The sequence analysis of EIN3 in *Arabidopsis thaliana*

习雨琳 郝冬冬 黄培鑫 马梦迪

2013.6.21

Contents

- Background
- Questions and approaches
- Results
- Discussion
- Acknowledgement

Biological function of Ethylene in plants
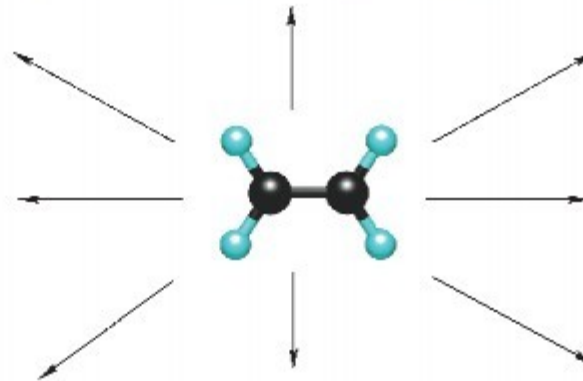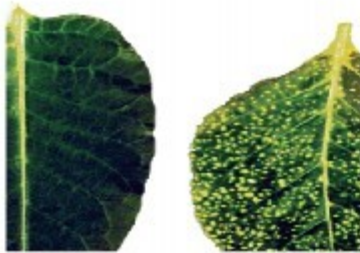
triple response

flower development

root elongation

defense

seed germination

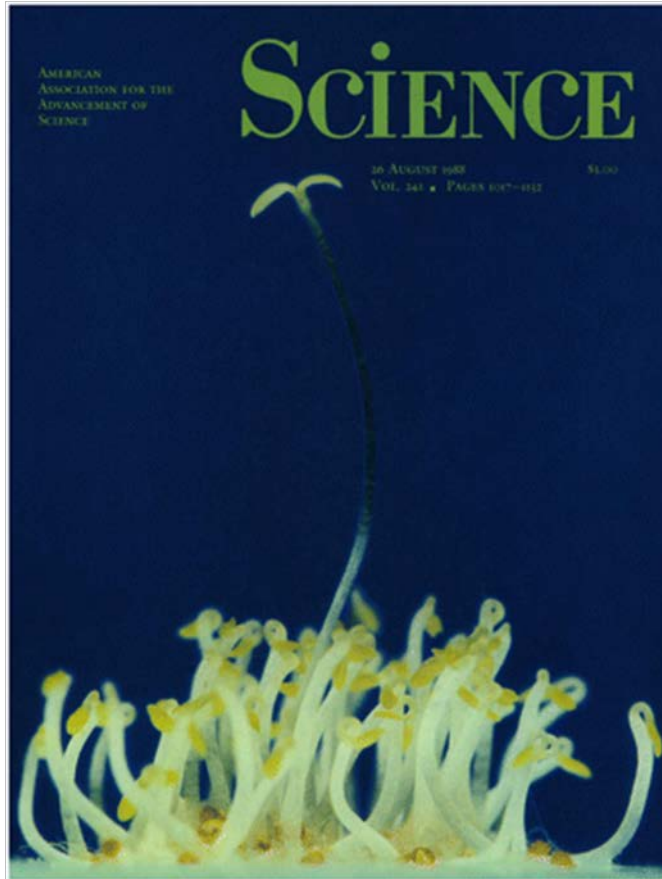**fruit ripening**

leaf senescence

root hair development

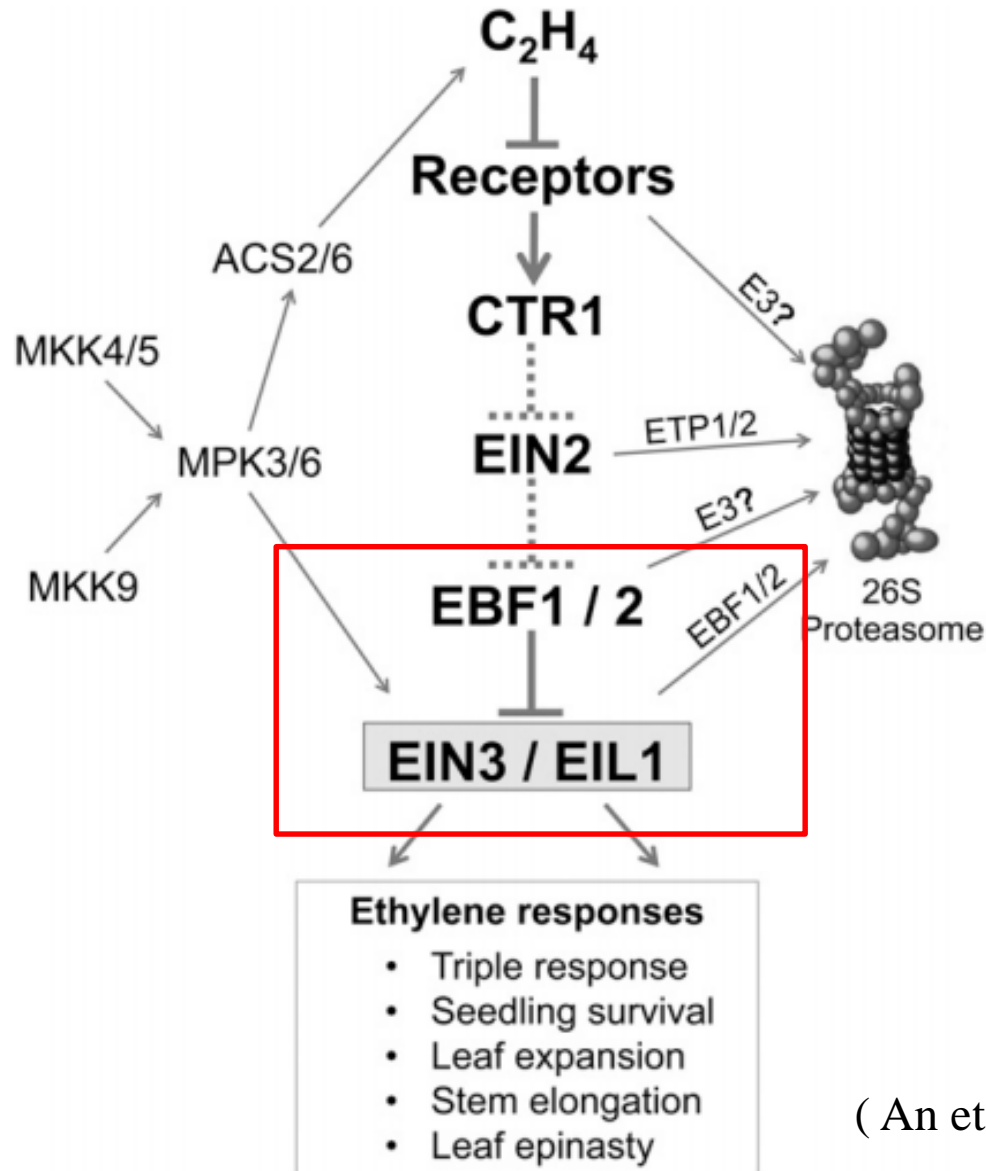# The initiation of research on Ethylene signaling

