



HAL
open science

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

► **To cite this version:**

Baihui Wang, Qiang Luo, Yingping Li, Kangxi Du, Zhen Wu, et al.. Structural insights into partner selection for MYB and bHLH transcription factor complexes. *Nature Plants*, 2022, 10.1038/s41477-022-01223-w . hal-03760096

HAL Id: hal-03760096

<https://hal.science/hal-03760096v1>

Submitted on 8 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Title:**

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

4

5 **Authors:**

6 Baihui Wang^{1,†}, Qiang Luo^{1,†}, Yingping Li^{1,†}, Kangxi Du¹, Zhen Wu¹, Tianyang Li¹,
7 Wen-Hui Shen², Chien-Hsun Huang^{3*}, Jianhua Gan^{4*}, Aiwu Dong^{1,*}

8

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

10 *Correspondence: aiwudong@fudan.edu.cn, ganjhh@fudan.edu.cn and
11 chhuang1981@139.com

12

13 **Affiliations:**

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

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

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

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

23 Collaborative Innovation Center of Genetics and Development, Department of
24 Biochemistry and Biophysics, School of Life Sciences, Fudan University, Shanghai
25 200438, P.R. China
26

27 **Abstract:**

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

39

40 **Introduction**

41 By recognizing specific DNA elements within the genome, transcription factors (TFs)
42 play central roles in gene regulatory networks in multicellular organisms. In higher
43 plants, MYB and basic helix-loop-helix (bHLH) TFs are among the largest groups of
44 TFs; for example, more than 300 MYB and 100 bHLH TFs are present in Arabidopsis^{1,2}.
45 MYB TFs contain a conserved DNA-binding domain (MYB domain) and extensive
46 intrinsically disordered regions (IDRs) at the C-terminus. MYB TFs are grouped into
47 different types on the basis of three imperfect repeats (R1, R2, and R3) within the MYB
48 domain of c-Myb, and further divided into subgroups based on the IDRs³. bHLH TFs
49 are defined by ~60 conserved amino acids, including a basic DNA-binding domain and
50 a dimerization domain⁴.

51 Formation of complexes between MYB and bHLH TFs is widespread and different
52 MYB–bHLH complexes regulate diverse physiological processes, including organ
53 development, metabolic pathways, and biotic and abiotic stress responses^{1-3,5,6}. For
54 instance, WEREWOLF (WER), a well-studied R2R3-type MYB in Arabidopsis,
55 interacts with the bHLH TFs GL3/EGL3, directly binds to and activates *GLABRA 2*
56 (*GL2*), the central regulator of epidermal cell fate determination, and leads to the non-
57 hair cell fate during root hair development⁷. CAPRICE (CPC), a R3-type MYB TF in
58 Arabidopsis, competes with WER to interact with GL3/EGL3 and inhibits *GL2*
59 expression, leading to the hair cell fate establishment⁸⁻¹⁰. The Arabidopsis bHLH TFs
60 MYC2/MYC3/MYC4 interact with MYB21/MYB24 to regulate stamen development
61 and seed production¹¹. MYC2/MYC3/MYC4 play crucial roles in jasmonate (JA)

62 signaling through interaction with jasmonate ZIM (zinc-finger inflorescence
63 meristem)-domain (JAZ) repressors and participate in both development and stress
64 responses^{12,13}. MYC2/MYC3/MYC4 also regulate glucosinolate biosynthesis by
65 interacting with glucosinolate-related MYBs, including MYB28, MYB29, MYB76,
66 MYB34, MYB51, and MYB122, which belong to MYB subgroup 12¹⁴. In addition,
67 many other MYB–bHLH complexes function in plant growth and development, stress
68 defense, and metabolism regulation⁶.

69 The co-crystal structure of MYC3 in complex with JAZ transcriptional repressors
70 was resolved previously¹³, and recently we resolved the complex structure of WER–
71 DNA¹⁵. The structure of MYB–bHLH complexes remains uncertain, which precludes
72 understanding the molecular mechanisms that regulate the interaction and partner
73 selection of MYB and bHLH TFs. Here we report the crystal structures of three MYB–
74 bHLH complexes, WER–EGL3, CPC–EGL3, and MYB29–MYC3, which reveals two
75 distinct MYB–bHLH interacting modes. The WER–EGL3 and MYB29–MYC3
76 binding modes are widely applied among a set of MYB – bHLH complexes in
77 Arabidopsis. Our results shed light on the mechanisms by which different MYB and
78 bHLH TFs select their partners to form complexes and indicate that the two types of
79 MYB–bHLH interaction modes are evolutionarily conserved but evolve independently
80 in plants.

81

82 **Results**

83 **Co-crystal structure of WER–EGL3 complex**

84 Following our previous study on the crystal structure of WER in complex with its target
85 DNA¹⁵, we selected WER and EGL3 as a model to analyze the interaction of MYB and
86 bHLH TFs. WER contains a R2R3-MYB domain at the N-terminus. EGL3 contains a
87 JAZ-interacting domain (JID) and transcriptional activation domain (TAD) at the N-
88 terminus, a bHLH domain in the central region, and a C-terminal domain (CTD) at the
89 C-terminus (Extended Data Figs. 1a, b). Yeast two-hybrid (Y2H) and size-exclusion
90 chromatography (SEC) experiments showed that WER interacted with EGL3 through
91 the R3 domain of WER (WER 67–120) and the N-terminus of EGL3 (EGL3 1–205)
92 (Extended Data Figs. 1c, d). To investigate the molecular basis underlying the
93 interaction between WER and EGL3, we determined the crystal structure of WER 67–
94 120 in complex with EGL3 1–205 (hereafter the WER–EGL3 complex). The complex
95 structure was refined to 2.90 Å resolution (Table 1). The β -strands β 1– β 5 of EGL3 1–
96 205 formed a flat β -sheet in the center, flanked by helices α 1 and α 6 on one side and
97 by α 3, α 4, and α 5 on the opposite side (Fig. 1a). The EGL3 JID domain consisted of
98 helices α 1 and α 2, and β -strands β 1, β 2, and α 2 located at the C-terminus of JID.
99 Although TAD domains are generally unstructured when not bound to their targets^{16–18},
100 the EGL3 TAD was well ordered. The EGL3 TAD was composed of two helices, α 3
101 and α 4, which packed against the JID domain and β 3– β 5, respectively. The R3 domain
102 of WER was composed of three helices α 1– α 3 linked by two short loops; the α 1 and
103 α 2 helices of WER packed against the α 2, α 3, and α 5 helices of EGL3 (Fig. 1a).
104 Structural superposition of WER–EGL3 and WER–DNA¹⁵ suggested that WER–EGL3
105 complex formation has no effect on WER–DNA interaction (Fig. 1b), which is

106 consistent with their inherent DNA binding and gene activation activities.

107 The detailed interaction between WER and EGL3 was shown in Fig. 1c. The
108 electrostatic surface potentials of the interfaces (Extended Data Figs. 2a,b) and the
109 electron density maps of the key residues involved in the interaction (Extended Data
110 Fig. 3a) of WER-EGL3 were also shown. WER Gln77 at the N-terminus of α 1 formed
111 hydrogen bonds (H-bonds) with the side chain of Ser153 and the main chains of Phe156
112 and Leu160 of EGL3. WER Leu81 nestled in the shallow hydrophobic cavity formed
113 by Val120, Leu160, and Leu161 of EGL3. Via the guanidine group, WER Arg84 formed
114 two H-bonds with EGL3 Ser123. Interestingly, the guanidine group of WER Arg84 also
115 formed an ion- π interaction with the side chain of EGL3 Phe124. Similar to Leu81,
116 WER Leu85 and Leu88 formed hydrophobic interactions with EGL3. The side chain
117 of Leu85 pointed toward Trp116 and Val120 of EGL3, whereas the side chain of Leu88
118 nestled in the pocket formed by Tyr81, Leu84, and Leu119 of EGL3. WER Arg99 at
119 the C-terminus of α 2 (residues 92–99) formed a salt bridge with Asp113 and an ion- π
120 interaction with Trp116 of EGL3 (Fig. 1c). The crucial residues of WER involved in
121 the interaction with EGL3 exactly overlapped with the previously identified R/B-like
122 bHLH TF binding (RB) motif, which is responsible for interactions between MYB and
123 R/B-like bHLH TFs ¹⁹. *In vitro* isothermal titration calorimetry (ITC) analysis showed
124 that mutations of all WER–EGL3 interacting residues weakened the binding affinity
125 between WER and EGL3 (Figs. 1d, e, **Supplementary Table 1**). Together, these results
126 indicated that EGL3 physically interacts with the Q77X3L81X2R84L85X2L88X10R99
127 signature motif of the WER R3 domain.

128

129 **Structural basis of CPC competing with WER to bind to EGL3**

130 Y2H assay showed that, similar to WER, CPC interacted with EGL3 1–205 through the
131 R3 domain of CPC (Extended Data Fig. 4a). ITC analysis indicated that full-length CPC
132 (K_d : 48.3 nM) and CPC 30–94 (CPC R3, K_d : 50 nM) displayed similar binding affinities
133 to EGL3, thus CPC 30–94 was used in subsequent analyses (Extended Data Fig. 4b,
134 **Supplementary Table 1**). We determined the crystal structure of CPC 30–94 in complex
135 with EGL3 1–205 (hereafter the CPC–EGL3 complex) at 1.80 Å resolution (Fig. 2a,
136 Extended Data Figs. 2c,d and 3b, **Table 1**). Structural superposition revealed that the
137 overall structure of CPC–EGL3 was similar to that of WER–EGL3 (Fig. 2b), and the
138 root mean square deviation (RMSD) between CPC–EGL3 and WER–EGL3 was only
139 0.70 Å, based on the superposition of 177 pairs of C α atoms. The low RMSD value
140 indicates that CPC and WER share the same binding mode to EGL3. The detailed CPC–
141 EGL3 interactions are depicted in Fig. 2c. Except for Glu41 and Met49 within CPC, all
142 residues involved in the CPC–EGL3 interaction were identical to those observed in the
143 WER–EGL3 complex (Fig. 1c). Glu41 and Met49 of CPC corresponded to Gln77 and
144 Leu85 of WER, respectively. Given that CPC Glu41 mainly formed weak H-bond
145 interaction (3.0 Å) with the side chain of EGL3 Ser153, and it is variable (Asp, Glu or
146 Gln) in other R2R3-MYB and R3-MYB TFs (Extended Data Fig. 5), thus we mainly
147 focused on CPC Met49 and WER Leu85 for further analysis (Figs. 2d, e). The S δ atom
148 of CPC Met49 substitutes the C δ 2 atom of WER Leu85, forming stronger (3.4 Å vs 3.7
149 Å) hydrophobic interactions with the side-chain C γ 2 atom of EGL3 Val120. The C ϵ

150 atom of CPC Met49 stretches more closely (3.7 Å vs 4.5 Å for the C δ 2 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 $\sim 90^\circ$ 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 PDBePISA²⁰, CPC Met49 occupies
156 an interface of $\sim 67 \text{ \AA}^2$, which was $\sim 17 \text{ \AA}^2$ broader than that of WER Leu85 ($\sim 50 \text{ \AA}^2$).
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

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

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

206

207 **Co-crystal structure of MYB29–MYC3 complex**

208 Recently, split-ubiquitin assays identified a new MYC-interaction motif (MIM) within
209 the subgroup 12 MYBs. The motif is located in the center of the non-MYB region and
210 is responsible for interaction with MYC2/MYC3/MYC4⁶, suggesting that the MYB29–
211 MYC3 complex may represent a different interaction mode from that of WER–EGL3
212 and CPC–EGL3. To verify this assumption, we co-expressed the MIM motif of MYB29
213 (MYB29 174–222) and the N-terminal non-DNA-binding region of MYC3 (MYC3 44–
214 238). By extensive crystallization trials, we solved the crystal structure of MYB29 174–
215 222 in complex with MYC3 44–238 (hereafter MYB29–MYC3) at 2.5 Å resolution

216 (Extended Data Figs. 2e,f and 3c, **Table 1**). The MYB29 184–203 region is folded into
217 an ordered α helix and a short loop, whereas other regions are disordered (Fig. 4a).
218 MYB29 fit into the deep groove and formed extensive interactions with the TAD and
219 JID domains of MYC3. MYB29 Thr186 formed an H-bond with TAD Glu148 of MYC3.
220 MYB29 Leu190 formed extensive hydrophobic interactions with TAD Phe151 and
221 Leu152 of MYC3. Via the side-chain OD1 and ND2 atoms, MYB29 Asn191 formed
222 two H-bonds with the main chain N- and O-atoms of MYC3 JID Tyr97, respectively.
223 The side chain of MYB29 Ala194 fit into a shallow concavity, formed by JID Trp92 as
224 well as TAD Phe151 and Met155 of MYC3 (Fig. 4b). In the MYB29–MYC3 complex
225 structure, Leu190 and Asn191 of MYB29 formed the core platform for MYC3 binding,
226 which is consistent with previous findings that either Leu190 or Asn191 mutation
227 within MYB29 abolishes the interaction between MYB29 and MYC3, and the mutated
228 MYB29 fails to rescue *myb29* mutant phenotypes⁶. MST analysis further showed that
229 all mutations of the interacting residues within MYC3 weakened the interaction
230 between MYB29 and MYC3 (**Supplementary Fig. 1**). To our surprise, sequence
231 similarity between MYB29 and JAZ proteins was very low (**Supplementary Fig. 2**), but
232 the proteins displayed the same binding mode to MYC3 (Fig. 4c). In the structure of
233 the MYB29–MYC3 and MYC3–JAZ1¹³ complexes, the MYC3 JID helices flipped
234 outward and attached loosely to the main body of MYC3, and the helices of MYB29 or
235 JAZ proteins fit into the grooves formed by TAD and JID of MYC3. Together, these
236 results showed that MYB29 interacts with MYC3 by its C-terminal MYC-interaction
237 motif, different from the binding modes of WER–EGL3 and CPC–EGL3.

238

239 Discussion

240 Two MYB–bHLH binding modes are widely applied in Arabidopsis

241 Sequence alignment revealed that the MYB29-interacting residues within MYC3 were
242 conserved in EGL3 (Supplementary Fig. 3). However, MYB29 does not interact with
243 GL3/EGL3⁶. To investigate the basis for the two MYB–MYC binding modes, we
244 carefully compared the structures of EGL3 and MYC3. The overall structures of EGL3
245 and MYC3 were similar but their JID helices behaved differently (Supplementary Fig.
246 4). The conformation of the EGL3 JID helix was stable and formed extensive
247 interactions with TAD and the other regions of JID (Fig. 5a, Supplementary Fig. 4a-c).
248 Presence or absence of WER had no effect on the conformation of the EGL3 JID helix
249 (Supplementary Fig. 4d). In contrast, the conformation of the MYC3 JID helix was
250 highly dynamic¹³ and underwent substantial conformational change to expose TAD and
251 other regions of JID when binding to MYB29 or JAZ1 (Fig. 5a, Supplementary Fig. 4e,
252 f). In EGL3, the corresponding interface for adopting MYB29 was buried by the stable
253 JID helix, which prevented MYB29 binding to EGL3 (Fig. 5a, Supplementary Fig. 4g).
254 Except for certain residues within the MYC3 TAD, most residues responsible for
255 EGL3–WER interaction were not conserved in MYC3 (Supplementary Fig. 3),
256 preventing the MYC3 surface ($\alpha 4$ and $\alpha 7$) from binding to WER. These results
257 suggested that the JID helices of EGL3 and MYC3 played critical roles in partner
258 selection. In detail, the conformational change of the MYC3 JID helix provided the
259 interface for binding to the subgroup 12 MYB TFs or JAZ repressors. The rigidity of

260 the EGL3 JID helix prevented an interface forming, but the $\alpha 3$ and $\alpha 5$ helices of EGL3
261 formed a novel interface that specifically recognized the RB motif in the R3 domain of
262 MYB TFs. Thus, we uncovered two MYB–bHLH interaction modes: in the WER–
263 EGL3 mode, $\alpha 3$ and $\alpha 5$ of EGL3 formed an interface and specifically interacted with
264 the R3 domain of WER; in the MYB29–MYC3 mode, the MIM motif of MYB29 fit
265 into the groove formed by the TAD and JID domains of MYC3 (Fig. 5b).

