Supplemental Material

The master regulator MAT1-1-1 of fungal mating binds to its targets via a conserved motif in the human pathogen *Aspergillus fumigatus*

Barbara Ramšak*, Jessica Markau*, Tobias Pazen†, Tim A. Dahlmann*, Sven Krappmann†, Ulrich Kück*[1]

*Allgemeine und Molekulare Botanik, Ruhr-Universität Bochum, 44780 Bochum, Germany

† Mikrobiologisches Institut – Klinische Mikrobiologie, Immunologie und Hygiene, Universitätsklinikum Erlangen und Friedrich-Alexander-Universität (FAU) Erlangen-Nürnberg, 91054 Erlangen, Germany

[1] Corresponding author:

Ulrich Kück

E-mail: ulrich.kueck@rub.de

Supplemental files:

File S1. Protein synthesis in *E. coli*.

E. coli cells were grown in 2-YT liquid-medium (1.6% Tryptone, 0.5% NaCl, 1% yeast extract, pH 7.0) containing 35 µg/ml chloramphenicol and 100 µg/ml ampicillin at 37°C. When the cells reached an optical density of A_{600} = 0.6, the temperature was lowered to 20°C and protein expression was induced using 0.5 mM *β*-D-1-thiogalactopyranoside (IPTG). Induced cultures were then incubated an additional 6 h at 20°C. Following incubation, cells were harvested and cell pellets were stored at -70°C until further use. For the cell lysate preparation, we resuspended cells in ice-cold phosphate buffered saline (PBS) buffer (10 mM Na_2HPO_4, 18 mM KH_2PO_4, 140 mM NaCl, 27 mM KCl) containing 5 mM DTT, 0.1 % Protease Inhibitor Cocktail IV (MiliporeSigma, Burlington, Massachusetts) and 1 mM phenylmethylsulfonyl fluoride (PMSF). Cells were then disrupted by 3 rounds of sonification pulses (10% amplitude) for 30 seconds with a pause interval of 30 seconds on ice. Cell lysates were cleared by centrifugation, and 6 µl of clarified lysate from each protein, mixed with 2x Laemmli buffer, were separated by 10% SDS-polyacrylamide gel electrophoresis (SDS-PAGE). Protein expression was analyzed by western blotting.

File S2. High-yield purification of the recombinant AfMAT1-1-1$_{78-235}$.

For high-yield purification of GST-AfMAT1-1-1$_{78-235}$, cleared lysates were prepared from 2 l of induced bacteria using lysis buffer (10 ml of buffer/gram cells) (500 mM NaCl, 27 mM KCl, 10 mM Na_2HPO_4, 18 mM KH_2PO_4) containing 5 mM DTT, 0.1% Protease Inhibitor Cocktail IV (MiliporeSigma, Burlington, Massachusetts) and 1 mM PMSF. Prior to cell disruption, 5 mg of DNase I (Sigma-Aldrich, St. Louis, Missouri) was added per 100 ml of cell suspension and the mixture was then disrupted with a microfluidizer M-100L (Microfluidics) at 13,000 psi while constantly cooling on ice. The cell suspension was centrifuged at 4°C to remove cell debris. The supernatant was filtered through a 0.22 µm filter and loaded onto a pre-equilibrated 20 ml GSTrap™ (GE Healthcare, Chicago, Illinois) glutathione agarose column. Unbound material was washed with the lysis buffer in the absence of protease inhibitors. Proteins were subsequently eluted using a glutathione elution buffer (50 mM Tris, 500 mM NaCl, 5 mM DTT, 20 mM L-glutathione reduced). Fractions of unbound proteins (washing step) and fractions containing the protein (elution step), were collected and analysed by 10% SDS-PAGE. Elution fractions with the highest protein amount were pooled together and concentrated using Amicon® Ultra centrifugal filter with a 10,000 Da molecular weight cut-off (MWCO). The GST-tag and AfMAT1-1-1$_{78-235}$ were separated overnight by TEV digestion using a protein to enzyme ratio of 50:1 (w:w), while dialyzing against dialysis buffer (50 mM HEPES [pH 7.5], 500 mM NaCl, 1 mM EDTA, 2 mM DTT) in a 6,000 – 8,000 Da (MWCO) tubing (Thermo Fisher Scientific, Waltham, Massachusetts). Next, GST-tag and GST-tagged TEV protease were removed using Glutathione Sepharose (GE Healthcare, Chicago, Illinois) beads. Finally, size-exclusion chromatography was performed using a Hiload^{TM} 16/60 Superdex^{TM} 75 pg column (GE Healthcare, Chicago, Illinois) in the dialysis buffer.

File S3. Protein quantification and western blotting.

