Short linear motifs - Underexplored players driving *Toxoplasma gondii* infection

Jesús Alvarado Valverde[1,2,3*], Karine Lapouge[4], Arne Boergel[4], Kim Remans[4], Katja Luck[3*], Toby J. Gibson[1*].

1 Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg 69117, Germany

2 Collaboration for a joint PhD degree between EMBL and Heidelberg University, Faculty of Biosciences

3 Institute of Molecular Biology (IMB) gGmbH, 55128 Mainz, Germany

4 Protein Expression and Purification Core Facility, European Molecular Biology Laboratory, Heidelberg 69117, Germany

* Correspondence should be addressed to j.alvarado@imb-mainz.de, k.luck@imb-mainz.de or toby.gibson@embl.de

Abstract

Pathogens infect hosts by interacting with host proteins and exploiting their functions to their advantage. Short linear motifs, small functional regions within intrinsically disordered protein regions, are common mediators of host-pathogen protein interactions. While motifs have been more extensively studied in viruses and bacteria, the extent to which eukaryotic unicellular parasites use motifs during infection remains unexplored. *Toxoplasma gondii* is a widespread intracellular Apicomplexan parasite capable of infecting all warm-blooded animals and invading any of their nucleated cells. *Toxoplasma's* secreted proteins are key in interacting with host proteins during infection, making them potential sources for motifs. To highlight the role of motifs in *Toxoplasma gondii* infection, we curated 21 known motif instances in Toxoplasma proteins from the scientific literature. To identify more motifs in Toxoplasma secreted proteins, we developed a computational pipeline that annotates putative motif matches with structural and functional features. Through this approach, we identified a set of 24,291 motif matches in 295 secreted proteins. We highlight strategies for further prioritisation of likely functional motif matches by focusing on integrin motifs, degrons and TRAF6-binding motifs. We subjected four predicted TRAF6-binding motifs to experimental validation, supporting the predicted motifs in the Toxoplasma proteins RON10 and GRA15. Our motif predictions provide a valuable resource for generating hypotheses and designing experiments to study infection mechanisms. The characterisation of motifs in *Toxoplasma* will be key to understanding the molecular principles underlying its broad host range and more comprehensive Apicomplexan infection strategies.

Importance

Toxoplasma gondii is a widely distributed intracellular parasite that achieves a successful infection by interacting with different host cell proteins. Short linear motifs are small functional modules found in unstructured protein regions and recognised by folded protein domains. Given that unstructured protein regions are a common feature of Toxoplasma's proteins, we hypothesise that motifs play important roles during its infection cycle. Here, we highlight the role of motifs during the Toxoplasma host cell invasion cycle through a curated set of motif examples. Through a computational pipeline, we predict thousands of motifs in secreted proteins, outline strategies for working with these predictions and finally experimentally test proteins containing a motif involved in the innate immune response, successfully showing the binding of two motifs. Our work provides a resource for further motif testing in Toxoplasma proteins, aiming at understanding the molecular mechanisms of its infection strategies and its broad host range.

Introduction

Intracellular pathogens, such as viruses, bacteria or eukaryotic parasites, interact with host cell components throughout their infection cycles. These pathogens can attach to the host membrane, and when internalised, they can interact with the host cytoskeleton and organelles, as well as with its export and translation machinery (1–3). By interfering with host biological processes, pathogens modulate the host immune response, replicate and disperse. Pathogen proteins play a central role in these interactions. A common mode of protein binding that mediates host-pathogen protein-protein interactions (PPIs) consists of a folded domain, e.g. in host proteins, and short linear motifs (hereafter referred to as motifs) in pathogen proteins (2–4). Motifs are