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

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

292

293 **Two MYB–bHLH interaction modes evolved independently**

294 To investigate the occurrence of the two types of MYB–bHLH interaction modes during
295 the evolution of land plants, we searched the genomes of representative species across
296 the plant kingdom (Supplementary Tables 2-4) for the genes encoding homologs of
297 EGL3, WER, CPC, MYC3, and MYB29 and reconstructed their phylogenies
298 (Supplementary Figs. 6-10). The major results were summarized in Fig. 5d, where
299 homologs of EGL3 and MYC3 of the bHLH family showed divergence as early as from
300 mosses during evolution, whereas homologs of WER, CPC and MYB29 of the MYB
301 family showed divergence later from gymnosperms during evolution. For the WER–
302 EGL3 interaction mode, we found that the crucial motifs in EGL3 homologs were
303 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

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

336 Members applying the two MYB–bHLH interaction modes also evolved diverged
337 biological functions. EGL3 and WER/CPC homologs are mostly involved in
338 development of hairs (root hairs or trichomes) and biosynthesis of secondary
339 metabolites associated with antioxidation (anthocyanin and flavonol) (From TAIR on
340 www.arabidopsis.org) (Supplementary Table 5), implying that the WER–EGL3
341 interaction mode might be implemented in stress resistance, for example to cold, heat,
342 drought or insects, to adapt the varied environments in ancestral seed plants. For the
343 MYB29–MYC3 interaction mode, MYB28, MYB29 and MYB76 are functional in
344 regulation of aliphatic glucosinolate biosynthesis²²⁻³⁰, while MYB34, MYB51 and
345 MYB122 modulate indolic glucosinolate biosynthesis^{27, 30-32}. Glucosinolates are
346 primarily found in Brassicales and can be modified into toxins against insect herbivory.
347 It is possible that the MYB29–MYC3 interaction mode might be evolved to allow

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

351

352 **Methods**

353 **Yeast two-hybrid (Y2H) assay**

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

359 **Protein expression and purification**

360 The DNA fragments encoding WER 12-120, EGL3 1-205, full-length CPC, CPC 30-
361 94 and their mutants were generated by PCR and subcloned into pSUMO vector,
362 respectively (Supplementary Table 6). All the recombinated proteins were expressed in
363 *Escherichia coli* strain BL21 (DE3) cells and induced with 0.2 mM isopropyl β -D-
364 thiogalactoside (IPTG) when the OD₆₀₀ reached ~0.8. The induced cultures were grown
365 at 18°C for an additional 16-18 hours. For purification of WER and its mutants, cells
366 were enriched and suspended in Ni-NTA (GE Healthcare) binding buffer (500 mM
367 NaCl, 20 mM Tris-HCl and 25 mM imidazole, pH 8.0). The cells were lysed by high-
368 pressure cell disruptor and then centrifuged at 34,000 g for 1 hour. The supernatant was
369 loaded onto a Ni-NTA column. The target proteins were gradually eluted using elution

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

376 For co-purification of WER 67-120/EGL3 1-205, CPC 30-94/EGL3 1-205 and
377 MYB29 174-222/MYC3 44-238, the DNA fragments encoding WER 67-120, CPC 30-
378 96 and MYB29 174-222 were generated by PCR or RT-PCR and constructed into
379 pCDF-duet vector, respectively (Supplementary Table 6). The DNA fragment encoding
380 MYC3 44-238 was generated by RT-PCR and subcloned into pSUMO vector
381 (Supplementary Table 6). The plasmids expressing WER 67-120 and EGL3 1-205, CPC
382 30-94 and EGL3 1-205, or MYB29 174-222 and MYC3 44-238 were co-transferred
383 into *Escherichia coli* strain BL21 (DE3) cells. Protein expression was induced by 0.2
384 mM IPTC when the OD₆₀₀ reached 0.8. The cells were collected and the target
385 complexes were purified using the same procedures as that of EGL3 1-205.

386 **Size-exclusion chromatography (SEC) analysis**

387 EGL3 1-205, WER 12-120, and CPC 30-94 proteins were respectively dissolved in
388 SEC buffer (20 mM Tris-HCl, 150 mM NaCl, 5 mM DTT, pH 8.0). To prepare WER–
389 EGL3 and CPC–EGL3 complexes, EGL3 1-205 was mixed with equivalent molar of
390 WER 12-120 and CPC 30-94, respectively and the mixtures were incubated on ice for
391 30 min. For the competition assay, equivalent molar of CPC 30-94 was added into the

392 pre-prepared WER–EGL3 complex and incubated on ice for 30 min. The concentration
393 of EGL3 1-205, WER 12-120, and CPC 30-94 were fixed at 50 μ M in all samples,
394 which were sequentially analyzed by SuperdexTM increase 200 10/300 GL column.

395 **Crystallization and structure determination**

396 The apo EGL3 1-205 crystals were grown at 18°C using the sitting-drop vapor diffusion
397 method. The drop contains 0.2 μ l 25 mg/ml EGL3 1-205 protein and 0.2 μ l reservoir
398 solution composed of 100 mM CAPS, 200 mM lithium sulfate, and 2 M ammonium
399 sulfate, pH 10.5. All crystals of protein complexes were grown at 18°C by using the
400 hanging-drop vapor diffusion method. For crystallization of WER 67-120 and EGL3 1-
401 205 complex, the co-purified protein complex was concentrated to 20 mg/ml and the
402 crystals were grown in well solution containing 20% PEG3350, 200 mM magnesium
403 formate. The crystals of CPC 30-94 and EGL3 1-205 complex (20 mg/ml) were grown
404 in well solution (0.1 M Tris-HCl, 20% PEG3350, 0.2 M ammonium sulfate, pH 6.5).
405 For crystallization of MYB29 174-222 and MYC3 44-238 complex, the co-purified
406 complex was concentrated to 20 mg/ml and the crystals were grown in well solution
407 (0.1 M HEPES, 20% (w/v) PEG 6000, 0.2 M Sodium chloride, pH 7.0).

408 The X-ray data were collected at BL17U and BL18U beamlines at Shanghai
409 Synchrotron Radiation Facility (SSRF). Diffractive data were indexed, integrated and
410 scaled with HKL3000 program³³. The apo EGL3 structure and WER–EGL3 complex
411 were all solved by molecular replacement (MR) method using the Phaser program of
412 CCP4i³⁴, using the MYC3 structure (PDB_ID: 4RRU) as the search model. The CPC–
413 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 COOT³⁵ and PHENIX³⁶. 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,

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

447 **Microscale thermophoresis (MST) analysis**

448 MST experiments were performed according to the published methods³⁷. All proteins
449 were dialyzed into MST reaction buffer (20 mM HEPES and 150 mM NaCl, pH 7.5),
450 labelled and purified using the Protein labelling kit RED-NHS (Nanotemper, cat. no.
451 L001) and the recommended procedures. WER protein was labeled with cy5. 60 nM
452 EGL3 1-205 and 30 nM cy5-labelled WER were mixed in reaction buffer (20 mM
453 HEPES, 150 mM NaCl and 0.03% tween-20, pH 7.5) and incubated on ice for 30 min.
454 Then diluted CPC or mutants (from 0.4 mM to 1.22×10^{-5} mM) were added into the
455 reaction. For the MYB29-MYC3 interaction, Fluorescein amidites (FAM) labeled
456 MYB29 (184-205) was purchased from Scilight-Peptide company ([http://www.scilight-
457 peptide.com/](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 μ M 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 \times His-SUMO tagged
466 WER 12-120, 6 \times His-SUMO tagged CPC 30-94, EGL3 1-205 and 6 \times 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 \times 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 methods³⁸⁻³⁹.

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

480 Genes encoding homologous proteins of EGL3 (AT1G63650), WER (AT5G14750),
481 CPC (AT2G46410), MYC3 (AT5G46760) and MYB29 (AT5G07690) were searched
482 by all-against-all BLASTP with the E-value of $1e^{-5}$, and the identity $>20\%$ in
483 *Arabidopsis thaliana*. Sequences were aligned by MAFFT⁴⁰ with accurate aligning
484 options “- maxiterate 1000 - localpair”, adjusted manually with the use of AliView⁴¹.
485 Those candidate homologs were carefully examined for their interaction motifs. The
486 homologs with opposite chemical characteristics in the key residues with our query
487 proteins were filtered out. Due to the great variability, the Ala197 and Ile203 in
488 MYB29s homologs were not selected during this step.

489 **Phylogenetic reconstruction and logo comparison**

490 We selected a total of 23 publicly-available genomes (Supplementary Table 2)
491 representing major plant lineages including seven eudicots, two monocots, two of ANA
492 clade, two gymnosperms, three moniliophytes/lycophytes, three
493 mosses/liverworts/hornworts, and four algae for phylogenetic reconstruction and logo
494 comparison of key motifs. Genomes were retrieved from the Phytozome
495 (<https://phytozome.jgi.doe.gov/pz/portal.html>), NCBI (<https://www.ncbi.nlm.nih.gov>),
496 FernBase (<https://www.fernbase.org>), GIGADB datasets (<http://gigadb.org/>), Spruce
497 Genome Project (<http://congenie.org/>), MarpolBase (<https://marchantia.info>) and from
498 files in previous studies⁴²⁻⁴⁴. Genes encoding homologous proteins of EGL3, WER,
499 CPC, MYC3 and MYB29 were searched and aligned as above, and then trimmed by
500 trimAl⁴⁵ with -gt 0.1. We used RAxML to reconstruct the phylogenetic trees⁴⁶ with
501 default settings (-m PROTGAMMAJTT). We examined the phylogenetic trees,

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

511

512 **Data availability**

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

517

518 **Acknowledgements**

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

523 the Microscale thermophoresis System at the National Facility for Protein Science in
524 Shanghai (NFPS), Zhangjiang Lab, Shanghai Advanced Research Institute, Chinese
525 Academy of Science, China for providing technical support and assistance in data
526 collection and analysis. We thank Robert McKenzie, PhD, from Liwen Bianji (Edanz)
527 (www.liwenbianji.cn/ac), for editing the English text of a draft of this manuscript. This
528 work was supported by the National Natural Science Foundation of China
529 (NSFC31930017) for A.D. and the National Basic Research Program of China
530 (2012CB910500) for W-H.S.

531

532 **Author Contributions**

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

537

538 **Competing Interests**

539 The authors declare no competing interests.

540

541 **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.

548 **Figure Legends**

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

561

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

575

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

578 which is further detected by SDS-PAGE. For CPC competition assay, an equivalent
579 concentration of CPC was incubated with the prepared WER-EGL3 complex on ice for
580 30 min and analyzed by SEC experiment. The SEC fraction of WER-EGL3 plus CPC
581 was analyzed by SDS-PAGE, which was repeated independently twice with similar
582 results. **b**, Split-luciferase assay to detect the competition of wild-type and mutated CPC
583 against WER to interact with EGL3. Tobacco leaves were co-infiltrated with
584 Agrobacterium containing EGL3-nLUC and cLUC-WER with or without CPC-YFP,
585 and the luminescence images were captured by a CCD imaging system. **c**, Quantitative
586 measurement of the competitive capabilities of CPC and its mutants against WER to
587 form a complex with EGL3 by a microscale thermophoresis (MST) assay, and the data
588 were presented as mean values \pm SD of three independent experiments (n=3).

589

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

599

600 **Figure 5. Two interaction modes of MYB–bHLH complexes.** **a**, Structural
601 superposition of the WER–EGL3 and MYB29–MYC3 complexes. WER and MYB29
602 are colored in magenta and brown, respectively. Both EGL3 and MYC3 are colored in
603 grey, whereas their JID helixes are colored in cyan and pink, respectively. **b**, The
604 different interaction modes of the WER–EGL3 and MYB29–MYC3 complexes. **c**,
605 Sequence alignment of homologs of EGL3, WER, CPC, MYC3, and MYB29 in
606 Arabidopsis. For each panel, the uppermost sequence is the logo comparison, with the
607 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

615 **References:**

- 616 1. Feller, A., Machemer, K., Braun, E. L. & Grotewold, E., Evolutionary and comparative analysis of
617 MYB and bHLH plant transcription factors. *Plant J.* **66**, 94-116 (2011).
- 618 2. Pireyre, M. & Burow, M., Regulation of MYB and bHLH transcription factors: a glance at the protein
619 level. *Mol. Plant* **8**, 378-388 (2015).
- 620 3. Dubos, C. *et al.*, MYB transcription factors in Arabidopsis. *Trends Plant Sci.* **15**, 573-581 (2010).
- 621 4. Heim, M. A. *et al.*, The basic helix-loop-helix transcription factor family in plants: a genome-wide
622 study of protein structure and functional diversity. *Mol. Biol. Evol.* **20**, 735-747 (2003).
- 623 5. Toledo-Ortiz, G., Huq, E. & Quail, P. H., The Arabidopsis basic/helix-loop-helix transcription factor
624 family. *Plant Cell* **15**, 1749-1770 (2003).
- 625 6. Millard, P. S., Weber, K., Kragelund, B. B. & Burow, M., Specificity of MYB interactions relies on
626 motifs in ordered and disordered contexts. *Nucleic Acids Res.* **47**, 9592-9608 (2019).
- 627 7. Schiefelbein, J., Kwak, S. H., Wiecekowski, Y., Barron, C. & Bruex, A., The gene regulatory network
628 for root epidermal cell-type pattern formation in Arabidopsis. *J. Exp. Bot.* **60**, 1515-1521 (2009).
- 629 8. Song, S. K. *et al.*, Cell fate in the Arabidopsis root epidermis is determined by competition between
630 WEREWOLF and CAPRICE. *Plant Physiol.* **157**, 1196-1208 (2011).
- 631 9. Kang, Y. H., Song, S. K., Schiefelbein, J. & Lee, M. M., Nuclear trapping controls the position-
632 dependent localization of CAPRICE in the root epidermis of Arabidopsis. *Plant Physiol.* **163**, 193-
633 204 (2013).
- 634 10. Tominaga-Wada, R. & Wada, T., Regulation of root hair cell differentiation by R3 MYB transcription
635 factors in tomato and Arabidopsis. *Front. Plant Sci.* **5**, 91 (2014).
- 636 11. Qi, T., Huang, H., Song, S. & Xie, D., Regulation of jasmonate-mediated stamen development and
637 seed production by a bHLH-MYB complex in Arabidopsis. *Plant Cell* **27**, 1620-1633 (2015).
- 638 12. Fernandez-Calvo, P. *et al.*, The Arabidopsis bHLH transcription factors MYC3 and MYC4 are targets
639 of JAZ repressors and act additively with MYC2 in the activation of jasmonate responses. *Plant*
640 *Cell* **23**, 701-715 (2011).
- 641 13. Zhang, F. *et al.*, Structural basis of JAZ repression of MYC transcription factors in jasmonate
642 signalling. *Nature* **525**, 269-273 (2015).
- 643 14. Schweizer, F. *et al.*, Arabidopsis basic helix-loop-helix transcription factors MYC2, MYC3, and
644 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*

- 675 glucosinolate biosynthesis regulators. *Plant Physiol.* **148**, 2021-2049 (2008).
- 676 28. Sonderby, I. E., Burow, M., Rowe, H. C., Kliebenstein, D. J. & Halkier, B. A. A complex interplay
677 of three R2R3 MYB transcription factors determines the profile of aliphatic glucosinolates in
678 Arabidopsis. *Plant Physiol.* **153**, 348-363 (2010).
- 679 29. Li, Y. et al. Novel insights into the function of Arabidopsis R2R3-MYB transcription factors
680 regulating aliphatic glucosinolate biosynthesis. *Plant Cell Physiol.* **54**, 1335-1344 (2013).
- 681 30. Frerigmann, H. & Gigolashvili, T. MYB34, MYB51, and MYB122 distinctly regulate indolic
682 glucosinolate biosynthesis in Arabidopsis thaliana. *Mol. Plant.* **7**, 814-828 (2014).
- 683 31. Celenza, J. L. et al. The Arabidopsis ATR1 Myb transcription factor controls indolic glucosinolate
684 homeostasis. *Plant Physiol.* **137**, 253-262 (2005).
- 685 32. Gigolashvili, T. et al. The transcription factor HIG1/MYB51 regulates indolic glucosinolate
686 biosynthesis in Arabidopsis thaliana. *Plant J.* **50**, 886-901 (2007).
- 687 33. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M., HKL-3000: the integration of data
688 reduction and structure solution-from diffraction images to an initial model in minutes. *Acta*
689 *Crystallogr. D Biol. Crystallogr.* **62**, 859-866 (2006).
- 690 34. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E., A graphical user interface to the CCP4
691 program suite. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1131-1137 (2003).
- 692 35. Emsley, P. & Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr. D*
693 *Biol. Crystallogr.* **60**, 2126-2132 (2004).
- 694 36. Adams, P. D. et al., PHENIX: a comprehensive Python-based system for macromolecular structure
695 solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-221 (2010).
- 696 37. Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P. & Duhr, S., Molecular interaction
697 studies using microscale thermophoresis. *Assay Drug Dev. Technol.* **9**, 342-353 (2011).
- 698 38. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* **596**, 583-589
699 (2021).
- 700 39. Evans, Richard, et al. Protein complex prediction with AlphaFold-Multimer. Preprint at <https://www.biorxiv.org/content/10.1101/2021.10.04.463034v1>
701
- 702 40. Katoh, K. & Standley, D. M., MAFFT multiple sequence alignment software version 7:
703 improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772-780 (2013).
- 704 41. Larsson, A., AliView: a fast and lightweight alignment viewer and editor for large datasets.

- 705 *Bioinformatics* **30**, 3276-3278 (2014).
- 706 42. Hori, K. *et al.*, Klebsormidium flaccidum genome reveals primary factors for plant terrestrial
707 adaptation. *Nat. Commun.* **5**, 3978 (2014).
- 708 43. Nishiyama, T. *et al.*, The chara genome: secondary complexity and implications for plant
709 terrestrialization. *Cell* **174**, 448-464 (2018).
- 710 44. Wan, T. *et al.*, A genome for gnetophytes and early evolution of seed plants. *Nat. Plants* **4**, 82-89
711 (2018).
- 712 45. Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T., trimAl: a tool for automated alignment
713 trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009).
- 714 46. Stamatakis, A., RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
715 phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).