The concentration of purified proteins was determined using the Bradford assay. Protein samples were boiled in 2x Laemmli buffer for 5 min before SDS-PAGE, then gels were equilibrated in blocking buffer (25 mM Tris, 190 mM Glycine and 0.1% SDS (w/v)) before western blotting onto PVDF membranes (GE Healthcare, Chicago, Illinois). Non-specific binding was prevented by incubating membranes in PBST blocking buffer (80 mM Na_2HPO_4, 21 mM NaH_2PO_4, 100 mM NaCl (pH 7.5), 0.1% Tween 20 (v/v)), containing 5% dried milk powder, for 1 h at room temperature, and proteins were detected using a 1:5000 dilution of an anti-GST HRP conjugate (RPN1236V) (GE Healthcare, Chicago, Illinois) in PBST buffer for 1 h at room temperature. Proteins were visualized using ChemiDoc XRS+ Imaging System (Bio-Rad, Hercules, California).

Supplemental tables:

Table S1. List of oligonucleotides used in this study. Oligonucleotide specificity and orientation are given (f: forward; r: reverse). The AfMAT1-1-1 binding motifs are underlined and the chromosomal positions of the oligonucleotides carrying corresponding binding motif are indicated.

Oligonucleotide	**Sequence (5' to 3')**	**Specificity**
Plasmids construction and sequencing		
AfMAT1-1_f	ATATACCATGGAAGCTGCAATCTCTCC	*A. fumigatus MAT1-1-1* gene; used for vectors pGEX-AfMAT1-1-1_F1, -F4, -F6; *Nco*I restriction site
AfMAT1-1_r	ATATAGAATTCTCAGACGTTGATGTATTGATCAATGTC	*A. fumigatus MAT1-1-1* gene; used for vectors pGEX-AfMAT1-1-1_F3; *Eco*RI restriction site
AfMAT1-1_F4_r	ATATAGAATTCTCACATTGTGGAAGTATTGTTGCCATTAC	*A. fumigatus MAT1-1-1* gene; used for vectors pGEX-AfMAT1-1-1_F4, -F5; *Eco*RI restriction site
AfMAT1-1_F6_r	ATATAGAATTCTCAAGTGATCTGAATGCCGTTCTC	*A. fumigatus MAT1-1-1* gene; used for vectors pGEX-AfMAT1-1-1_F6, -F7; *Eco*RI restriction site
AfMAT1-1_d_f	ATATACCATGGCCAAACGCACCCAG	*A. fumigatus MAT1-1-1* gene; used for vectors pGEX-AfMAT1-1-1_F2, -F3, -F5; *Nco*I restriction site
AfMAT1-1_d_r	ATATAGAATTCTCAATCGGCTTCAGGAATTGTCG	*A. fumigatus MAT1-1-1* gene; used for vectors pGEX-AfMAT1-1-1_F1, -F2; *Eco*RI restriction site
pGex_f	ATAGCATGGCCTTTGCAG	pGEX-4T-1-TEV backbone; used for sequencing
pGex_r	GAGCTGCATGTGTCAGAG	pGEX-4T-1-TEV backbone; used for sequencing
qRT-PCR		
ES077	ACACATTACGCCGTGGTGCC	*ppgA*
ES078	ATCGGCAGAGCGCTTCACCT	*ppgA*
TP033	AACGTCATTTGCGTTGGTGC	*tomA*
TP034	TTCTTGGAGTTCGGTGCTTCG	*tomA*
TP039	TCGTTGCCCGATAGTCCTAC	*cp9*
TP040	CCCATTGCCTGTCATCAACC	*cp9*
TP043	AACAATCAGGCGTGGGATGA	*AFUB_070880*
TP044	CGCATTGGGAATTGGGATGG	*AFUB_070880*

EMSA		
tom1-2_f	TCACGTGATCTCTATTGAGAACAATAGAA	chr2:3627755-3627783
tom1-2_r	CTTCTATTGTTCTCAATAGAGATCACGTG	
ppg1-2_f	CAGTTCTCAATAGGAATCTTATTGACCGA	chr1:278174-278202
ppg1-2_r	CGATGGAATGAATGAGTTTATTATTGATC	
pre1-2_f	AGATCAATAATAAACTCATTCATTCCATC	chr1:3250686-3250714
pre1-2_r	CGATGGAATGAATGAGTTTATTATTGATC	
tomA-1_f	TCCTATTCTGCCTATTGATGGGAAAGAAC	chr4:1566229-1566257
tomA-1_r	TGTTCTTTCCCATCAATAGGCAGAATAGG	
tomA-2_f	AACAATACCAACTCAATAAGGAGGCCGGG	chr4:1566283-1566256
tomA-2_r	CCCCGGCCTCCTTATTGAGTTGGTATTGT	
ppgA-1_f	AAGGCTACCTCTTATTGAGAAGGTGATTT	chr6:1366658-1366630
ppgA-1_r	AAAATCACCTTCTCAATAAGAGGTAGCCT	
ppgA-2_f	TAAAAATGTGGATCAATAACACAAAGGCT	chr6:1366652-1366680
ppgA-2_r	TAGCCTTTGTGTTATTGATCCACATTTTT	
preA-1_f	AAACAGCGGGCTTATTGACACCCAGAAAG	chr5:1975777-1975805
preA-1_r	CCTTTCTGGGTGTCAATAAGCCCGCTGATT	
preA-2_f	AAGTCGATCTTCTCAATAGTCAGGTGATG	chr5:1975984-1975920
preA-2_r	TCATCACCTGACTATTGAGAAGATCGACT	
preA-3_f	CATGGATGCAATCATTGACTGAATTACTT	chr5:1975999-1976027
preA-3_r	CAAGTAATTCAGTCAATGATTGCATCCAT	
tomA-1_m1_f	TCCTATTCTGCCT**g**TTGATGGGAAAGAAC	chr4:1566229-1566257; A → G at 1566242
tomA-1_m1_r	TGTTCTTTCCCATCAACAGGCAGAATAGG	
tomA-1_m2_f	TCCTATTCTGCCT**g**TT**a**ATGGGAAAGAAC	chr4:1566229-1566257; A → G at 1566242 and G → A at 1566245

