER R2 and R3 domains are indicated by arrows. **c,** Interactions between WER and EGL3. H-bonds are indicated by dashed black lines. For clarity, the detailed hydrophobic interactions are not shown. **d,** ITC results showing the binding affinities between EGL3 and wild-type or mutated WER proteins. **e,** ITC results showing the binding affinities between WER and wild-type or mutated EGL3 proteins.

Figure 2. **Co-crystal structure of the CPC–EGL3 complex. a,** Overall structure of the CPC–EGL3 complex, in which the JID and TAD domains of EGL3 are shown in cyan and green, respectively, and CPC is shown in purple. Invisible regions (aa 55-63 and 87-104) of EGL3 are indicated by dashed grey lines. **b,** Structure superimposition of the CPC–EGL3 complex with the WER–EGL3 complex. The CPC-EGL3 complex is colored in grey and purple, whereas the WER-EGL3 complex is colored in yellow and magenta. **c,** Interactions between CPC and EGL3. H-bonds are indicated by dashed black lines. For clarity, the detailed hydrophobic interactions are not shown. **d,** Structural superimposition of WER R3 and CPC R3, which are colored in magenta and purple, respectively. Residues involved in EGL3 interaction are shown, and CPC M49 and WER L85 are highlighted by a black dashed cycle. **e,** Close-up view of the crucial amino acids differing between WER and CPC, and the Van der Waals surfaces are shown by dots at the bottom.

Figure 3. **Functional importance of Met49 of CPC in competition with WER to bind to EGL3. a,** CPC peels WER from the WER–EGL3 complex in a SEC analysis,

which is further detected by SDS-PAGE. For CPC competition assay, an equivalent concentration of CPC was incubated with the prepared WER-EGL3 complex on ice for 30 min and analyzed by SEC experiment. The SEC fraction of WER-EGL3 plus CPC was analyzed by SDS-PAGE, which was repeated independently twice with similar results. **b,** Split-luciferase assay to detect the competition of wild-type and mutated CPC against WER to interact with EGL3. Tobacco leaves were co-infiltrated with Agrobacterium containing EGL3-nLUC and cLUC-WER with or without CPC-YFP, and the luminescence images were captured by a CCD imaging system. **c,** Quantitative measurement of the competitive capabilities of CPC and its mutants against WER to form a complex with EGL3 by a microscale thermophoresis (MST) assay, and the data were presented as mean values ± SD of three independent experiments (n=3).

Figure 4. **Co-crystal structure of the MYB29–MYC3 complex. a,** Overall structure of the MYB29–MYC3 complex. The JID and TAD domains of MYC3 are shown in cyan and green, respectively, and MYB29 is shown in brown. **b,** Interactions between MYB29 and MYC3. H-bonds are indicated by dashed black lines. For clarity, the detailed hydrophobic interactions are not shown. **c,** Structural superposition of the MYB29–MYC3 complex with the JAZ1–MYC3 complex (PDB:4YZ6). MYB29 and MYC3 in the MYB29–MYC3 complex are colored in brown and grey, respectively. JAZ1 and MYC3 in the JAZ1–MYC3 complex are colored in blue and cyan, respectively.

Figure 5. **Two interaction modes of MYB–bHLH complexes. a,** Structural superposition of the WER–EGL3 and MYB29–MYC3 complexes. WER and MYB29 are colored in magenta and brown, respectively. Both EGL3 and MYC3 are colored in grey, whereas their JID helixes are colored in cyan and pink, respectively. **b,** The different interaction modes of the WER–EGL3 and MYB29–MYC3 complexes. **c,** Sequence alignment of homologs of EGL3, WER, CPC, MYC3, and MYB29 in Arabidopsis. For each panel, the uppermost sequence is the logo comparison, with the *x*-axis scaled to the position of amino acids of EGL3, WER, CPC, MYC3, and MYB29,

608 respectively. The hydrophilic, neutral, and hydrophobic nature for the amino acids is
609 indicated in blue, green, and black, respectively. Stars indicate the residues crucial for
610 MYB–bHLH interaction. **d,** Evolution of *EGL3*, *WER*, *CPC*, *MYC3*, and *MYB29*
611 among land plants. Genes able to utilize the EGL3–WER interaction mode are shown
612 in green, and those capable of the MYC3–MYB29 interaction mode are shown in
613 orange.
614

References:

1. Feller, A., Machemer, K., Braun, E. L. & Grotewold, E., Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J.* **66**, 94-116 (2011).

2. Pireyre, M. & Burow, M., Regulation of MYB and bHLH transcription factors: a glance at the protein level. *Mol. Plant* **8**, 378-388 (2015).

3. Dubos, C. *et al.*, MYB transcription factors in Arabidopsis. *Trends Plant Sci.* **15**, 573-581 (2010).

4. Heim, M. A. *et al.*, The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. *Mol. Biol. Evol.* **20**, 735-747 (2003).

5. Toledo-Ortiz, G., Huq, E. & Quail, P. H., The Arabidopsis basic/helix-loop-helix transcription factor family. *Plant Cell* **15**, 1749-1770 (2003).

6. 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.* **47**, 9592-9608 (2019).

7. Schiefelbein, J., Kwak, S. H., Wieckowski, Y., Barron, C. & Bruex, A., The gene regulatory network for root epidermal cell-type pattern formation in Arabidopsis. *J. Exp. Bot.* **60**, 1515-1521 (2009).

8. Song, S. K. *et al.*, Cell fate in the Arabidopsis root epidermis is determined by competition between WEREWOLF and CAPRICE. *Plant Physiol.* **157**, 1196-1208 (2011).

9. Kang, Y. H., Song, S. K., Schiefelbein, J. & Lee, M. M., Nuclear trapping controls the position-dependent localization of CAPRICE in the root epidermis of Arabidopsis. *Plant Physiol.* **163**, 193-204 (2013).

10. Tominaga-Wada, R. & Wada, T., Regulation of root hair cell differentiation by R3 MYB transcription factors in tomato and Arabidopsis. *Front. Plant Sci.* **5**, 91 (2014).

11. Qi, T., Huang, H., Song, S. & Xie, D., Regulation of jasmonate-mediated stamen development and seed production by a bHLH-MYB complex in Arabidopsis. *Plant Cell* **27**, 1620-1633 (2015).

12. Fernandez-Calvo, P. *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* **23**, 701-715 (2011).

13. Zhang, F. *et al.*, Structural basis of JAZ repression of MYC transcription factors in jasmonate signalling. *Nature* **525**, 269-273 (2015).

14. Schweizer, F. *et al.*, Arabidopsis basic helix-loop-helix transcription factors MYC2, MYC3, and MYC4 regulate glucosinolate biosynthesis, insect performance, and feeding behavior. *Plant Cell*

645 **25**, 3117-3132 (2013).
646 15. Wang, B. *et al.*, Structural insights into target DNA recognition by R2R3-MYB transcription factors.
647 *Nucleic Acids Res.* **48**, 460-471 (2020).
648 16. Triezenberg, S. J., Structure and function of transcriptional activation domains. *Curr. Opin. Genet.*
649 *Dev.* **5**, 190-196 (1995).
650 17. Melcher, K., The strength of acidic activation domains correlates with their affinity for both
651 transcriptional and non-transcriptional proteins. *J. Mol. Biol.* **301**, 1097-1112 (2000).
652 18. Sun, X., Rikkerink, E. H., Jones, W. T. & Uversky, V. N., Multifarious roles of intrinsic disorder in
653 proteins illustrate its broad impact on plant biology. *Plant Cell* **25**, 38-55 (2013).
654 19. Zimmermann, I. M., Heim, M. A., Weisshaar, B. & Uhrig, J. F., Comprehensive identification of
655 Arabidopsis thaliana MYB transcription factors interacting with R/B-like BHLH proteins. *Plant J.*
656 **40**, 22-34 (2004).
657 20. Krissinel, E. & Henrick, K., Inference of macromolecular assemblies from crystalline state. *J. Mol.*
658 *Biol.* **372**, 774-797 (2007).
659 21. Frerigmann, H., Berger, B. & Gigolashvili, T., bHLH05 is an interaction partner of MYB51 and a
660 novel regulator of glucosinolate biosynthesis in Arabidopsis. *Plant Physiol.* **166**, 349-369 (2014).
661 22. Gigolashvili, T., Yatusevich, R., Berger, B., Muller, C. & Flugge, U. I. The R2R3-MYB
662 transcription factor HAG1/MYB28 is a regulator of methionine-derived glucosinolate biosynthesis
663 in Arabidopsis thaliana. *Plant J.* **51**, 247-261 (2007).
664 23. Hirai, M. Y. et al. Omics-based identification of Arabidopsis Myb transcription factors regulating
665 aliphatic glucosinolate biosynthesis. *Proc Natl Acad Sci U S A.* **104**, 6478-6483 (2007).
666 24. Sonderby, I. E. et al. A systems biology approach identifies a R2R3 MYB gene subfamily with
667 distinct and overlapping functions in regulation of aliphatic glucosinolates. *PLoS One.* **2**, e1322
668 (2007).
669 25. Beekwilder, J. et al. The impact of the absence of aliphatic glucosinolates on insect herbivory in
670 Arabidopsis. *PLoS One*. **3**, e2068 (2008).
671 26. Gigolashvili, T., Engqvist, M., Yatusevich, R., Muller, C. & Flugge, U. I. HAG2/MYB76 and
672 HAG3/MYB29 exert a specific and coordinated control on the regulation of aliphatic glucosinolate
673 biosynthesis in Arabidopsis thaliana. *New Phytol*. **177**, 627-642 (2008).
674 27. Malitsky, S. et al. The transcript and metabolite networks affected by the two clades of Arabidopsis

glucosinolate biosynthesis regulators. *Plant Physiol.* **148**, 2021-2049 (2008).

28. Sonderby, I. E., Burow, M., Rowe, H. C., Kliebenstein, D. J. & Halkier, B. A. A complex interplay of three R2R3 MYB transcription factors determines the profile of aliphatic glucosinolates in Arabidopsis. *Plant Physiol.* **153**, 348-363 (2010).

29. Li, Y. et al. Novel insights into the function of Arabidopsis R2R3-MYB transcription factors regulating aliphatic glucosinolate biosynthesis. *Plant Cell Physiol.* **54**, 1335-1344 (2013).

30. Frerigmann, H. & Gigolashvili, T. MYB34, MYB51, and MYB122 distinctly regulate indolic glucosinolate biosynthesis in Arabidopsis thaliana. *Mol. Plant*. **7**, 814-828 (2014).

31. Celenza, J. L. et al. The Arabidopsis ATR1 Myb transcription factor controls indolic glucosinolate homeostasis. *Plant Physiol.* **137**, 253-262 (2005).

32. Gigolashvili, T. et al. The transcription factor HIG1/MYB51 regulates indolic glucosinolate biosynthesis in Arabidopsis thaliana. *Plant J.* **50**, 886-901 (2007).

33. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M., HKL-3000: the integration of data reduction and structure solution-from diffraction images to an initial model in minutes. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 859-866 (2006).

34. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E., A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1131-1137 (2003).

35. Emsley, P. & Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126-2132 (2004).

36. Adams, P. D. *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-221 (2010).

37. Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P. & Duhr, S., Molecular interaction studies using microscale thermophoresis. *Assay Drug Dev. Technol.* **9**, 342-353 (2011).

38. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. **596**, 583-589 (2021).

39. Evans, Richard, et al. Protein complex prediction with AlphaFold-Multimer. Preprint at https: //www.biorxiv.org/content/10.1101/2021.10.04.463034v1

40. Katoh, K. & Standley, D. M., MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772-780 (2013).

41. Larsson, A., AliView: a fast and lightweight alignment viewer and editor for large datasets.

Bioinformatics **30**, 3276-3278 (2014).

42. Hori, K. *et al.*, Klebsormidium flaccidum genome reveals primary factors for plant terrestrial adaptation. *Nat. Commun.* **5**, 3978 (2014).

43. Nishiyama, T. *et al.*, The chara genome: secondary complexity and implications for plant terrestrialization. *Cell* **174**, 448-464 (2018).

44. Wan, T. *et al.*, A genome for gnetophytes and early evolution of seed plants. *Nat. Plants* **4**, 82-89 (2018).

45. Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T., trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009).

46. Stamatakis, A., RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).

Title:

Structural insights into partner selection for MYB and bHLH transcription factor complexes

Authors:

Baihui Wang[1,†], Qiang Luo[1,†], Yingping Li[1,†], Kangxi Du[1], Zhen Wu[1], Tianyang Li[1], Wen-Hui Shen[2], Chien-Hsun Huang[3*], Jianhua Gan[4*], Aiwu Dong[1,*]

[†]B.W., Q.L. and Y.L. contributed equally to this work.

[*]Correspondence: aiwudong@fudan.edu.cn, ganjhh@fudan.edu.cn and chhuang1981@139.com

Affiliations:

[1]State Key Laboratory of Genetic Engineering, Collaborative Innovation Center for Genetics and Development, Department of Biochemistry and Biophysics, Institute of Plant Biology, School of Life Sciences, Fudan University, Shanghai 200438, P.R. China

[2]Institut de Biologie Moléculaire des Plantes, CNRS, Université de Strasbourg, 12 rue du Général Zimmer, 67084 Strasbourg Cédex, France

[3]State Key Laboratory of Genetic Engineering, Center for Evolutionary Biology, Institute of Plant Biology, School of Life Sciences, Fudan University, Shanghai 200438, P.R. China

[4]Shanghai Public Health Clinical Center, State Key Laboratory of Genetic Engineering,

Collaborative Innovation Center of Genetics and Development, Department of Biochemistry and Biophysics, School of Life Sciences, Fudan University, Shanghai 200438, P.R. China

Abstract:

MYB and bHLH transcription factors form complexes to regulate diverse metabolic and developmental processes in plants. However, the molecular mechanisms responsible for MYB–bHLH interaction and partner selection remain unclear. Here, we report the crystal structures of three MYB–bHLH complexes (WER–EGL3, CPC–EGL3, and MYB29–MYC3), uncovering two MYB–bHLH interaction modes. WER and CPC are R2R3- and R3-type MYBs, respectively, but interact with EGL3 through their N-terminal R3 domain in a similar mode. A single amino acid of CPC, Met49, is crucial for competition with WER to interact with EGL3. MYB29, a R2R3-type MYB TF, interacts with MYC3 by its C-terminal MYC-interaction motif. The WER–EGL3 and MYB29–MYC3 binding modes are widely applied among MYB–bHLH complexes in Arabidopsis and evolve independently in plants.

Introduction

By recognizing specific DNA elements within the genome, transcription factors (TFs) play central roles in gene regulatory networks in multicellular organisms. In higher plants, MYB and basic helix-loop-helix (bHLH) TFs are among the largest groups of TFs; for example, more than 300 MYB and 100 bHLH TFs are present in Arabidopsis[1,2]. MYB TFs contain a conserved DNA-binding domain (MYB domain) and extensive intrinsically disordered regions (IDRs) at the C-terminus. MYB TFs are grouped into different types on the basis of three imperfect repeats (R1, R2, and R3) within the MYB domain of c-Myb, and further divided into subgroups based on the IDRs[3]. bHLH TFs are defined by ~60 conserved amino acids, including a basic DNA-binding domain and a dimerization domain[4].

Formation of complexes between MYB and bHLH TFs is widespread and different MYB–bHLH complexes regulate diverse physiological processes, including organ development, metabolic pathways, and biotic and abiotic stress responses[1-3,5,6]. For instance, WEREWOLF (WER), a well-studied R2R3-type MYB in Arabidopsis, interacts with the bHLH TFs GL3/EGL3, directly binds to and activates *GLABRA 2* (*GL2*), the central regulator of epidermal cell fate determination, and leads to the non-hair cell fate during root hair development[7]. CAPRICE (CPC), a R3-type MYB TF in Arabidopsis, competes with WER to interact with GL3/EGL3 and inhibits *GL2* expression, leading to the hair cell fate establishment[8-10]. The Arabidopsis bHLH TFs MYC2/MYC3/MYC4 interact with MYB21/MYB24 to regulate stamen development and seed production[11]. MYC2/MYC3/MYC4 play crucial roles in jasmonate (JA)

signaling through interaction with jasmonate ZIM (zinc-finger inflorescence meristem)-domain (JAZ) repressors and participate in both development and stress responses[12,13]. MYC2/MYC3/MYC4 also regulate glucosinolate biosynthesis by interacting with glucosinolate-related MYBs, including MYB28, MYB29, MYB76, MYB34, MYB51, and MYB122, which belong to MYB subgroup 12[14]. In addition, many other MYB–bHLH complexes function in plant growth and development, stress defense, and metabolism regulation[6].

The co-crystal structure of MYC3 in complex with JAZ transcriptional repressors was resolved previously[13], and recently we resolved the complex structure of WER–DNA[15]. The structure of MYB–bHLH complexes remains uncertain, which precludes understanding the molecular mechanisms that regulate the interaction and partner selection of MYB and bHLH TFs. Here we report the crystal structures of three MYB–bHLH complexes, WER–EGL3, CPC–EGL3, and MYB29–MYC3, which reveals two distinct MYB–bHLH interacting modes. The WER–EGL3 and MYB29–MYC3 binding modes are widely applied among a set of MYB‒bHLH complexes in Arabidopsis. Our results shed light on the mechanisms by which different MYB and bHLH TFs select their partners to form complexes and indicate that the two types of MYB–bHLH interaction modes are evolutionarily conserved but evolve independently in plants.