Air   ET

Triple response of **etiolated seedlings:**

Short, thickened root and hypocotyl

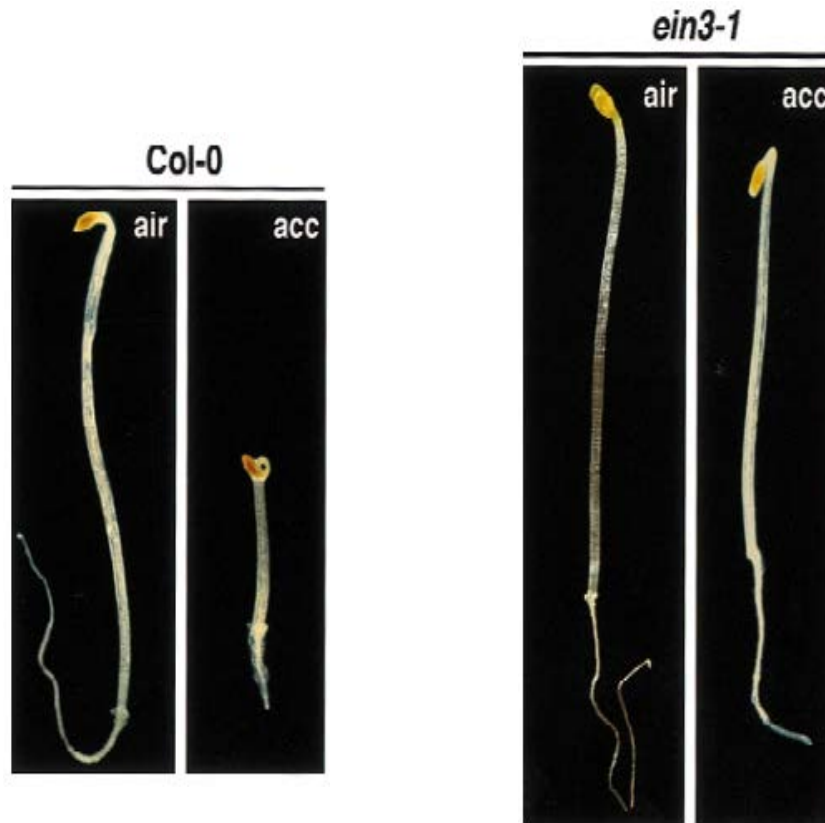**Exaggerated** curvature **of the apical hook**

(Bleecker et al., *Science*, 1988)

Ethylene signaling pathway in *Arabidopsis thaliana*

(An et al., *Plant Cell*, 2010)

The phenotype of *ein3(ethylene-insensitive 3)*

(Solano et al., *G&D*, 1998)

Networks through EIN3/EIL1
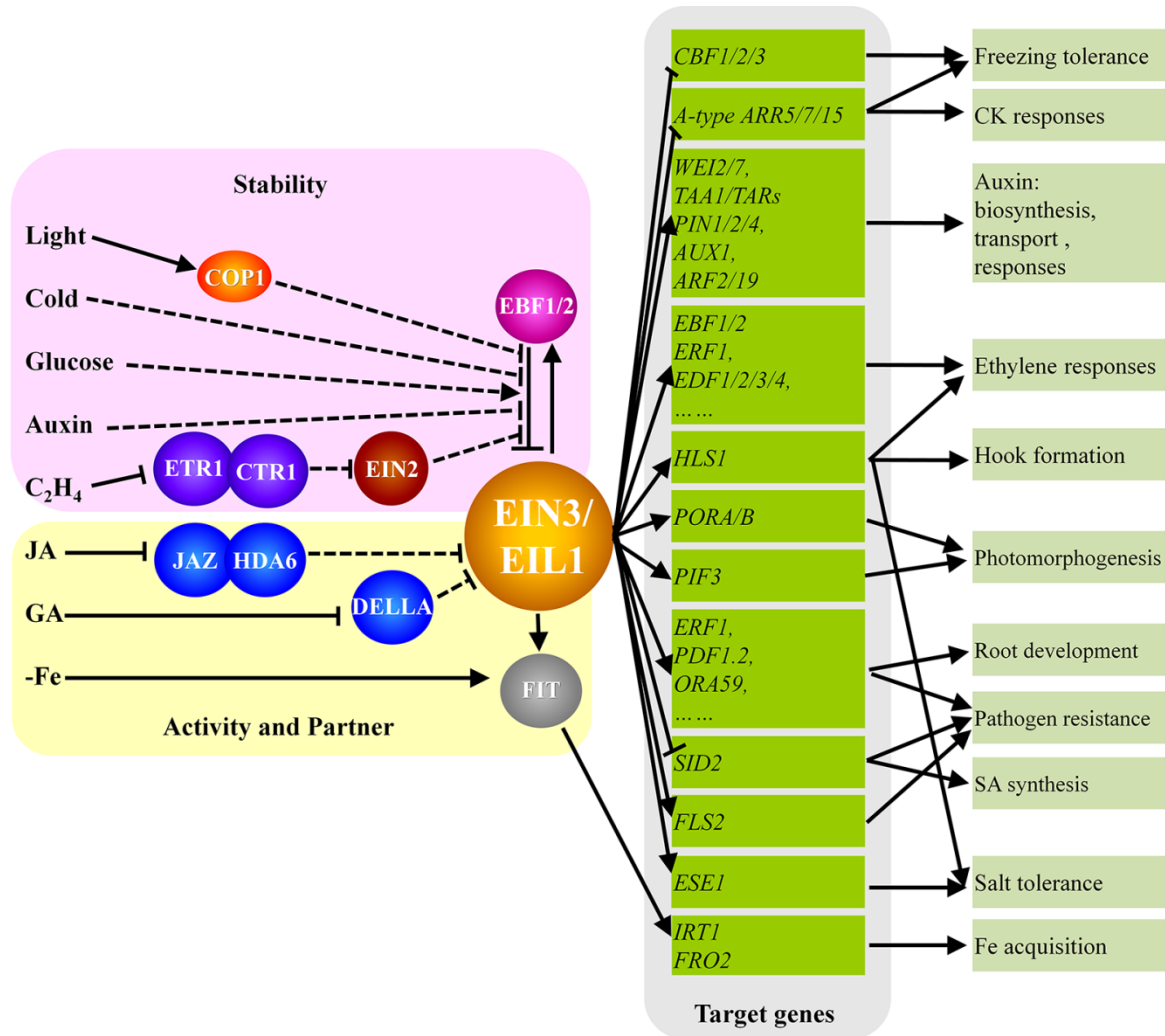
(Li et al., unpublished)

Questions & Approaches

- Questions: The function of EIN3 is quiet clear, but less is known about the relationship between its sequence and function.

- Approaches: Use bioinformatic tools to predict the conserved motifs and post-translational modifications