1 **Title:**

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

4

5 **Authors:**

6 Baihui Wang^{1,†}, Qiang Luo^{1,†}, Yingping Li^{1,†}, Kangxi Du¹, Zhen Wu¹, Tianyang Li¹,
7 Wen-Hui Shen², Chien-Hsun Huang^{3*}, Jianhua Gan^{4*}, Aiwu Dong^{1,*}

8

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

10 *Correspondence: aiwudong@fudan.edu.cn, ganjhh@fudan.edu.cn and
11 chhuang1981@139.com

12

13 **Affiliations:**

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

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

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

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

23 Collaborative Innovation Center of Genetics and Development, Department of
24 Biochemistry and Biophysics, School of Life Sciences, Fudan University, Shanghai
25 200438, P.R. China
26

27 **Abstract:**

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

39

40 **Introduction**

41 By recognizing specific DNA elements within the genome, transcription factors (TFs)
42 play central roles in gene regulatory networks in multicellular organisms. In higher
43 plants, MYB and basic helix-loop-helix (bHLH) TFs are among the largest groups of
44 TFs; for example, more than 300 MYB and 100 bHLH TFs are present in Arabidopsis^{1,2}.
45 MYB TFs contain a conserved DNA-binding domain (MYB domain) and extensive
46 intrinsically disordered regions (IDRs) at the C-terminus. MYB TFs are grouped into
47 different types on the basis of three imperfect repeats (R1, R2, and R3) within the MYB
48 domain of c-Myb, and further divided into subgroups based on the IDRs³. bHLH TFs
49 are defined by ~60 conserved amino acids, including a basic DNA-binding domain and
50 a dimerization domain⁴.

51 Formation of complexes between MYB and bHLH TFs is widespread and different
52 MYB–bHLH complexes regulate diverse physiological processes, including organ
53 development, metabolic pathways, and biotic and abiotic stress responses^{1-3,5,6}. For
54 instance, WEREWOLF (WER), a well-studied R2R3-type MYB in Arabidopsis,
55 interacts with the bHLH TFs GL3/EGL3, directly binds to and activates *GLABRA 2*
56 (*GL2*), the central regulator of epidermal cell fate determination, and leads to the non-
57 hair cell fate during root hair development⁷. CAPRICE (CPC), a R3-type MYB TF in
58 Arabidopsis, competes with WER to interact with GL3/EGL3 and inhibits *GL2*
59 expression, leading to the hair cell fate establishment⁸⁻¹⁰. The Arabidopsis bHLH TFs
60 MYC2/MYC3/MYC4 interact with MYB21/MYB24 to regulate stamen development
61 and seed production¹¹. MYC2/MYC3/MYC4 play crucial roles in jasmonate (JA)

62 signaling through interaction with jasmonate ZIM (zinc-finger inflorescence
63 meristem)-domain (JAZ) repressors and participate in both development and stress
64 responses^{12,13}. MYC2/MYC3/MYC4 also regulate glucosinolate biosynthesis by
65 interacting with glucosinolate-related MYBs, including MYB28, MYB29, MYB76,
66 MYB34, MYB51, and MYB122, which belong to MYB subgroup 12¹⁴. In addition,
67 many other MYB–bHLH complexes function in plant growth and development, stress
68 defense, and metabolism regulation⁶.

69 The co-crystal structure of MYC3 in complex with JAZ transcriptional repressors
70 was resolved previously¹³, and recently we resolved the complex structure of WER–
71 DNA¹⁵. The structure of MYB–bHLH complexes remains uncertain, which precludes
72 understanding the molecular mechanisms that regulate the interaction and partner
73 selection of MYB and bHLH TFs. Here we report the crystal structures of three MYB–
74 bHLH complexes, WER–EGL3, CPC–EGL3, and MYB29–MYC3, which reveals two
75 distinct MYB–bHLH interacting modes. The WER–EGL3 and MYB29–MYC3
76 binding modes are widely applied among a set of MYB – bHLH complexes in
77 Arabidopsis. Our results shed light on the mechanisms by which different MYB and
78 bHLH TFs select their partners to form complexes and indicate that the two types of
79 MYB–bHLH interaction modes are evolutionarily conserved but evolve independently
80 in plants.

81

82 **Results**

83 **Co-crystal structure of WER–EGL3 complex**

84 Following our previous study on the crystal structure of WER in complex with its target
85 DNA¹⁵, we selected WER and EGL3 as a model to analyze the interaction of MYB and
86 bHLH TFs. WER contains a R2R3-MYB domain at the N-terminus. EGL3 contains a
87 JAZ-interacting domain (JID) and transcriptional activation domain (TAD) at the N-
88 terminus, a bHLH domain in the central region, and a C-terminal domain (CTD) at the
89 C-terminus (Extended Data Figs. 1a, b). Yeast two-hybrid (Y2H) and size-exclusion
90 chromatography (SEC) experiments showed that WER interacted with EGL3 through
91 the R3 domain of WER (WER 67–120) and the N-terminus of EGL3 (EGL3 1–205)
92 (Extended Data Figs. 1c, d). To investigate the molecular basis underlying the
93 interaction between WER and EGL3, we determined the crystal structure of WER 67–
94 120 in complex with EGL3 1–205 (hereafter the WER–EGL3 complex). The complex
95 structure was refined to 2.90 Å resolution (Table 1). The β -strands β 1– β 5 of EGL3 1–
96 205 formed a flat β -sheet in the center, flanked by helices α 1 and α 6 on one side and
97 by α 3, α 4, and α 5 on the opposite side (Fig. 1a). The EGL3 JID domain consisted of
98 helices α 1 and α 2, and β -strands β 1, β 2, and α 2 located at the C-terminus of JID.
99 Although TAD domains are generally unstructured when not bound to their targets^{16–18},
100 the EGL3 TAD was well ordered. The EGL3 TAD was composed of two helices, α 3
101 and α 4, which packed against the JID domain and β 3– β 5, respectively. The R3 domain
102 of WER was composed of three helices α 1– α 3 linked by two short loops; the α 1 and
103 α 2 helices of WER packed against the α 2, α 3, and α 5 helices of EGL3 (Fig. 1a).
104 Structural superposition of WER–EGL3 and WER–DNA¹⁵ suggested that WER–EGL3
105 complex formation has no effect on WER–DNA interaction (Fig. 1b), which is

106 consistent with their inherent DNA binding and gene activation activities.

107 The detailed interaction between WER and EGL3 was shown in Fig. 1c. The
108 electrostatic surface potentials of the interfaces (Extended Data Figs. 2a,b) and the
109 electron density maps of the key residues involved in the interaction (Extended Data
110 Fig. 3a) of WER-EGL3 were also shown. WER Gln77 at the N-terminus of α 1 formed
111 hydrogen bonds (H-bonds) with the side chain of Ser153 and the main chains of Phe156
112 and Leu160 of EGL3. WER Leu81 nestled in the shallow hydrophobic cavity formed
113 by Val120, Leu160, and Leu161 of EGL3. Via the guanidine group, WER Arg84 formed
114 two H-bonds with EGL3 Ser123. Interestingly, the guanidine group of WER Arg84 also
115 formed an ion- π interaction with the side chain of EGL3 Phe124. Similar to Leu81,
116 WER Leu85 and Leu88 formed hydrophobic interactions with EGL3. The side chain
117 of Leu85 pointed toward Trp116 and Val120 of EGL3, whereas the side chain of Leu88
118 nestled in the pocket formed by Tyr81, Leu84, and Leu119 of EGL3. WER Arg99 at
119 the C-terminus of α 2 (residues 92–99) formed a salt bridge with Asp113 and an ion- π
120 interaction with Trp116 of EGL3 (Fig. 1c). The crucial residues of WER involved in
121 the interaction with EGL3 exactly overlapped with the previously identified R/B-like
122 bHLH TF binding (RB) motif, which is responsible for interactions between MYB and
123 R/B-like bHLH TFs ¹⁹. *In vitro* isothermal titration calorimetry (ITC) analysis showed
124 that mutations of all WER–EGL3 interacting residues weakened the binding affinity
125 between WER and EGL3 (Figs. 1d, e, Supplementary Table 1). Together, these results
126 indicated that EGL3 physically interacts with the Q77X3L81X2R84L85X2L88X10R99
127 signature motif of the WER R3 domain.

128

129 **Structural basis of CPC competing with WER to bind to EGL3**

130 Y2H assay showed that, similar to WER, CPC interacted with EGL3 1–205 through the
131 R3 domain of CPC (Extended Data Fig. 4a). ITC analysis indicated that full-length CPC
132 (K_d : 48.3 nM) and CPC 30–94 (CPC R3, K_d : 50 nM) displayed similar binding affinities
133 to EGL3, thus CPC 30–94 was used in subsequent analyses (Extended Data Fig. 4b,
134 Supplementary Table 1). We determined the crystal structure of CPC 30–94 in complex
135 with EGL3 1–205 (hereafter the CPC–EGL3 complex) at 1.80 Å resolution (Fig. 2a,
136 Extended Data Figs. 2c,d and 3b, Table 1). Structural superposition revealed that the
137 overall structure of CPC–EGL3 was similar to that of WER–EGL3 (Fig. 2b), and the
138 root mean square deviation (RMSD) between CPC–EGL3 and WER–EGL3 was only
139 0.70 Å, based on the superposition of 177 pairs of C α atoms. The low RMSD value
140 indicates that CPC and WER share the same binding mode to EGL3. The detailed CPC–
141 EGL3 interactions are depicted in Fig. 2c. Except for Glu41 and Met49 within CPC, all
142 residues involved in the CPC–EGL3 interaction were identical to those observed in the
143 WER–EGL3 complex (Fig. 1c). Glu41 and Met49 of CPC corresponded to Gln77 and
144 Leu85 of WER, respectively. Given that CPC Glu41 mainly formed weak H-bond
145 interaction (3.0 Å) with the side chain of EGL3 Ser153, and it is variable (Asp, Glu or
146 Gln) in other R2R3-MYB and R3-MYB TFs (Extended Data Fig. 5), thus we mainly
147 focused on CPC Met49 and WER Leu85 for further analysis (Figs. 2d, e). The S δ atom
148 of CPC Met49 substitutes the C δ 2 atom of WER Leu85, forming stronger (3.4 Å vs 3.7
149 Å) hydrophobic interactions with the side-chain C γ 2 atom of EGL3 Val120. The C ϵ

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 PDBePISA²⁰, CPC Met49 occupies
156 an interface of ~67 Å², which was ~17 Å² broader than that of WER Leu85 (~50 Å²).
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

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

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

206

207 **Co-crystal structure of MYB29–MYC3 complex**

208 Recently, split-ubiquitin assays identified a new MYC-interaction motif (MIM) within
209 the subgroup 12 MYBs. The motif is located in the center of the non-MYB region and
210 is responsible for interaction with MYC2/MYC3/MYC4⁶, suggesting that the MYB29–
211 MYC3 complex may represent a different interaction mode from that of WER–EGL3
212 and CPC–EGL3. To verify this assumption, we co-expressed the MIM motif of MYB29
213 (MYB29 174–222) and the N-terminal non-DNA-binding region of MYC3 (MYC3 44–
214 238). By extensive crystallization trials, we solved the crystal structure of MYB29 174–
215 222 in complex with MYC3 44–238 (hereafter MYB29–MYC3) at 2.5 Å resolution

216 (Extended Data Figs. 2e,f and 3c, Table 1). The MYB29 184–203 region is folded into
217 an ordered α helix and a short loop, whereas other regions are disordered (Fig. 4a).
218 MYB29 fit into the deep groove and formed extensive interactions with the TAD and
219 JID domains of MYC3. MYB29 Thr186 formed an H-bond with TAD Glu148 of MYC3.
220 MYB29 Leu190 formed extensive hydrophobic interactions with TAD Phe151 and
221 Leu152 of MYC3. Via the side-chain OD1 and ND2 atoms, MYB29 Asn191 formed
222 two H-bonds with the main chain N- and O-atoms of MYC3 JID Tyr97, respectively.
223 The side chain of MYB29 Ala194 fit into a shallow concavity, formed by JID Trp92 as
224 well as TAD Phe151 and Met155 of MYC3 (Fig. 4b). In the MYB29–MYC3 complex
225 structure, Leu190 and Asn191 of MYB29 formed the core platform for MYC3 binding,
226 which is consistent with previous findings that either Leu190 or Asn191 mutation
227 within MYB29 abolishes the interaction between MYB29 and MYC3, and the mutated
228 MYB29 fails to rescue *myb29* mutant phenotypes⁶. MST analysis further showed that
229 all mutations of the interacting residues within MYC3 weakened the interaction
230 between MYB29 and MYC3 (Supplementary Fig. 1). To our surprise, sequence
231 similarity between MYB29 and JAZ proteins was very low (Supplementary Fig. 2), but
232 the proteins displayed the same binding mode to MYC3 (Fig. 4c). In the structure of
233 the MYB29–MYC3 and MYC3–JAZ1¹³ complexes, the MYC3 JID helices flipped
234 outward and attached loosely to the main body of MYC3, and the helices of MYB29 or
235 JAZ proteins fit into the grooves formed by TAD and JID of MYC3. Together, these
236 results showed that MYB29 interacts with MYC3 by its C-terminal MYC-interaction
237 motif, different from the binding modes of WER–EGL3 and CPC–EGL3.

238

239 **Discussion**

240 **Two MYB–bHLH binding modes are widely applied in Arabidopsis**

241 Sequence alignment revealed that the MYB29-interacting residues within MYC3 were
242 conserved in EGL3 (Supplementary Fig. 3). However, MYB29 does not interact with
243 GL3/EGL3⁶. To investigate the basis for the two MYB–MYC binding modes, we
244 carefully compared the structures of EGL3 and MYC3. The overall structures of EGL3
245 and MYC3 were similar but their JID helices behaved differently (Supplementary Fig.
246 4). The conformation of the EGL3 JID helix was stable and formed extensive
247 interactions with TAD and the other regions of JID (Fig. 5a, Supplementary Fig. 4a-c).
248 Presence or absence of WER had no effect on the conformation of the EGL3 JID helix
249 (Supplementary Fig. 4d). In contrast, the conformation of the MYC3 JID helix was
250 highly dynamic¹³ and underwent substantial conformational change to expose TAD and
251 other regions of JID when binding to MYB29 or JAZ1 (Fig. 5a, Supplementary Fig. 4e,
252 f). In EGL3, the corresponding interface for adopting MYB29 was buried by the stable
253 JID helix, which prevented MYB29 binding to EGL3 (Fig. 5a, Supplementary Fig. 4g).
254 Except for certain residues within the MYC3 TAD, most residues responsible for
255 EGL3–WER interaction were not conserved in MYC3 (Supplementary Fig. 3),
256 preventing the MYC3 surface ($\alpha 4$ and $\alpha 7$) from binding to WER. These results
257 suggested that the JID helices of EGL3 and MYC3 played critical roles in partner
258 selection. In detail, the conformational change of the MYC3 JID helix provided the
259 interface for binding to the subgroup 12 MYB TFs or JAZ repressors. The rigidity of

260 the EGL3 JID helix prevented an interface forming, but the $\alpha 3$ and $\alpha 5$ helices of EGL3
261 formed a novel interface that specifically recognized the RB motif in the R3 domain of
262 MYB TFs. Thus, we uncovered two MYB–bHLH interaction modes: in the WER–
263 EGL3 mode, $\alpha 3$ and $\alpha 5$ of EGL3 formed an interface and specifically interacted with
264 the R3 domain of WER; in the MYB29–MYC3 mode, the MIM motif of MYB29 fit
265 into the groove formed by the TAD and JID domains of MYC3 (Fig. 5b).

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

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

292

293 **Two MYB–bHLH interaction modes evolved independently**

294 To investigate the occurrence of the two types of MYB–bHLH interaction modes during
295 the evolution of land plants, we searched the genomes of representative species across
296 the plant kingdom (Supplementary Tables 2-4) for the genes encoding homologs of
297 EGL3, WER, CPC, MYC3, and MYB29 and reconstructed their phylogenies
298 (Supplementary Figs. 6-10). The major results were summarized in Fig. 5d, where
299 homologs of EGL3 and MYC3 of the bHLH family showed divergence as early as from
300 mosses during evolution, whereas homologs of WER, CPC and MYB29 of the MYB
301 family showed divergence later from gymnosperms during evolution. For the WER–
302 EGL3 interaction mode, we found that the crucial motifs in EGL3 homologs were
303 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

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

336 Members applying the two MYB–bHLH interaction modes also evolved diverged
337 biological functions. EGL3 and WER/CPC homologs are mostly involved in
338 development of hairs (root hairs or trichomes) and biosynthesis of secondary
339 metabolites associated with antioxidation (anthocyanin and flavonol) (From TAIR on
340 www.arabidopsis.org) (Supplementary Table 5), implying that the WER–EGL3
341 interaction mode might be implemented in stress resistance, for example to cold, heat,
342 drought or insects, to adapt the varied environments in ancestral seed plants. For the
343 MYB29–MYC3 interaction mode, MYB28, MYB29 and MYB76 are functional in
344 regulation of aliphatic glucosinolate biosynthesis²²⁻³⁰, while MYB34, MYB51 and
345 MYB122 modulate indolic glucosinolate biosynthesis^{27, 30-32}. Glucosinolates are
346 primarily found in Brassicales and can be modified into toxins against insect herbivory.
347 It is possible that the MYB29–MYC3 interaction mode might be evolved to allow

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

351

352 **Methods**

353 **Yeast two-hybrid (Y2H) assay**

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

359 **Protein expression and purification**

360 The DNA fragments encoding WER 12-120, EGL3 1-205, full-length CPC, CPC 30-
361 94 and their mutants were generated by PCR and subcloned into pSUMO vector,
362 respectively (Supplementary Table 6). All the recombinated proteins were expressed in
363 *Escherichia coli* strain BL21 (DE3) cells and induced with 0.2 mM isopropyl β -D-
364 thiogalactoside (IPTG) when the OD₆₀₀ reached ~0.8. The induced cultures were grown
365 at 18°C for an additional 16-18 hours. For purification of WER and its mutants, cells
366 were enriched and suspended in Ni-NTA (GE Healthcare) binding buffer (500 mM
367 NaCl, 20 mM Tris-HCl and 25 mM imidazole, pH 8.0). The cells were lysed by high-
368 pressure cell disruptor and then centrifuged at 34,000 g for 1 hour. The supernatant was
369 loaded onto a Ni-NTA column. The target proteins were gradually eluted using elution

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

376 For co-purification of WER 67-120/EGL3 1-205, CPC 30-94/EGL3 1-205 and
377 MYB29 174-222/MYC3 44-238, the DNA fragments encoding WER 67-120, CPC 30-
378 96 and MYB29 174-222 were generated by PCR or RT-PCR and constructed into
379 pCDF-duet vector, respectively (Supplementary Table 6). The DNA fragment encoding
380 MYC3 44-238 was generated by RT-PCR and subcloned into pSUMO vector
381 (Supplementary Table 6). The plasmids expressing WER 67-120 and EGL3 1-205, CPC
382 30-94 and EGL3 1-205, or MYB29 174-222 and MYC3 44-238 were co-transferred
383 into *Escherichia coli* strain BL21 (DE3) cells. Protein expression was induced by 0.2
384 mM IPTC when the OD₆₀₀ reached 0.8. The cells were collected and the target
385 complexes were purified using the same procedures as that of EGL3 1-205.