Results

Co-crystal structure of WER–EGL3 complex

Following our previous study on the crystal structure of WER in complex with its target DNA[15], we selected WER and EGL3 as a model to analyze the interaction of MYB and bHLH TFs. WER contains a R2R3-MYB domain at the N-terminus. EGL3 contains a JAZ-interacting domain (JID) and transcriptional activation domain (TAD) at the N-terminus, a bHLH domain in the central region, and a C-terminal domain (CTD) at the C-terminus (Extended Data Figs. 1a, b). Yeast two-hybrid (Y2H) and size-exclusion chromatography (SEC) experiments showed that WER interacted with EGL3 through the R3 domain of WER (WER 67–120) and the N-terminus of EGL3 (EGL3 1–205) (Extended Data Figs. 1c, d). To investigate the molecular basis underlying the interaction between WER and EGL3, we determined the crystal structure of WER 67–120 in complex with EGL3 1–205 (hereafter the WER–EGL3 complex). The complex structure was refined to 2.90 Å resolution (Table 1). The β-strands β1-β5 of EGL3 1–205 formed a flat β-sheet in the center, flanked by helices α1 and α6 on one side and by α3, α4, and α5 on the opposite side (Fig. 1a). The EGL3 JID domain consisted of helices α1 and α2, and β-strands β1, β2, and α2 located at the C-terminus of JID. Although TAD domains are generally unstructured when not bound to their targets[16-18], the EGL3 TAD was well ordered. The EGL3 TAD was composed of two helices, α3 and α4, which packed against the JID domain and β3-β5, respectively. The R3 domain of WER was composed of three helices α1-α3 linked by two short loops; the α1 and α2 helices of WER packed against the α2, α3, and α5 helices of EGL3 (Fig. 1a). Structural superposition of WER–EGL3 and WER–DNA[15] suggested that WER–EGL3 complex formation has no effect on WER–DNA interaction (Fig. 1b), which is

consistent with their inherent DNA binding and gene activation activities.

The detailed interaction between WER and EGL3 was shown in Fig. 1c. The electrostatic surface potentials of the interfaces (Extended Data Figs. 2a,b) and the electron density maps of the key residues involved in the interaction (Extended Data Fig. 3a) of WER-EGL3 were also shown. WER Gln77 at the N-terminus of α1 formed hydrogen bonds (H-bonds) with the side chain of Ser153 and the main chains of Phe156 and Leu160 of EGL3. WER Leu81 nestled in the shallow hydrophobic cavity formed by Val120, Leu160, and Leu161 of EGL3. Via the guanidine group, WER Arg84 formed two H-bonds with EGL3 Ser123. Interestingly, the guanidine group of WER Arg84 also formed an ion-π interaction with the side chain of EGL3 Phe124. Similar to Leu81, WER Leu85 and Leu88 formed hydrophobic interactions with EGL3. The side chain of Leu85 pointed toward Trp116 and Val120 of EGL3, whereas the side chain of Leu88 nestled in the pocket formed by Tyr81, Leu84, and Leu119 of EGL3. WER Arg99 at the C-terminus of α2 (residues 92–99) formed a salt bridge with Asp113 and an ion-π interaction with Trp116 of EGL3 (Fig. 1c). The crucial residues of WER involved in the interaction with EGL3 exactly overlapped with the previously identified R/B-like bHLH TF binding (RB) motif, which is responsible for interactions between MYB and R/B-like bHLH TFs [19]. *In vitro* isothermal titration calorimetry (ITC) analysis showed that mutations of all WER–EGL3 interacting residues weakened the binding affinity between WER and EGL3 (Figs. 1d, e, Supplementary Table 1). Together, these results indicated that EGL3 physically interacts with the $Q_{77}X_3L_{81}X_2R_{84}L_{85}X_2L_{88}X_{10}R_{99}$ signature motif of the WER R3 domain.

Structural basis of CPC competing with WER to bind to EGL3

Y2H assay showed that, similar to WER, CPC interacted with EGL3 1–205 through the R3 domain of CPC (Extended Data Fig. 4a). ITC analysis indicated that full-length CPC (K_d: 48.3 nM) and CPC 30–94 (CPC R3, K_d: 50 nM) displayed similar binding affinities to EGL3, thus CPC 30–94 was used in subsequent analyses (Extended Data Fig. 4b, Supplementary Table 1). We determined the crystal structure of CPC 30–94 in complex with EGL3 1–205 (hereafter the CPC–EGL3 complex) at 1.80 Å resolution (Fig. 2a, Extended Data Figs. 2c,d and 3b, Table 1). Structural superposition revealed that the overall structure of CPC–EGL3 was similar to that of WER–EGL3 (Fig. 2b), and the root mean square deviation (RMSD) between CPC–EGL3 and WER–EGL3 was only 0.70 Å, based on the superposition of 177 pairs of Cα atoms. The low RMSD value indicates that CPC and WER share the same binding mode to EGL3. The detailed CPC–EGL3 interactions are depicted in Fig. 2c. Except for Glu41 and Met49 within CPC, all residues involved in the CPC–EGL3 interaction were identical to those observed in the WER–EGL3 complex (Fig. 1c). Glu41 and Met49 of CPC corresponded to Gln77 and Leu85 of WER, respectively. Given that CPC Glu41 mainly formed weak H-bond interaction (3.0 Å) with the side chain of EGL3 Ser153, and it is variable (Asp, Glu or Gln) in other R2R3-MYB and R3-MYB TFs (Extended Data Fig. 5), thus we mainly focused on CPC Met49 and WER Leu85 for further analysis (Figs. 2d, e). The Sδ atom of CPC Met49 substitutes the Cδ2 atom of WER Leu85, forming stronger (3.4 Å vs 3.7 Å) hydrophobic interactions with the side-chain Cγ2 atom of EGL3 Val120. The Cε

150 atom of CPC Met49 stretches more closely (3.7 Å vs 4.5 Å for the Cδ2 atom of WER)
151 to the indole ring of EGL3 Trp116. In addition, the side chain of CPC Met49 attracts
152 the side chain of CPC Arg63. Compared with WER Arg99 in the WER–EGL3 complex,
153 the guanidine group of CPC Arg63 is rotated ~90° in the CPC–EGL3 complex, leading
154 to strong Van der Waals contacts with the indole ring of EGL3 Trp116 and H-bond
155 interaction with EGL3 Asp113. As calculated by PDBePISA[20], CPC Met49 occupies
156 an interface of ~67 Å^2, which was ~17 Å^2 broader than that of WER Leu85 (~50 Å^2).
157 Compared with WER Leu85, CPC Met49 stretches more closely to the indole ring of
158 EGL3 Trp116, probably forming a more stable hydrophobic interaction with EGL3 (Fig.
159 2e).

160 To validate the importance of CPC Met49, we first analyzed competition between
161 WER and CPC. *In vitro* SEC experiments (Fig. 3a) showed that adding CPC to the
162 WER–EGL3 complex resulted in a shifted elution peak from 15.75 mL (corresponding
163 to the WER–EGL3 complex) to 15.95 mL (corresponding to the CPC–EGL3 complex)
164 and an additional elution peak at 18 mL (corresponding to free WER). A SDS-PAGE
165 analysis confirmed the SEC results (Fig. 3a, Extended Data Fig. 6). But a weak band of
166 WER was still detected in the CPC–EGL3 peak (Extended Data Fig. 6a). Split-
167 luciferase assays showed that CPC inhibited the interaction between WER and EGL3
168 *in vivo* (Fig. 3b, upper panels). Together, the SEC and split-luciferase experiments
169 indicated that CPC competes with WER to interact with EGL3 *in vitro* and *in vivo*. To
170 test whether Met49 contributes to the competition between CPC and WER, we
171 substituted Met49 with Leu in CPC R3 (CPC M49L) because CPC M49L mimicked

WER R3. ITC analysis showed that the EGL3-binding affinity of CPC M49L (K_d: 137 nM) (Extended Data Fig. 7a, Supplementary Table 1) was distinctly weaker than that of wild-type CPC R3 (K_d: 50 nM) (Extended Data Fig. 4b), but comparable to that of wild-type WER R3 (K_d: 118.6 nM) (Fig. 1d). The mutant CPC M49A, a substitution with Ala, showed a much lower EGL3-binding affinity (K_d: 1000 nM; Extended Data Fig. 7a, Supplementary Table 1) compared with that of wild-type CPC, suggesting that Met49 is important for CPC interaction with EGL3. In addition to CPC mutants, we also constructed two WER mutants, WER L85M and WER Q77E/L85M, which mimicked CPC R3. ITC analysis showed that the EGL3-binding affinities of WER L85M (K_d: 72.5 nM) and WER Q77E/L85M (K_d: 73.0 nM) (Extended Data Fig. 7b, Supplementary Table 1) were similarly increased compared with that of wild-type WER (K_d: 118.6 nM) (Fig. 1d), further supporting that Met49 not Glu41 of CPC is crucial for competition with WER to interact with EGL3. Split-luciferase experiments *in planta* confirmed the importance of CPC Met49: wild-type CPC abolished the interaction between WER and EGL3 (Fig. 3b, upper panels) at the molecular ratio of 1:1:1, whereas CPC M49L and CPC M49A failed to block WER–EGL3 complex formation (Fig. 3b, lower panels), although similar amounts of wild-type and mutated CPC as well as WER and EGL3 proteins were expressed in the split-luciferase assays (Extended Data Fig. 8). In addition, we performed microscale thermophoresis (MST) analysis to quantify the competitive capability of CPC against WER. CPC peeled WER from the WER–EGL3 complex with an EC50 value of ~6.03×10^{-6} M, whereas the competitive capability of CPC M49L was reduced to ~5-fold to an EC50 value of ~3.01×10^{-5} M and the EC50

value of CPC M49A was undetectable (Fig. 3c). Consistently, split-luciferase experiments showed that WER failed to affect CPC-EGL3 complex formation at the molecular ratio of 1:1:1 (Extended Data Fig. 9a), and MST analysis showed that the competitive capability of WER against CPC is reduced to ~52-fold to an EC50 value of ~3.19×10^{-4} M compared with that of CPC against WER (Extended Data Fig. 9b). Since ITC results cannot well explain the strong competitive capability of CPC against WER, we performed Bio-Layer Interferometry (BLI) experiments to further analyze the binding affinities of CPC-EGL3 and WER-EGL3. BLI results (Extended Data Fig. 10) showed that the K_d value of CPC-EGL3 complex (K_d: 11.9 nM) was ~18-fold lower than that of WER-EGL3 complex (K_d: 222 nM), supporting our SEC, split-luciferase and MST results (Fig. 3). Together, these results proved that the Met49 residue within CPC plays a vital role in competition with WER to interact with EGL3.

Co-crystal structure of MYB29–MYC3 complex

Recently, split-ubiquitin assays identified a new MYC-interaction motif (MIM) within the subgroup 12 MYBs. The motif is located in the center of the non-MYB region and is responsible for interaction with MYC2/MYC3/MYC4[6], suggesting that the MYB29–MYC3 complex may represent a different interaction mode from that of WER–EGL3 and CPC–EGL3. To verify this assumption, we co-expressed the MIM motif of MYB29 (MYB29 174-222) and the N-terminal non-DNA-binding region of MYC3 (MYC3 44–238). By extensive crystallization trials, we solved the crystal structure of MYB29 174–222 in complex with MYC3 44–238 (hereafter MYB29–MYC3) at 2.5 Å resolution

(Extended Data Figs. 2e,f and 3c, Table 1). The MYB29 184–203 region is folded into an ordered α helix and a short loop, whereas other regions are disordered (Fig. 4a). MYB29 fit into the deep groove and formed extensive interactions with the TAD and JID domains of MYC3. MYB29 Thr186 formed an H-bond with TAD Glu148 of MYC3. MYB29 Leu190 formed extensive hydrophobic interactions with TAD Phe151 and Leu152 of MYC3. Via the side-chain OD1 and ND2 atoms, MYB29 Asn191 formed two H-bonds with the main chain N- and O-atoms of MYC3 JID Tyr97, respectively. The side chain of MYB29 Ala194 fit into a shallow concavity, formed by JID Trp92 as well as TAD Phe151 and Met155 of MYC3 (Fig. 4b). In the MYB29–MYC3 complex structure, Leu190 and Asn191 of MYB29 formed the core platform for MYC3 binding, which is consistent with previous findings that either Leu190 or Asn191 mutation within MYB29 abolishes the interaction between MYB29 and MYC3, and the mutated MYB29 fails to rescue *myb29* mutant phenotypes[6]. MST analysis further showed that all mutations of the interacting residues within MYC3 weakened the interaction between MYB29 and MYC3 (Supplementary Fig. 1). To our surprise, sequence similarity between MYB29 and JAZ proteins was very low (Supplementary Fig. 2), but the proteins displayed the same binding mode to MYC3 (Fig. 4c). In the structure of the MYB29–MYC3 and MYC3–JAZ1[13] complexes, the MYC3 JID helices flipped outward and attached loosely to the main body of MYC3, and the helices of MYB29 or JAZ proteins fit into the grooves formed by TAD and JID of MYC3. Together, these results showed that MYB29 interacts with MYC3 by its C-terminal MYC-interaction motif, different from the binding modes of WER–EGL3 and CPC–EGL3.

Discussion

Two MYB–bHLH binding modes are widely applied in Arabidopsis

Sequence alignment revealed that the MYB29-interacting residues within MYC3 were conserved in EGL3 (Supplementary Fig. 3). However, MYB29 does not interact with GL3/EGL3[6]. To investigate the basis for the two MYB–MYC binding modes, we carefully compared the structures of EGL3 and MYC3. The overall structures of EGL3 and MYC3 were similar but their JID helices behaved differently (Supplementary Fig. 4). The conformation of the EGL3 JID helix was stable and formed extensive interactions with TAD and the other regions of JID (Fig. 5a, Supplementary Fig. 4a-c). Presence or absence of WER had no effect on the conformation of the EGL3 JID helix (Supplementary Fig. 4d). In contrast, the conformation of the MYC3 JID helix was highly dynamic[13] and underwent substantial conformational change to expose TAD and other regions of JID when binding to MYB29 or JAZ1 (Fig. 5a, Supplementary Fig. 4e, f). In EGL3, the corresponding interface for adopting MYB29 was buried by the stable JID helix, which prevented MYB29 binding to EGL3 (Fig. 5a, Supplementary Fig. 4g). Except for certain residues within the MYC3 TAD, most residues responsible for EGL3–WER interaction were not conserved in MYC3 (Supplementary Fig. 3), preventing the MYC3 surface (α4 and α7) from binding to WER. These results suggested that the JID helices of EGL3 and MYC3 played critical roles in partner selection. In detail, the conformational change of the MYC3 JID helix provided the interface for binding to the subgroup 12 MYB TFs or JAZ repressors. The rigidity of

the EGL3 JID helix prevented an interface forming, but the α3 and α5 helices of EGL3 formed a novel interface that specifically recognized the RB motif in the R3 domain of MYB TFs. Thus, we uncovered two MYB–bHLH interaction modes: in the WER–EGL3 mode, α3 and α5 of EGL3 formed an interface and specifically interacted with the R3 domain of WER; in the MYB29–MYC3 mode, the MIM motif of MYB29 fit into the groove formed by the TAD and JID domains of MYC3 (Fig. 5b).

To determine the number of MYB and bHLH TFs that may utilize the two MYB–bHLH interaction modes, we searched the Arabidopsis genome for MYB and bHLH proteins that shared the crucial interaction motifs (Fig. 5c). Fifteen WER- and seven CPC-type MYBs and four EGL3-type MYCs were detected (Fig. 5c). All crucial residues within the RB motifs were highly conserved among WER- and CPC-type MYBs, and the crucial residues, especially those within the TAD domains, were conserved among the EGL3-type MYCs, indicating that many MYB–bHLH complexes likely share the same interaction mode as WER–EGL3. Similarly, eight MYB29-type MYBs and seven MYC-type MYCs were detected in the Arabidopsis genome. The crucial residues within the JID and TAD domains were highly conserved among MYC2, MYC3, MYC4, MYC5, bHLH13, AIB (bHLH17) and bHJH3. For MYB29-type MYBs, the crucial residues within the core MIM domains were highly conserved (Fig. 5c), whereas those in the loop region (corresponding to Ala197 and Ile203 in MYB29) were conserved in MYB28, MYB29, and MYB76 but varied in the other MYBs. However, physical interaction between the eight MYBs and MYC2/3/4 has been reported previously[6,14,21], indicating that MYB29-type MYBs contain a functional MIM domain

to apply the same interaction mode as MYB29–MYC3. To prove the utilization of the two MYB–bHLH interaction modes in Arabidopsis, we chose GL1 and MYB28, homologs of WER and MYB29, respectively, and predicted the structures of GL1-EGL3 and MYB28-MYC3 complexes by AlphaFold2. Structure superposition showed that the interaction modes of GL1-EGL3 and MYB28-MYC3 are exactly similar to those of WER-EGL3 and MYB29-MYC3, respectively (Supplementary Fig. 5), supporting that the interaction modes revealed by this study are conserved for other homologous MYB-bHLH complexes. Taken together, our results indicated that the two types of MYB–bHLH binding modes are widely applied in MYB–bHLH complexes in Arabidopsis.