Results

- Conserved motifs of EIN3

- Probable post-translational modification sites

Phylogeny of EIN3/EILs
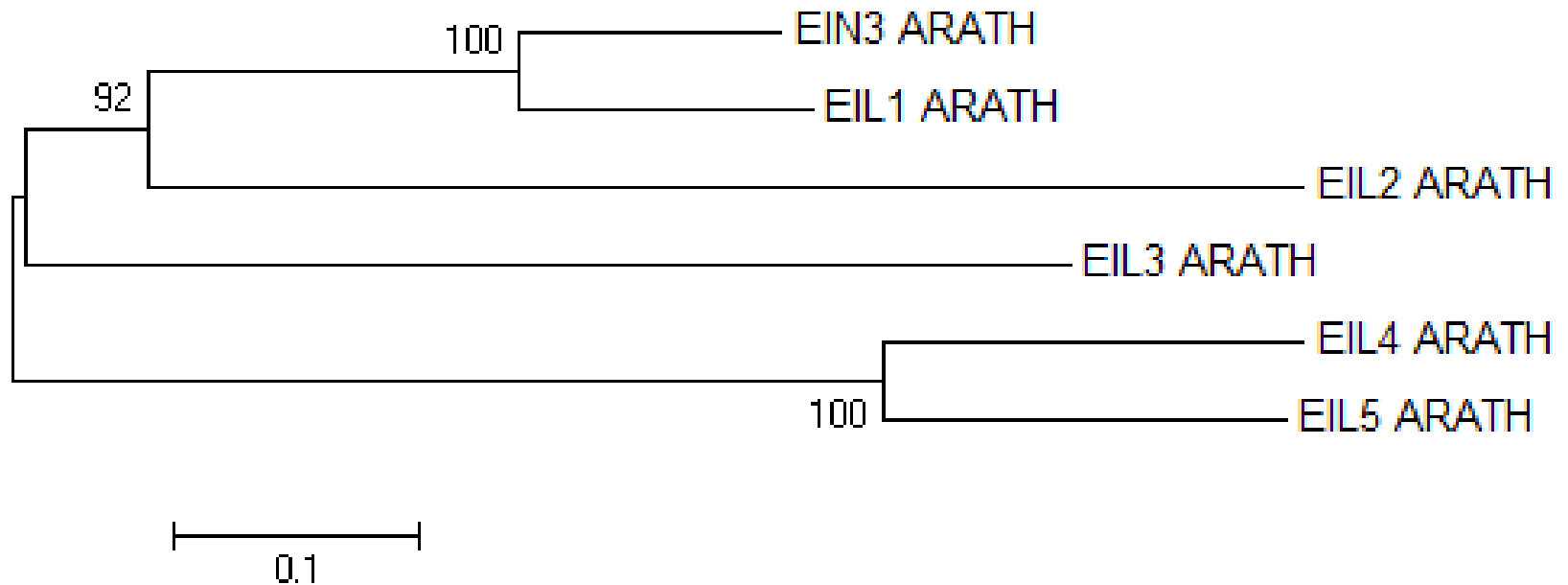
EIN3 and EIL polypeptides showing their similarities
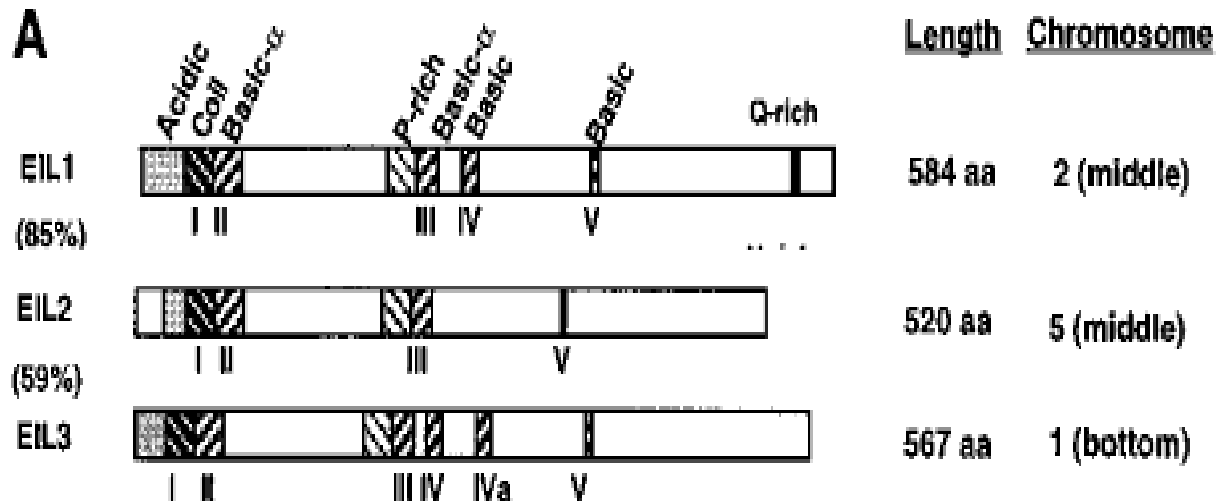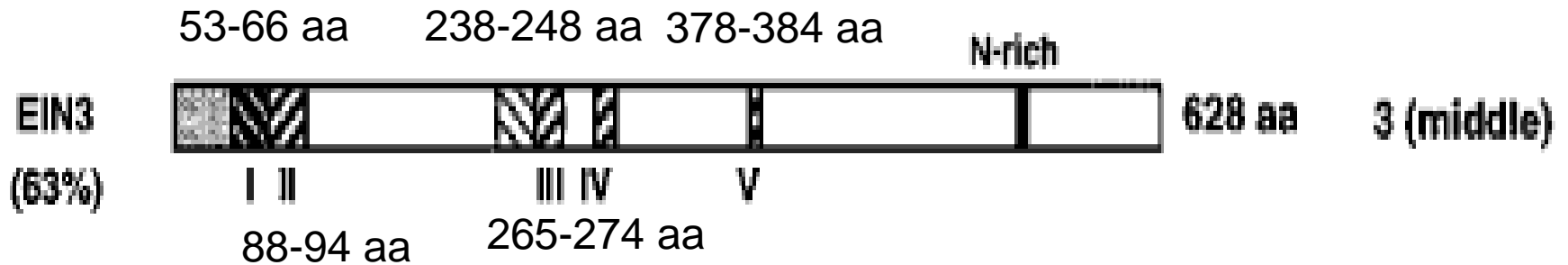
(Chao *et al.*, *Cell*, 1997)

Prediction of Arabidopsis EIN3 domain
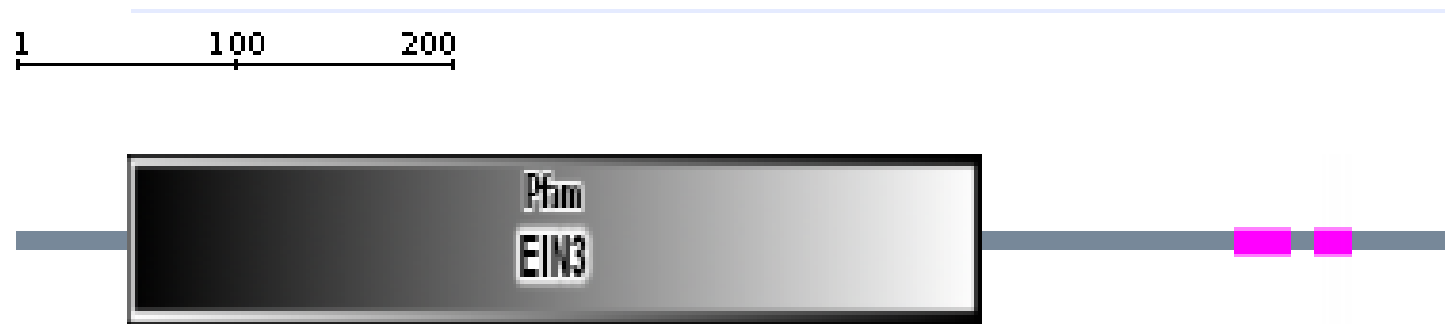
Confidently predicted domains, repeats, motifs and features:

Name	Start ▲	End	E-value
Pfam:EIN3	49	422	2.7e-149
low complexity	533	558	3
low complexity	567	583	3