386 **Size-exclusion chromatography (SEC) analysis**

387 EGL3 1-205, WER 12-120, and CPC 30-94 proteins were respectively dissolved in
388 SEC buffer (20 mM Tris-HCl, 150 mM NaCl, 5 mM DTT, pH 8.0). To prepare WER-
389 EGL3 and CPC-EGL3 complexes, EGL3 1-205 was mixed with equivalent molar of
390 WER 12-120 and CPC 30-94, respectively and the mixtures were incubated on ice for
391 30 min. For the competition assay, equivalent molar of CPC 30-94 was added into the

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

395 **Crystallization and structure determination**

396 The apo EGL3 1-205 crystals were grown at 18°C using the sitting-drop vapor diffusion
397 method. The drop contains 0.2 μ l 25 mg/ml EGL3 1-205 protein and 0.2 μ l reservoir
398 solution composed of 100 mM CAPS, 200 mM lithium sulfate, and 2 M ammonium
399 sulfate, pH 10.5. All crystals of protein complexes were grown at 18°C by using the
400 hanging-drop vapor diffusion method. For crystallization of WER 67-120 and EGL3 1-
401 205 complex, the co-purified protein complex was concentrated to 20 mg/ml and the
402 crystals were grown in well solution containing 20% PEG3350, 200 mM magnesium
403 formate. The crystals of CPC 30-94 and EGL3 1-205 complex (20 mg/ml) were grown
404 in well solution (0.1 M Tris-HCl, 20% PEG3350, 0.2 M ammonium sulfate, pH 6.5).
405 For crystallization of MYB29 174-222 and MYC3 44-238 complex, the co-purified
406 complex was concentrated to 20 mg/ml and the crystals were grown in well solution
407 (0.1 M HEPES, 20% (w/v) PEG 6000, 0.2 M Sodium chloride, pH 7.0).

408 The X-ray data were collected at BL17U and BL18U beamlines at Shanghai
409 Synchrotron Radiation Facility (SSRF). Diffractive data were indexed, integrated and
410 scaled with HKL3000 program³³. The apo EGL3 structure and WER–EGL3 complex
411 were all solved by molecular replacement (MR) method using the Phaser program of
412 CCP4i³⁴, using the MYC3 structure (PDB_ID: 4RRU) as the search model. The CPC–
413 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 COOT³⁵ and PHENIX³⁶. 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,

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

447 **Microscale thermophoresis (MST) analysis**

448 MST experiments were performed according to the published methods³⁷. All proteins
449 were dialyzed into MST reaction buffer (20 mM HEPES and 150 mM NaCl, pH 7.5),
450 labelled and purified using the Protein labelling kit RED-NHS (Nanotemper, cat. no.
451 L001) and the recommended procedures. WER protein was labeled with cy5. 60 nM
452 EGL3 1-205 and 30 nM cy5-labelled WER were mixed in reaction buffer (20 mM
453 HEPES, 150 mM NaCl and 0.03% tween-20, pH 7.5) and incubated on ice for 30 min.
454 Then diluted CPC or mutants (from 0.4 mM to 1.22×10^{-5} mM) were added into the
455 reaction. For the MYB29-MYC3 interaction, Fluorescein amidites (FAM) labeled
456 MYB29 (184-205) was purchased from Scilight-Peptide company ([http://www.scilight-
457 peptide.com/](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 μ M 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 \times His-SUMO tagged
466 WER 12-120, 6 \times His-SUMO tagged CPC 30-94, EGL3 1-205 and 6 \times 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 \times 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 methods³⁸⁻³⁹.

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

480 Genes encoding homologous proteins of EGL3 (AT1G63650), WER (AT5G14750),
481 CPC (AT2G46410), MYC3 (AT5G46760) and MYB29 (AT5G07690) were searched
482 by all-against-all BLASTP with the E-value of $1e^{-5}$, and the identity $>20\%$ in
483 *Arabidopsis thaliana*. Sequences were aligned by MAFFT⁴⁰ with accurate aligning
484 options “- maxiterate 1000 - localpair”, adjusted manually with the use of AliView⁴¹.
485 Those candidate homologs were carefully examined for their interaction motifs. The
486 homologs with opposite chemical characteristics in the key residues with our query
487 proteins were filtered out. Due to the great variability, the Ala197 and Ile203 in
488 MYB29s homologs were not selected during this step.

489 **Phylogenetic reconstruction and logo comparison**

490 We selected a total of 23 publicly-available genomes (Supplementary Table 2)
491 representing major plant lineages including seven eudicots, two monocots, two of ANA
492 clade, two gymnosperms, three moniliophytes/lycophytes, three
493 mosses/liverworts/hornworts, and four algae for phylogenetic reconstruction and logo
494 comparison of key motifs. Genomes were retrieved from the Phytozome
495 (<https://phytozome.jgi.doe.gov/pz/portal.html>), NCBI (<https://www.ncbi.nlm.nih.gov>),
496 FernBase (<https://www.fernbase.org>), GIGADB datasets (<http://gigadb.org/>), Spruce
497 Genome Project (<http://congenie.org/>), MarpolBase (<https://marchantia.info>) and from
498 files in previous studies⁴²⁻⁴⁴. Genes encoding homologous proteins of EGL3, WER,
499 CPC, MYC3 and MYB29 were searched and aligned as above, and then trimmed by
500 trimAl⁴⁵ with -gt 0.1. We used RAxML to reconstruct the phylogenetic trees⁴⁶ with
501 default settings (-m PROTGAMMAJTT). We examined the phylogenetic trees,

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

511

512 **Data availability**

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

517

518 **Acknowledgements**

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

523 the Microscale thermophoresis System at the National Facility for Protein Science in
524 Shanghai (NFPS), Zhangjiang Lab, Shanghai Advanced Research Institute, Chinese
525 Academy of Science, China for providing technical support and assistance in data
526 collection and analysis. We thank Robert McKenzie, PhD, from Liwen Bianji (Edanz)
527 (www.liwenbianji.cn/ac), for editing the English text of a draft of this manuscript. This
528 work was supported by the National Natural Science Foundation of China
529 (NSFC31930017) for A.D. and the National Basic Research Program of China
530 (2012CB910500) for W-H.S..

531

532 **Author Contributions**

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

537

538 **Competing Interests**

539 The authors declare no competing interests.

540

541 **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

548 **Figure Legends**

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

561

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

575

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

578 which is further detected by SDS-PAGE. For CPC competition assay, an equivalent
579 concentration of CPC was incubated with the prepared WER-EGL3 complex on ice for
580 30 min and analyzed by SEC experiment. The SEC fraction of WER-EGL3 plus CPC
581 was analyzed by SDS-PAGE, which was repeated independently twice with similar
582 results. **b**, Split-luciferase assay to detect the competition of wild-type and mutated CPC
583 against WER to interact with EGL3. Tobacco leaves were co-infiltrated with
584 Agrobacterium containing EGL3-nLUC and cLUC-WER with or without CPC-YFP,
585 and the luminescence images were captured by a CCD imaging system. **c**, Quantitative
586 measurement of the competitive capabilities of CPC and its mutants against WER to
587 form a complex with EGL3 by a microscale thermophoresis (MST) assay, and the data
588 were presented as mean values \pm SD of three independent experiments (n=3).

589

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

599

600 **Figure 5. Two interaction modes of MYB–bHLH complexes.** **a**, Structural
601 superposition of the WER–EGL3 and MYB29–MYC3 complexes. WER and MYB29
602 are colored in magenta and brown, respectively. Both EGL3 and MYC3 are colored in
603 grey, whereas their JID helixes are colored in cyan and pink, respectively. **b**, The
604 different interaction modes of the WER–EGL3 and MYB29–MYC3 complexes. **c**,
605 Sequence alignment of homologs of EGL3, WER, CPC, MYC3, and MYB29 in
606 Arabidopsis. For each panel, the uppermost sequence is the logo comparison, with the
607 *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

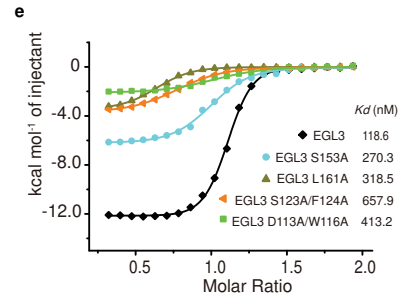
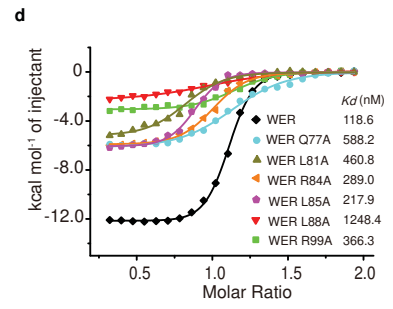
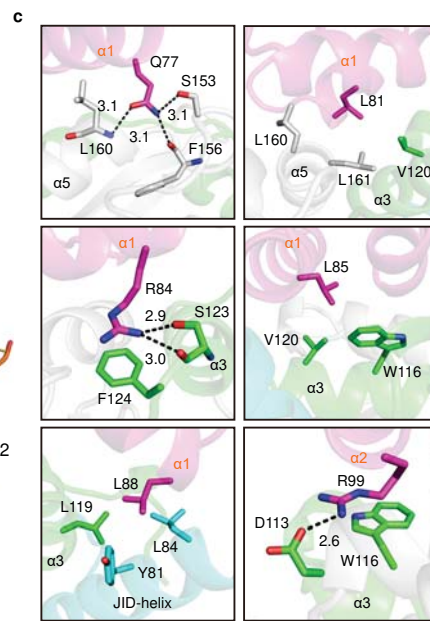
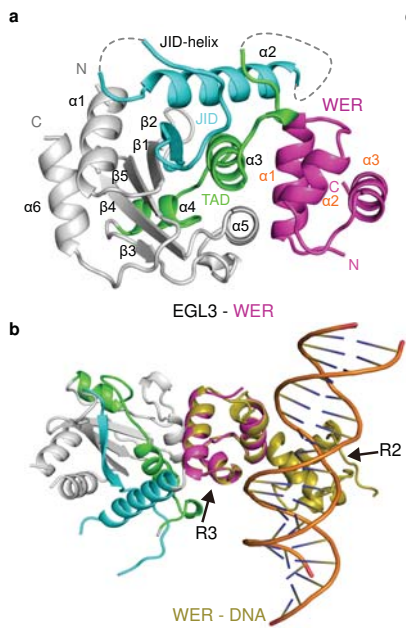
615 **References:**

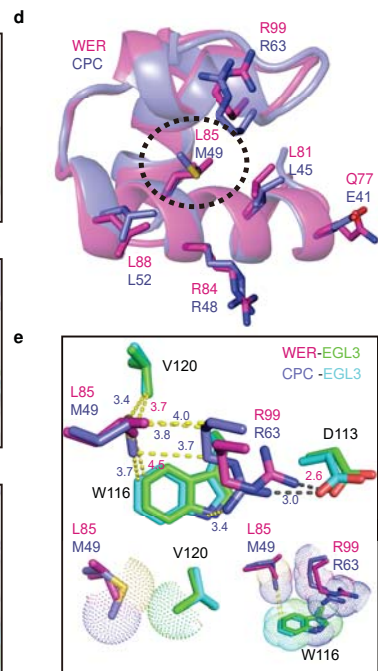
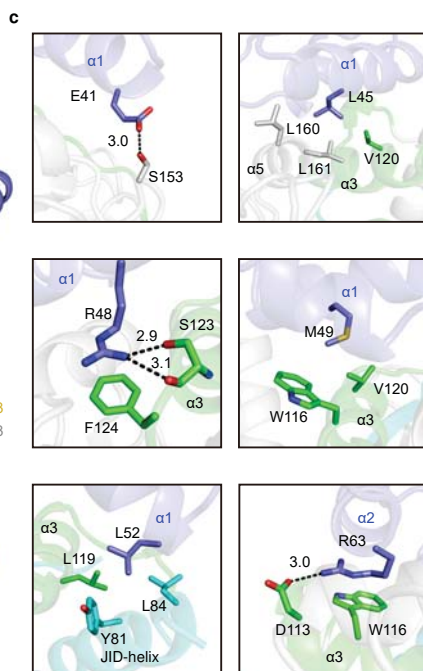
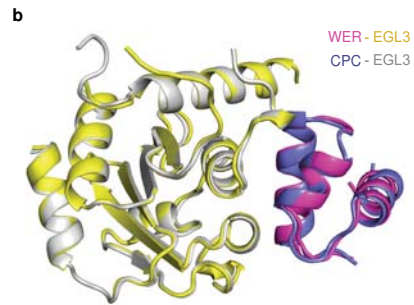
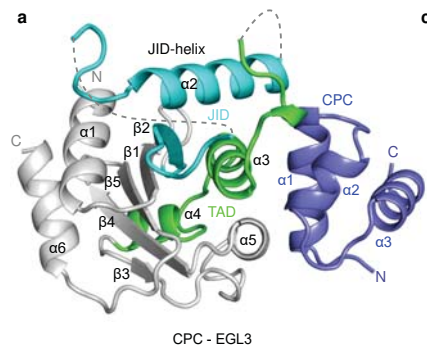
- 616 1. Feller, A., Machemer, K., Braun, E. L. & Grotewold, E., Evolutionary and comparative analysis of
617 MYB and bHLH plant transcription factors. *Plant J.* **66**, 94-116 (2011).
- 618 2. Pireyre, M. & Burow, M., Regulation of MYB and bHLH transcription factors: a glance at the protein
619 level. *Mol. Plant* **8**, 378-388 (2015).
- 620 3. Dubos, C. *et al.*, MYB transcription factors in Arabidopsis. *Trends Plant Sci.* **15**, 573-581 (2010).
- 621 4. Heim, M. A. *et al.*, The basic helix-loop-helix transcription factor family in plants: a genome-wide
622 study of protein structure and functional diversity. *Mol. Biol. Evol.* **20**, 735-747 (2003).
- 623 5. Toledo-Ortiz, G., Huq, E. & Quail, P. H., The Arabidopsis basic/helix-loop-helix transcription factor
624 family. *Plant Cell* **15**, 1749-1770 (2003).
- 625 6. Millard, P. S., Weber, K., Kragelund, B. B. & Burow, M., Specificity of MYB interactions relies on
626 motifs in ordered and disordered contexts. *Nucleic Acids Res.* **47**, 9592-9608 (2019).
- 627 7. Schiefelbein, J., Kwak, S. H., Wiecekowski, Y., Barron, C. & Bruex, A., The gene regulatory network
628 for root epidermal cell-type pattern formation in Arabidopsis. *J. Exp. Bot.* **60**, 1515-1521 (2009).
- 629 8. Song, S. K. *et al.*, Cell fate in the Arabidopsis root epidermis is determined by competition between
630 WEREWOLF and CAPRICE. *Plant Physiol.* **157**, 1196-1208 (2011).
- 631 9. Kang, Y. H., Song, S. K., Schiefelbein, J. & Lee, M. M., Nuclear trapping controls the position-
632 dependent localization of CAPRICE in the root epidermis of Arabidopsis. *Plant Physiol.* **163**, 193-
633 204 (2013).
- 634 10. Tominaga-Wada, R. & Wada, T., Regulation of root hair cell differentiation by R3 MYB transcription
635 factors in tomato and Arabidopsis. *Front. Plant Sci.* **5**, 91 (2014).
- 636 11. Qi, T., Huang, H., Song, S. & Xie, D., Regulation of jasmonate-mediated stamen development and
637 seed production by a bHLH-MYB complex in Arabidopsis. *Plant Cell* **27**, 1620-1633 (2015).
- 638 12. Fernandez-Calvo, P. *et al.*, The Arabidopsis bHLH transcription factors MYC3 and MYC4 are targets
639 of JAZ repressors and act additively with MYC2 in the activation of jasmonate responses. *Plant*
640 *Cell* **23**, 701-715 (2011).
- 641 13. Zhang, F. *et al.*, Structural basis of JAZ repression of MYC transcription factors in jasmonate
642 signalling. *Nature* **525**, 269-273 (2015).
- 643 14. Schweizer, F. *et al.*, Arabidopsis basic helix-loop-helix transcription factors MYC2, MYC3, and
644 MYC4 regulate glucosinolate biosynthesis, insect performance, and feeding behavior. *Plant Cell*