Two MYB–bHLH interaction modes evolved independently

To investigate the occurrence of the two types of MYB–bHLH interaction modes during the evolution of land plants, we searched the genomes of representative species across the plant kingdom (Supplementary Tables 2-4) for the genes encoding homologs of EGL3, WER, CPC, MYC3, and MYB29 and reconstructed their phylogenies (Supplementary Figs. 6-10). The major results were summarized in Fig. 5d, where homologs of EGL3 and MYC3 of the bHLH family showed divergence as early as from mosses during evolution, whereas homologs of WER, CPC and MYB29 of the MYB family showed divergence later from gymnosperms during evolution. For the WER–EGL3 interaction mode, we found that the crucial motifs in EGL3 homologs were conserved in land plants (Supplementary Figs. 6 and 11), whereas those in WER

304 homologs were conserved in gymnosperms (Supplementary Figs. 7 and 12), which
305 indicates that a functional WER–EGL3 interaction mode has co-evolved from the
306 formation of motifs responsible for interacting with EGL3 in WER homologs in seed
307 plants. For the MYB29–MYC3 interaction mode, the crucial motifs in MYC3 homologs
308 were conserved in land plants (Supplementary Figs. 9 and 13), whereas the MIM motif
309 of MYB29 was only detected in Arabidopsis among the 23 representative plant species
310 (Supplementary Figs. 10 and 14, Supplementary Table 2). To further trace the
311 origination of the MIM motif, we selected additional 20 species of eudicots
312 (Supplementary Fig. 15, Supplementary Table 4), and found that the MIM motif was
313 highly conserved in Brassicales species (Supplementary Fig. 16), indicating that the
314 MYB29–MYC3 interaction mode is functional since the occurrence of the Brassicales.
315 To further dissect the evolution of the two interaction modes, we examined the amino
316 acid residues of their motifs to assess whether the interaction may be substituted by
317 other proteins involved in a different mode. From sequence alignment, we found that
318 the residues corresponding to Tyr81, Ser123, Phe124, Leu160 and Leu161 in EGL3
319 have been replaced by amino acids with distinctive chemical characterizations in
320 MYC3 homologs (Supplementary Fig. 11), which would hinder the potential for MYC3
321 interaction with WER/CPC homologs. Regarding the JID domain, in the major regions
322 that distinguish the two MYB–bHLH interaction modes, the residues corresponding to
323 Ser293 (neutral) and Tyr300 (hydrophobic) of EGL3 were replaced by Lys/Arg
324 (hydrophilic) and Gln/His/Asn (hydrophilic or neutral), respectively, in MYC3
325 homologs (Supplementary Fig. 17). These results suggested that at an early stage of

land plant evolution (probably before the emergence of the mosses), the EGL3 and MYC3 clades have already evolved distinctive MYB-binding motifs. By examining the key residues in the alignment of all homologs of WER and MYB29, the crucial residues corresponding to Leu81 (hydrophobic) and Arg99 (hydrophilic) in the RB motif of WER were replaced by Thr (neutral) and His (neutral), respectively, in MYB29 homologs (Supplementary Fig. 12), and the crucial residues within the MIM motif of MYB29 were highly variable among WER homologs (Supplementary Fig. 14). These observations suggest that the two MYC-binding motifs had diverged prior to the separation of the WER and MYB29 clades in gymnosperms. Taken together, the two MYB–bHLH interacting modes are interpreted to have evolved independently.

Members applying the two MYB–bHLH interaction modes also evolved diverged biological functions. EGL3 and WER/CPC homologs are mostly involved in development of hairs (root hairs or trichomes) and biosynthesis of secondary metabolites associated with antioxidation (anthocyanin and flavonol) (From TAIR on www.arabidopsis.org) (Supplementary Table 5), implying that the WER–EGL3 interaction mode might be implemented in stress resistance, for example to cold, heat, drought or insects, to adapt the varied environments in ancestral seed plants. For the MYB29–MYC3 interaction mode, MYB28, MYB29 and MYB76 are functional in regulation of aliphatic glucosinolate biosynthesis[22-30], while MYB34, MYB51 and MYB122 modulate indolic glucosinolate biosynthesis[27, 30-32]. Glucosinolates are primarily found in Brassicales and can be modified into toxins against insect herbivory. It is possible that the MYB29–MYC3 interaction mode might be evolved to allow

higher plants to resist to insect attacks. Taken together, probably due to the adaptation to different environmental stress, the two MYB–bHLH interaction modes evolved independently during plant evolution.

Methods

Yeast two-hybrid (Y2H) assay

The coding regions of full-length or truncated *WER*, *CPC* and *EGL3* were PCR- or RT-PCR-amplified and cloned into pGADT7 or pGBKT7 vectors (Clontech) (Supplementary Table 6). The yeast two hybrid (Y2H) assay was performed following the manufacturer's protocol (Clontech) and protein-protein interaction was detected on media lacking leucine (Leu), tryptophan (Trp) and histidine (His).

Protein expression and purification

The DNA fragments encoding WER 12-120, EGL3 1-205, full-length CPC, CPC 30-94 and their mutants were generated by PCR and subcloned into pSUMO vector, respectively (Supplementary Table 6). All the recombined proteins were expressed in *Escherichia coli* strain BL21 (DE3) cells and induced with 0.2 mM isopropyl β-D-thiogalactoside (IPTG) when the OD_{600} reached ~0.8. The induced cultures were grown at 18°C for an additional 16-18 hours. For purification of WER and its mutants, cells were enriched and suspended in Ni-NTA (GE Healthcare) binding buffer (500 mM NaCl, 20 mM Tris-HCl and 25 mM imidazole, pH 8.0). The cells were lysed by high-pressure cell disruptor and then centrifuged at 34,000 g for 1 hour. The supernatant was loaded onto a Ni-NTA column. The target proteins were gradually eluted using elusion

buffer (500 mM NaCl, 20 mM Tris-HCl and 500 mM imidazole, pH 8.0). The proteins were treated by ULP1 protease for 3 hours. The cleaved proteins were diluted to 200 mM NaCl and 20 mM Tris-HCl (pH 8.0), loaded onto HiTrap S column (GE Healthcare) and eluted by elution buffer (1 M NaCl and 20 mM Tris-HCl, pH 8.0). The eluted proteins were concentrated and applied to a pre-treated HiLoad 16/60 Superdex 75 gel filtration column in buffer containing 300 mM NaCl and 20 mM Tris-HCl, pH 8.0.

For co-purification of WER 67-120/EGL3 1-205, CPC 30-94/EGL3 1-205 and MYB29 174-222/MYC3 44-238, the DNA fragments encoding WER 67-120, CPC 30-96 and MYB29 174-222 were generated by PCR or RT-PCR and constructed into pCDF-duet vector, respectively (Supplementary Table 6). The DNA fragment encoding MYC3 44-238 was generated by RT-PCR and subcloned into pSUMO vector (Supplementary Table 6). The plasmids expressing WER 67-120 and EGL3 1-205, CPC 30-94 and EGL3 1-205, or MYB29 174-222 and MYC3 44-238 were co-transferred into *Escherichia coli* strain BL21 (DE3) cells. Protein expression was induced by 0.2 mM IPTC when the OD_{600} reached 0.8. The cells were collected and the target complexes were purified using the same procedures as that of EGL3 1-205.

Size-exclusion chromatography (SEC) analysis

EGL3 1-205, WER 12-120, and CPC 30-94 proteins were respectively dissolved in SEC buffer (20 mM Tris-HCl, 150 mM NaCl, 5 mM DTT, pH 8.0). To prepare WER–EGL3 and CPC–EGL3 complexes, EGL3 1-205 was mixed with equivalent molar of WER 12-120 and CPC 30-94, respectively and the mixtures were incubated on ice for 30 min. For the competition assay, equivalent molar of CPC 30-94 was added into the

pre-prepared WER–EGL3 complex and incubated on ice for 30 min. The concentration of EGL3 1-205, WER 12-120, and CPC 30-94 were fixed at 50 μM in all samples, which were sequentially analyzed by Superdex^{TM} increase 200 10/300 GL column.

Crystallization and structure determination

The apo EGL3 1-205 crystals were grown at 18°C using the sitting-drop vapor diffusion method. The drop contains 0.2 μl 25 mg/ml EGL3 1-205 protein and 0.2 μl reservoir solution composed of 100 mM CAPS, 200 mM lithium sulfate, and 2 M ammonium sulfate, pH 10.5. All crystals of protein complexes were grown at 18°C by using the hanging-drop vapor diffusion method. For crystallization of WER 67-120 and EGL3 1-205 complex, the co-purified protein complex was concentrated to 20 mg/ml and the crystals were grown in well solution containing 20% PEG3350, 200 mM magnesium formate. The crystals of CPC 30-94 and EGL3 1-205 complex (20 mg/ml) were grown in well solution (0.1 M Tris-HCl, 20% PEG3350, 0.2 M ammonium sulfate, pH 6.5). For crystallization of MYB29 174-222 and MYC3 44-238 complex, the co-purified complex was concentrated to 20 mg/ml and the crystals were grown in well solution (0.1 M HEPES, 20% (w/v) PEG 6000, 0.2 M Sodium chloride, pH 7.0).

The X-ray data were collected at BL17U and BL18U beamlines at Shanghai Synchrotron Radiation Facility (SSRF). Diffractive data were indexed, integrated and scaled with HKL3000 program[33]. The apo EGL3 structure and WER–EGL3 complex were all solved by molecular replacement (MR) method using the Phaser program of CCP4i[34], using the MYC3 structure (PDB_ID：4RRU) as the search model. The CPC–EGL3 complex structure was solved by MR method using the apo EGL3 structure as

the search model. The MYB29–MYC3 complex structure was solved by MR method using the MYC3 structure (PDB_ID：4RRU) as the search model. The model building and refinement were performed with COOT[35] and PHENIX[36]. All the structural images were generated by using the PyMOL program (http://www.pymol.org).

Isothermal titration calorimetry (ITC) assays

All ITC experiments were performed using an iTC200 MicroCalorimeter from MicroCal. Interaction was performed in a buffer (20 mM Tris-HCl, 100 mM NaCl, pH8.0) at 25°C. 38 μl of EGL3 1-205 (200 μM) was titrated into the cell containing 200 μl wild-type or mutated WER 12-120 (20 μM). A total of 25 injections (each of 1.5 μl) were performed. Binding curves were generated by plotting the heat change of the binding reaction, and the data were fitted using one-site binding model with Origin 7.0 (Supplementary Table 1).

Spilt luciferase assays

The DNA fragment encoding C-terminus of luciferase was fused with the DNA fragment encoding EGL3 1-205 to create p35S::3×Flag-EGL3 1-205-nLUC construct (Supplementary Table 6). The DNA fragments encoding WER 12-120 and CPC 30-94 were respectively fused with the DNA fragment encoding N-terminus of luciferase to create p35S::cLUC-WER 12-120-4×Myc and p35S::cLUC-CPC 30-94-4×Myc (Supplementary Table 6). The DNA fragments encoding CPC, CPC M49A, CPC M49L and WER were combined with DNA fragment encoding YFP and cloned into pCAMBIA 1300 to create p35S::CPC 30-94-YFP, p35S::CPC 30-94 M49A-YFP, p35S::CPC 30-94 M49L-YFP and p35S::WER 12-120-YFP fusion constructs,

respectively (Supplementary Table 6). Plasmids were transformed into Agrobacterium strain *Agrobacterium tumefaciens* GV3101 with different transgenic genes and were cultivated in LB medium at 28°C overnight. Different combinations were then co-infiltrated into young leaves of *N. benthamiana*. After treated in darkness for 1 day, the plants were exposed to light for 2 days. Luciferin were injected into the Agrobacterium-infiltrated positions and luciferase activity was measured. To verify the expression level of proteins, equal-size tobacco leaves were cut into small size, cells were lysed and total proteins were extracted and analyzed by Western blotting using anti-MYC (M20002L, Abmart, 1:5000 dilution), anti-GFP (M20004L, Abmart, 1:5000 dilution), anti-Actin (M20009L, Abmart, 1:5000 dilution), and anti-FLAG (F1804, Sigma, 1:2000 dilution) antibodies, respectively.

Microscale thermophoresis (MST) analysis

MST experiments were performed according to the published methods [37]. All proteins were dialyzed into MST reaction buffer (20 mM HEPES and 150 mM NaCl, pH 7.5), labelled and purified using the Protein labelling kit RED-NHS (Nanotemper, cat. no. L001) and the recommended procedures. WER protein was labeled with cy5. 60 nM EGL3 1-205 and 30 nM cy5-labelled WER were mixed in reaction buffer (20 mM HEPES, 150 mM NaCl and 0.03% tween-20, pH 7.5) and incubated on ice for 30 min. Then diluted CPC or mutants (from 0.4 mM to 1.22×10^{-5} mM) were added into the reaction. For the MYB29-MYC3 interaction, Fluorescein amidites (FAM) labeled MYB29 (184-205) was purchased from Scilight-Peptide company (http://www.scilight-peptide.com/) and were dialyzed into MST reaction buffer before reaction. 50 nM

458 FAM-MYB29 were incubated with MYC3 (from 25 μM to 0.763 nM) in reaction buffer
459 and incubated on ice for 30 min. The MST experiments were conducted with Monolith
460 NT.115 instrument (NanoTemper Technologies) and the data were collected under 100%
461 infrared laser power and 20% light-emitting diode power at 25°C. The data were
462 analyzed by MO.Affinity Analysis v2.3 software and the EC50 determined.

463 **Bio-Layer Interferometry (BLI) analysis**

464 BLI (OctetRed96) assay was performed for measuring the binding affinities and kinetic
465 parameters of WER-EGL3 and CPC-EGL3 complexes, by using 6×His-SUMO tagged
466 WER 12-120, 6×His-SUMO tagged CPC 30-94, EGL3 1-205 and 6×His-SUMO as a
467 negative control with Octet RED96 instrument (ForteBio). Experiments were
468 conducted at 30 °C with a shaking speed of 1000 rpm. Nitrilotriacetic acid (NTA)
469 capture tips were soaked into the kinetics buffer (20 mM Tris, 100 mM NaCl,
470 0.02%Tween 20, pH 8.0) for 10 min to baseline equilibration. 6×His-SUMO tagged
471 WER 12-120 or CPC 30-94 was immobilized on the biosensor tips for 300 s, then
472 washed by kinetics buffer till all the parameters were stable. Next, the association of
473 EGL3 1-205 was determined at 500, 125, 31.3, 15.6 nM for 300 s, followed by 300 s
474 of dissociation in kinetics buffer. Affinity constants were calculated by a 1:1 global fit
475 model via ForteBio 10.0 data analysis software.

476 **Structure prediction by Alphafold**

477 Alphafold 2.0.1 was used to predict the structures of GL1-EGL3 and MYB28-MYC3
478 complexes according to the published methods[38-39].

479 **Homologs searching in *Arabidopsis thaliana***

Genes encoding homologous proteins of EGL3 (AT1G63650), WER (AT5G14750), CPC (AT2G46410), MYC3 (AT5G46760) and MYB29 (AT5G07690) were searched by all-against-all BLASTP with the E-value of 1e-^{5}, and the identity >20% in *Arabidopsis thaliana*. Sequences were aligned by MAFFT[40] with accurate aligning options "- maxiterate 1000 - localpair", adjusted manually with the use of AliView[41]. Those candidate homologs were carefully examined for their interaction motifs. The homologs with opposite chemical characteristics in the key residues with our query proteins were filtered out. Due to the great variability, the Ala197 and Ile203 in MYB29s homologs were not selected during this step.

Phylogenetic reconstruction and logo comparison

We selected a total of 23 publicly-available genomes (Supplementary Table 2) representing major plant lineages including seven eudicots, two monocots, two of ANA clade, two gymnosperms, three moniliophytes/lycophytes, three mosses/liverworts/hormworts, and four algae for phylogenetic reconstruction and logo comparison of key motifs. Genomes were retrieved from the Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html), NCBI (https://www.ncbi.nlm.nih.gov), FernBase (https://www.fernbase.org), GIGADB datasets (http://gigadb.org/), Spruce Genome Project (http://congenie.org/), MarpolBase (https://marchantia.info) and from files in previous studies[42-44]. Genes encoding homologous proteins of EGL3, WER, CPC, MYC3 and MYB29 were searched and aligned as above, and then trimmed by trimAl[45] with -gt 0.1. We used RAxML to reconstruct the phylogenetic trees[46] with default settings (-m PROTGAMMAJTT). We examined the phylogenetic trees,

extracted the clades containing the query sequences (*EGL3*, *WER*, *CPC*, *MYC3* and *MYB29*) with bootstrap value > 90, and repeated the steps of alignment and phylogenetic reconstruction. By repeating these steps, the sizes of the trees were gradually reduced until each tree contains only the clade including the query sequences and its sister clade. In order to reveal the evolutionary history of *CPC* and *MYB29*, 43 and 21 species were selected for further analysis, respectively (Supplementary Table 3, 4). The logo comparison analyses were performed on WebLogo3 (http://weblogo.threeplusone.com/create.cgi). The same method was also applied for homologs in Arabidopsis for Fig. 5d.

Data availability

Structural factors and coordinates have been deposited in the Protein Data Bank (PDB) under accession codes 7FDL, 7FDM, 7FDN and 7FDO for WER–EGL3, MYB29–MYC3, EGL3, and CPC–EGL3. And the structure of WER-DNA complex is available in the PDB by accession code 6KKS.

Acknowledgements

We thank Drs. Wei Yang, Hong Ma and Xiaoya Chen for a critical reading of the manuscript. We thank Dr. Yu Ding for help in BLI and ITC experiments. We thank the staff of beamlines BL17U1, BL18U1, and BL19U1 at the Shanghai Synchrotron Radiation Facility for assistance with data collection. We thank the staff members of

the Microscale thermophoresis System at the National Facility for Protein Science in Shanghai (NFPS), Zhangjiang Lab, Shanghai Advanced Research Institute, Chinese Academy of Science, China for providing technical support and assistance in data collection and analysis. We thank Robert McKenzie, PhD, from Liwen Bianji (Edanz) (www.liwenbianji.cn/ac), for editing the English text of a draft of this manuscript. This work was supported by the National Natural Science Foundation of China (NSFC31930017) for A.D. and the National Basic Research Program of China (2012CB910500) for W-H.S..

Author Contributions

A.D. conceived and designed the research. B.W., Q.L., Y.L., K.D., Z.W., and T.L. performed the experiments. Q.L., B.W., J.G., and C.H. analyzed the data. Q.L., B.W., W-H.S., C.H., J.G., and A.D. wrote the manuscript. All authors read, revised, and approved the manuscript.