tomA-1_m2_r	TGTTCTTTCCCATTAACAGGCAGAATAGG	
ppgA-1_m1_f	AAGGCTACCTCTT**g**TTGAGAAGGTGATTT	chr6:1366658-1366630; A → G at 1366645
ppgA-1_m1_r	AAAATCACCTTCTCAACAAGAGGTAGCCT	
ppgA-1_m2_f	AAGGCTACCTCTT**g**TT**a**AGAAGGTGATTT	chr6:1366658-1366630; A → G at 1366645 and G → A at 1366642
ppgA-1_m2_r	AAAATCACCTTCTTAACAAGAGGTAGCCT	

Table S2. List of plasmids used in this study.

Name	Description	Source
pUC19-AfMAT1-1	General cloning vector containing *MAT1-1-1* cDNA (1-1107 bp) from *A. fumigatus*	Krappmann, unpublished
pGEX-4T-1-TEV	overexpression of the recombinant proteins in *E.coli* strain BL21 (DE3) pLysS	Laboratory stock
pGEX-MAT1	*MAT1-1-1* cDNA from *P. chrysogenum*; overexpression of a full-length GST-MAT1-1-1 (1-342 aa)	(Becker *et al.* 2015)
pGEX-AfMAT1-1-1	*MAT1-1-1* cDNA (1-1107 bp) from *A. fumigatus*; overexpression of a GST-MAT1-1-1 (1-368 aa)	This study
pGEX-AfMAT1-1-1_F1	*MAT1-1-1* cDNA (1-591 bp) from *A. fumigatus*; overexpression of a GST-MAT1-1-1_{1-197} (1-197 aa)	This study
pGEX-AfMAT1-1-1_F2	*MAT1-1-1* cDNA (232-591 bp) from *A. fumigatus*; overexpression of a GST-MAT1-1-1_{78-197} (78-197 aa)	This study
pGEX-AfMAT1-1-1_F3	*MAT1-1-1* cDNA (232-1107 bp) from *A. fumigatus*; overexpression of a GST-MAT1-1-1_{78-368} (78-368 aa)	This study
pGEX-AfMAT1-1-1_F4	*MAT1-1-1* cDNA (1-705 bp) from *A. fumigatus*; overexpression of a GST-MAT1-1-1_{1-235} (1-235 aa)	This study
pGEX-AfMAT1-1-1_F5	*MAT1-1-1* cDNA (232-705 bp) from *A. fumigatus*; overexpression of a GST-MAT1-1-1_{78-235} (78-235 aa)	This study
pGEX-AfMAT1-1-1_F6	*MAT1-1-1* cDNA (1-762 bp) from *A. fumigatus*; overexpression of a GST-MAT1-1-1_{1-254} (1-254 aa)	This study
pGEX-AfMAT1-1-1_F7	*MAT1-1-1* cDNA (232-762 bp) from *A. fumigatus*; overexpression of a GST-MAT1-1-1_{78-254} (78-254 aa)	This study

Table S3. Theoretical molecular weights (MWs) of *A. fumigatus* MAT1-1-1 proteins.

Protein construct	**Amino acid position**	**Theoretical [a, b] MW (Daltons)**
GST-AfMAT1-1-1[c]	1-368	40,800
GST-AfMAT1-1-1$_{1\text{-}197}$	1-197	22,300
GST-AfMAT1-1-1$_{78\text{-}197}$	78-197	14,000
GST-AfMAT1-1-1$_{78\text{-}368}$	78-368	32,400
GST-AfMAT1-1-1$_{1\text{-}235}$	1-235	26,300
GST-AfMAT1-1-1$_{78\text{-}235}$	78-235	18,000
GST-AfMAT1-1-1$_{1\text{-}254}$	1-254	28,300
GST-AfMAT1-1-1$_{78\text{-}254}$	78-254	20,000

[a] the theoretical molecular weight predicted by ProtParam tool (Gasteiger *et al.* 2005).
[b] MWs of the proteins without N-terminal GST-tag (26,000 Da).
[c]full-length AfMAT1-1-1.