Conserved Motifs in EIN3/EILs
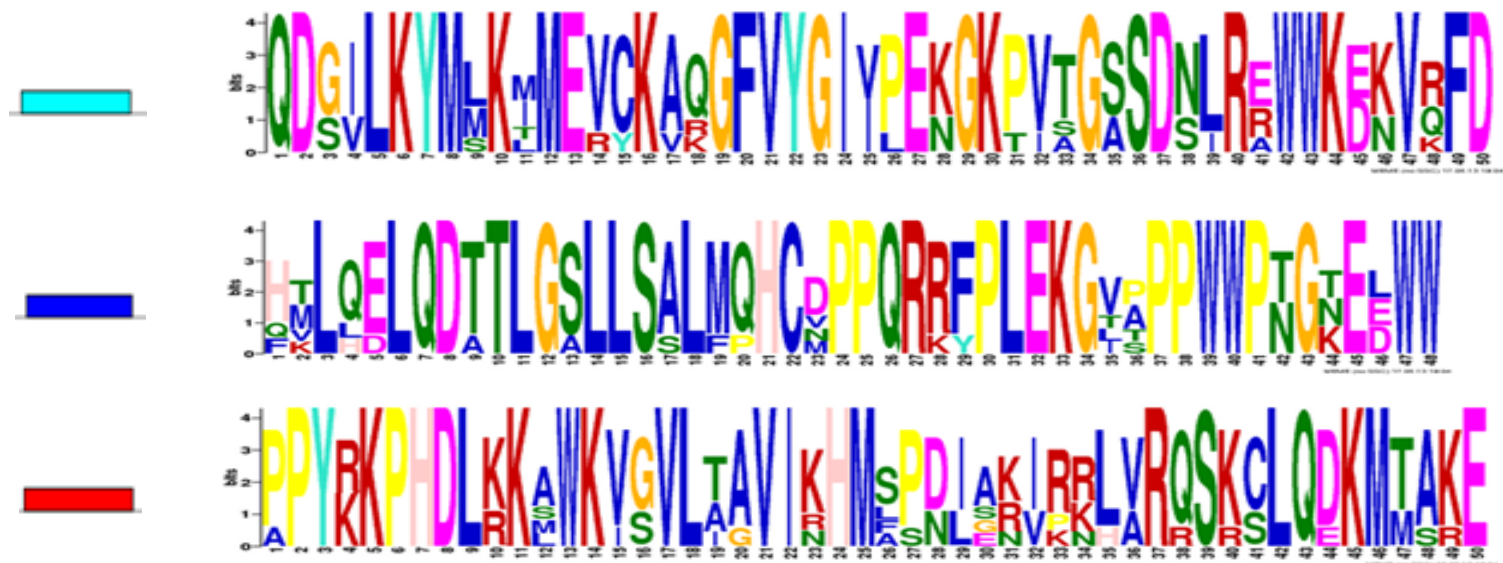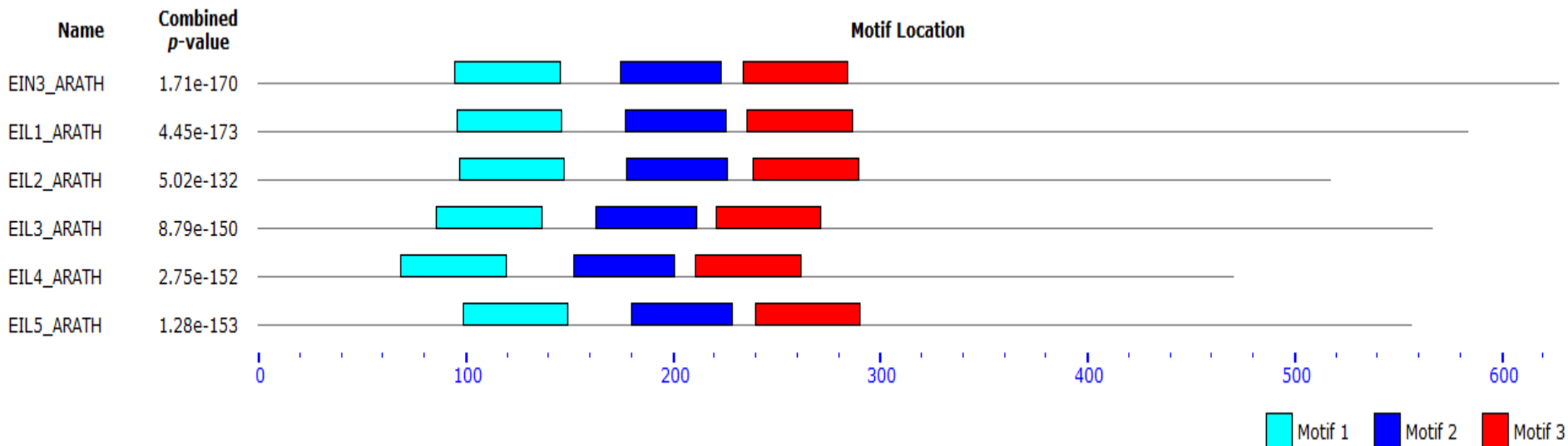
The structure of EIL3-DBD
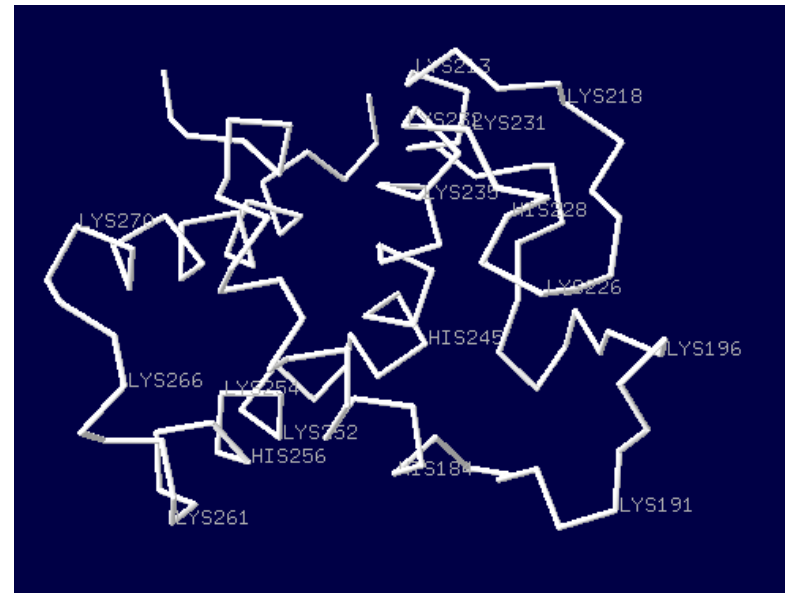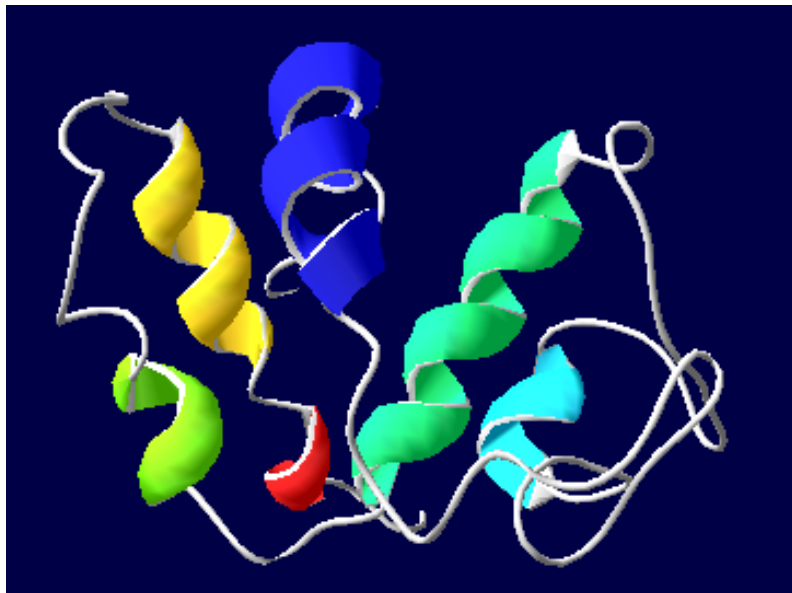
Homology Modeling-Swissmodel
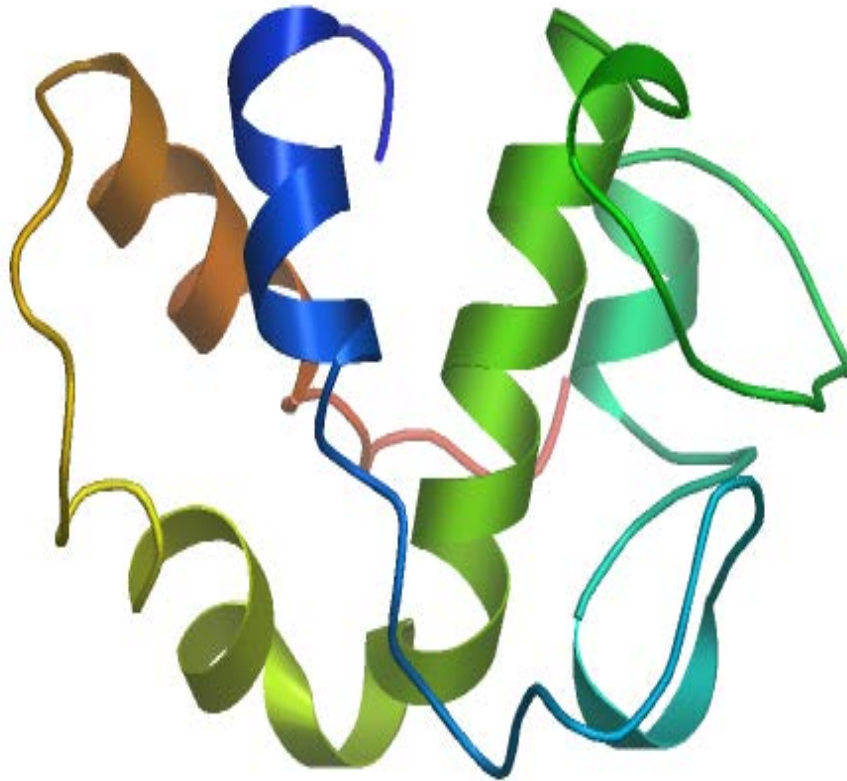

```
TARGET     1          LQEL QDTTLGSLLS ALMQHCDPPQ RRFPLEKGVP PPWWPNGKED
1wij_1#1  162   sqfv--lqdl qdatlgslls slmqhcdppq rkyplekgtp ppwwptgnee

TARGET                       hhhhhhhhh
1wij_1#1                     hhhhhhhhh h                               h


TARGET    45   WWPQLGLPKD QGPAPYKKPH DLKKAWKVGV LTAVIKHMFP DIAKIRKLVR
1wij_1#1  210   wwvklglpks qsp-pyrkph dlkkmwkvgv ltavinhmlp diakikrhvr

TARGET         hhhh                    hhhhhhh hhhhhh        hhhhh
1wij_1#1       hhhh                    hhhhhhh hhhhhh        hhhhh


TARGET    95   QSKCLQDKMT AKESATWLAI INQEESLARE
1wij_1#1  259   qskclqdkmt akesaiwlav lnqeesliqq -

TARGET                      hhhhhhhh    hhh
1wij_1#1                     hhhhhhhh    hhh
```