Competing Interests

The authors declare no competing interests.

Additional Information

542 Supplementary Information is available for this paper. Correspondence and requests

543 for materials should be addressed to Aiwu Dong.

544

545 **Table 1**. **Data collection and refinement statistics.**

	EGL3	WER-EGL3	CPC-EGL3	MYB29-MYC3
Data collection				
Space group	P32	C2221	C121	P32
Cell parameter				
a (Å)	70.4	75.9	86.0	85.2
b (Å)	70.4	193.2	50.3	85.2
c (Å)	88.5	224.6	81.6	57.0
α (°)	90.0	90.0	90.0	90.0
β (°)	90.0	90.0	97.2	90.0
γ (°)	120.0	90.0	90.0	120.0
Wavelength(Å)	0.97930	0.97930	0.97915	0.97930
Resolution (Å)	30.0-1.90	30.0-2.90	30.0-1.80	30.0-2.50
Last shell (Å)	1.97-1.90	3.0-2.90	1.89-1.80	2.59-2.50
Completeness (%)	99.8(99.2)	99.7(98.4)	97.5(98.8)	97.5(91.3)
Redundancy	9.7(7.7)	8.2(6.1)	6.1(5.7)	6.2(3.2)
I/σ(I)	37.9(2.0)	12.4(2.0)	60.8(7.5)	32.1(2.0)
Rmerge (%)	5.9(95.2)	14.8(70.8)	2.4(18.2)	4.1(42.1)
CC(1/2)	0.998(0.876)	0.994(0.114)	0.997(0.983)	0.999(0.114)
Refinement				
Resolution (Å)	30.0-1.90	30.0-2.90	30.0-1.80	30.0-2.50
R_{work} (%) / R_{free} (%)	18.5/22.7	25.4/29.9	18.8/21.5	23.0/28.4
No. of atoms				
Protein	2671	8858	1715	2840
Ligand/ion	25	0	5	0
Water	126	0	139	14
R.m.s. deviations				
Bond length (Å)	0.012	0.009	0.013	0.006
Bond angle (°)	1.689	1.061	1.557	0.935
Ramachandran plot (%)				
Most favored	97.54	95.15	95.0	97.71
Additional allowed	2.46	4.85	5.0	2.29
PDB number	7FDN	7FDL	7FDO	7FDM

546 *Values in parentheses are for the highest-resolution shell.

547

Figure Legends

Figure 1. **Co-crystal structure of the WER–EGL3 complex. a,** Overall structure of the WER–EGL3 complex. The JID and TAD domains of EGL3 are shown in cyan and green, respectively, and WER is shown in pink. The invisible regions (aa 55-66 and 87-96) of EGL3 in electron density map are indicated by dashed grey lines. **b,** Structural superposition of the WER–EGL3 complex with the WER–DNA complex (PDB:6KKS). The WER–EGL3 complex is colored as in **a**. DNA and WER in the WER-DNA complex are colored in brown and yellow, respectively. WER R2 and R3 domains are indicated by arrows. **c,** Interactions between WER and EGL3. H-bonds are indicated by dashed black lines. For clarity, the detailed hydrophobic interactions are not shown. **d,** ITC results showing the binding affinities between EGL3 and wild-type or mutated WER proteins. **e,** ITC results showing the binding affinities between WER and wild-type or mutated EGL3 proteins.

Figure 2. **Co-crystal structure of the CPC–EGL3 complex. a,** Overall structure of the CPC–EGL3 complex, in which the JID and TAD domains of EGL3 are shown in cyan and green, respectively, and CPC is shown in purple. Invisible regions (aa 55-63 and 87-104) of EGL3 are indicated by dashed grey lines. **b,** Structure superimposition of the CPC–EGL3 complex with the WER–EGL3 complex. The CPC-EGL3 complex is colored in grey and purple, whereas the WER-EGL3 complex is colored in yellow and magenta. **c,** Interactions between CPC and EGL3. H-bonds are indicated by dashed black lines. For clarity, the detailed hydrophobic interactions are not shown. **d,** Structural superimposition of WER R3 and CPC R3, which are colored in magenta and purple, respectively. Residues involved in EGL3 interaction are shown, and CPC M49 and WER L85 are highlighted by a black dashed cycle. **e,** Close-up view of the crucial amino acids differing between WER and CPC, and the Van der Waals surfaces are shown by dots at the bottom.

Figure 3. **Functional importance of Met49 of CPC in competition with WER to bind to EGL3. a,** CPC peels WER from the WER–EGL3 complex in a SEC analysis,

which is further detected by SDS-PAGE. For CPC competition assay, an equivalent concentration of CPC was incubated with the prepared WER-EGL3 complex on ice for 30 min and analyzed by SEC experiment. The SEC fraction of WER-EGL3 plus CPC was analyzed by SDS-PAGE, which was repeated independently twice with similar results. **b,** Split-luciferase assay to detect the competition of wild-type and mutated CPC against WER to interact with EGL3. Tobacco leaves were co-infiltrated with Agrobacterium containing EGL3-nLUC and cLUC-WER with or without CPC-YFP, and the luminescence images were captured by a CCD imaging system. **c,** Quantitative measurement of the competitive capabilities of CPC and its mutants against WER to form a complex with EGL3 by a microscale thermophoresis (MST) assay, and the data were presented as mean values ± SD of three independent experiments (n=3).

Figure 4. **Co-crystal structure of the MYB29–MYC3 complex. a,** Overall structure of the MYB29–MYC3 complex. The JID and TAD domains of MYC3 are shown in cyan and green, respectively, and MYB29 is shown in brown. **b,** Interactions between MYB29 and MYC3. H-bonds are indicated by dashed black lines. For clarity, the detailed hydrophobic interactions are not shown. **c,** Structural superposition of the MYB29–MYC3 complex with the JAZ1–MYC3 complex (PDB:4YZ6). MYB29 and MYC3 in the MYB29–MYC3 complex are colored in brown and grey, respectively. JAZ1 and MYC3 in the JAZ1–MYC3 complex are colored in blue and cyan, respectively.

Figure 5. **Two interaction modes of MYB–bHLH complexes. a,** Structural superposition of the WER–EGL3 and MYB29–MYC3 complexes. WER and MYB29 are colored in magenta and brown, respectively. Both EGL3 and MYC3 are colored in grey, whereas their JID helixes are colored in cyan and pink, respectively. **b,** The different interaction modes of the WER–EGL3 and MYB29–MYC3 complexes. **c,** Sequence alignment of homologs of EGL3, WER, CPC, MYC3, and MYB29 in Arabidopsis. For each panel, the uppermost sequence is the logo comparison, with the *x*-axis scaled to the position of amino acids of EGL3, WER, CPC, MYC3, and MYB29,

608 respectively. The hydrophilic, neutral, and hydrophobic nature for the amino acids is
609 indicated in blue, green, and black, respectively. Stars indicate the residues crucial for
610 MYB–bHLH interaction. **d,** Evolution of *EGL3*, *WER*, *CPC*, *MYC3*, and *MYB29*
611 among land plants. Genes able to utilize the EGL3–WER interaction mode are shown
612 in green, and those capable of the MYC3–MYB29 interaction mode are shown in
613 orange.
614

References:

1. Feller, A., Machemer, K., Braun, E. L. & Grotewold, E., Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J.* **66**, 94-116 (2011).

2. Pireyre, M. & Burow, M., Regulation of MYB and bHLH transcription factors: a glance at the protein level. *Mol. Plant* **8**, 378-388 (2015).

3. Dubos, C. *et al.*, MYB transcription factors in Arabidopsis. *Trends Plant Sci.* **15**, 573-581 (2010).

4. Heim, M. A. *et al.*, The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. *Mol. Biol. Evol.* **20**, 735-747 (2003).

5. Toledo-Ortiz, G., Huq, E. & Quail, P. H., The Arabidopsis basic/helix-loop-helix transcription factor family. *Plant Cell* **15**, 1749-1770 (2003).

6. 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.* **47**, 9592-9608 (2019).

7. Schiefelbein, J., Kwak, S. H., Wieckowski, Y., Barron, C. & Bruex, A., The gene regulatory network for root epidermal cell-type pattern formation in Arabidopsis. *J. Exp. Bot.* **60**, 1515-1521 (2009).

8. Song, S. K. *et al.*, Cell fate in the Arabidopsis root epidermis is determined by competition between WEREWOLF and CAPRICE. *Plant Physiol.* **157**, 1196-1208 (2011).

9. Kang, Y. H., Song, S. K., Schiefelbein, J. & Lee, M. M., Nuclear trapping controls the position-dependent localization of CAPRICE in the root epidermis of Arabidopsis. *Plant Physiol.* **163**, 193-204 (2013).

10. Tominaga-Wada, R. & Wada, T., Regulation of root hair cell differentiation by R3 MYB transcription factors in tomato and Arabidopsis. *Front. Plant Sci.* **5**, 91 (2014).

11. Qi, T., Huang, H., Song, S. & Xie, D., Regulation of jasmonate-mediated stamen development and seed production by a bHLH-MYB complex in Arabidopsis. *Plant Cell* **27**, 1620-1633 (2015).

12. Fernandez-Calvo, P. *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* **23**, 701-715 (2011).

13. Zhang, F. *et al.*, Structural basis of JAZ repression of MYC transcription factors in jasmonate signalling. *Nature* **525**, 269-273 (2015).

14. Schweizer, F. *et al.*, Arabidopsis basic helix-loop-helix transcription factors MYC2, MYC3, and MYC4 regulate glucosinolate biosynthesis, insect performance, and feeding behavior. *Plant Cell*

645 **25**, 3117-3132 (2013).
646 15. Wang, B. *et al.*, Structural insights into target DNA recognition by R2R3-MYB transcription factors.
647 *Nucleic Acids Res.* **48**, 460-471 (2020).
648 16. Triezenberg, S. J., Structure and function of transcriptional activation domains. *Curr. Opin. Genet.*
649 *Dev.* **5**, 190-196 (1995).
650 17. Melcher, K., The strength of acidic activation domains correlates with their affinity for both
651 transcriptional and non-transcriptional proteins. *J. Mol. Biol.* **301**, 1097-1112 (2000).
652 18. Sun, X., Rikkerink, E. H., Jones, W. T. & Uversky, V. N., Multifarious roles of intrinsic disorder in
653 proteins illustrate its broad impact on plant biology. *Plant Cell* **25**, 38-55 (2013).
654 19. Zimmermann, I. M., Heim, M. A., Weisshaar, B. & Uhrig, J. F., Comprehensive identification of
655 Arabidopsis thaliana MYB transcription factors interacting with R/B-like BHLH proteins. *Plant J.*
656 **40**, 22-34 (2004).
657 20. Krissinel, E. & Henrick, K., Inference of macromolecular assemblies from crystalline state. *J. Mol.*
658 *Biol.* **372**, 774-797 (2007).
659 21. Frerigmann, H., Berger, B. & Gigolashvili, T., bHLH05 is an interaction partner of MYB51 and a
660 novel regulator of glucosinolate biosynthesis in Arabidopsis. *Plant Physiol.* **166**, 349-369 (2014).
661 22. Gigolashvili, T., Yatusevich, R., Berger, B., Muller, C. & Flugge, U. I. The R2R3-MYB
662 transcription factor HAG1/MYB28 is a regulator of methionine-derived glucosinolate biosynthesis
663 in Arabidopsis thaliana. *Plant J.* **51**, 247-261 (2007).
664 23. Hirai, M. Y. et al. Omics-based identification of Arabidopsis Myb transcription factors regulating
665 aliphatic glucosinolate biosynthesis. *Proc Natl Acad Sci U S A.* **104**, 6478-6483 (2007).
666 24. Sonderby, I. E. et al. A systems biology approach identifies a R2R3 MYB gene subfamily with
667 distinct and overlapping functions in regulation of aliphatic glucosinolates. *PLoS One.* **2**, e1322
668 (2007).
669 25. Beekwilder, J. et al. The impact of the absence of aliphatic glucosinolates on insect herbivory in
670 Arabidopsis. *PLoS One*. **3**, e2068 (2008).
671 26. Gigolashvili, T., Engqvist, M., Yatusevich, R., Muller, C. & Flugge, U. I. HAG2/MYB76 and
672 HAG3/MYB29 exert a specific and coordinated control on the regulation of aliphatic glucosinolate
673 biosynthesis in Arabidopsis thaliana. *New Phytol*. **177**, 627-642 (2008).
674 27. Malitsky, S. et al. The transcript and metabolite networks affected by the two clades of Arabidopsis

glucosinolate biosynthesis regulators. *Plant Physiol.* **148**, 2021-2049 (2008).

28. Sonderby, I. E., Burow, M., Rowe, H. C., Kliebenstein, D. J. & Halkier, B. A. A complex interplay of three R2R3 MYB transcription factors determines the profile of aliphatic glucosinolates in Arabidopsis. *Plant Physiol.* **153**, 348-363 (2010).

29. Li, Y. et al. Novel insights into the function of Arabidopsis R2R3-MYB transcription factors regulating aliphatic glucosinolate biosynthesis. *Plant Cell Physiol.* **54**, 1335-1344 (2013).

30. Frerigmann, H. & Gigolashvili, T. MYB34, MYB51, and MYB122 distinctly regulate indolic glucosinolate biosynthesis in Arabidopsis thaliana. *Mol. Plant*. **7**, 814-828 (2014).

31. Celenza, J. L. et al. The Arabidopsis ATR1 Myb transcription factor controls indolic glucosinolate homeostasis. *Plant Physiol.* **137**, 253-262 (2005).

32. Gigolashvili, T. et al. The transcription factor HIG1/MYB51 regulates indolic glucosinolate biosynthesis in Arabidopsis thaliana. *Plant J.* **50**, 886-901 (2007).

33. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M., HKL-3000: the integration of data reduction and structure solution-from diffraction images to an initial model in minutes. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 859-866 (2006).

34. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E., A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1131-1137 (2003).

35. Emsley, P. & Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126-2132 (2004).

36. Adams, P. D. *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-221 (2010).

37. Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P. & Duhr, S., Molecular interaction studies using microscale thermophoresis. *Assay Drug Dev. Technol.* **9**, 342-353 (2011).

38. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. **596**, 583-589 (2021).

39. Evans, Richard, et al. Protein complex prediction with AlphaFold-Multimer. Preprint at https: //www.biorxiv.org/content/10.1101/2021.10.04.463034v1

40. Katoh, K. & Standley, D. M., MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772-780 (2013).

41. Larsson, A., AliView: a fast and lightweight alignment viewer and editor for large datasets.

Bioinformatics **30**, 3276-3278 (2014).

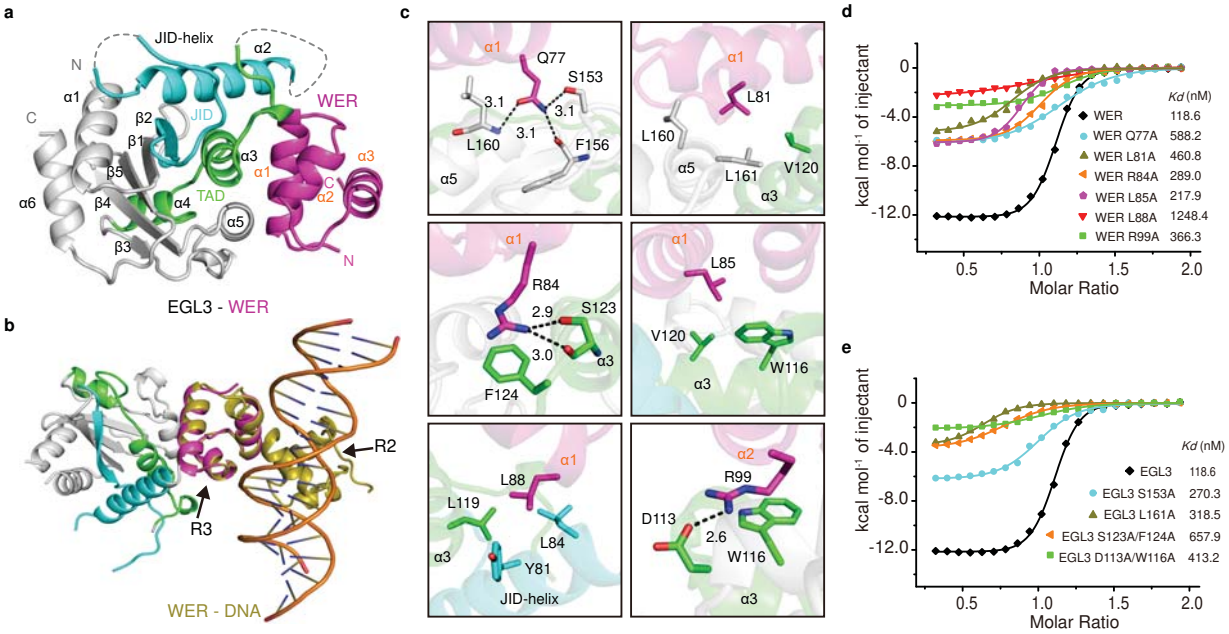
42. Hori, K. *et al.*, Klebsormidium flaccidum genome reveals primary factors for plant terrestrial adaptation. *Nat. Commun.* **5**, 3978 (2014).