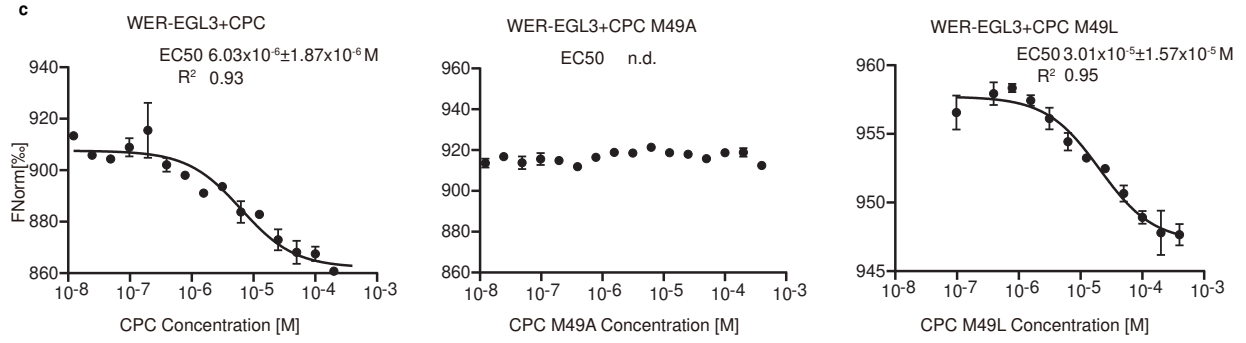
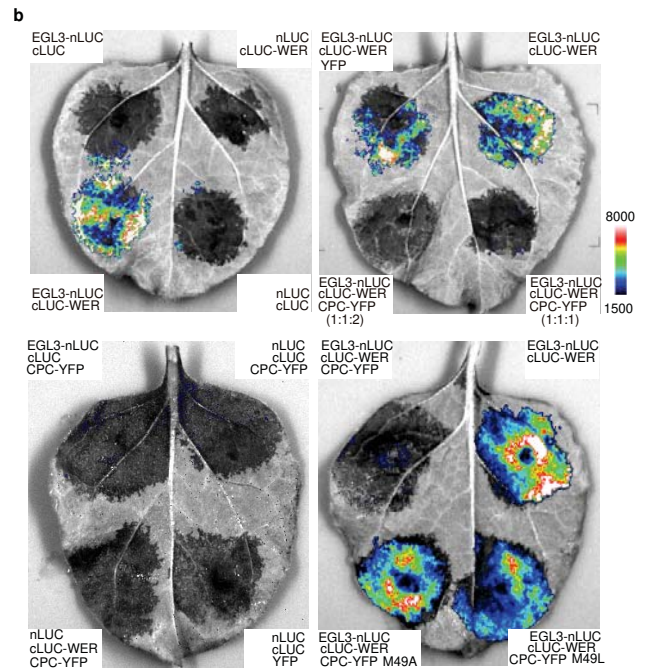
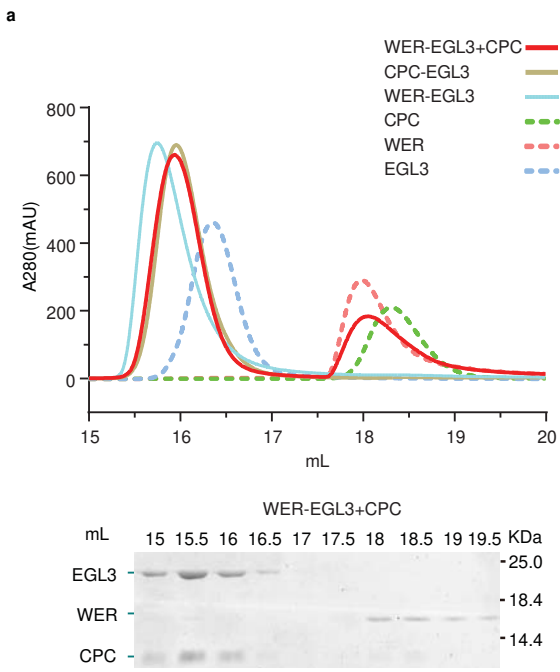
- 645 **25**, 3117-3132 (2013).
- 646 15. Wang, B. *et al.*, Structural insights into target DNA recognition by R2R3-MYB transcription factors.
647 *Nucleic Acids Res.* **48**, 460-471 (2020).
- 648 16. Triezenberg, S. J., Structure and function of transcriptional activation domains. *Curr. Opin. Genet.*
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*

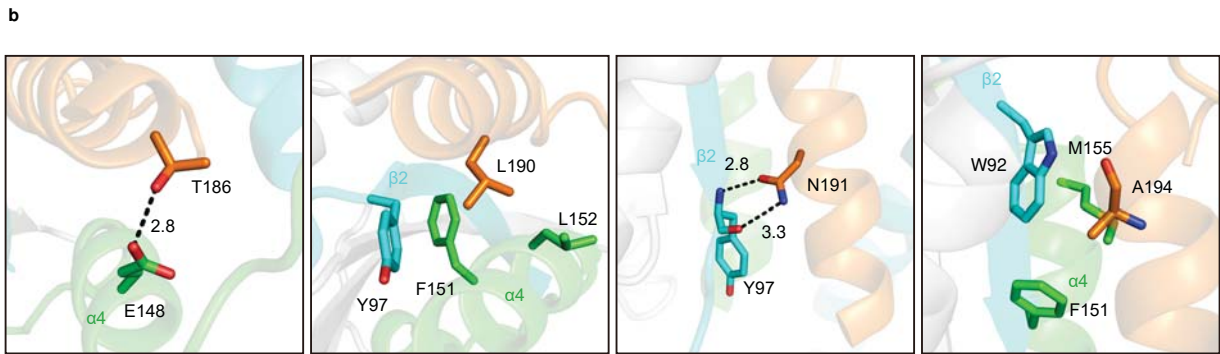
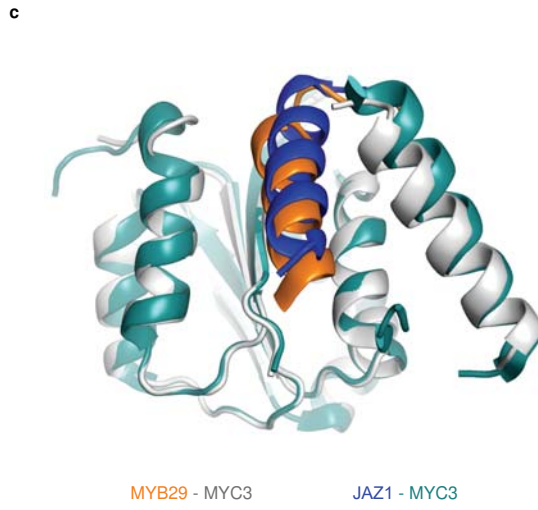
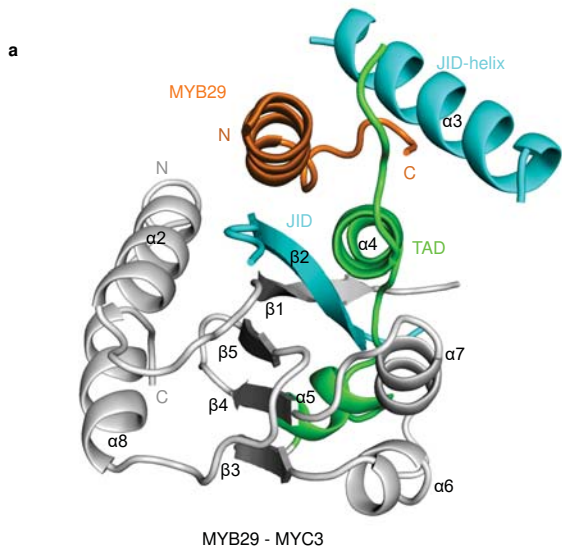
- 675 glucosinolate biosynthesis regulators. *Plant Physiol.* **148**, 2021-2049 (2008).
- 676 28. Sonderby, I. E., Burow, M., Rowe, H. C., Kliebenstein, D. J. & Halkier, B. A. A complex interplay
677 of three R2R3 MYB transcription factors determines the profile of aliphatic glucosinolates in
678 Arabidopsis. *Plant Physiol.* **153**, 348-363 (2010).
- 679 29. Li, Y. et al. Novel insights into the function of Arabidopsis R2R3-MYB transcription factors
680 regulating aliphatic glucosinolate biosynthesis. *Plant Cell Physiol.* **54**, 1335-1344 (2013).
- 681 30. Frerigmann, H. & Gigolashvili, T. MYB34, MYB51, and MYB122 distinctly regulate indolic
682 glucosinolate biosynthesis in Arabidopsis thaliana. *Mol. Plant.* **7**, 814-828 (2014).
- 683 31. Celenza, J. L. et al. The Arabidopsis ATR1 Myb transcription factor controls indolic glucosinolate
684 homeostasis. *Plant Physiol.* **137**, 253-262 (2005).
- 685 32. Gigolashvili, T. et al. The transcription factor HIG1/MYB51 regulates indolic glucosinolate
686 biosynthesis in Arabidopsis thaliana. *Plant J.* **50**, 886-901 (2007).
- 687 33. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M., HKL-3000: the integration of data
688 reduction and structure solution-from diffraction images to an initial model in minutes. *Acta*
689 *Crystallogr. D Biol. Crystallogr.* **62**, 859-866 (2006).
- 690 34. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E., A graphical user interface to the CCP4
691 program suite. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1131-1137 (2003).
- 692 35. Emsley, P. & Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr. D*
693 *Biol. Crystallogr.* **60**, 2126-2132 (2004).
- 694 36. Adams, P. D. et al., PHENIX: a comprehensive Python-based system for macromolecular structure
695 solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-221 (2010).
- 696 37. Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P. & Duhr, S., Molecular interaction
697 studies using microscale thermophoresis. *Assay Drug Dev. Technol.* **9**, 342-353 (2011).
- 698 38. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* **596**, 583-589
699 (2021).
- 700 39. Evans, Richard, et al. Protein complex prediction with AlphaFold-Multimer. Preprint at <https://www.biorxiv.org/content/10.1101/2021.10.04.463034v1>
701
- 702 40. Katoh, K. & Standley, D. M., MAFFT multiple sequence alignment software version 7:
703 improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772-780 (2013).
- 704 41. Larsson, A., AliView: a fast and lightweight alignment viewer and editor for large datasets.

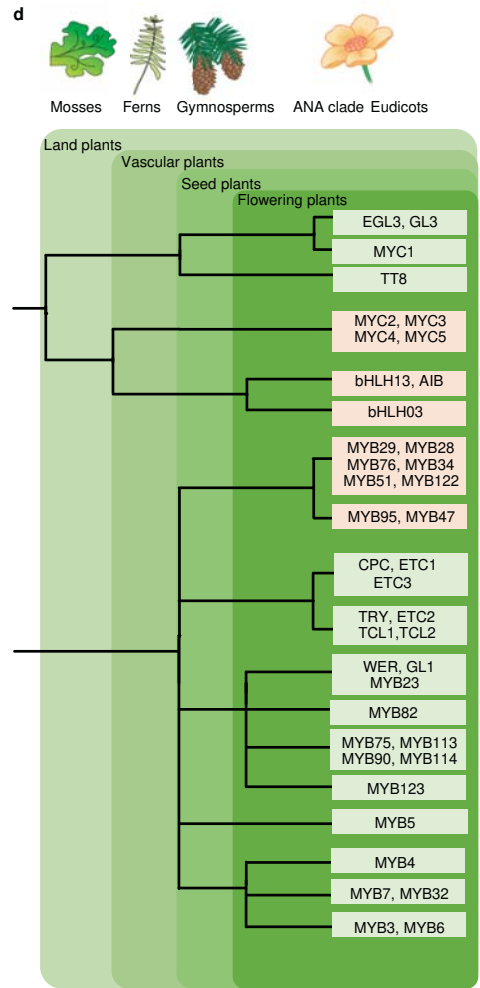
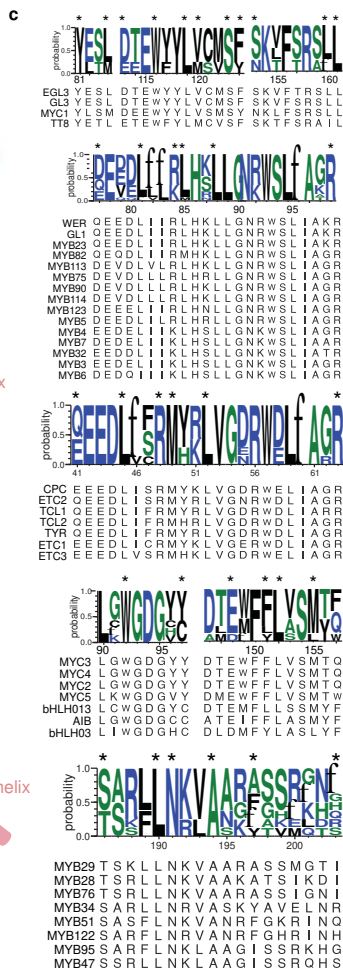
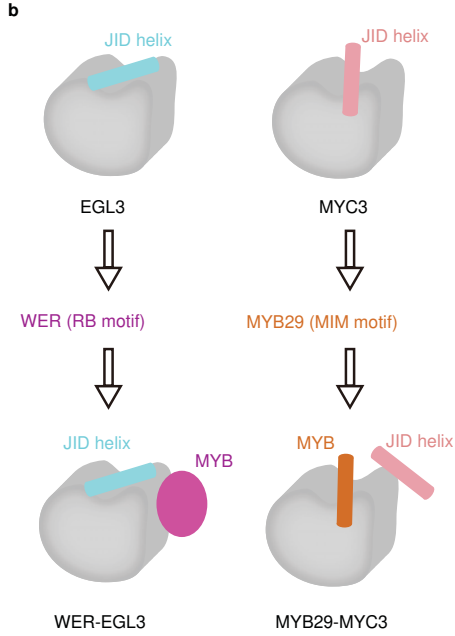
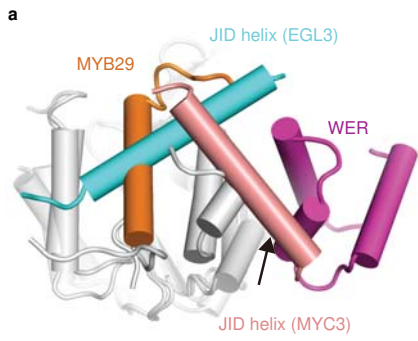
- 705 *Bioinformatics* **30**, 3276-3278 (2014).
- 706 42. Hori, K. *et al.*, Klebsormidium flaccidum genome reveals primary factors for plant terrestrial
707 adaptation. *Nat. Commun.* **5**, 3978 (2014).
- 708 43. Nishiyama, T. *et al.*, The chara genome: secondary complexity and implications for plant
709 terrestrialization. *Cell* **174**, 448-464 (2018).
- 710 44. Wan, T. *et al.*, A genome for gnetophytes and early evolution of seed plants. *Nat. Plants* **4**, 82-89
711 (2018).
- 712 45. Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T., trimAl: a tool for automated alignment
713 trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009).
- 714 46. Stamatakis, A., RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
715 phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).