Supplemental figures:

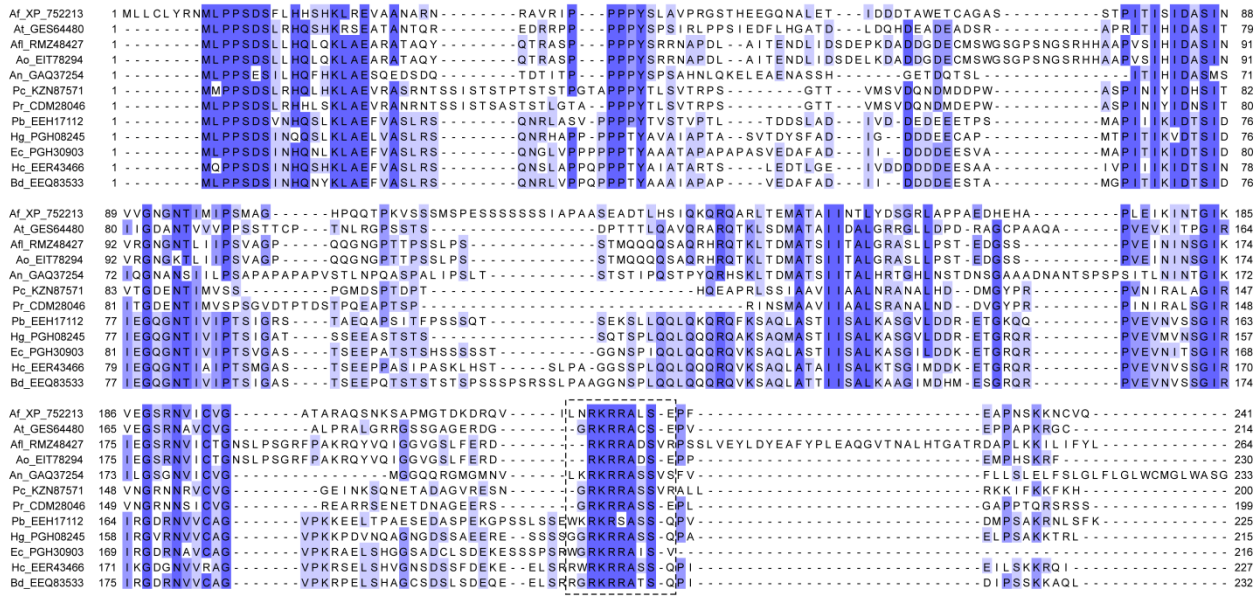

Figure S1. Multiple alignments of TomA orthologs. The *tomA* gene is found exclusively in the subclass of Eurotiomycetes, Eurotiomycetidae. The amino acid sequence of TomA from *A. fumigatus* (Af_XP_752213) was aligned with identified orthologs *Aspergillus terreus* (At_GES64480), *Aspergillus flavus* (Afl_RMZ48427); *Aspergillus oryzae* (Ao_EIT78294); *Aspergillus niger* (An_ GAQ37254); *Penicillium chrysogenum* (Pc_KZN87571); *Penicillium roqueforti* (Pr_CDM28046); *Paracoccidioides brasiliensis* (Pb_EEH17112); *Helicocarpus griseus* (Hg_PGH08245); *Emmonsia crescens* (Ec_PGH30903); *Histoplasma capsulatum* (Hc_EER43466); *Blastomyces dermatitidis* (Bd_EEQ83533). Identical and conserved residues are shaded in dark blue and light blue, respectively. Potential nuclear localization site (NLS) (dashed rectangle) was predicted using cNLS Mapper (Kosugi *et al.* 2009).

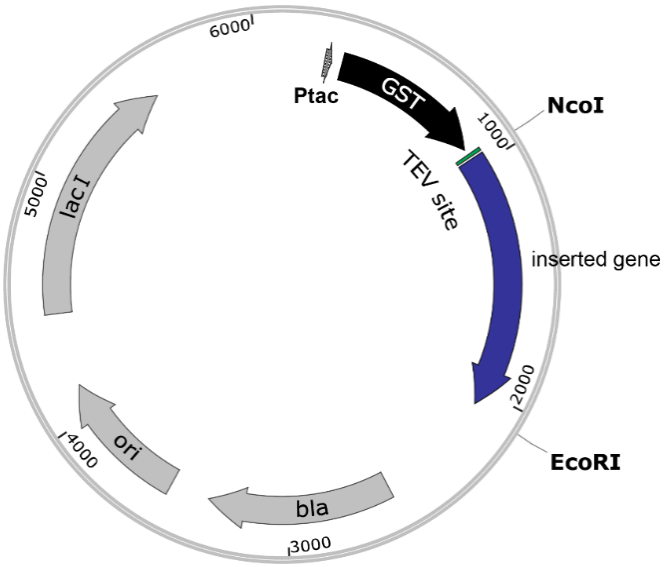

Figure S2. Schematic diagram of the pGEX-4T-1-TEV bacterial expression vector. The pGEX-4T-1-TEV vector was used for protein expression of the GST-AfMAT1-1-1 and its truncated derivatives. The expression of inserted gene versions was controlled by the inducible tac promoter (Ptac). GST-tag, indicated in black, may be cleaved off at the tobacco etch virus (TEV) protease recognition site, shown in green. Truncated *AfMAT1-1-1* gene versions were inserted at the indicated position.