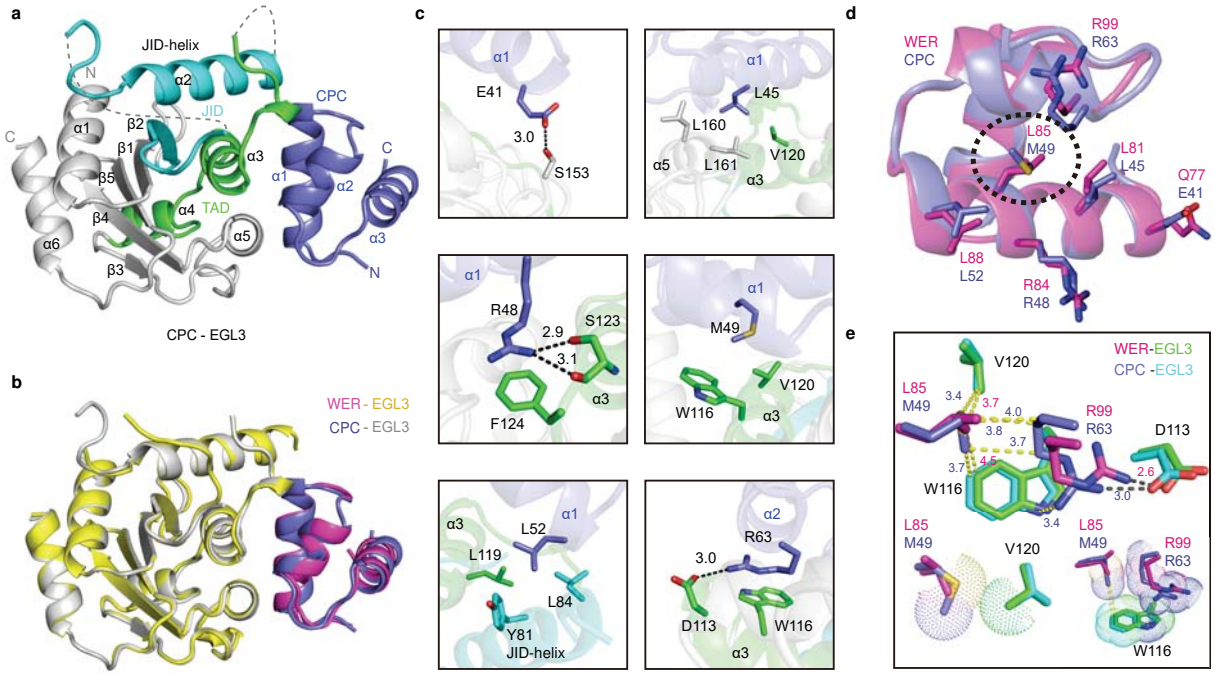
43. Nishiyama, T. *et al.*, The chara genome: secondary complexity and implications for plant terrestrialization. *Cell* **174**, 448-464 (2018).

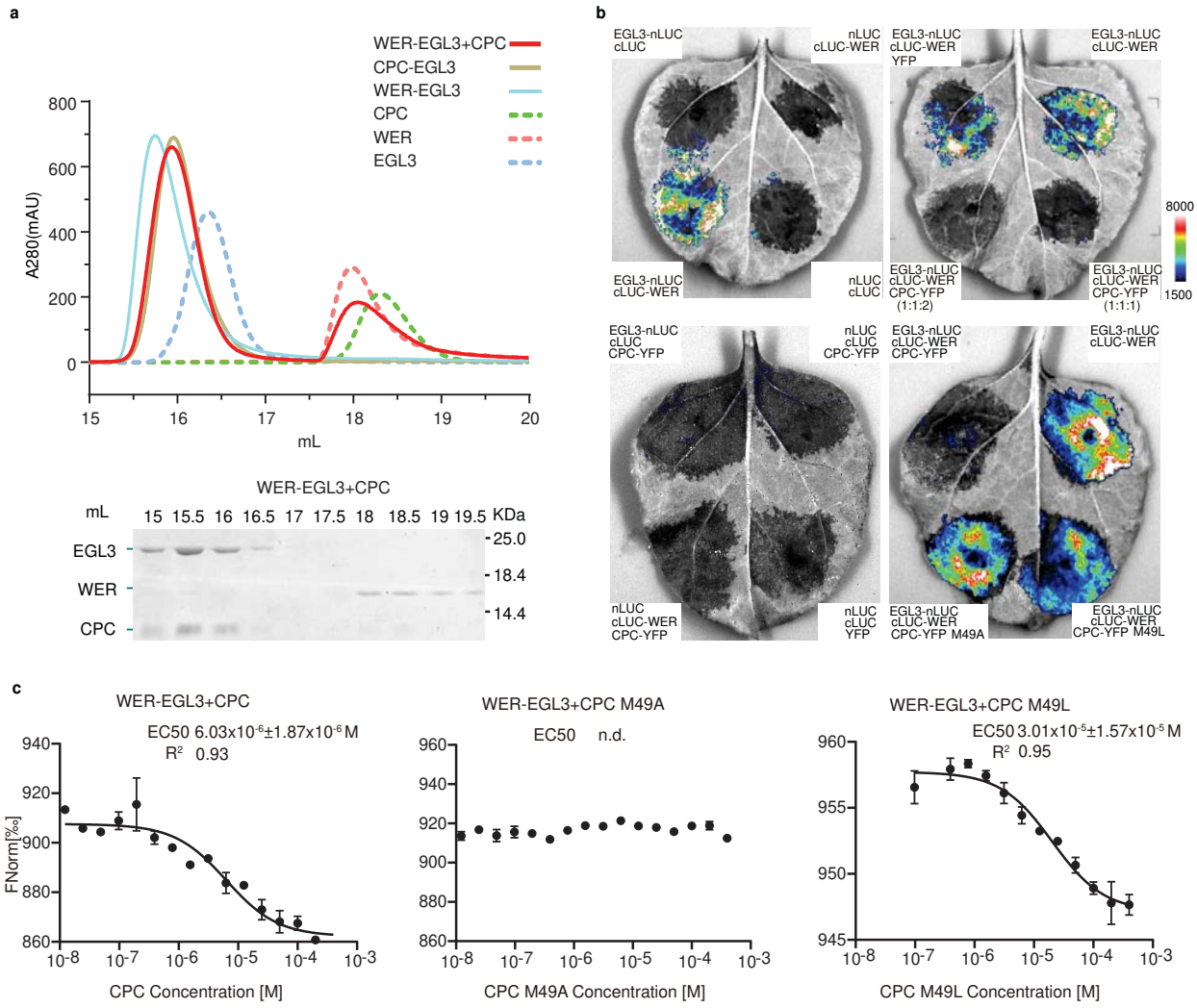
44. Wan, T. *et al.*, A genome for gnetophytes and early evolution of seed plants. *Nat. Plants* **4**, 82-89 (2018).

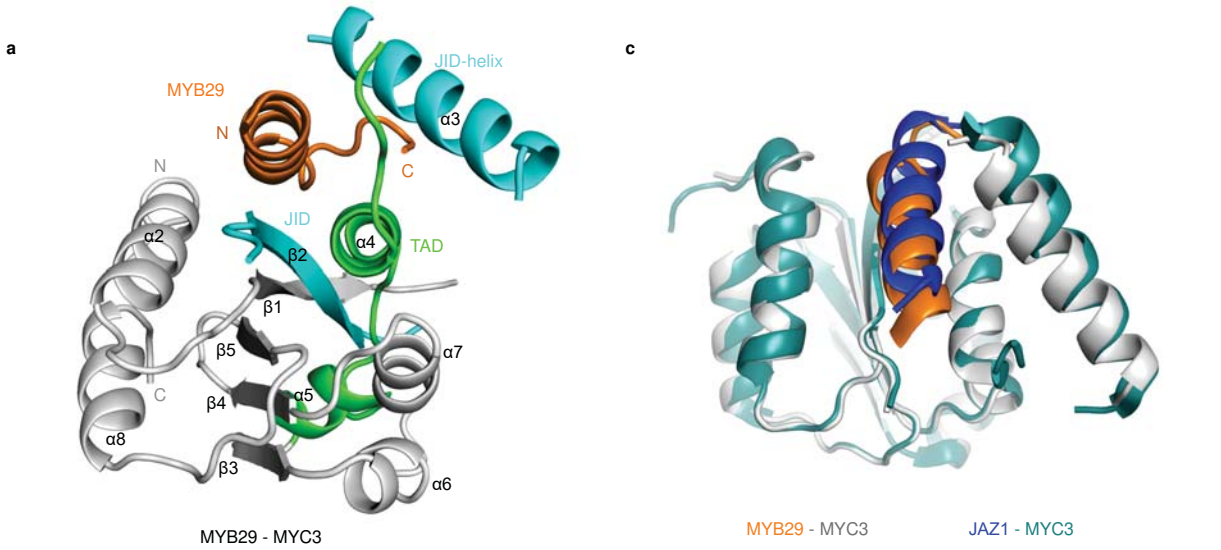
45. Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T., trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009).

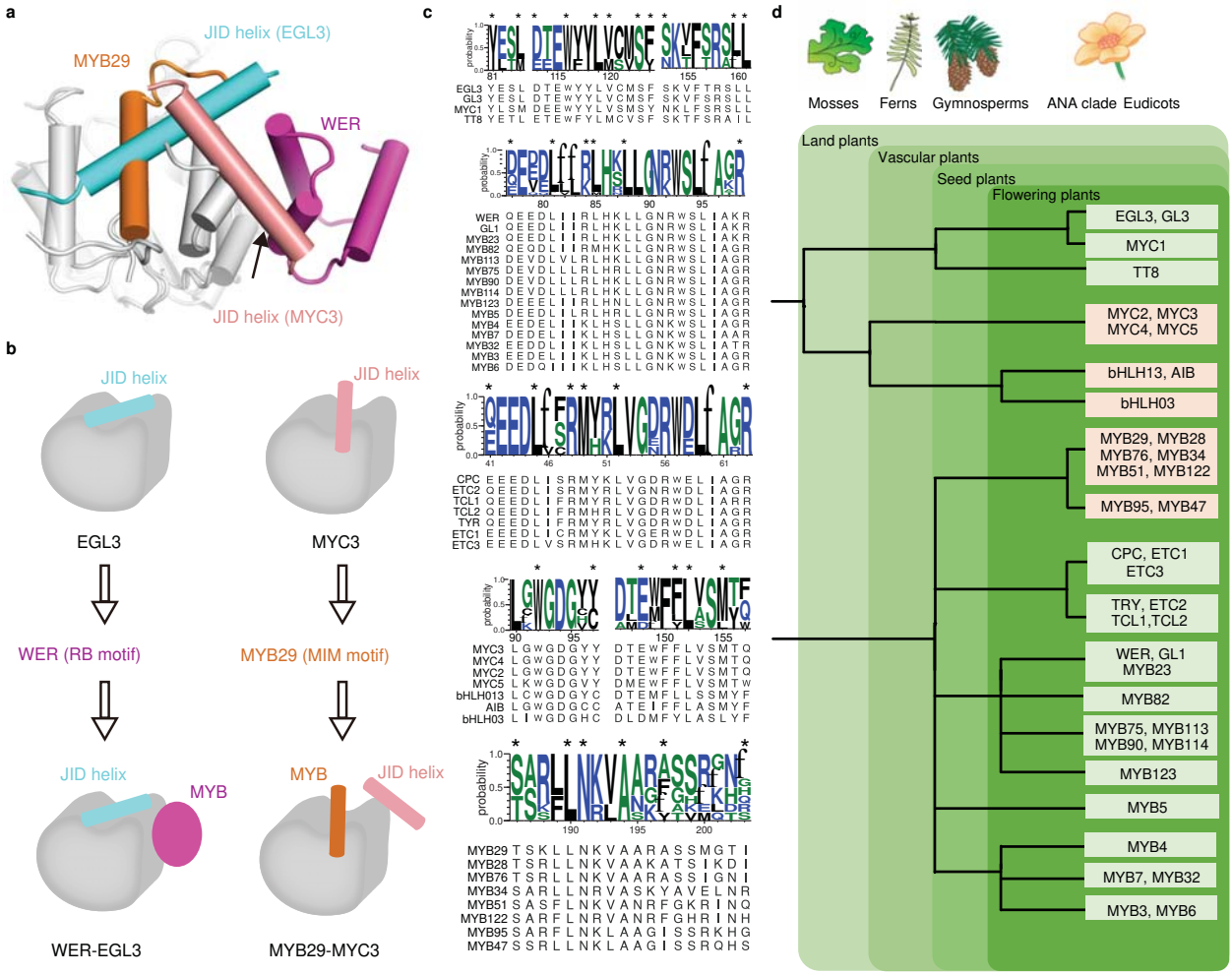
46. Stamatakis, A., RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).

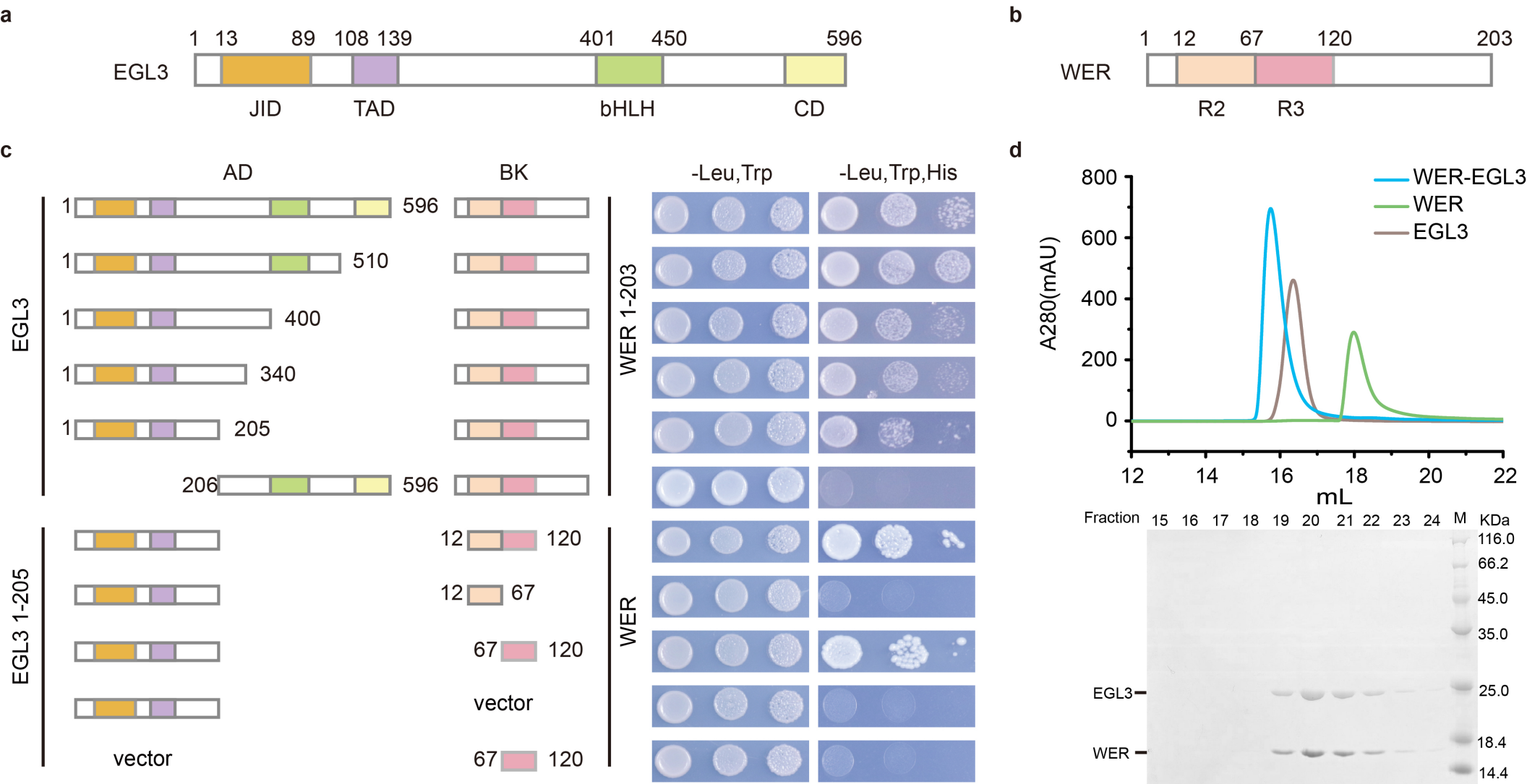

a
JID-helix
α2
N
α1
C
β2
JID
WER
β1
α3
α3
β5
α1
C
α6
β4
α4
TAD
α2
α5
β3
N
EGL3 - WER
b
R2
R3
WER - DNA
c
α1
Q77
S153
3.1
3.1
3.1
L160
F156
α5
α1
L81
L160
α5
L161
V120
α3
α1
R84
2.9
S123
3.0
α3
F124
α1
L85
V120
W116
α3
α1
L88
L119
L84
α3
Y81
JID-helix
α2
R99
D113
2.6
W116
α3
d
kcal mol-1 of injectant
0
-4.0
-8.0
-12.0
0.5
1.0
1.5
2.0
Molar Ratio
Kd (nM)
WER 118.6
WER Q77A 588.2
WER L81A 460.8
WER R84A 289.0
WER L85A 217.9
WER L88A 1248.4
WER R99A 366.3
e
kcal mol-1 of injectant
0
-4.0
-8.0
-12.0
0.5
1.0
1.5
2.0
Molar Ratio
Kd (nM)
EGL3 118.6
EGL3 S153A 270.3
EGL3 L161A 318.5
EGL3 S123A/F124A 657.9
EGL3 D113A/W116A 413.2