Summary 1

➢49-422 aa may be the functional domain of EIN3.

➢178-301 aa is the predicted DNA binding domain.

Results

- Conserved motifs of EIN3

- Probable post-translational modification sites

The previous studies on EIN3 modifications

nature

ARTICLES

Dual control of nuclear EIN3 by bifurcate MAPK cascades in C_2H_4 signalling

Proteomic analyses identify a diverse array of nuclear processes affected by small ubiquitin-like modifier conjugation in *Arabidopsis*

Marcus J. Miller[a], Gregory A. Barrett-Wilt[b], Zhihua Hua[a], and Richard D. Vierstra[a,1]

www.jxb.oxfordjournals.org

RESEARCH PAPER

A conserved phosphorylation site regulates the transcriptional function of ETHYLENE-INSENSITIVE3-like1 in tomato

What we focused on

- ## Phosphorylation
  - Scansite, Netphos, PhosPhAt, DisPhos

- ## Sumoylation(Small ubiquitin-like modifier)
  - SUMOsp, SUMOplot

- ## Ubiquitination
  - Ubipred, Ubperd, BDM-PUB

Scansite

Acid_ST_kin
S19

Acid_ST_kin
T42

Predicted
Sites

(Domains not requested)

Surface
Accessibility
1.0

100   200   300   400   500   600   628 AA
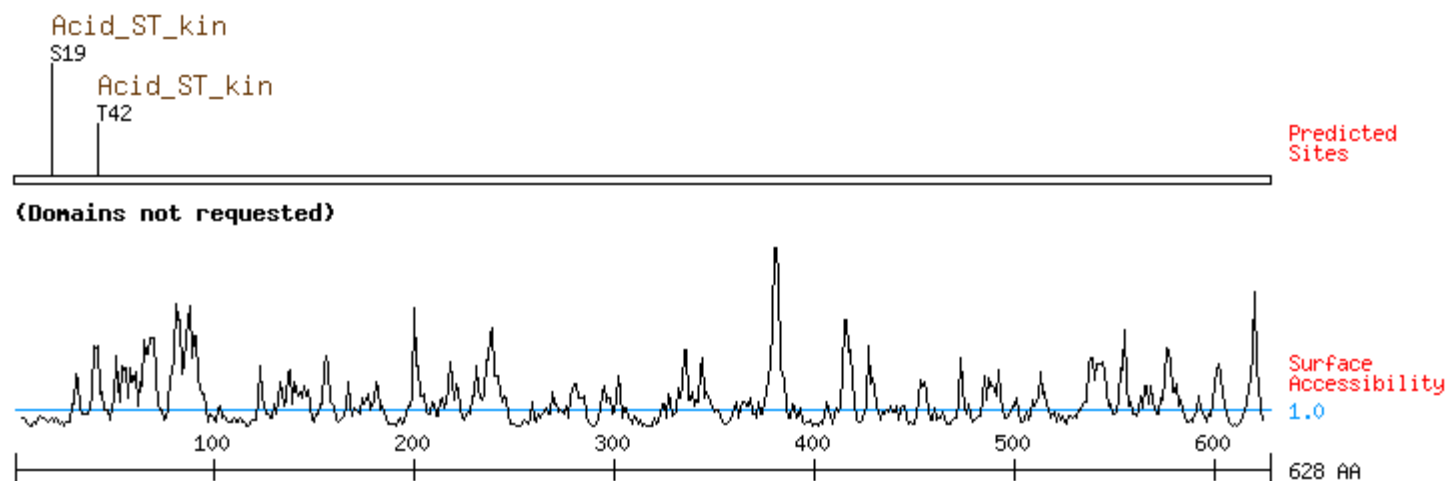
Acidophilic serine/threonine kinase group (Acid_ST_kin)				
Casein Kinase 1			Gene Card **CSNK1G2**	
Site	Score	Percentile	Sequence	SA
S19	0.3608	0.118 %	MDFFSSGSLGEVDFC	0.576
Casein Kinase 2			Gene Card **CSNK2B**	
Site	Score	Percentile	Sequence	SA
T42	0.2932	0.028 %	SIVEDDYTDDEIDVD	4.180

NetPhos 2.0 Server - prediction results

Technical University of Denmark

Phosphorylation sites predicted: Ser: 13 Thr: 3 Tyr: 4

NetPhos 2.0: predicted phosphorylation sites in seq.1
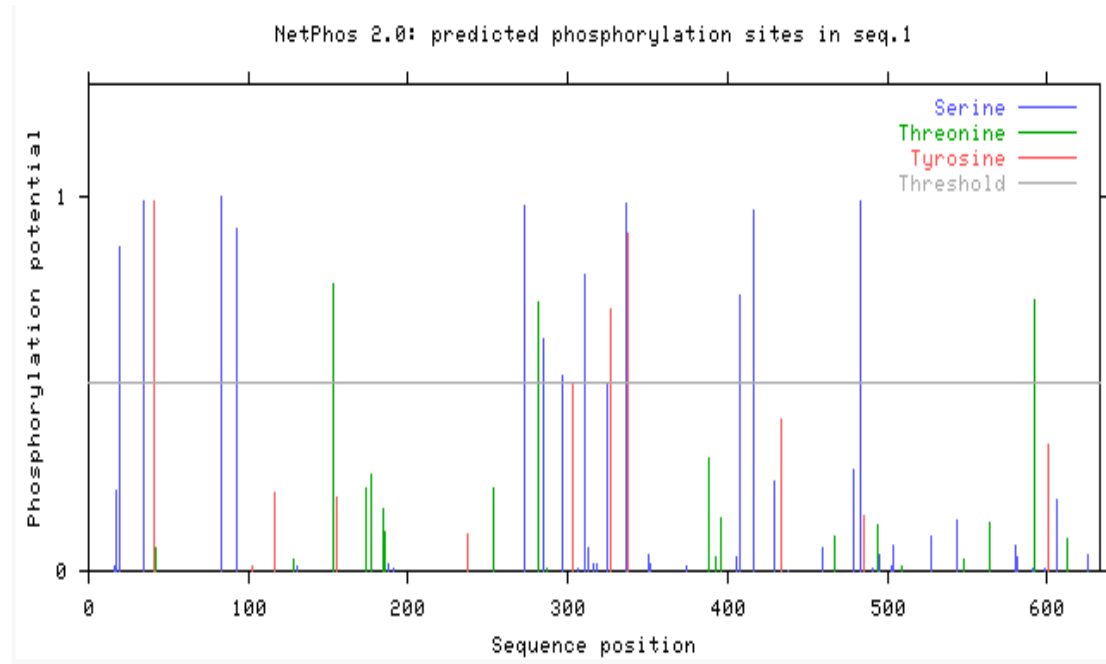
NetPhos 2.0 Server - prediction results