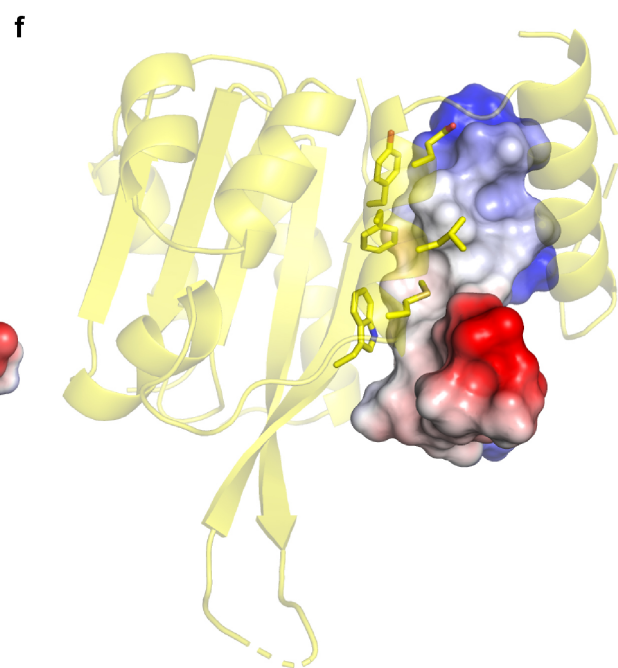
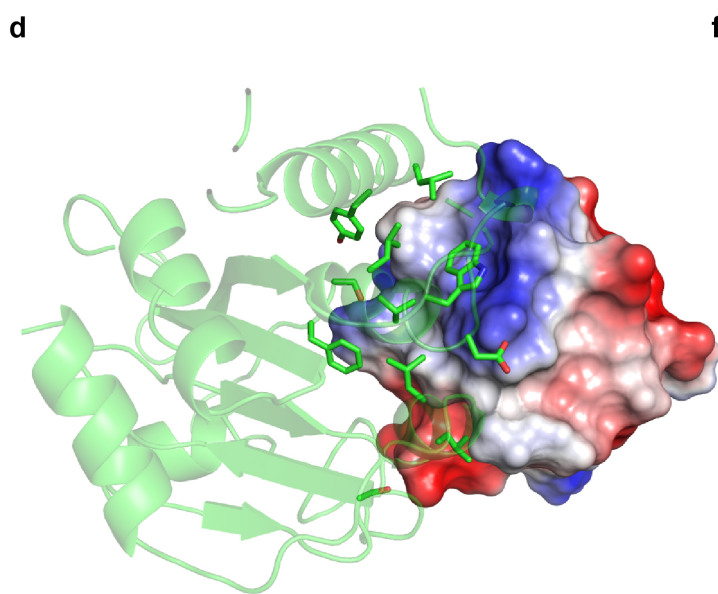
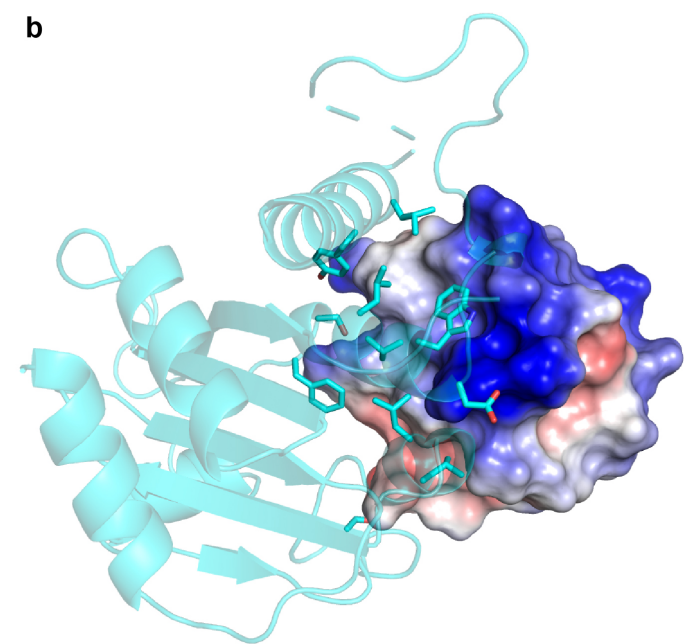
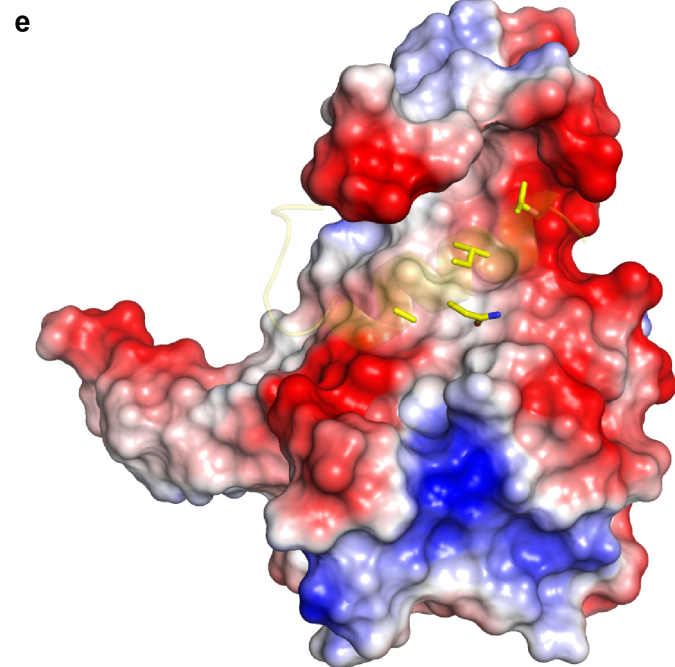
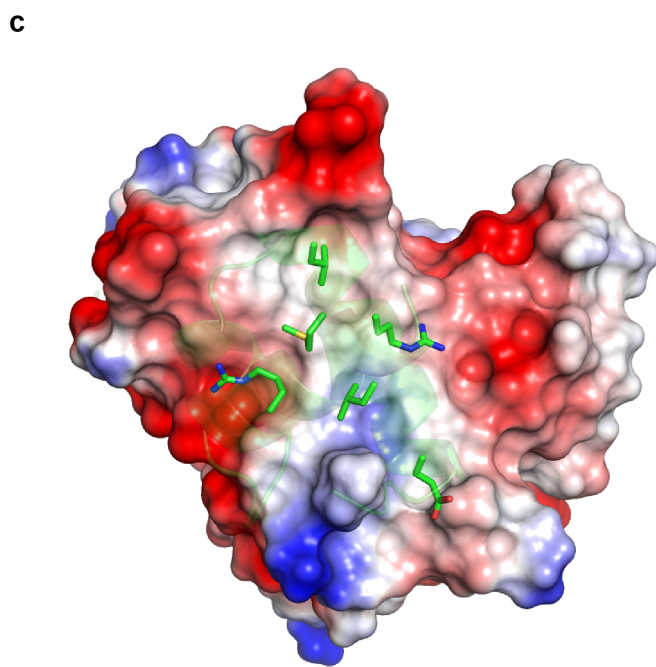
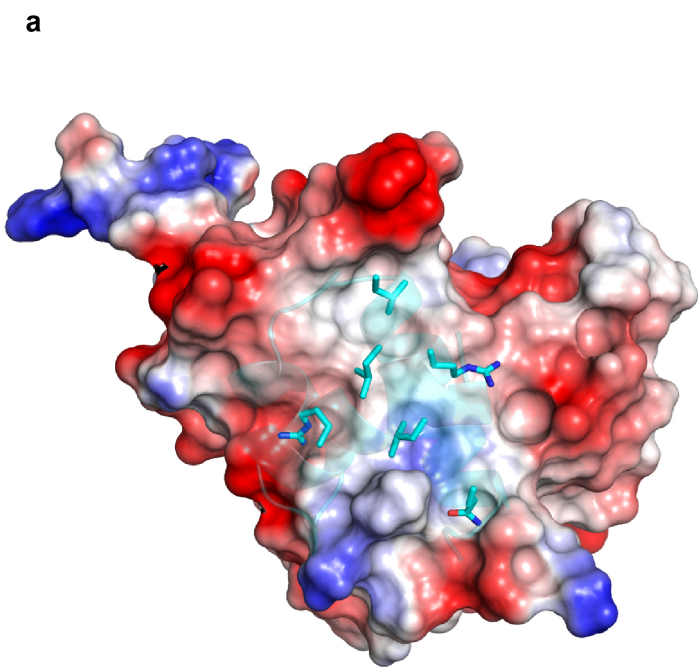


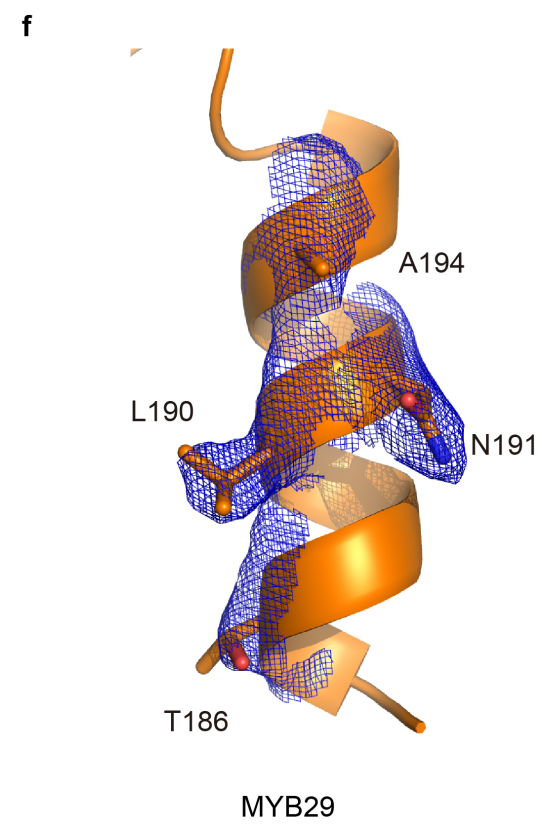
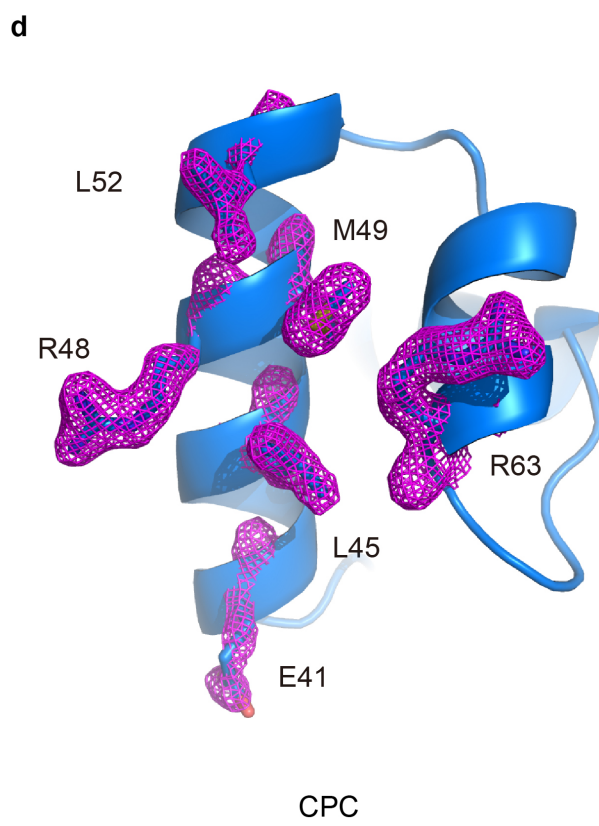
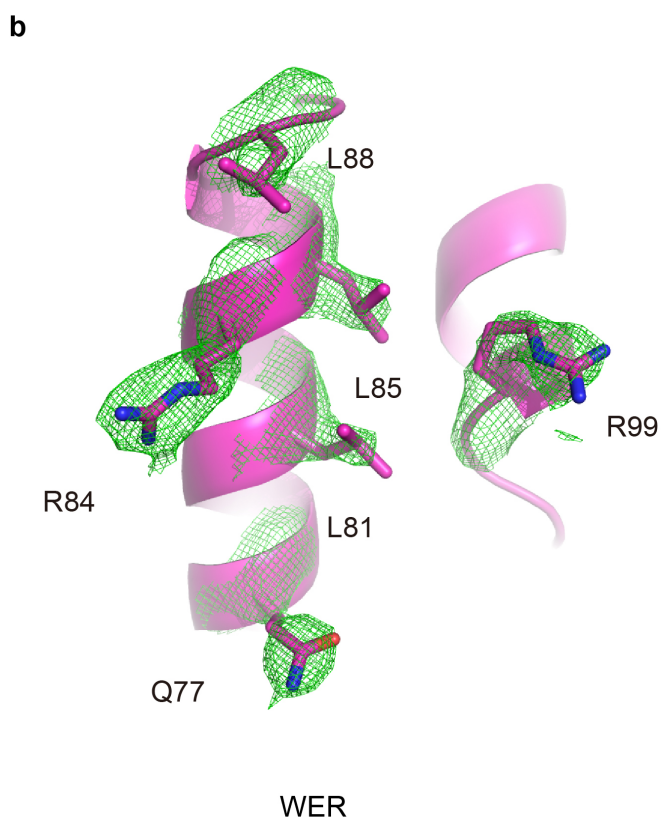
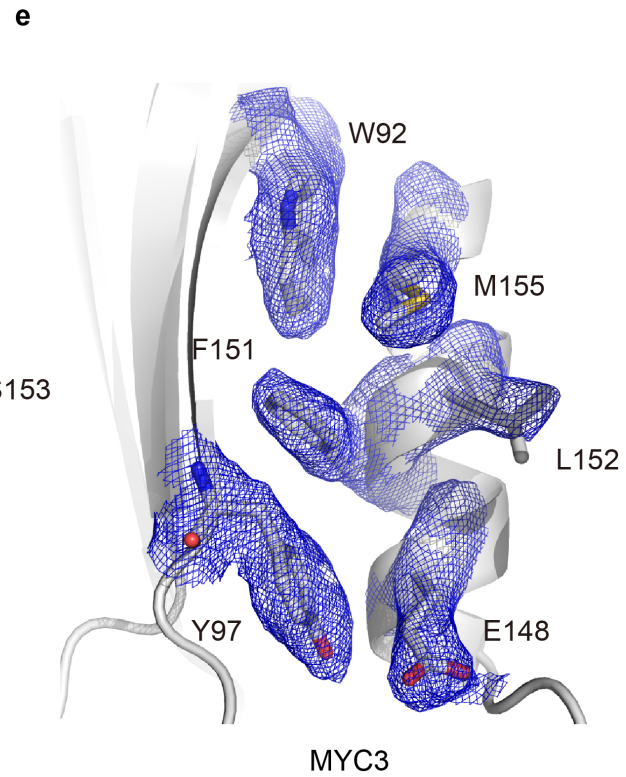
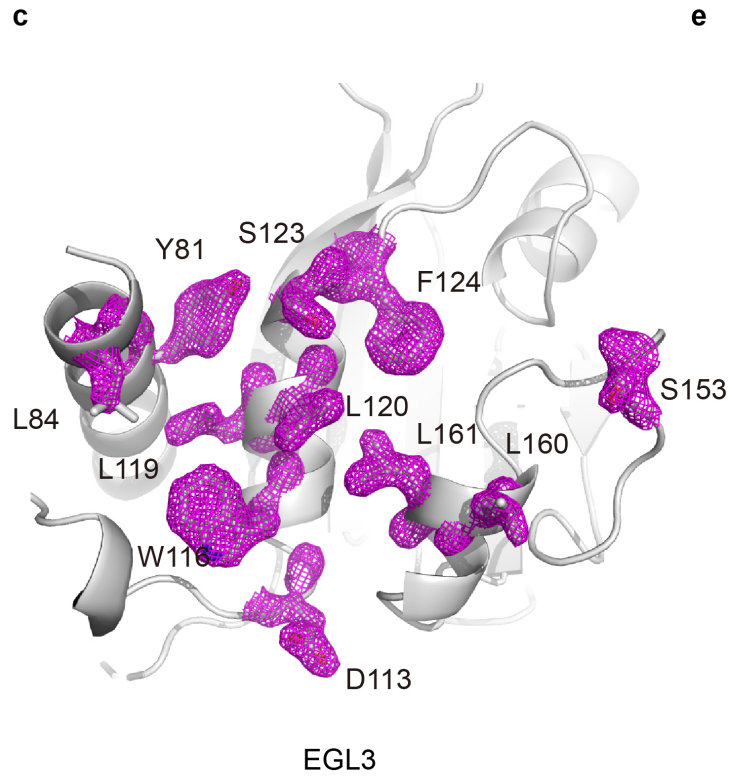
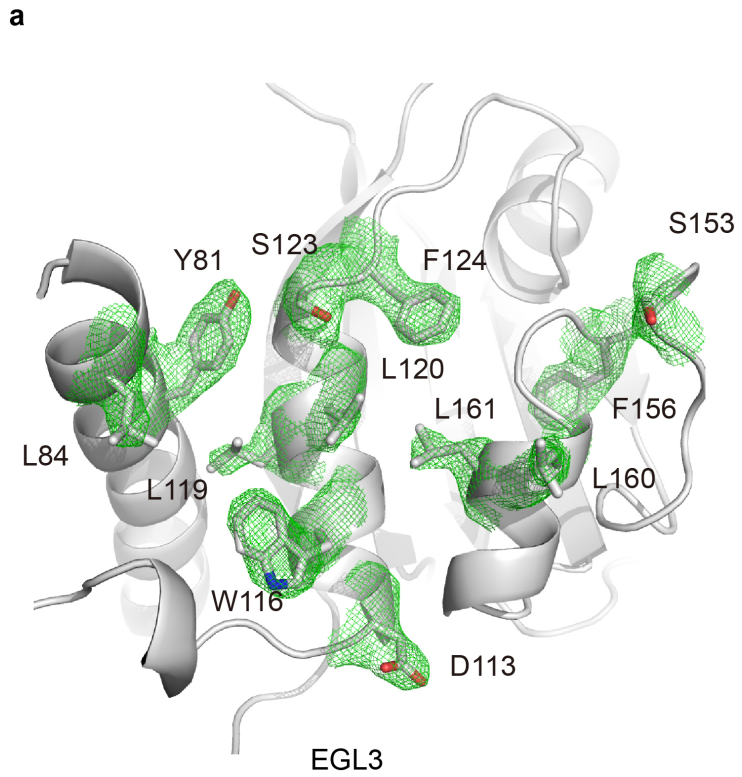