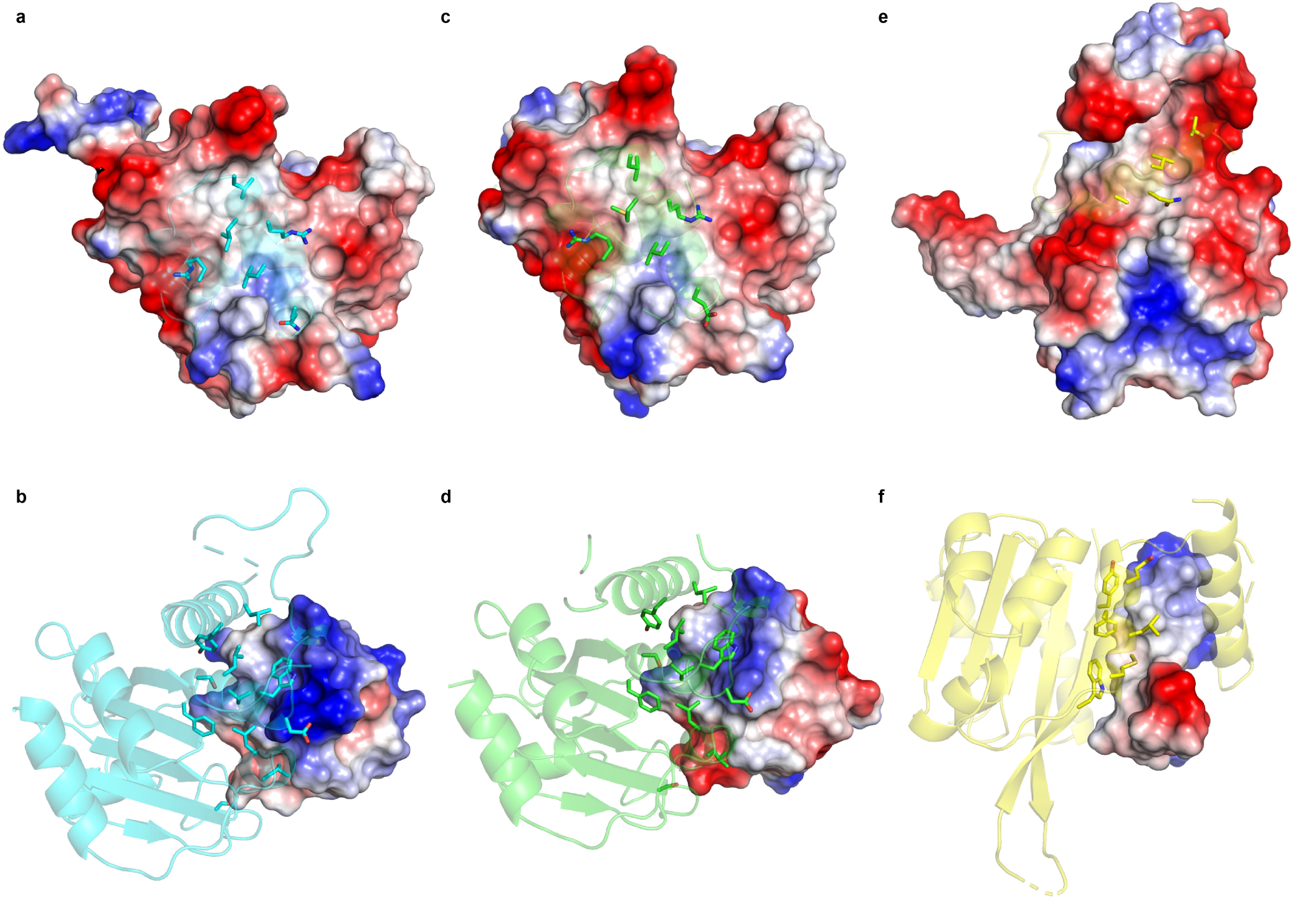

a
JID-helix
N
α2
JID
CPC
α1
β2
β1
α3
C
β5
α1
α2
C
β4
α4
TAD
α5
α3
α6
β3
N
CPC - EGL3
b
WER - EGL3
CPC - EGL3
c
α1
E41
3.0
S153
α1
L45
L160
α5
L161
V120
α3
α1
R48
2.9
S123
3.1
α3
F124
α1
M49
V120
W116
α3
α3
L52
α1
L119
L84
Y81
JID-helix
α2
R63
3.0
D113
W116
α3
d
WER
CPC
R99
R63
L85
M49
L81
L45
Q77
E41
L88
L52
R84
R48
e
V120
WER-EGL3
CPC -EGL3
L85
M49
3.4
3.7
4.0
R99
R63
D113
3.8
3.7
3.7
4.5
2.6
W116
3.0
3.4
L85
M49
V120
L85
M49
R99
R63
W116

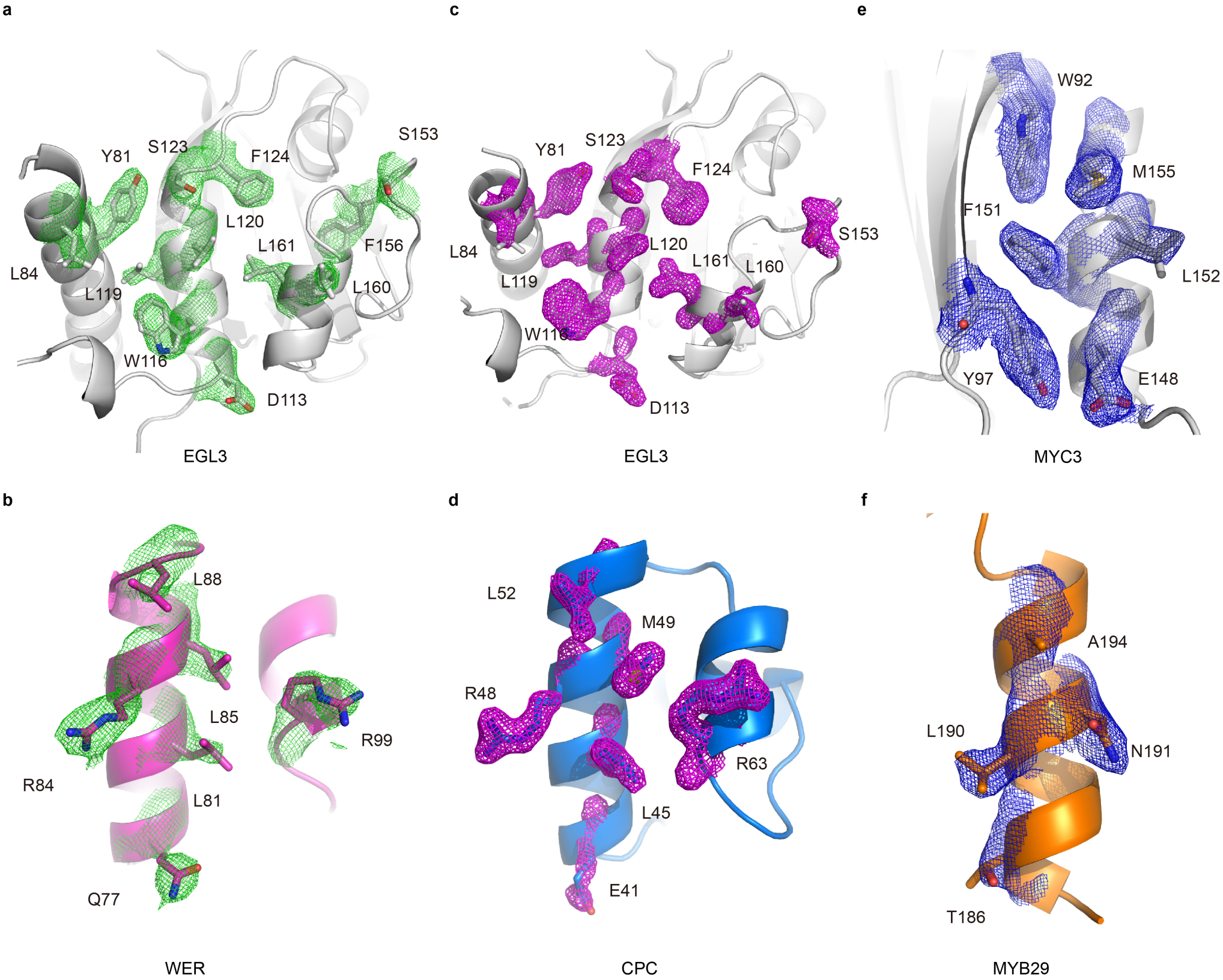

a
WER-EGL3+CPC
CPC-EGL3
WER-EGL3
CPC
WER
EGL3
A280(mAU)
mL
WER-EGL3+CPC
mL 15 15.5 16 16.5 17 17.5 18 18.5 19 19.5 KDa
EGL3
WER
CPC
25.0
18.4
14.4
b
EGL3-nLUC cLUC
nLUC cLUC-WER
EGL3-nLUC cLUC-WER YFP
EGL3-nLUC cLUC-WER
EGL3-nLUC cLUC-WER
nLUC cLUC
EGL3-nLUC cLUC-WER CPC-YFP (1:1:2)
EGL3-nLUC cLUC-WER CPC-YFP (1:1:1)
8000
1500
EGL3-nLUC cLUC CPC-YFP
nLUC cLUC CPC-YFP
EGL3-nLUC cLUC-WER CPC-YFP
EGL3-nLUC cLUC-WER
nLUC cLUC-WER CPC-YFP
nLUC cLUC YFP
EGL3-nLUC cLUC-WER CPC-YFP M49A
EGL3-nLUC cLUC-WER CPC-YFP M49L
c
WER-EGL3+CPC
EC50 6.03x10-6±1.87x10-6 M
R2 0.93
FNorm[‰]
CPC Concentration [M]
WER-EGL3+CPC M49A
EC50 n.d.
CPC M49A Concentration [M]
WER-EGL3+CPC M49L
EC50 3.01x10-5±1.57x10-5 M
R2 0.95
CPC M49L Concentration [M]

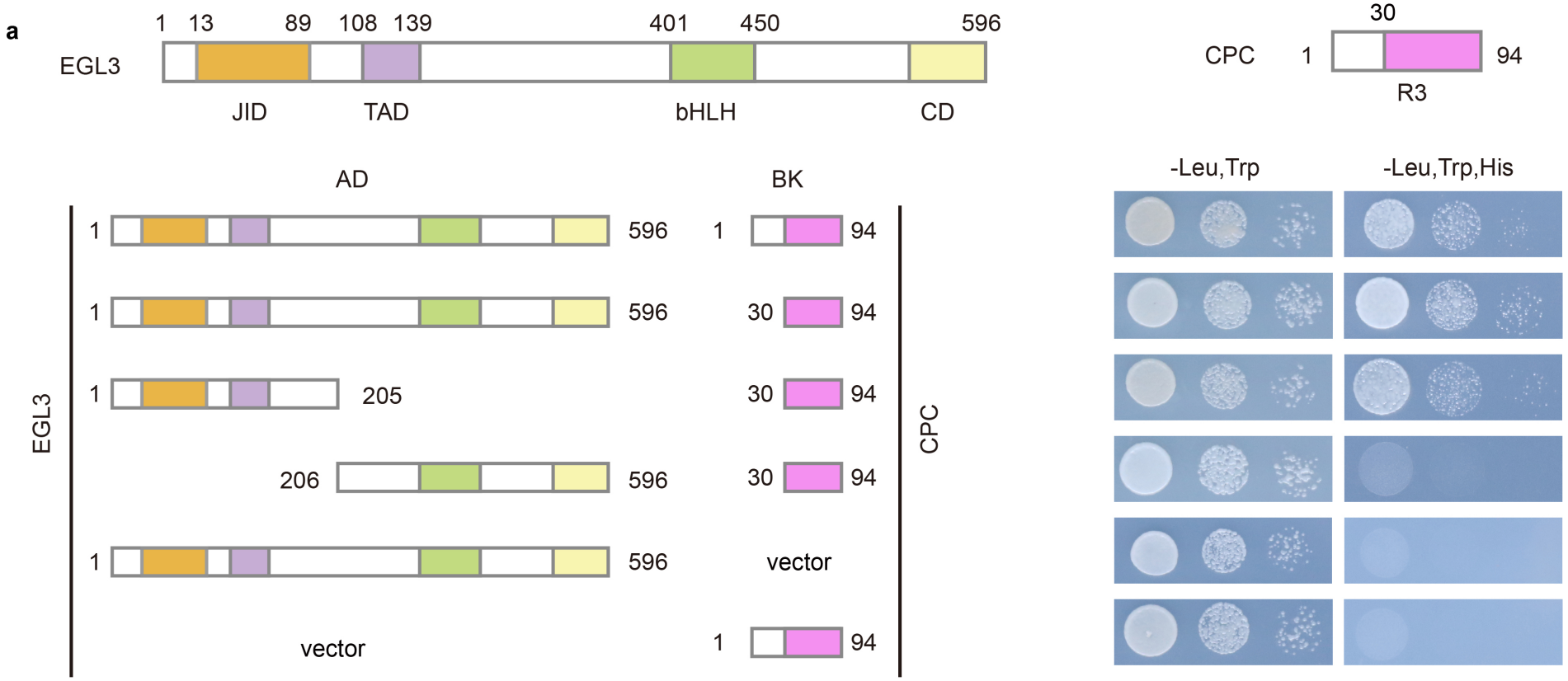

a
MYB29
N
JID-helix
α3
C
N
JID
α2
β2
α4
TAD
β1
β5
α7
C
β4
α5
α8
β3
α6
MYB29 - MYC3
c
MYB29 - MYC3
JAZ1 - MYC3
b
T186
2.8
α4
E148
β2
L190
L152
Y97
F151
α4
β2
2.8
N191
3.3
Y97
β2
W92
M155
A194
α4
F151

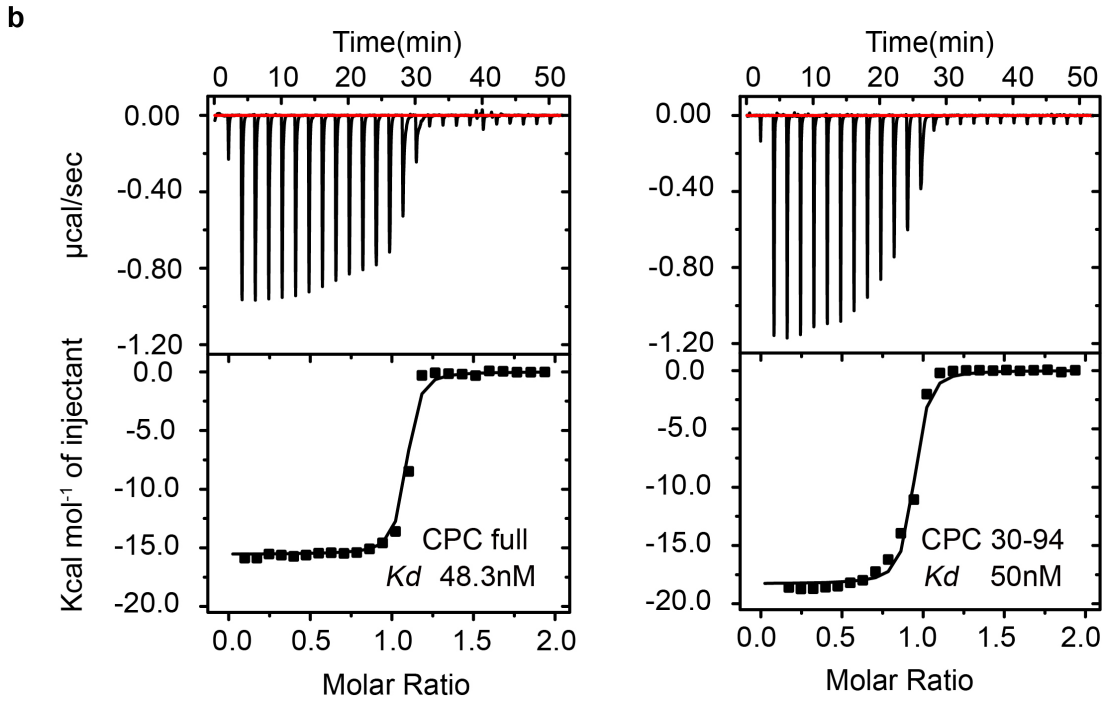

a
MYB29
JID helix (EGL3)
WER
JID helix (MYC3)
b
JID helix
EGL3
JID helix
MYC3
WER (RB motif)
MYB29 (MIM motif)
JID helix
MYB
WER-EGL3
MYB
JID helix
MYB29-MYC3
c
probability
EGL3 Y E S L D T E W Y Y L V C M S F S K V F T R S L L
GL3 Y E S L D T E W Y Y L V C M S F S K V F S R S L L
MYC1 Y L S M D E E W Y Y L V S M S Y N K L F S R S L L
TT8 Y E T L E T E W F Y L M C V S F S K T F S R A I L
WER Q E E D L I I R L H K L L G N R W S L I A K R
GL1 Q E E D L I I R L H K L L G N R W S L I A K R
MYB23 Q E E D L I I R L H K L L G N R W S L I A K R
MYB82 Q E E D L I I R M H K L L G N R W S L I A G R
MYB113 D E V D L V L R L H K L L G N R W S L I A G R
MYB75 D E V D L L L R L H R L L G N R W S L I A G R
MYB90 D E V D L L L R L H K L L G N R W S L I A G R
MYB114 D E V D L L L R L H K L L G N R W S L I A G R
MYB123 D E E E L I I R L H N L L G N R W S L I A G R
MYB5 D E E D L I I R L H R L L G N R W S L I A G R
MYB4 E E D E L I I K L H S L L G N K W S L I A G R
MYB7 D E D E L I I K L H S L L G N K W S L I A A R
MYB32 E E D D L I I K L H S L L G N K W S L I A T R
MYB3 E E D E L I I K L H S L L G N K W S L I A G R
MYB6 D E D Q I I I K L H S L L G N K W S L I A G R
CPC E E E D L I S R M Y K L V G D R W E L I A G R
ETC2 Q E E D L I S R M Y R L V G N R W D L I A G R
TCL1 Q E E D L I F R M Y R L V G D R W D L I A R R
TCL2 Q E E D L I F R M H R L V G D R W D L I A G R
TYR Q E E D L I F R M Y R L V G D R W D L I A G R
ETC1 E E E D L I C R M Y K L V G E R W D L I A G R
ETC3 E E E D L V S R M H K L V G D R W E L I A G R
MYC3 L G W G D G Y Y D T E W F F L V S M T Q
MYC4 L G W G D G Y Y D T E W F F L V S M T Q
MYC2 L G W G D G Y Y D T E W F F L V S M T Q
MYC5 L K W G D G V Y D M E W F F L V S M T W
bHLH013 L C W G D G Y C D T E M F L L S S M Y F
AIB L G W G D G C C A T E I F F L A S M Y F
bHLH03 L I W G D G H C D L D M F Y L A S L Y F
MYB29 T S K L L N K V A A R A S S M G T I
MYB28 T S R L L N K V A A K A T S I K D I
MYB76 T S R L L N K V A A R A S S I G N I
MYB34 S A R L L N R V A S K Y A V E L N R
MYB51 S A S F L N K V A N R F G K R I N Q
MYB122 S A R F L N R V A N R F G H R I N H
MYB95 S A R F L N K L A A G I S S R K H G
MYB47 S S R L L N K L A A G I S S R Q H S
d
Mosses
Ferns
Gymnosperms
ANA clade
Eudicots
Land plants
Vascular plants
Seed plants
Flowering plants
EGL3, GL3
MYC1
TT8
MYC2, MYC3
MYC4, MYC5
bHLH13, AIB
bHLH03
MYB29, MYB28
MYB76, MYB34
MYB51, MYB122
MYB95, MYB47
CPC, ETC1
ETC3
TRY, ETC2
TCL1,TCL2
WER, GL1
MYB23
MYB82
MYB75, MYB113
MYB90, MYB114
MYB123
MYB5
MYB4
MYB7, MYB32
MYB3, MYB6