Technical University of Denmark

Phosphorylation sites predicted: Ser: 13 Thr: 3 Tyr: 4

Serine predictions

Pos	Context	Score	Pred
		v	
16	MDFFSSGSL	0.012	.
17	DFFSSGSLG	0.217	.
19	FSSGSLGEV	0.863	*S*
35	AEPDSIVED	0.985	*S*
83	KQRQSQEQA	0.997	*S*
93	RKKMSRAQD	0.915	*S*
131	VTGASDNLR	0.013	.
188	TTLGSLLSA	0.016	.
191	GSLLSALMQ	0.008	.
273	LVRQSKCLQ	0.974	*S*
285	TAKESATWL	0.618	*S*
297	NQEESLARE	0.520	*S*
306	LYPESCPPL	0.005	.
311	CPPLSLSGG	0.793	*S*
313	PLSLSGGSC	0.064	.
316	LSGGSCSLL	0.019	.
318	GGSCSLLMN	0.020	.
325	MNDCSQYDV	0.501	*S*
336	FEKESHYEV	0.982	*S*
351	KVMNSSNFG	0.045	.
352	VMNSSNFGM	0.019	.
374	PAGNSEFMR	0.013	.
405	GCAHSEISR	0.034	.
408	HSEISRGFL	0.737	*S*
416	LDRNSRDNH	0.964	*S*
429	PHRDSRLPY	0.237	.
438	GAAPSRFHV	0.003	.
459	RPVNSVAQP	0.059	.
479	QKMISELMS	0.271	.
483	SELMSMYDR	0.985	*S*

Threonine predictions

Pos	Context	Score	Pred
		v	
42	EDDYTDDEI	0.060	.
128	GKPVTGASD	0.033	.
153	PAAITKYQA	0.768	*T*
174	PIGPTPHTL	0.223	.
177	PTPHTLQEL	0.256	.
184	ELQDTTLGS	0.166	.
185	LQDTTLGSL	0.107	.
253	VGVLTAVIK	0.219	.
281	QDKMTAKES	0.715	*T*
287	KESATWLAI	0.007	.
388	RDLNTIMDR	0.298	.
393	IMDRTVFTC	0.037	.
396	RTVFTCENL	0.139	.
467	PIDLTGIVP	0.089	.
494	QSNQTSMVM	0.123	.
509	LLQPTVHNH	0.012	.
548	SNNQTFFQG	0.028	.
564	FKFDTADHN	0.128	.
592	VFDSTPFDM	0.723	*T*
613	GVVGTMDGM	0.083	.

Tyrosine predictions

Pos	Context	Score	Pred
		v	
41	VEDDYTDDE	0.987	*Y*
102	GILKYMLKM	0.015	.
117	QGFVYGIIP	0.208	.
155	AITKYQAEN	0.195	.
237	GPAPYKKPH	0.096	.
303	ARELYPESC	0.502	*Y*
327	DCSQYDVEG	0.696	*Y*
338	KESHYEVEE	0.902	*Y*
433	SRLPYGAAP	0.406	.
485	LMSMYDRNV	0.149	.
601	ASFDYRDDM	0.338	.

PhosPhAt
The Arabidopsis Protein Phosphorylation Site Database

| Species:	Arabidopsis thaliana
Protein:	AT3G20770.1
Description:	EIN3, AtEIN3; Ethylene insensitive 3 family protein
MapMan:	27.3.19 RNA.regulation of transcription.EIN3-like(EIL) transcription factor family
Substrate for Kinase:	AT1G73500.1 - ATMKK9, MKK9; MAP kinase kinase 9
	AT2G43790.1 - ATMPK6, MPK6, MAPK6, ATMAPK6; MAP kinase 6
	AT3G45640.1 - ATMPK3, MPK3, ATMAPK3; mitogen-activated protein kinase 3
	AT5G03730.1 - CTR1, SIS1, AtCTR1; Protein kinase superfamily protein