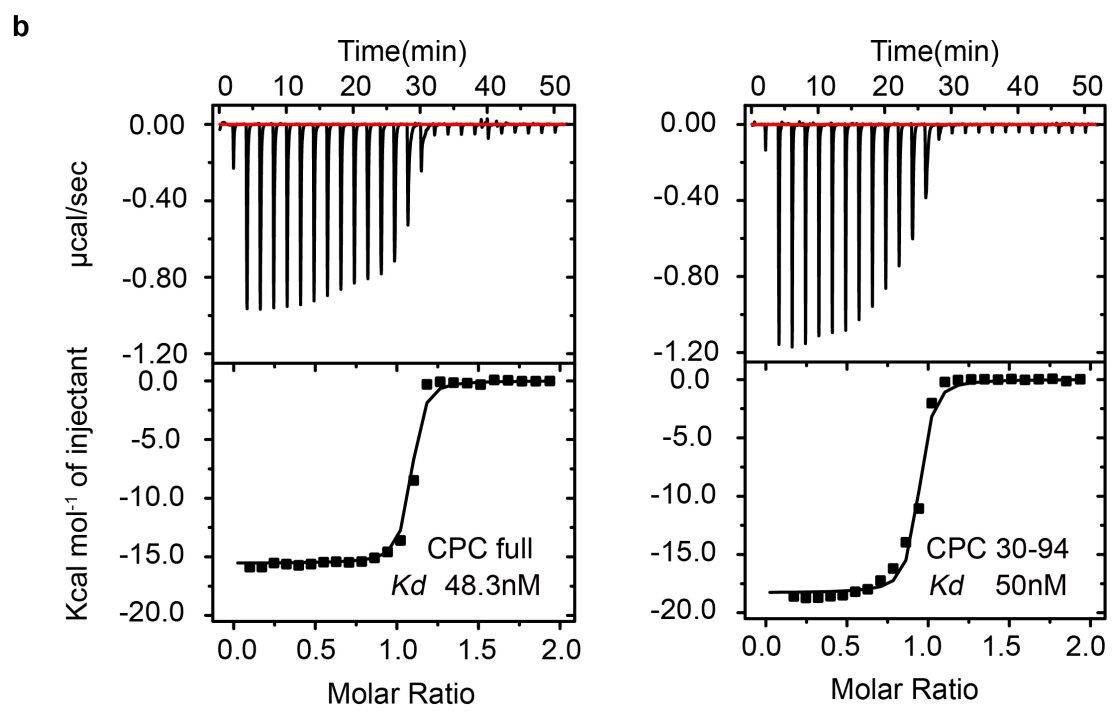
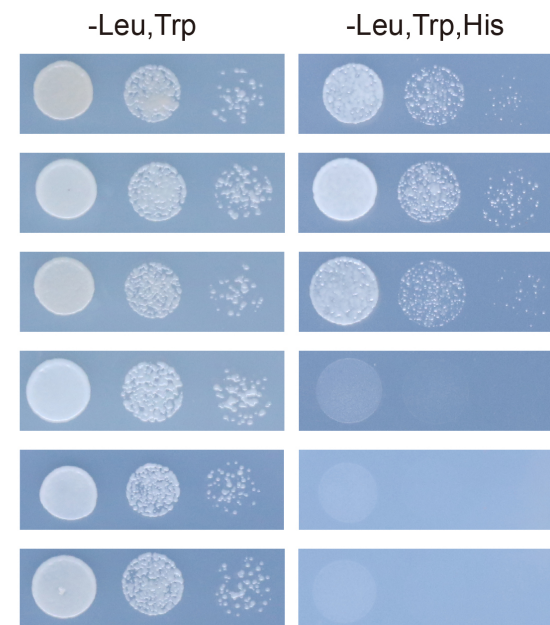
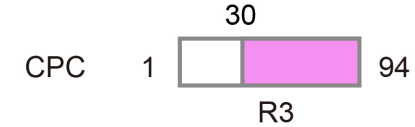
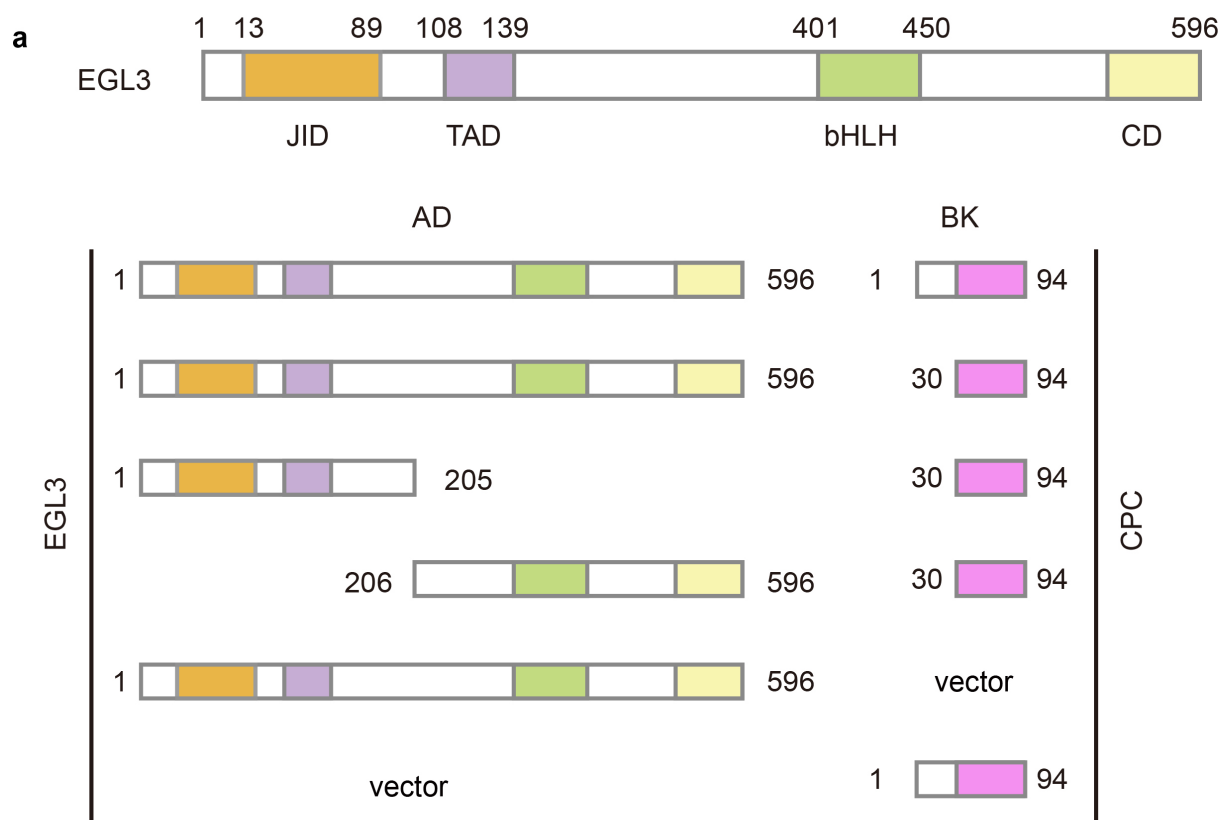


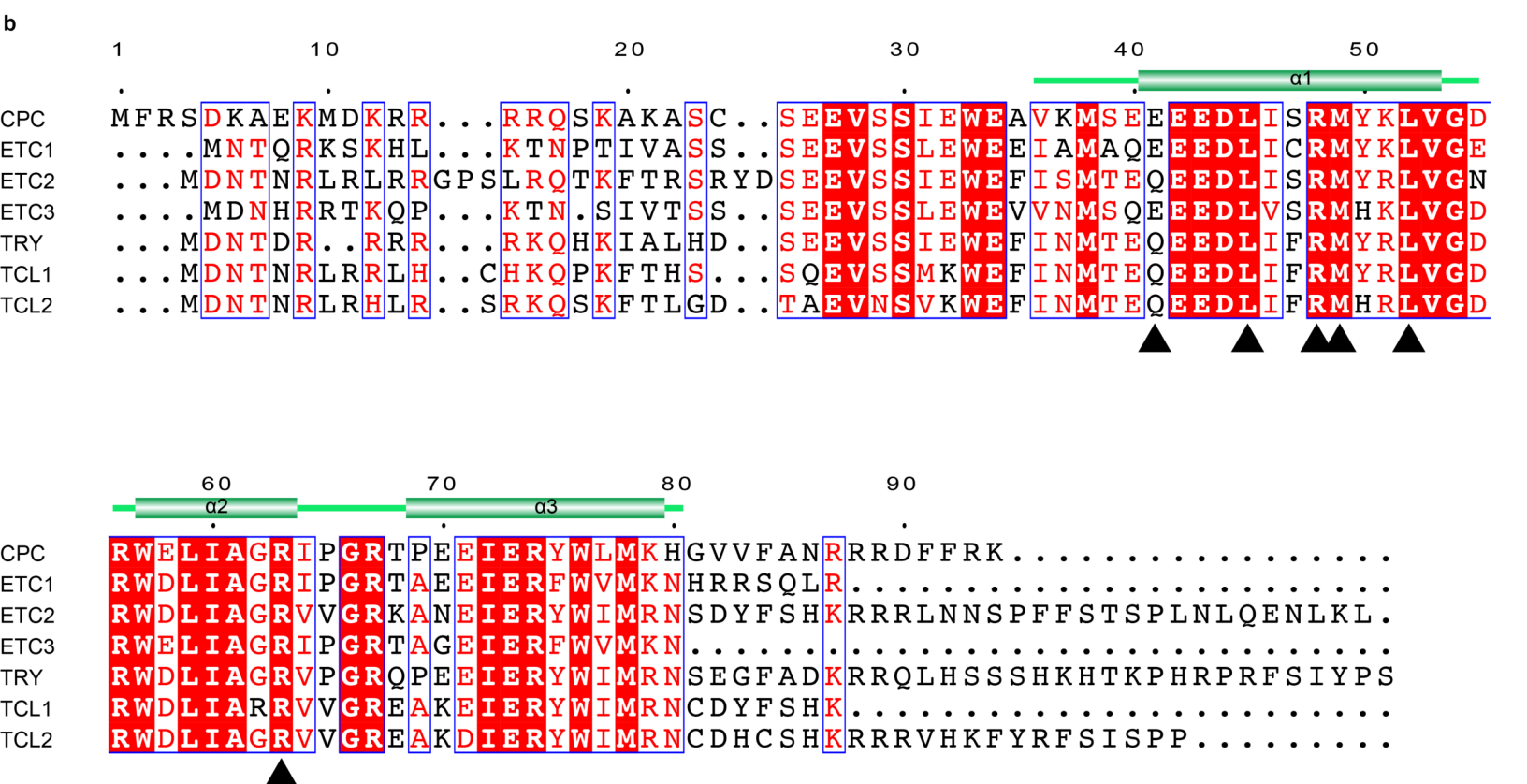
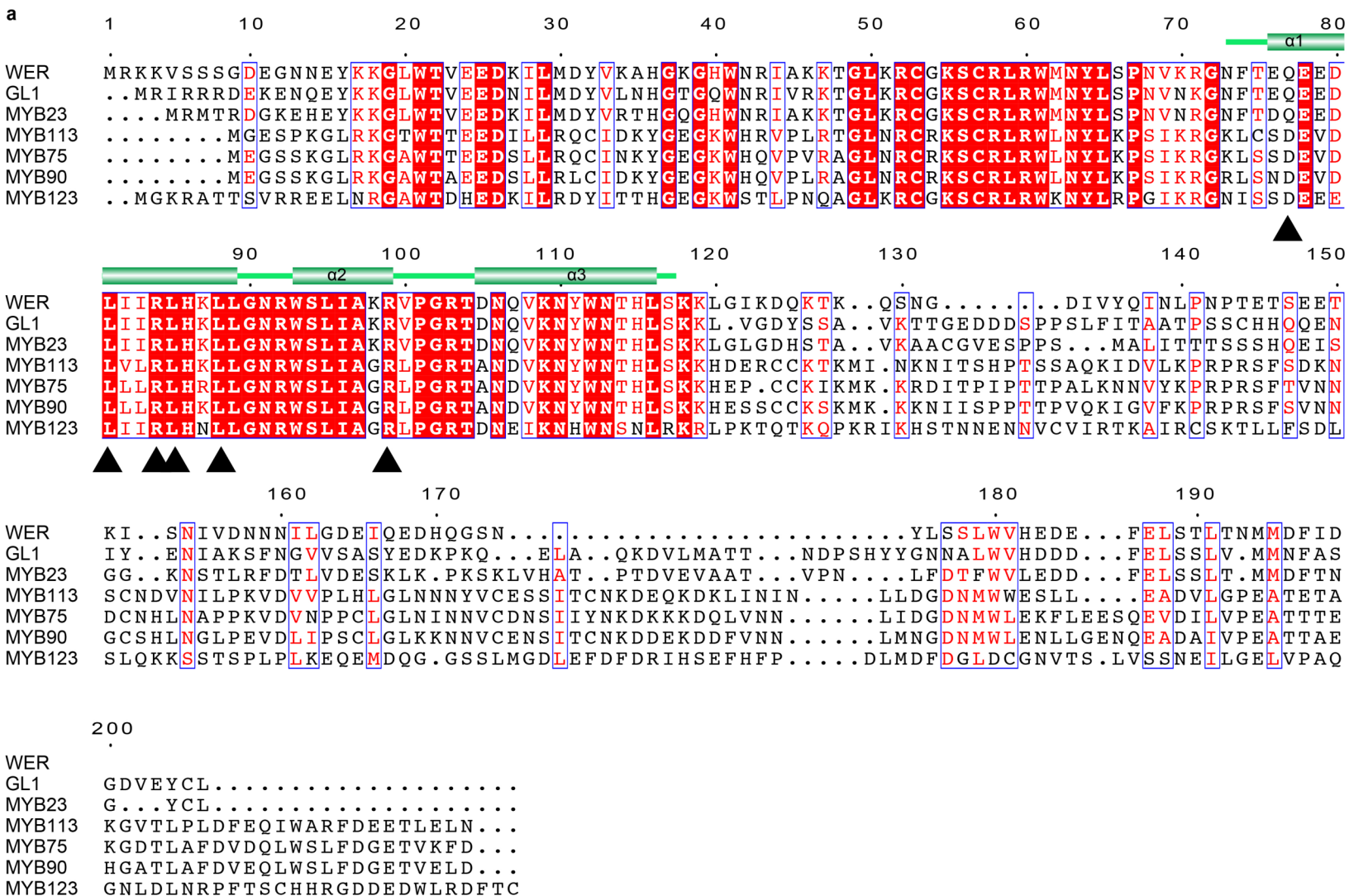