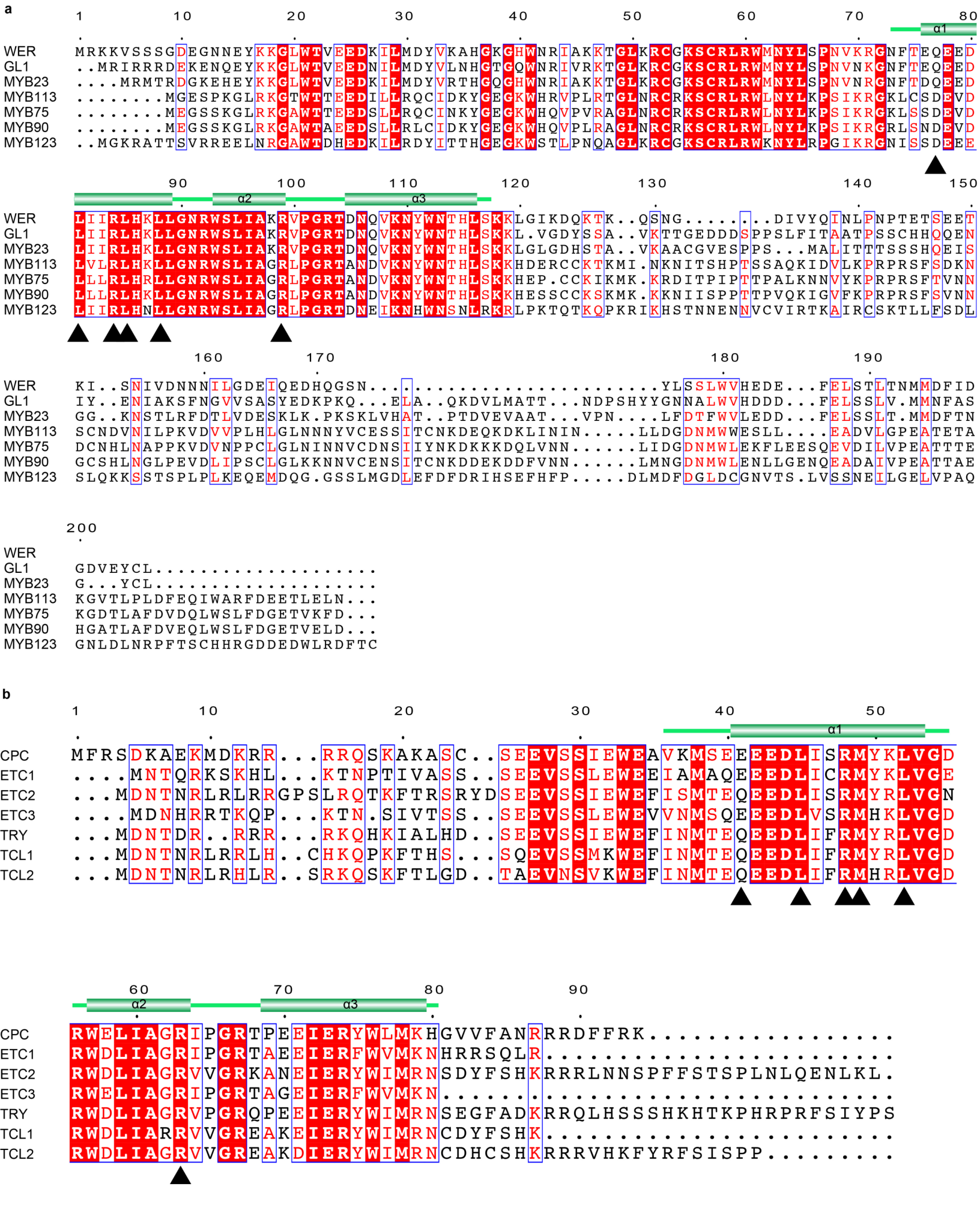

a
EGL3
1
13
89
108
139
401
450
596
JID
TAD
bHLH
CD
b
WER
1
12
67
120
203
R2
R3
c
AD
BK
-Leu,Trp
-Leu,Trp,His
EGL3
EGL3 1-205
WER 1-203
WER
vector
d
WER-EGL3
WER
EGL3
A280(mAU)
mL
Fraction
KDa
M
116.0
66.2
45.0
35.0
25.0
18.4
14.4

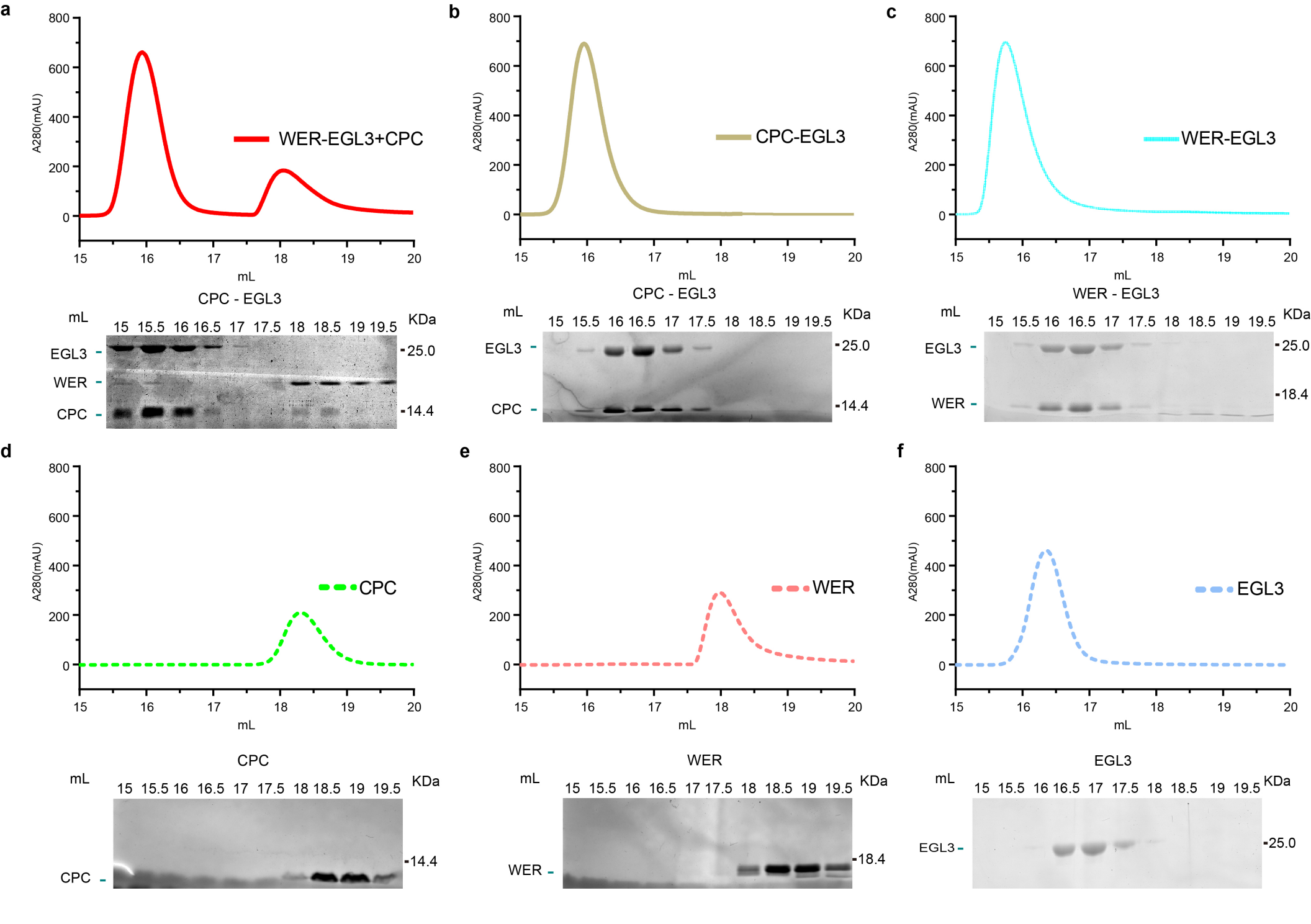

a
c
e
b
d
f

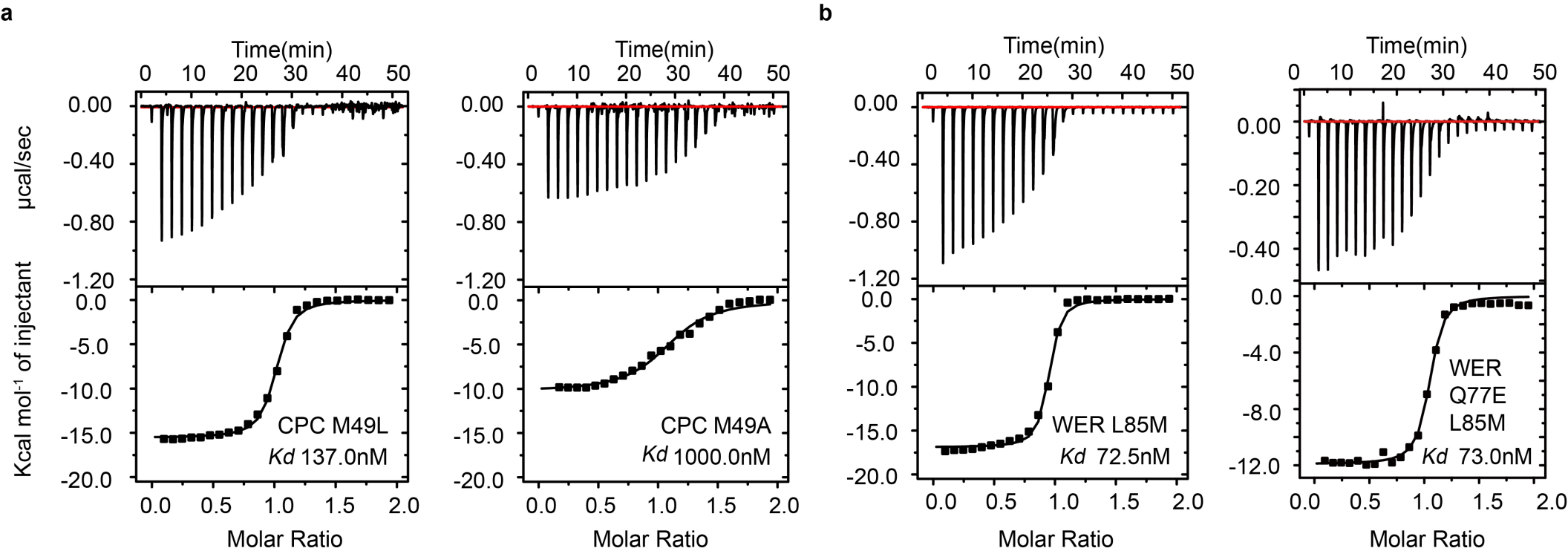

a
Y81
S123
F124
S153
L120
L161
F156
L84
L160
L119
W116
D113
EGL3
c
Y81
S123
F124
L84
S153
L120
L161
L160
L119
W116
D113
EGL3
e
W92
M155
F151
L152
Y97
E148
MYC3
b
L88
L85
R99
R84
L81
Q77
WER
d
L52
M49
R48
R63
L45
E41
CPC
f
A194
L190
N191
T186
MYB29

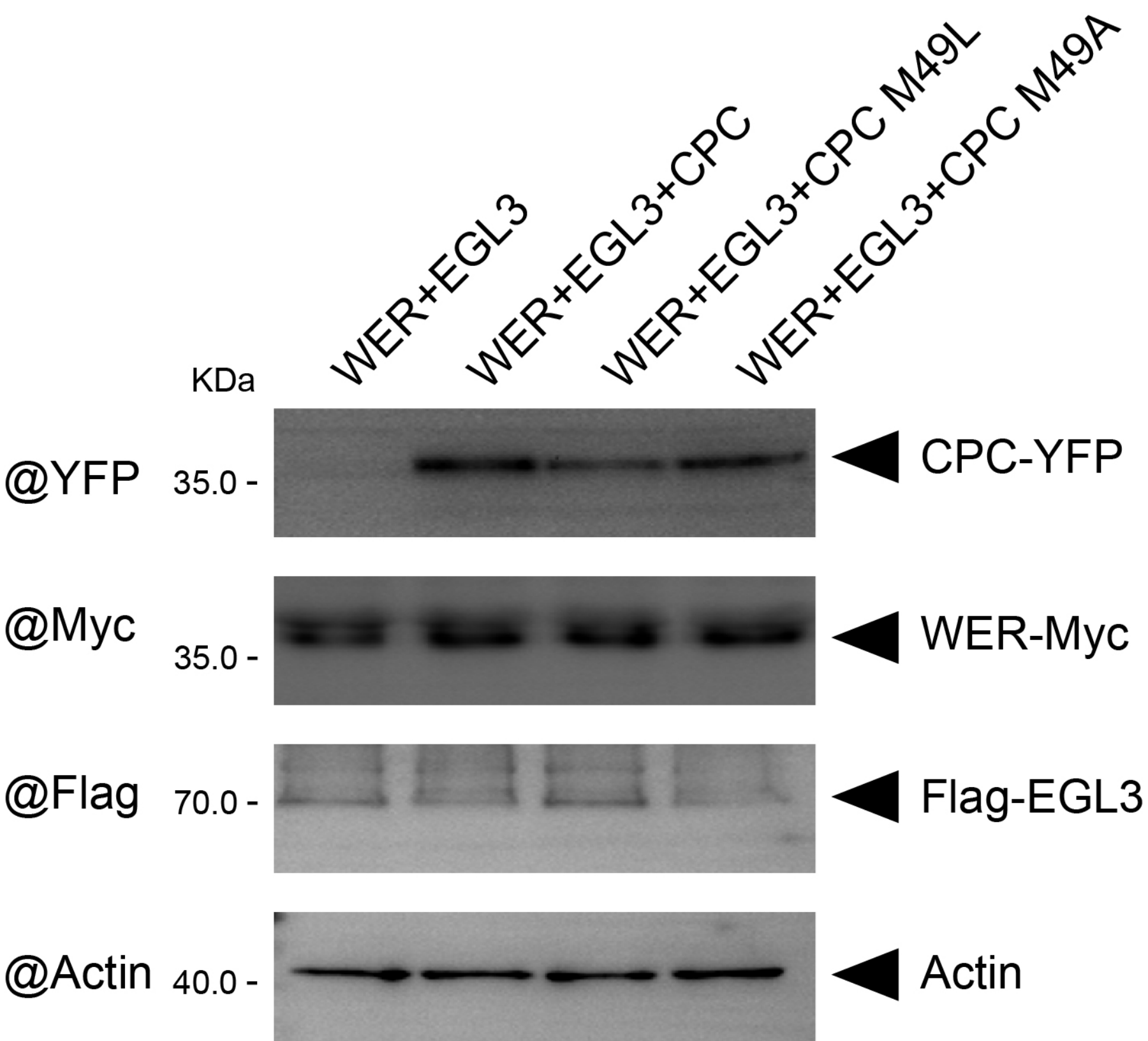

a
EGL3
1
13
89
108
139
401
450
596
JID
TAD
bHLH
CD
CPC
30
1
94
R3
AD
BK
EGL3
CPC
205
206
vector
-Leu,Trp
-Leu,Trp,His


b
Time(min)
µcal/sec
Kcal mol-1 of injectant
CPC full
Kd 48.3nM
CPC 30-94
Kd 50nM
Molar Ratio