usually between 3 and 10 amino acids long and occur in intrinsically disordered regions of proteins (IDRs)(5,6). Motifs mediate protein interactions that primarily participate in regulatory and cell signalling processes. Motifs define sites for post-translational modifications, subcellular protein localisation, as well as substrate recognition signals for enzymes such as E3 ligases, proteases, phosphatases, etc (5). Motifs are unstructured in their unbound state and often adopt an alpha-helix or beta-strand upon binding to partner proteins (6). Because only a few mutations are needed to create a motif, they can evolve in much shorter evolutionary timescales compared to folded protein regions (7). Thus, pathogens have evolved to exploit these host motifs to infect their hosts, probably because motifs are short and easily emerge in proteins (7), but also because of their many functional roles in cell signalling. Many instances of host motif mimicry have been characterised for viruses and bacteria (2,3,8–11). Although there have been some studies focusing on intracellular eukaryotic parasites (12,13), the extent to which parasites use motifs is less understood. *Toxoplasma gondii,* a member of the Apicomplexa phylum of unicellular parasites, has one of the proteomes with higher levels of IDRs among Eukaryotes (14,15), making it a good candidate to discover and characterise the role of motifs in its proteins during parasitic infection.

Toxoplasma is a successful eukaryotic unicellular parasite, as it can infect any warm-blooded animal and proliferate within any of its nucleated cells. It is estimated that a third of the human population has been infected with Toxoplasma (16). Even though most of the infections in humans are asymptomatic, immunocompromised individuals and susceptible populations can develop different forms of Toxoplasmosis, from influenza-like symptoms to severe phenotypes such as encephalitis, making

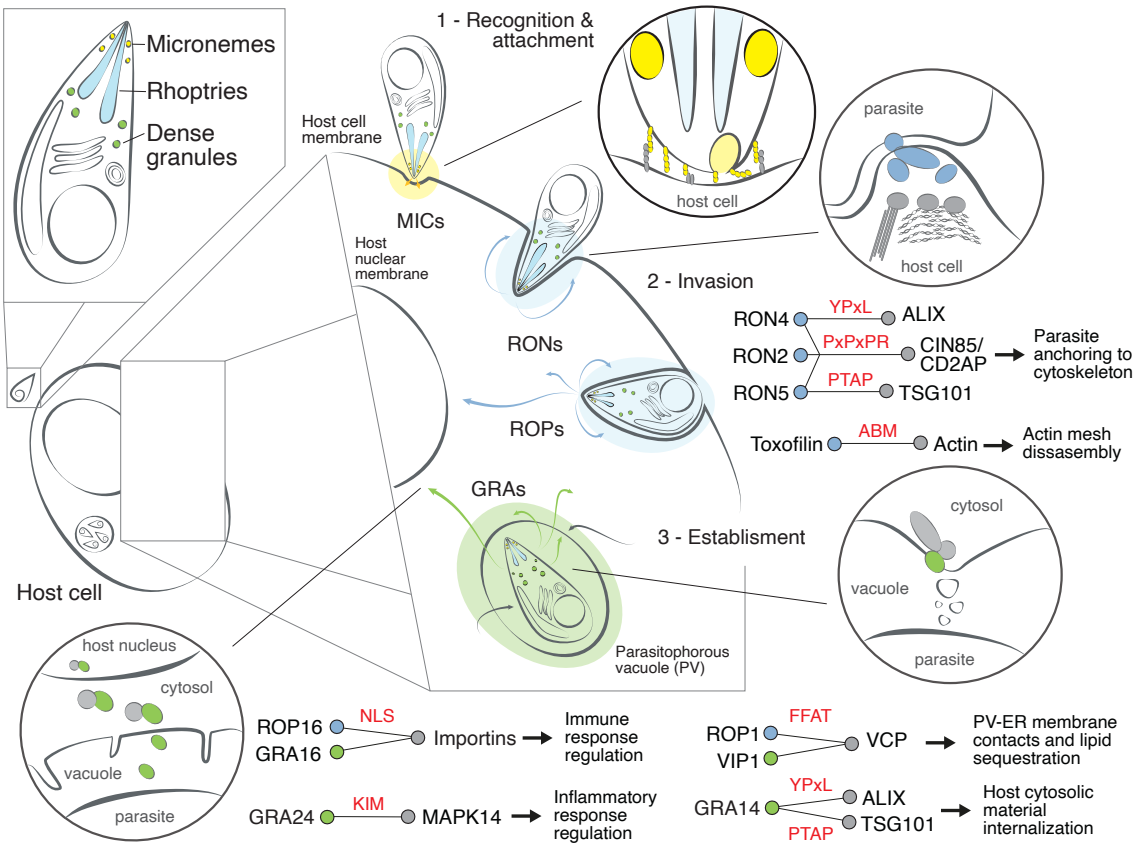
102 *Toxoplasma* a global health burden (16,17). *Toxoplasma's* definitive hosts are cats
103 and all members of the feline family, where it reproduces sexually in their gut to then
104 disseminate in the environment. In all other hosts, such as all mammals and birds,
105 *Toxoplasma* reproduces asexually in different cell types to multiply its numbers and
106 disseminate along the food chain (18). Having multiple hosts means that *Toxoplasma*
107 not only has the molecular tools to infect different host cells but also to deal with
108 diverse cell programmes and immune responses, to thrive in each of its hosts. Motifs
109 in Toxoplasma proteins might play important roles in mediating the adaptation to the
110 many different host environments this parasite encounters.
111
112 Once consumed by its host, *T. gondii* cell invasion cycles depend on the sequential
113 secretion of proteins contained in its specialised infection organelles: the micronemes,
114 rhoptries and dense granules (Fig. 1A). Microneme proteins, or MICs, are targeted to
115 the parasite cell surface and are involved in host recognition and attachment, while
116 rhoptry proteins, termed RONs and ROPs, are secreted into the host cytoplasm and
117 are involved in host invasion and niche establishment (Fig. 1B). During the invasion
118 process, a parasitophorous vacuole (PV) is created from the invagination of the host
119 cell membrane (19). From within this space, *Toxoplasma* creates a suitable
120 environment and continues to secrete proteins from membraneless organelles called
121 dense granules. These proteins, mostly known as GRAs, are secreted either into the
122 PV lumen, inserted into its membrane (PVM) or into the host cytoplasm and even into
123 the host nucleus, helping the parasite in modulating the host cell's regulatory
124 processes to proliferate (20). All these secreted proteins interact with host proteins
125 throughout the infection cycle, including host cell receptors, cytoskeletal proteins and
126 transcription factors (21,22). These host-parasite protein interactions are the subject