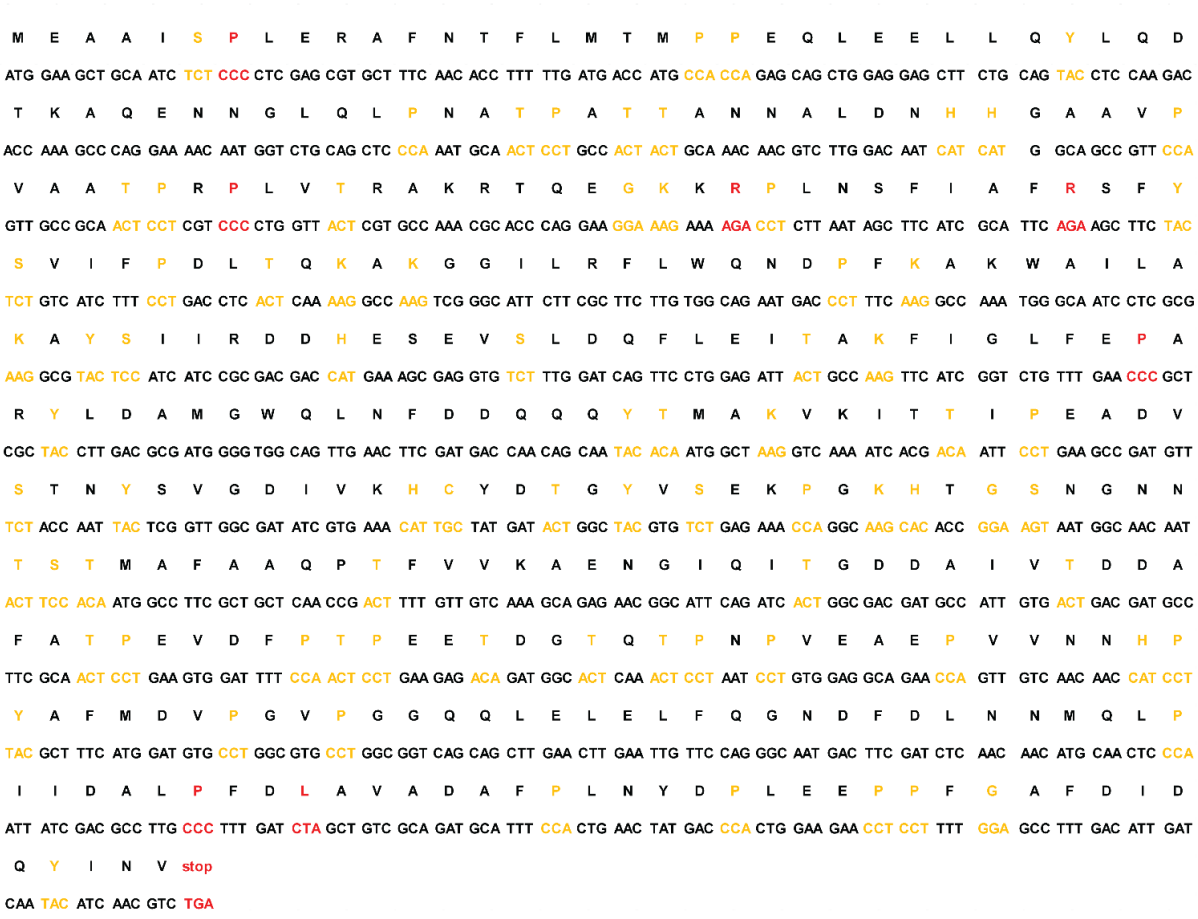

Figure S3. Distribution of rare (orange) and highly rare codons (red) in the coding sequence of *AfMAT1-1-1*. The codon optimization tool, freely available online (Daniel *et al.* 2015) was used to identify codons that are rarely used by *E. coli*. The *E. coli* strain B codon usage table was acquired from the codon usage database: https://www.kazusa.or.jp/codon/.