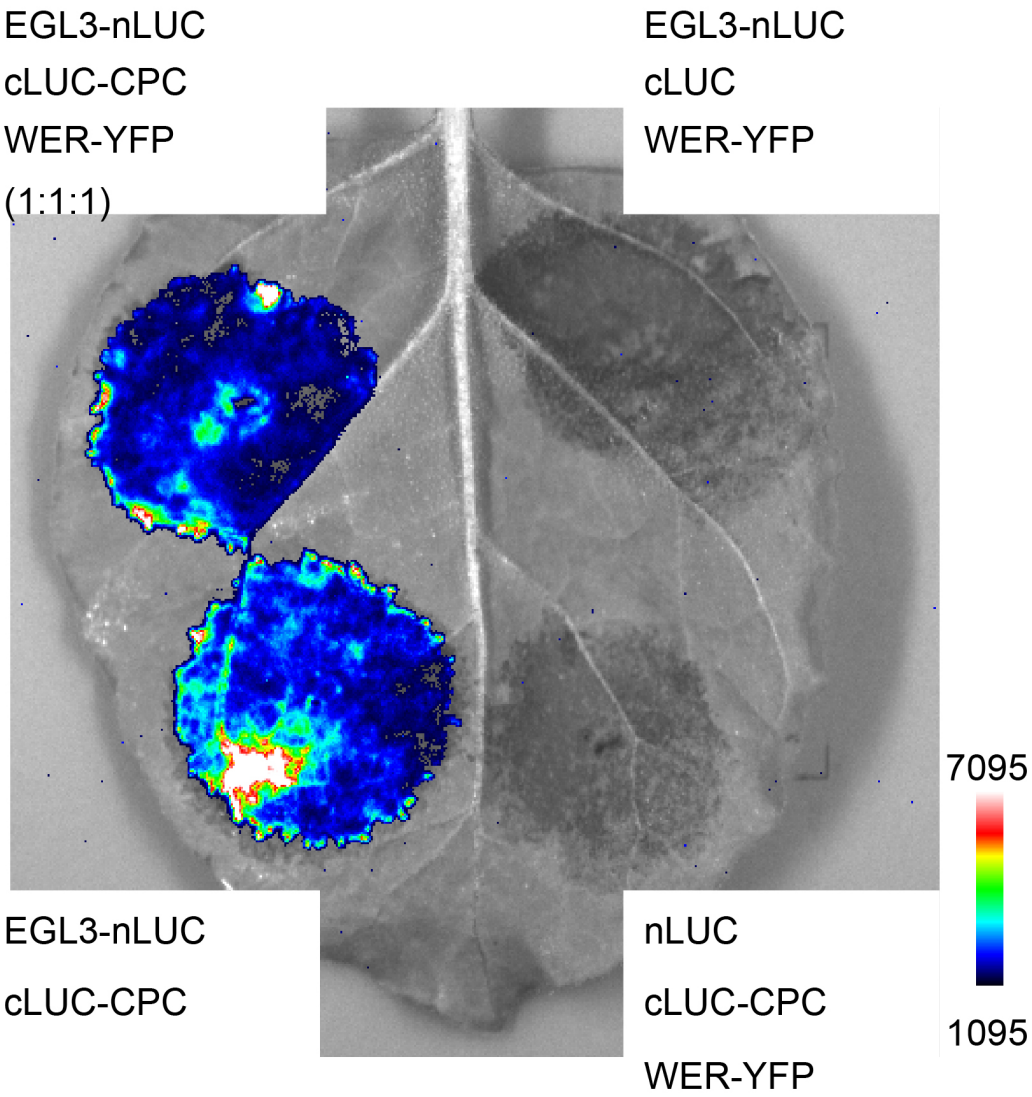

a
1 10 20 30 40 50 60 70 80
α1
WER MRKKVSSSGDEGNNEYKKGLWTVEEDKILMDYVKAHGKGHWNRIAKKTGLKRCGKSCRLRWMNYLSPNVKRGNFTEQEED
GL1 ..MRIRRRDEKENQEYKKGLWTVEEDNILMDYVLNHGTGQWNRIVRKTGLKRCGKSCRLRWMNYLSPNVNKGNFTEQEED
MYB23 ....MRMTRDGKEHEYKKGLWTVEEDKILMDYVRTHGQGHWNRIAKKTGLKRCGKSCRLRWMNYLSPNVNRGNFTDQEED
MYB113 ........MGESPKGLRKGTWTTEEDILLRQCIDKYGEGKWHRVPLRTGLNRCRKSCRLRWLNYLKPSIKRGKLCSDEVD
MYB75 ........MEGSSKGLRKGAWTTEEDSLLRQCINKYGEGKWHQVPVRAGLNRCRKSCRLRWLNYLKPSIKRGKLSSDEVD
MYB90 ........MEGSSKGLRKGAWTAEEDSLLRLCIDKYGEGKWHQVPLRAGLNRCRKSCRLRWLNYLKPSIKRGRLSNDEVD
MYB123 ..MGKRATTSVRREELNRGAWTDHEDKILRDYITTHGEGKWSTLPNQAGLKRCGKSCRLRWKNYLRPGIKRGNISSDEEE
90 100 110 120 130 140 150
α2 α3
WER LIIRLHKLLGNRWSLIAKRVPGRTDNQVKNYWNTHLSKKLGIKDQKTK..QSNG.....DIVYQINLPNPTETSEETT
GL1 LIIRLHKLLGNRWSLIAKRVPGRTDNQVKNYWNTHLSKKL.VGDYSSA..VKTTGEDDDSPPSLFITAATPSSCHHQQEN
MYB23 LIIRLHKLLGNRWSLIAKRVPGRTDNQVKNYWNTHLSKKLGLGDHSTA..VKAACGVESPPS...MALITTTSSSHQEIS
MYB113 LVLRLHKLLGNRWSLIAGRLPGRTANDVKNYWNTHLSKKHDERCCKTKMI.NKNITSHPTSSAQKIDVLKPRPRSFSDKN
MYB75 LLLRLHRLLGNRWSLIAGRLPGRTANDVKNYWNTHLSKKHEP.CCKIKMK.KRDITPIPTTPALKNNVYKPRPRSFTVNN
MYB90 LLLRLHKLLGNRWSLIAGRLPGRTANDVKNYWNTHLSKKHESSCCKSKMK.KKNIISPPTTPVQKIGVFKPRPRSFSVNN
MYB123 LIIRLHNLLGNRWSLIAGRLPGRTDNEIKNHWNSNLRKRLPKTQTKQPKRIKHSTNNENNVCVIRTKAIRCSKTLLFSDL
160 170 180 190
WER KI..SNIVDNNNILGDEIQEDHQGSN.....YLSSLWVHEDE...FELSTLTNMMDFID
GL1 IY..ENIAKSFNGVVSASYEDKPKQ...ELA..QKDVLMATT...NDPSHYYGNNALWVHDDD...FELSSLV.MMNFAS
MYB23 GG..KNSTLRFDTLVDESKLK.PKSKLVHAT..PTDVEVAAT...VPN....LFDTFWVLEDD...FELSSLT.MMDFTN
MYB113 SCNDVNILPKVDVVPLHLGLNNNYVCESSITCNKDEQKDKLININ.....LLDGDNMWWESLL....EADVLGPEATETA
MYB75 DCNHLNAPPKVDVNPPCLGLNINNVCDNSIIYNKDKKKDQLVNN.....LIDGDNMWLEKFLEESQEVDILVPEATTTE
MYB90 GCSHLNGLPEVDLIPSCLGLKKNNVCENSITCNKDDEKDDFVNN.....LMNGDNMWLENLLGENQEADAIVPEATTAE
MYB123 SLQKKSSTSPLPLKEQEMDQG.GSSLMGDLEFDFDRIHSEFHFP.....DLMDFDGLDCGNVTS.LVSSNEILGELVPAQ
200
WER
GL1 GDVEYCL.....
MYB23 G...YCL.....
MYB113 KGVTLPLDFEQIWARFDEETLELN...
MYB75 KGDTLAFDVDQLWSLFDGETVKFD...
MYB90 HGATLAFDVEQLWSLFDGETVELD...
MYB123 GNLDLNRPFTSCHHRGDDEDWLRDFTC
b
1 10 20 30 40 50
α1
CPC MFRSDKAEKMDKRR...RRQSKAKASC..SEEVSSIEWEAVKMSEEEEDLISRMYKLVGD
ETC1 ....MNTQRKSKHL...KTNPTIVASS..SEEVSSLEWEEIAMAQEEEDLICRMYKLVGE
ETC2 ...MDNTNRLRLRRGPSLRQTKFTRSRYDSEEVSSIEWEFISMTEQEEDLISRMYRLVGN
ETC3 ....MDNHRRTKQP...KTN.SIVTSS..SEEVSSLEWEVVNMSQEEEDLVSRMHKLVGD
TRY ...MDNTDR..RRR...RKQHKIALHD..SEEVSSIEWEFINMTEQEEDLIFRMYRLVGD
TCL1 ...MDNTNRLRRLH..CHKQPKFTHS...SQEVSSMKWEFINMTEQEEDLIFRMYRLVGD
TCL2 ...MDNTNRLRHLR..SRKQSKFTLGD..TAEVNSVKWEFINMTEQEEDLIFRMHRLVGD
60 70 80 90
α2 α3
CPC RWELIAGRIPGRTPEEIERYWLMKHGVVFANRRRDFFRK.....
ETC1 RWDLIAGRIPGRTAEEIERFWVMKNHRRSQLR.....
ETC2 RWDLIAGRVVGRKANEIERYWIMRNSDYFSHKRRRLNNSPFFSTSPLNLQENLKL.
ETC3 RWELIAGRIPGRTAGEIERFWVMKN.....
TRY RWDLIAGRVPGRQPEEIERYWIMRNSEGFADKRRQLHSSSHKHTKPHRPRFSIYPS
TCL1 RWDLIARRVVGREAKEIERYWIMRNCDYFSHK.....
TCL2 RWDLIAGRVVGREAKDIERYWIMRNCDHCSHKRRRVHKFYRFSISPP.....


a
b
c
d
e
f
A280(mAU)
mL
WER-EGL3+CPC
CPC-EGL3
WER-EGL3
CPC
WER
EGL3
CPC - EGL3
WER - EGL3
KDa
25.0
18.4
14.4

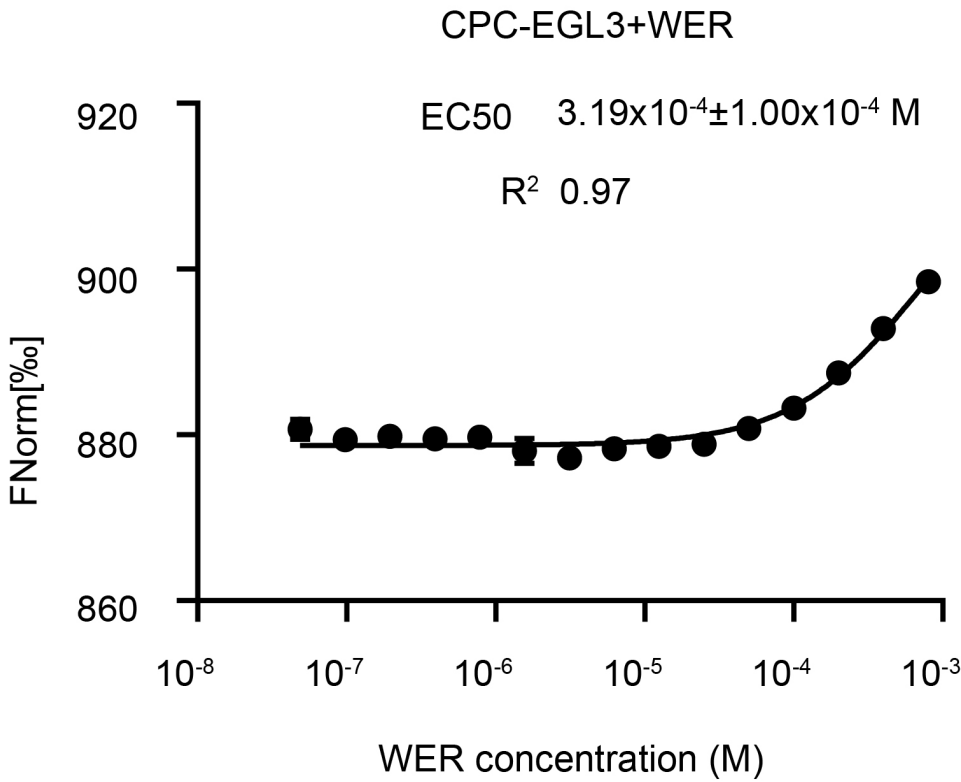

a
b
Time(min)
µcal/sec
Kcal mol-1 of injectant
CPC M49L
Kd 137.0nM
CPC M49A
Kd 1000.0nM
WER L85M
Kd 72.5nM
WER
Q77E
L85M
Kd 73.0nM
Molar Ratio

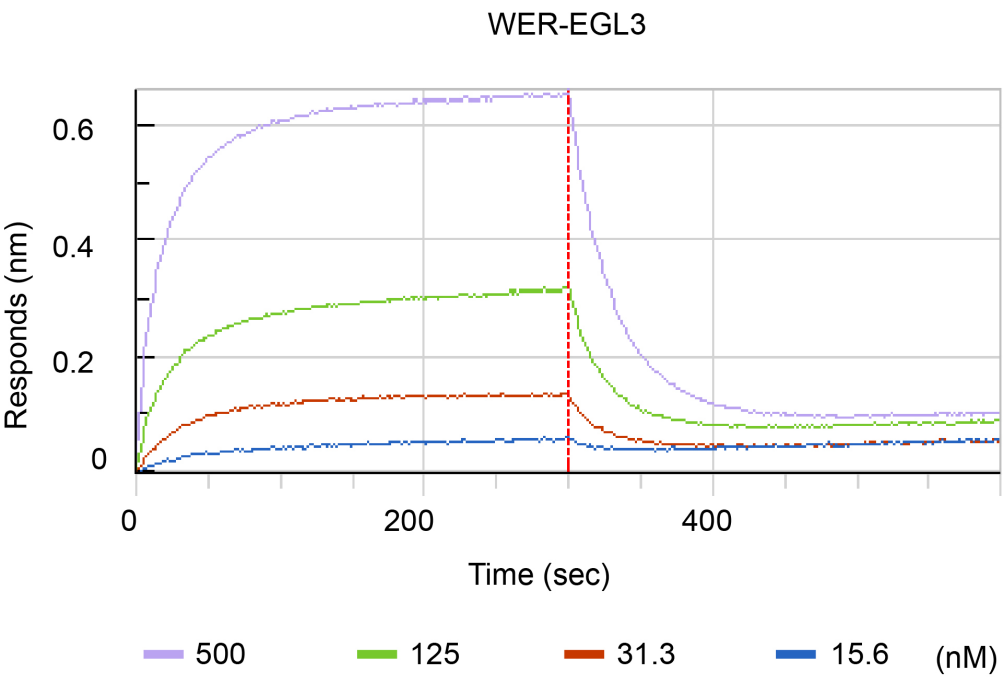

WER+EGL3
WER+EGL3+CPC
WER+EGL3+CPC M49L
WER+EGL3+CPC M49A
KDa
@YFP
35.0 -
CPC-YFP
@Myc
35.0 -
WER-Myc
@Flag
70.0 -
Flag-EGL3
@Actin
40.0 -
Actin

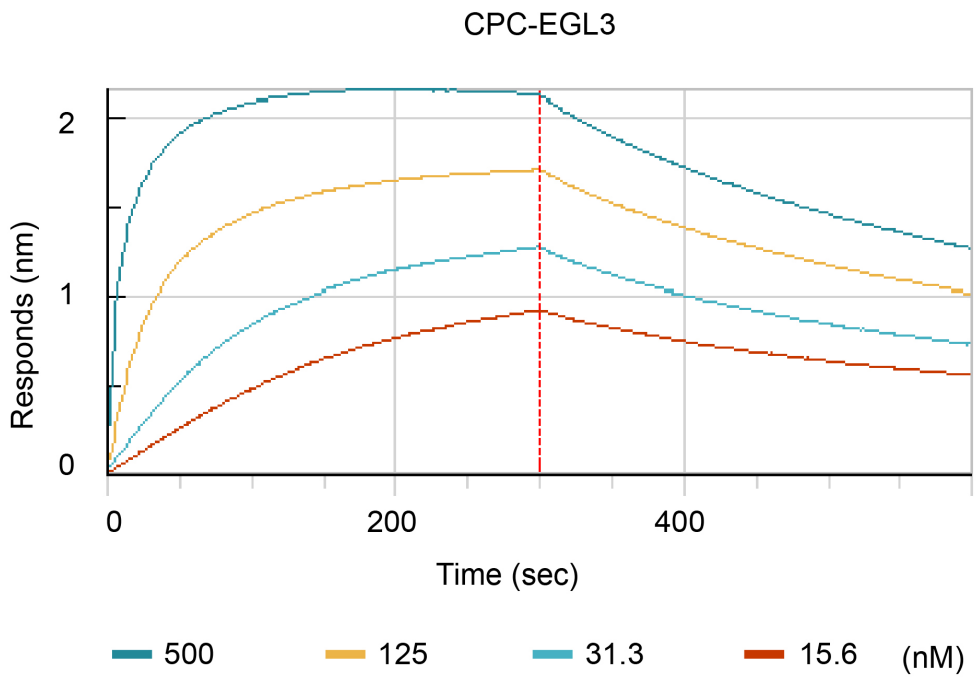
a


EGL3-nLUC
cLUC-CPC
WER-YFP
(1:1:1)
EGL3-nLUC
cLUC
WER-YFP
EGL3-nLUC
cLUC-CPC
nLUC
cLUC-CPC
WER-YFP
7095
1095

b


CPC-EGL3+WER
EC50 3.19x10-4±1.00x10-4 M
R2 0.97
920
900
880
860
FNorm[‰]
10-8
10-7
10-6
10-5
10-4
10-3
WER concentration (M)





Target	KD(M)	KD Error	kon(1/Ms)	kon Error	koff(1/s)	koff Error
WER	2.22E-07	4.86E-09	6.82E+04	1.26E+03	1.51E-02	1.76E-04
CPC	1.19E-08	1.39E-10	1.63E+05	1.16E+03	1.95E-03	1.80E-05