Sequence

```
          1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
  1...50   M  M  F  N  E  M  G  M  C  G  N  M  D  F  F  S  S  G  S  L  G  E  V  D  F  C  P  V  P  Q  A  E  P  D  S  I  V  E  D  D  Y  T  D  D  E  I  D  V  D  E
 51...100  L  E  R  R  M  W  R  D  K  M  R  L  K  R  L  K  E  Q  D  K  G  K  E  G  V  D  A  A  K  Q  R  Q  S  Q  E  Q  A  R  R  K  K  M  S  R  A  Q  D  G  I  L
101...150  K  Y  M  L  K  M  M  E  V  C  K  A  Q  G  F  V  Y  G  I  I  P  E  N  G  K  P  V  T  G  A  S  D  N  L  R  E  W  W  K  D  K  V  R  F  D  R  N  G  P  A
151...200  A  I  T  K  Y  Q  A  E  N  N  I  P  G  I  H  E  G  N  N  P  I  G  P  T  P  H  T  L  Q  E  L  Q  D  T  T  L  G  S  L  L  S  A  L  M  Q  H  C  D  P  P
201...250  Q  R  R  F  P  L  E  K  G  V  P  P  P  W  W  P  N  G  K  E  D  W  W  P  Q  L  G  L  P  K  D  Q  G  P  A  P  Y  K  K  P  H  D  L  K  K  A  W  K  V  G
251...300  V  L  T  A  V  I  K  H  M  F  P  D  I  A  K  I  R  K  L  V  R  Q  S  K  C  L  Q  D  K  M  T  A  K  E  S  A  T  W  L  A  I  I  N  Q  E  E  S  L  A  R
301...350  E  L  Y  P  E  S  C  P  P  L  S  L  S  G  G  S  C  S  L  L  M  N  D  C  S  Q  Y  D  V  E  G  F  E  K  E  S  H  Y  E  V  E  E  L  K  P  E  K  V  M  N
351...400  S  S  N  F  G  M  V  A  K  M  H  D  F  P  V  K  E  E  V  P  A  G  N  S  E  F  M  R  K  R  K  P  N  R  D  L  N  T  I  M  D  R  T  V  F  T  C  E  N  L
401...450  G  C  A  H  S  E  I  S  R  G  F  L  D  R  N  S  R  D  N  H  Q  L  A  C  P  H  R  D  S  R  L  P  Y  G  A  A  P  S  R  F  H  V  N  E  V  K  P  V  V  G
451...500  F  P  Q  P  R  P  V  N  S  V  A  Q  P  I  D  L  T  G  I  V  P  E  D  G  Q  K  M  I  S  E  L  M  S  M  Y  D  R  N  V  Q  S  N  Q  T  S  M  V  M  E  N
501...550  Q  S  V  S  L  L  Q  P  T  V  H  N  H  Q  E  H  L  Q  F  P  G  N  M  V  E  G  S  F  F  E  D  L  N  I  P  N  R  A  N  N  N  N  S  S  N  N  Q  T  F  F
551...600  Q  G  N  N  N  N  N  V  F  K  F  D  T  A  D  H  N  N  F  E  A  A  H  N  N  N  N  N  S  S  G  N  R  F  Q  L  V  F  D  S  T  P  F  D  M  A  S  F  D
601...650  Y  R  D  D  M  S  M  P  G  V  V  G  T  M  D  G  M  Q  Q  K  Q  Q  D  V  S  I  W  F
```
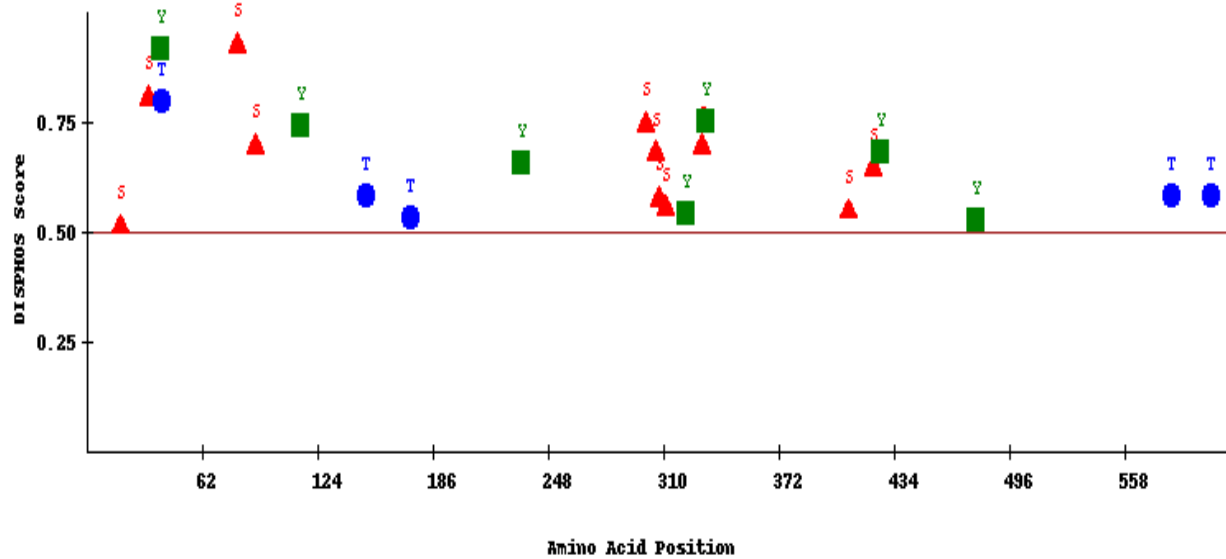
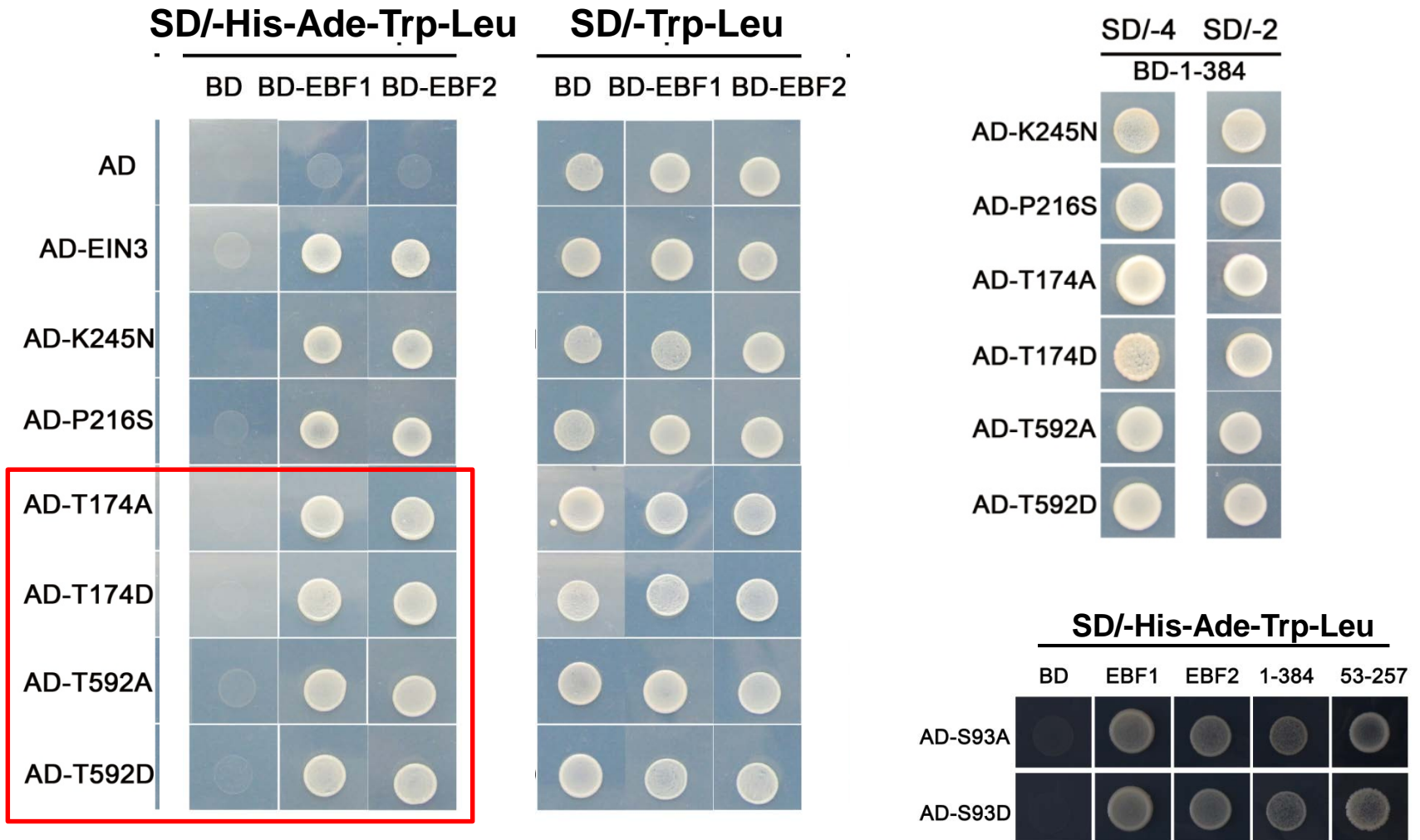# Common predicted phosphorylation sites

- Ser: 19, 83, 93

- Thr: 42, 592

- Tyr: 41

Some of these point mutations didn't affect the interaction with EBF1/EBF2 and EIN3_(1-384) in yeast

Sumoylation site prediction using SUMOsp2 0:

Predicted Sites

Position	Peptide	Score	Cutoff	Type
344	EEL**K**PEK	1.256	0.13	Typel: Ψ-K-X-E
366	FPV**K**EEV	1.322	0.13	Typel: Ψ-K-X-E

SUMOplot™ Analysis Program

Protein ID:	gi|37078133|sp|O24606.1|EIN3_ARATH
Defintion:	RecName: Full=Protein ETHYLENE INSENSITIVE 3
Length:	628 aa

```
  1 MMFNEMGMCG NMDFFSSGSL GEVDFCPVPQ AEPDSIVEDD YTDDEIDVDE
 51 LERRMWRDKM RLKRLKEQDK GKEGVDAAKQ RQSQEQARRK KMSRAQDGIL
101 KYMLKMMEVC KAQGFVYGII PENGKPVTGA SDNLREWWKD KVRFDRNGPA
151 AITKYQAENN IPGIHEGNNP IGPTPHTLQE LQDTTLGSLL SALMQHCDPP
201 QRRFPLEKGV PPPWWPNGKE DWWPQLGLPK DQGPAPYKKP HDLKKAWKVG
251 VLTAVIKHMF PDIAKIRKLV RQSKCLQDKM TAKESATWLA IINQEESLAR
301 ELYPESCPPL SLSGGSCSLL MNDCSQYDVE GFEKESHYEV EELKPEKVMN
351 SSNFGMVAKM HDFPVKEEVP AGNSEFMRKR KPNRDLNTIM DRTVFTCENL
401 GCAHSEISRG FLDRNSRDNH QLACPHRDSR LPYGAAPSRF HVNEVKPVVG
451 FPQPRPVNSV AQPIDLTGIV PEDGQKMISE LMSMYDRNVQ SNQTSMVMEN
501 QSVSLLQPTV HNHQEHLQFP GNMVEGSFFE DLNIPNRANN NNSSNNQTFF
551 QGNNNNNNVF KFDTADHNNF EAAHNNNNNS SGNRFQLVFD STPFDMASFD
601 YRDDMSMPGV VGTMDGMQQK QQDVSIWF
```

■ Motifs with high probability (red)

■ Motifs with low probability (blue)

■ Overlapping Motifs (teal)