127 of continuous study, serve as a reference to study other Apicomplexans, and are
128 promising targets for the development of therapies. Knowledge about motifs in
129 *Toxoplasma* secreted proteins would increase our mechanistic understanding of
130 Toxoplasma infection strategies.
131
132 To identify motifs in *Toxoplasma gondii* proteins, we collected motif instances in *T.*
133 *gondii* from the literature, predicted occurrences of known motif types in its secreted
134 proteins through the development of an integrative discovery pipeline, and tested the
135 binding of predicted TRAF6 binding motifs *in vitro,* which might have relevance in the
136 regulation of innate immune response. The available motif predictions along with
137 annotations will motivate further functional studies in *Toxoplasma* and other
138 Apicomplexan parasites to understand mechanisms of infection and provide further
139 examples of motif repurposing among intracellular pathogens.
140

141 **Results**

142 Curation of known motifs in *Toxoplasma gondii* secreted proteins
143 We collected 21 experimentally validated motif instances from 11 Toxoplasma proteins
144 from the scientific literature that serve as a reference of the current motif knowledge
145 in *Toxoplasma gondii* (Table 1, Fig. 1). The Eukaryotic Linear Motif (ELM) resource is
146 a reference point for curated eukaryotic motifs from the scientific literature. Based on
147 motif sequence patterns, structures, binder information, and functional annotations,
148 curators define motif types and group motif types into broader motif classes (23). Two
149 of the Toxoplasma motif instances were previously annotated as substrate recognition
150 sites in ELM (DOC). Eleven motif instances were novel with respect to ELM content
151 and were added to known motif classes of scaffolding classes in ELM (LIG). The

152 remaining eight novel motif instances are similar but were not able to be directly
153 annotated to ELM. Those correspond to one scaffolding class (LIG) and to three
154 subcellular localisation classes (TRG in ELM). We did not annotate motif instances in
155 Toxoplasma secreted proteins reported as motifs of cleavage sites (CLV in ELM),
156 degradation signals (DEG in ELM) or sites for post-translational modifications (MOD
157 in ELM). As detailed below, the curated 21 motif instances function throughout parasite
158 invasion and establishment (Fig. 1).
159



160

161 **Figure 1. Motif use and hijacking during the *Toxoplasma* infection process.** *Toxoplasma*
162 *gondii* secretory organelles and infection cycle illustrating different host-pathogen interaction
163 contexts, known motif instances and their infection outcomes. Motif names are highlighted in
164 red, proteins from micronemes are highlighted in yellow, those from rhoptries in blue, and
165 those from dense granules in green.