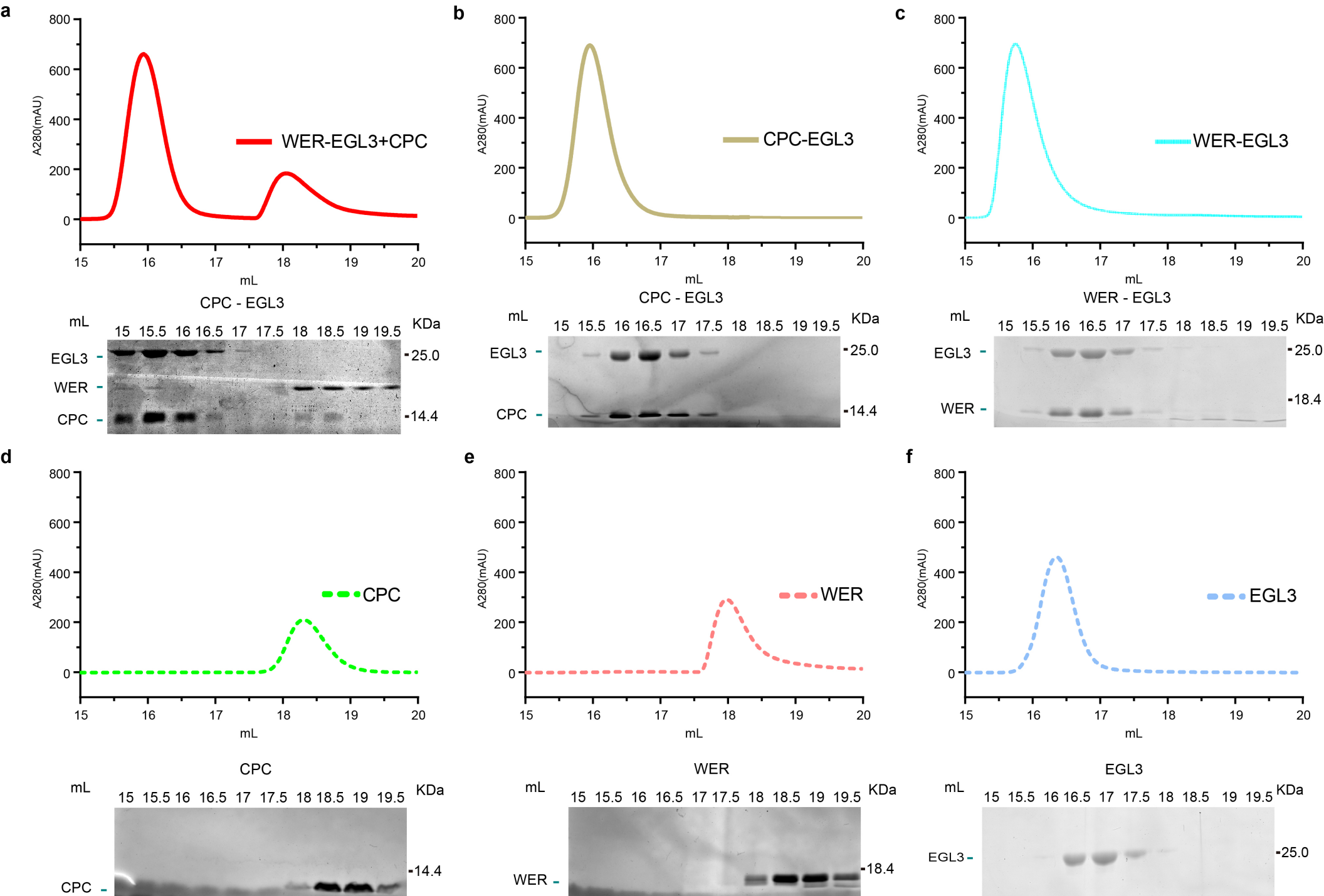


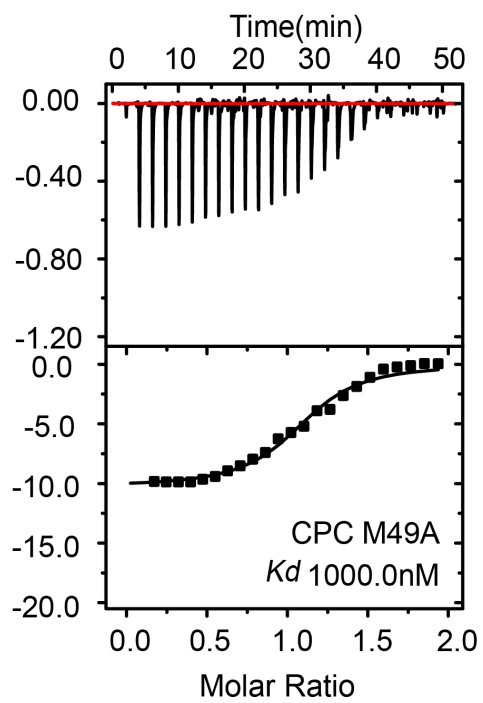
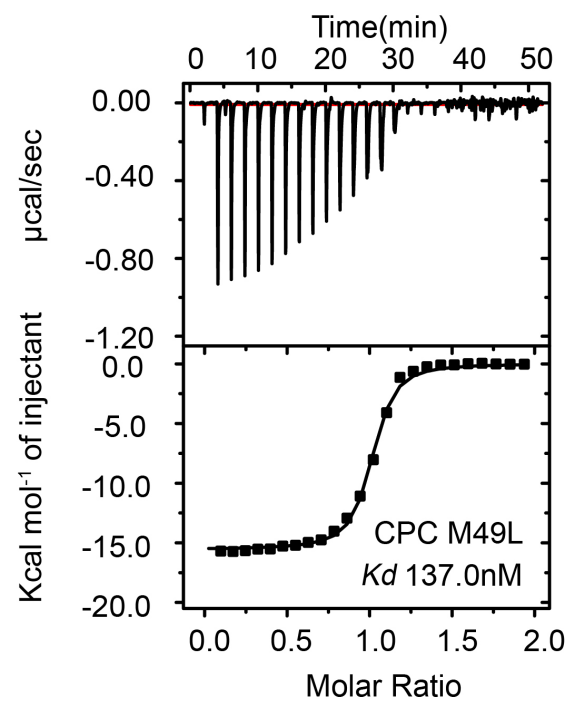
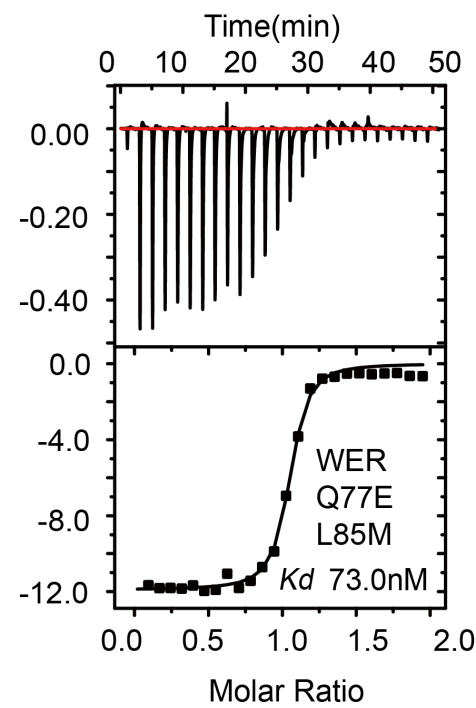
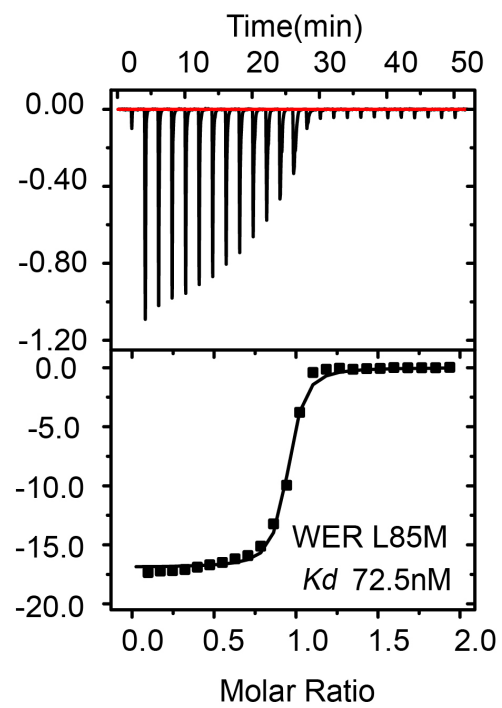


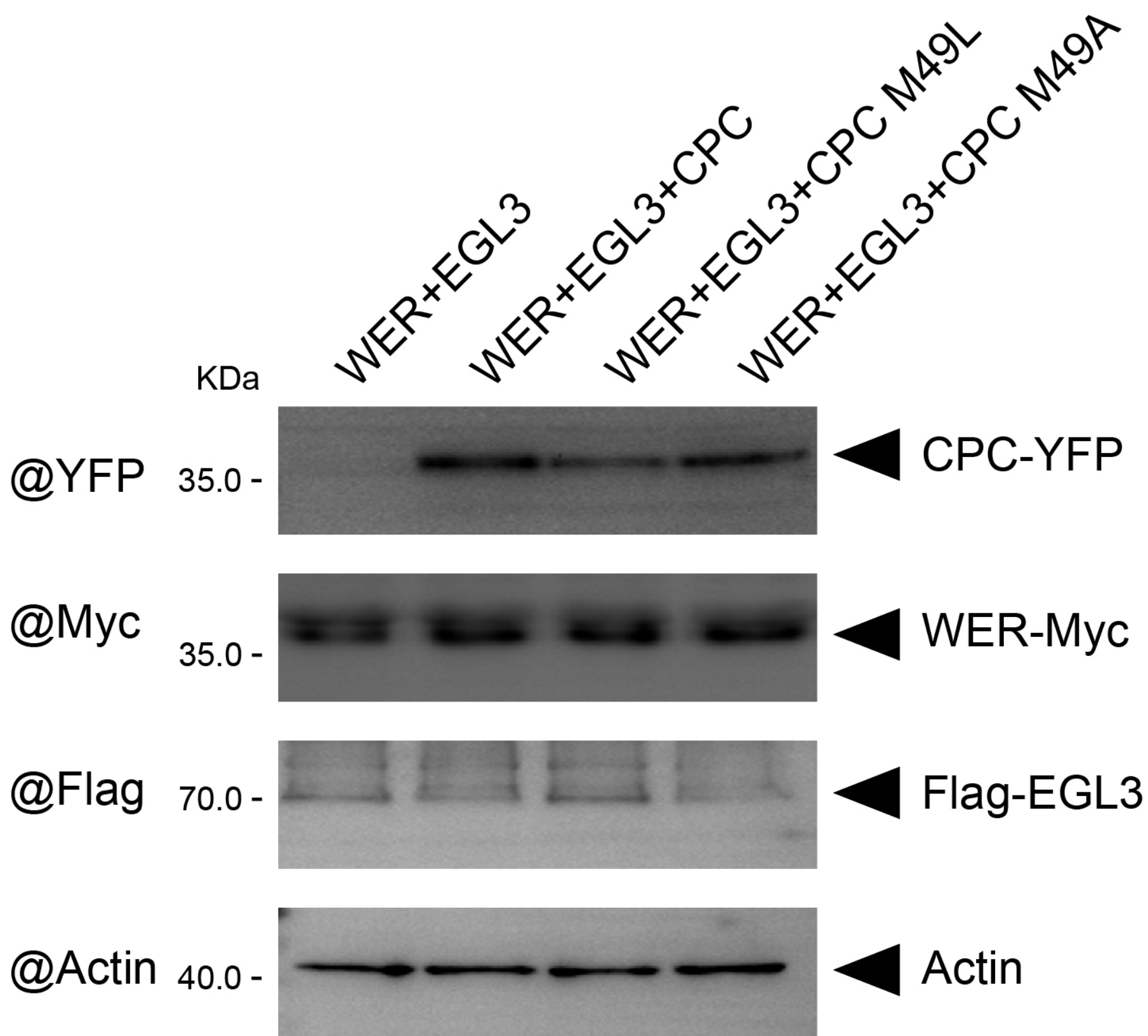








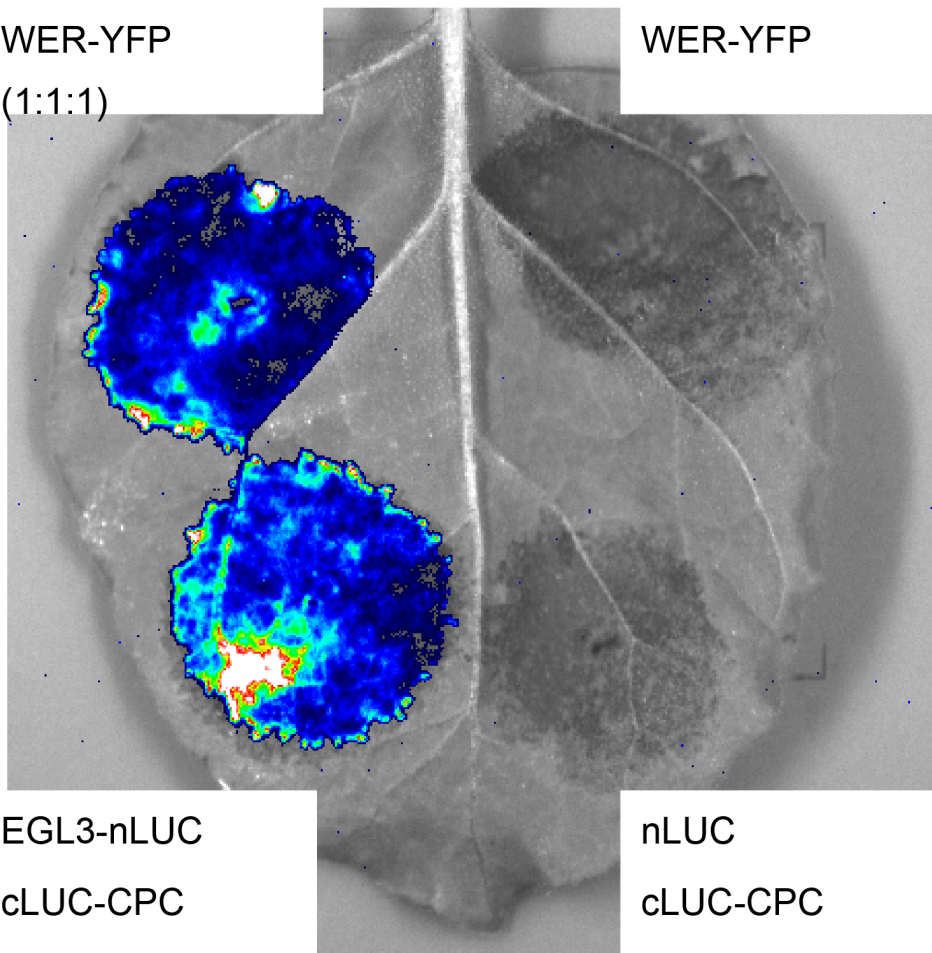
a**b**



a

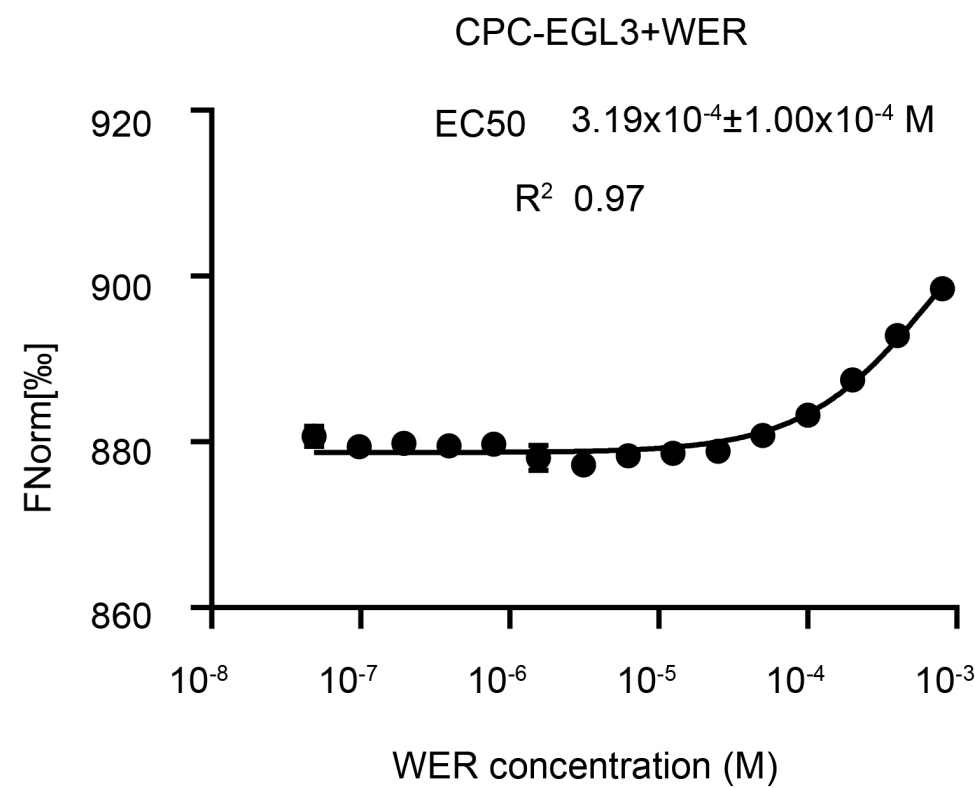
EGL3-nLUC
cLUC-CPC
WER-YFP
(1:1:1)

EGL3-nLUC
cLUC
WER-YFP

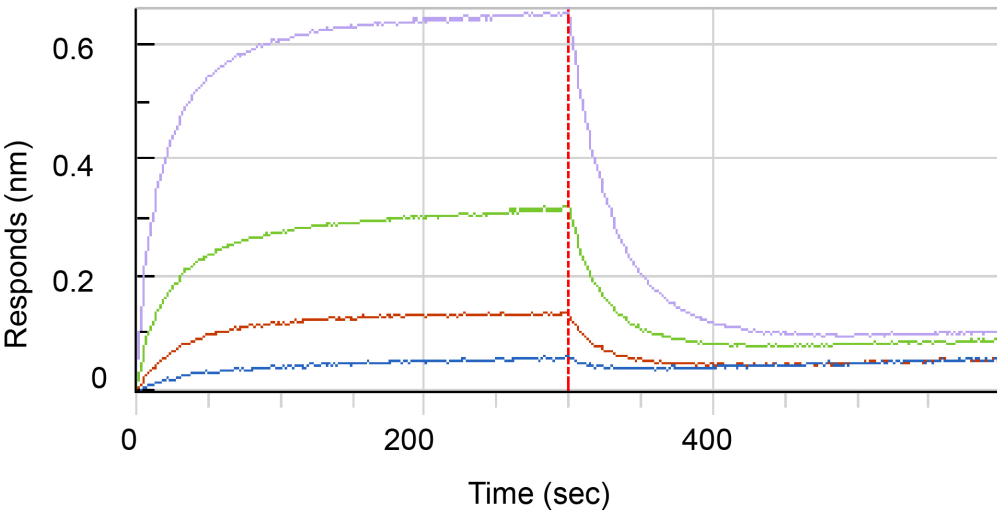


EGL3-nLUC
cLUC-CPC

nLUC
cLUC-CPC
WER-YFP

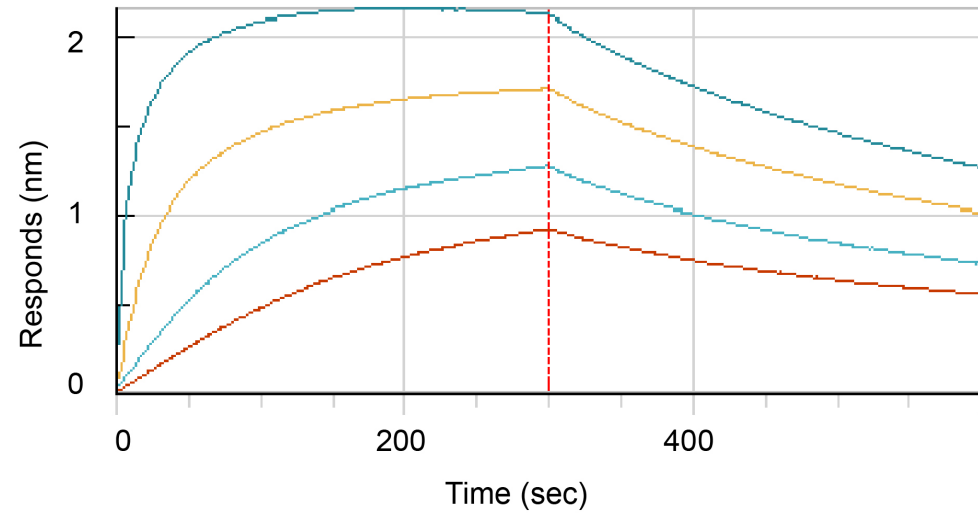
b

WER-EGL3



500 125 31.3 15.6 (nM)

CPC-EGL3



500 125 31.3 15.6 (nM)

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