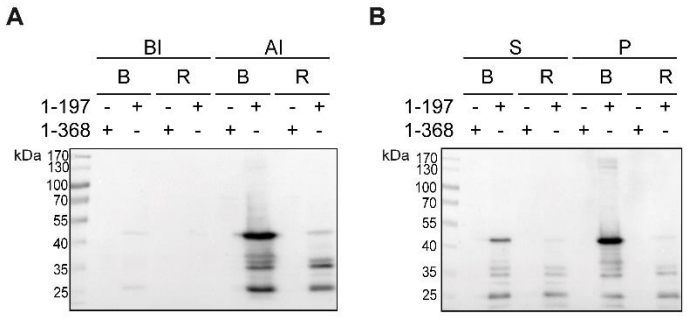

Figure S4. SDS-PAGE analysis of full-length GST-AfMAT1-1-1 and shortened variant GST-AfMAT1-1-1$_{1-197}$ expression in *E. coli*. (A) *E.coli* BL21 pLysS (B) and Rosetta (R) cells expressing GST-tagged full-length GST-AfMAT1-1-1 (1-368) and GST-AfMAT1-1-1$_{1-197}$ (1-197) were induced (AI) as described in Materials and Methods section. Protein fractions from uninduced cells (BI) are shown. (B) Cell lysate fractions supernatant (S) and pellet (P) were examined for GST-AfMAT1-1-1 (1-368) and GST-AfMAT1-1-1$_{1-197}$ (1-197) protein yields in *E.coli* BL21 pLysS (B) and Rosetta (R) cells. Predicted MW for GST-AfMAT1-1-1 (1-368) and GST-AfMAT1-1-1$_{1-197}$ (1-197) are 66.8 and 48.3 kDa, respectively.