No.	Pos.	Group	Score	No.	Pos.	Group	Score
1	K366	MHDFP **VKEE** VPAGN	0.93	4	K219	PWWPN **GKED** WWPQL	0.67
2	K344	YEVEE **LKPE** KVMNS	0.91	5	K72	KEQDK **GKEG** VDAAK	0.50
3	K561	NNNNV **FKFD** TADHN	0.85	6	K248	DLKKA **WKVG** VLTAV	0.47

Common predicted sumoylation sites

- K344

- K366

UbiPred: Prediction of ubiquitylation sites

Position	Sequence	Ubiquitylation	Score
59	DELERRIIRDKIRLKRLKEQD	Y	0.54
63	RRMWRDKMRLKRLKEQDKGKE	N	0.43
66	WRDKMRLKRLKEQDKGKEGVD	N	0.41
70	MRLKRLKEQDKGKEGVDAAKQ	N	0.26
72	LKRLKEQDKGKEGVDAAKQRQ	N	0.18
79	DKGKEGVDAAKQRQSQEQARR	N	0.26
90	QRQSQEQARRKKMSRAQDGIL	N	0.33
91	RQSQEQARRKKMSRAQDGILK	N	0.28
101	KIISRAQDGILKYILKIIEVCK	Y	0.63
105	AQDGILKYMLKMMEVCKAQGF	N	0.44
111	KYMLKMMEVCKAQGFVYGIIP	N	0.33
125	FVYGIIPENGKPVTGASDNLR	N	0.44
139	GASDNLREWWKDKVRFDRNGP	Y	0.82
141	SDNLREWWKDKVRFDRNGPAA	Y	0.83
154	FDRNGPAAITKYQAENNIPGI	N	0.44
208	DPPQRRFPLEKGVPPPWWPNG	N	0.42
219	GVPPPWWPNGKEDWWPQLGLP	N	0.44
230	EDWWPQLGLPKDQGPAPYKKP	N	0.48
238	LPKDQGPAPYKKPHDLKKAWK	N	0.33
239	PKDQGPAPYKKPHDLKKAWKV	N	0.37
244	PAPYKKPHDLKKAWKVGVLTA	Y	0.53
245	APYKKPHDLKKAWKVGVLTAV	Y	0.66
248	KKPHDLKKAWKVGVLTAVIKH	Y	0.58
257	WKVGVLTAVIKHMFPDIAKIR	N	0.23
265	VIKHMFPDIAKIRKLVRQSKC	N	0.39
268	HIIFPDIAKIRKLVRQSKCLQD	Y	0.63
274	AKIRKLVRQSKCLQDKMTAKE	N	0.41
279	LVRQSKCLQDKMTAKESATWL	N	0.32
283	SKCLQDKMTAKESATWLAIIN	N	0.41
334	CSQYDVEGFEKESHYEVEELK	N	0.37
344	KESHYEVEELKPEKVINSSNF	Y	0.60
347	HYEVEELKPEKVMNSSNFGMV	N	0.38
359	MNSSNFGMVAKMHDFPVKEEV	N	0.34
366	MVAKMHDFPVKEEVPAGNSEF	N	0.41
379	VPAGNSEFIRKRKPNRDLNTI	Y	0.58
381	AGNSEFIRKRKPNRDLNTIID	Y	0.68
446	APSRFHVNEVKPVVGFPQPRP	N	0.49
476	LTGIVPEDGQKMISELMSMYD	N	0.16
561	QGNNNNNNVFKFDTADHNNFE	Y	0.51
620	VVGTIIDGIQQKQQDVSIWF	Y	0.51

UbPred: predictor of protein ubiquitination sites

Residue	Score	Ubiquitinated
59	0.50	No
63	0.28	No
66	0.31	No
70	0.31	No
72	0.42	No
79	0.80	**Yes** Medium confidence
90	0.35	No
91	0.35	No
101	0.34	No
105	0.38	No
111	0.48	No
125	0.58	No
139	0.46	No
141	0.34	No
154	0.55	No
208	0.70	**Yes** Medium confidence
219	0.55	No
230	0.59	No
238	0.32	No
239	0.23	No

244	0.10	No
245	0.12	No
248	0.14	No
257	0.27	No
265	0.18	No
268	0.13	No
274	0.13	No
279	0.39	No
283	0.64	**Yes** Low confidence
334	0.92	**Yes** High confidence
344	0.94	**Yes** High confidence
347	0.87	**Yes** High confidence
359	0.75	**Yes** Medium confidence
366	0.78	**Yes** Medium confidence
379	0.15	No
381	0.43	No
446	0.49	No
476	0.54	No
561	0.81	**Yes** Medium confidence
620	0.62	**Yes** Low confidence

※ BDM-PUB: Prediction of Ubiquitination sites with Bayesian Discriminant Method

Go back to BDM-PUB prediction page

Predicted Ubiquitination sites:

Peptide	Position	Score	Threshold
ERRMWRDKMRLKRLK	59	0.78	0.3
WRDKMRLKRLKEQDK	63	1.42	0.3
KMRLKRLKEQDKGKE	66	1.84	0.3
KRLKEQDKGKEGVDA	70	0.83	0.3
LKEQDKGKEGVDAAK	72	0.80	0.3
KEGVDAAKQRQSQEQ	79	1.33	0.3
SQEQARRKKMSRAQD	90	1.54	0.3
QEQARRKKMSRAQDG	91	2.38	0.3
GIIPENGKPVTGASD	125	1.05	0.3
NGPAAITKYQAENNI	154	0.75	0.3
DQGPAPYKKPHDLKK	238	1.37	0.3
QGPAPYKKPHDLKKA	239	2.43	0.3
YKKPHDLKKAWKVGV	244	0.78	0.3
KKPHDLKKAWKVGVL	245	0.41	0.3
HDLKKAWKVGVLTAV	248	2.61	0.3
GVLTAVIKHMFPDIA	257	1.12	0.3
PDIAKIRKLVRQSKC	268	1.65	0.3
RKLVRQSKCLQDKMT	274	0.40	0.3
QSKCLQDKMTAKESA	279	2.77	0.3
SEFMRKRKPNRDLNT	381	1.37	0.3

Common predicted ubiquitination sites

Lysine

- 59, 79, 244, 245, 248, 268, 344, 561, 620

245K⟶N is the *ein3-3* mutation site
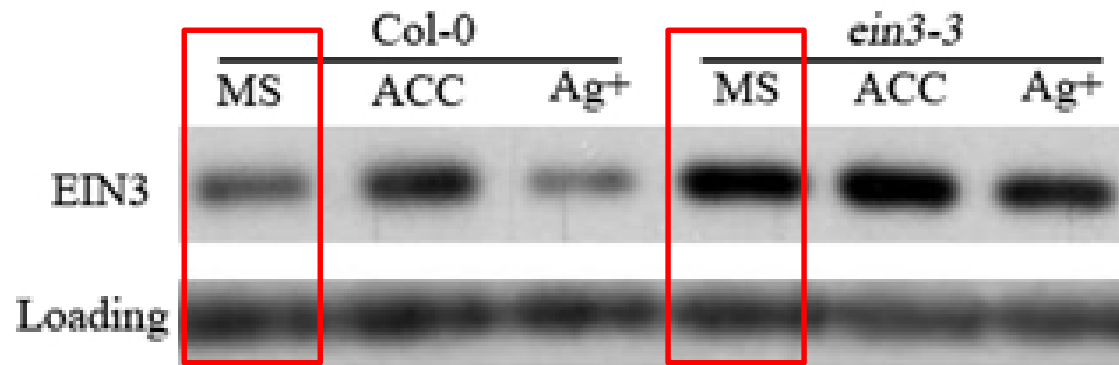
This mutation could stabilize EIN3 protein

Discussion

- Hint from the bioinformatic analysis:

  The conserved motifs and some sites may be necessary for the EIN3 function.

- More experimental evidences are needed:

  Construction of truncation and point mutations, then observe the phenotype and test the protein interaction

Acknowledgement

- Prof. Luo

- Prof. Guo

- All the members in the class

- Our labmates

We enjoy both life and science!