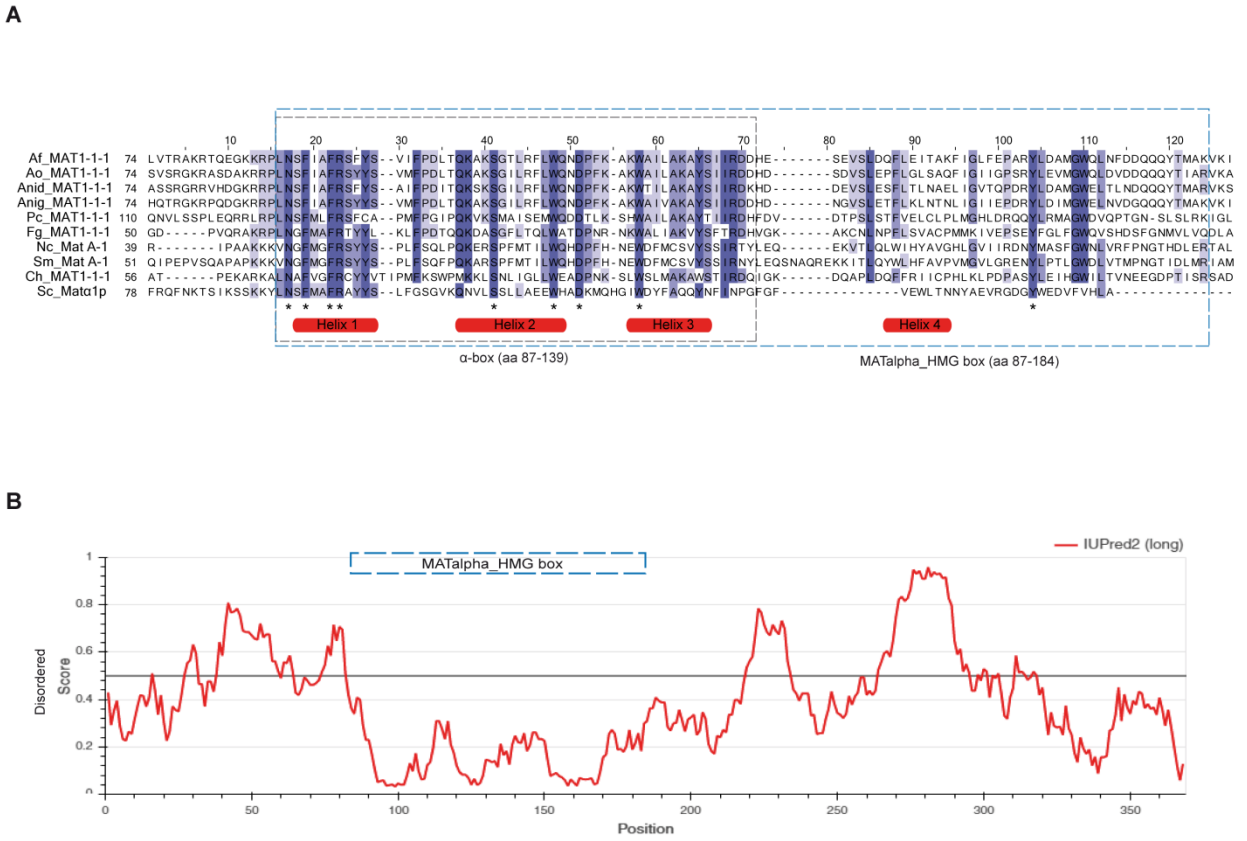

Figure S5. Multiple sequence alignment of selected alpha-box domains from representative species of Ascomycota. (A) The sequences of MAT1-1-1 proteins from *Aspergillus fumigatus* (Af, AAXB3123), *Aspergillus oryzae* (Ao, XP_001824461), *Aspergillus nidulans* (Anid, XP_660359), *Aspergillus niger* (Anig, XP_001394976), *Penicillium chrysogenum* (Pc, KZN86912), *Fusarium graminearum* (Fg, CEF77353), *Neurospora crassa* (Nc, P19392), *Sordaria macrospora* (Sm, KAA8630680), *Cochliobolus heterostrophus* (Ch, Q02990), *Saccharomyces cerevisiae* (Sc, ONH75384) were retrieved from the National Center for Biotechnology Information (NCBI) database. The *A. fumigatus* sequence was used as a template in a NCBI-BLASTP suite (Camacho *et al.* 2009). Conserved alpha-box domain (~55 amino acids) is indicated by the black dashed rectangle and the domain; as defined in the PFAM database (PF04769) MATalpha_HMG box, is indicated by the dashed blue rectangle. Conserved residues are highlighted in blue, and the fully conserved residues are marked by an asterisk (*) under the alignment. The numbering indicates the amino acid position relative to the N-terminus in each predicted protein. Secondary structure elements, predicted by Jpred server (Drozdetskiy *et al.* 2015) are shown under the alignment. **(B)** Disorder of full-length AfMAT1-1-1 as predicted by the IUPred2A server (Mészáros *et al.* 2018). Residues above threshold 0.5 are considered disordered. The position of the MATalpha_HMG box domain is indicated.

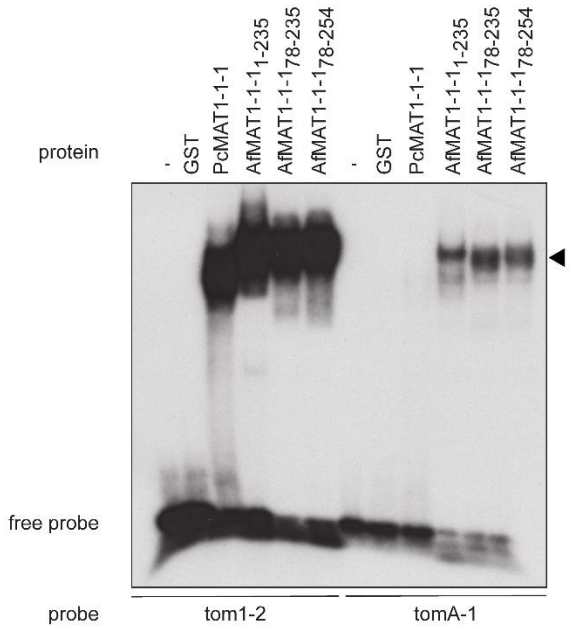

Figure S6. Binding of various shortened variants of AfMAT1-1-1 to the tom1-2 and tomA-1 dsDNA probes. 10 µg of each purified variant of AfMAT1-1-1: AfMAT1-1-1$_{1-235}$, AfMAT1-1-1$_{78-235}$, and AfMAT1-1-1$_{78-254}$ were incubated with radiolabeled DNA probes originating from *tom1* promoter from *P. chrysogenum* (tom1-2) and *tomA* promoter from *A. fumigatus* (tomA-1). The lanes are labeled as follows: "–", dsDNA probe without protein; "GST", GST protein; "AfMAT1-1-1$_{1-235}$", GST-tagged AfMAT1-1-1$_{1-235}$; "AfMAT1-1-1$_{78-235}$", GST-tagged AfMAT1-1-1$_{78-235}$; "AfMAT1-1-1$_{78-254}$", GST-tagged AfMAT1-1-1$_{78-254}$. Black arrow head indicates formation of the protein-DNA complexes**.**

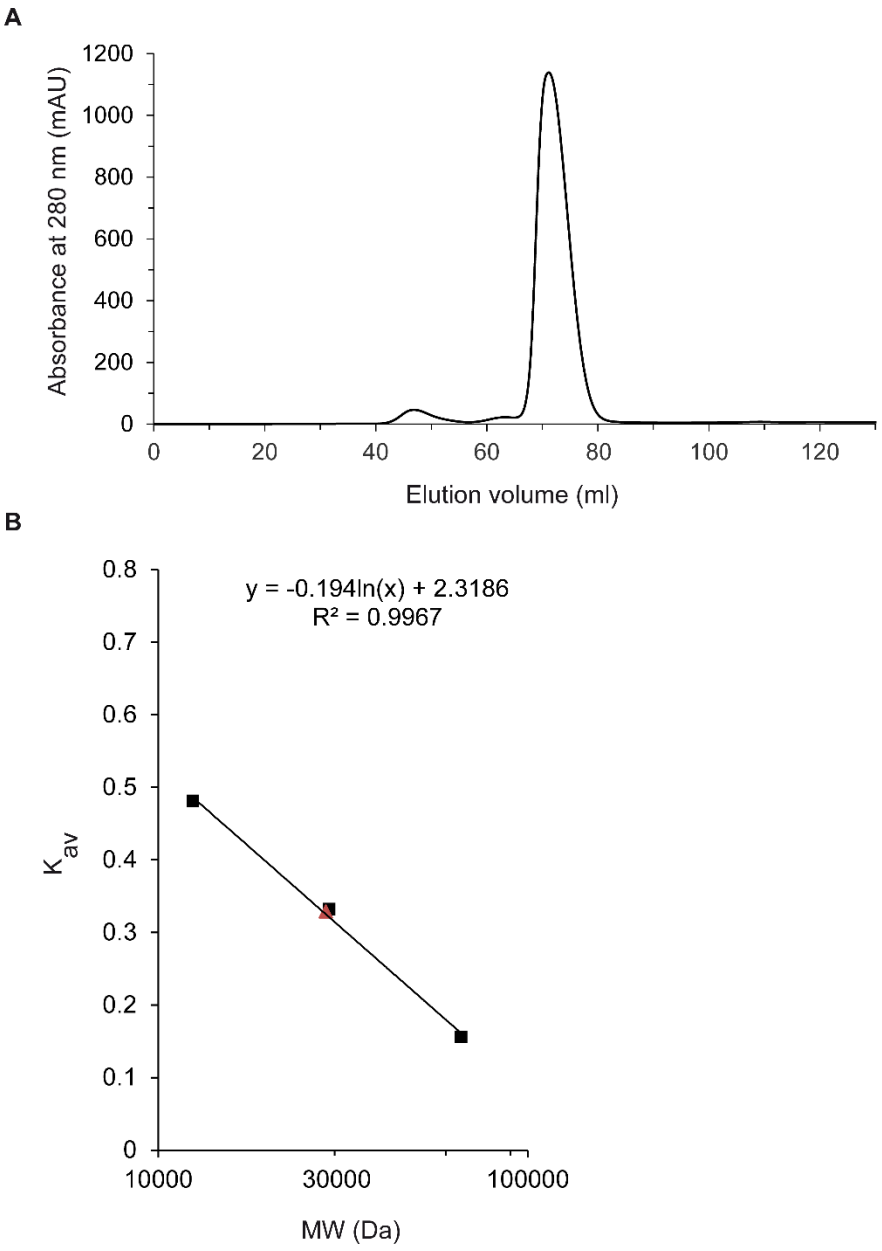

Figure S7. Size-exclusion chromatography (SEC) elution profile of AfMAT1-1-1$_{78-235}$. (A) Eluate fractions containing pure AfMAT1-1-1$_{78-235}$ (*Figure 3E*) were pooled together concentrated and loaded on the Superdex 75 HiLoad 16/600 *prep grade* column. The SEC chromatogram shows that AfMAT1-1-1$_{78-235}$ eluted at the volume of 71.1 ml. **(B)** Calibration curve of the Superdex 75 HiLoad 16/600 *prep grade* column using following molecular weight (MW) protein standards: albumin (66 kDa), carbonic anhydrase (29 kDa) and cytochrome C (12.4 kDa). The corresponding elution volume of AfMAT1-1-1$_{78-235}$ is indicated in red. The theoretical MW for AfMAT1-1-1$_{78-235}$ is 18 kDa. Based on the calibration curve, the estimated MW for AfMAT1-1-1$_{78-235}$ is 28.4 kDa, indicating that AfMAT1-1-1$_{78-235}$ most likely forms a dimer.

References

Becker K., C. Beer, M. Freitag, and U. Kück, 2015 Genome-wide identification of target genes of a mating-type α-domain transcription factor reveals functions beyond sexual development. Mol. Microbiol. 96: 1002–1022. https://doi.org/10.1111/mmi.12987

Camacho C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, *et al.*, 2009 BLAST+: architecture and applications. BMC Bioinformatics 10: 421. https://doi.org/10.1186/1471-2105-10-421

Daniel E., G. U. Onwukwe, R. K. Wierenga, S. E. Quaggin, S. J. Vainio, *et al.*, 2015 ATGme: Open-source web application for rare codon identification and custom DNA sequence optimization. BMC Bioinformatics 16: 303. https://doi.org/10.1186/s12859-015-0743-5

Drozdetskiy A., C. Cole, J. Procter, and G. J. Barton, 2015 JPred4: a protein secondary structure prediction server. Nucleic Acids Res. 43: W389–W394. https://doi.org/10.1093/nar/gkv332

Gasteiger E., C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, *et al.*, 2005 *Protein Analysis Tools on the ExPASy Server 571 571 From: The Proteomics Protocols Handbook Protein Identification and Analysis Tools on the ExPASy Server*.

Kosugi S., M. Hasebe, M. Tomita, and H. Yanagawa, 2009 Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. Proc. Natl. Acad. Sci. U. S. A. 106: 10171–10176. https://doi.org/10.1073/pnas.0900604106

Mészáros B., G. Erdős, and Z. Dosztányi, 2018 IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res. 46: W329–W337. https://doi.org/10.1093/nar/gky384