Protein name	ToxoDB (UniProt)	Motif name (ELM ID)	Host protein (Domain name)	Instances	Ref.
Rhoptry proteins					
RON2	TGME49_300100 (B6KV60)	PxPxPR (LIG_SH3_CIN85_PxpxPR_1)	CIN85 or CD2AP (SH3)	1	(24)
RON4	TGME49_229010 (S8F0B5)	PxPxPR (LIG_SH3_CIN85_PxpxPR_1)	CIN85 or CD2AP (SH3)	3	(24)
		LYPxL (LIG_LYPXL_S_1)	Alix (V-shaped)	2	(24)
RON5	TGME49_311470 (S8ERS3)	PxPxPR (LIG_SH3_CIN85_PxpxPR_1)	CIN85 or CD2AP (SH3)	1	(24)
		PTAP (LIG_PTAP_UEV_1)	Tsg101 (UEV)	2	(24)
ROP1	TGME49_309590 (A0A125YP48)	FFAT (TRG_ER_FFAT_1 like)	VAP (MSP)	1	(30)
ROP16	TGME49_262730 (Q3YJR5)	NLS (TRG_NLS_MonoExtN_4)	Exportin (Arm)	1	(37)
Toxofilin	TGME49_214080 (S8G4J8)	Actin-binding motif (LIG_Actin_WH2_1 like)	Actin	1	(28)
Dense granule proteins					
GRA14	TGME49_239740 (B5TVE8)	LYPxL (LIG_LYPXL_S_1)	Alix (V-shaped)	1	(32)
		PTAP (LIG_PTAP_UEV_1)	Tsg101 (UEV)	1	(32)
GRA16	TGME49_208830 (Q1JSN3)	Similar to PEXEL (TRG_Pf-PMV_PEXEL_1)	-	1	(36)
		NLS (TRG_NLS_MonoExtN_4)	Exportin (Arm)	2	(35)
GRA24	TGME49_230180 (A0A125YG37)	KIM (DOC_MAPK_GRA24_9)	MAPK (Kinase)	2	(33)
MYR1	TGME49_254470 (S8GT06)	Similar to PEXEL (TRG_Pf-PMV_PEXEL_1)	-	1	(36)
VIP1	TgME49_233220 (S8F2E2)	FFAT (TRG_ER_FFAT_1 like)	VAP (MSP)	1	(31)

166 **Table 1. Experimentally validated linear motif instances**. Known *Toxoplasma* motif
167 instances from the literature and interacting host protein domains. Protein IDs correspond to
168 the ToxoDB and UniProt accessions, while motif IDs correspond to the ELM motif classes.
169
170 Rhoptry proteins RON2, RON4 and RON5 each contain PxpxPR motifs. During host
171 cell invasion, these motifs help the parasite connect its tight junction (TJ) complex to

172 the host cytoskeleton (24) (Fig. 1). The PxpxPR motifs bind to the SH3 domain of host
173 adaptor proteins CIN85 or CD2AP, which are involved in cytoskeletal rearrangements,
174 vesicle trafficking and endocytosis (25,26). Additionally, RON4 and RON5 contain
175 instances of the motif types LYPxL and PTAP. The LYPxL motif mediates the
176 interaction of proteins with ALIX, while the PTAP motif binds to the UEV domain of
177 class E vacuolar sorting protein TSG101. ALIX and TSG101 are both key components
178 of the ESCRT membrane remodelling system, and during *Toxoplasma* cell invasion,
179 they are also recruited to the TJ for an efficient cell entry (24,27). This process is also
180 aided by parasite proteins such as Toxifilin, which binds to host actin through an actin-
181 binding motif (ABM) that is similar in sequence pattern to the annotated WH2-binding
182 motif in ELM (28). Toxofilin sequesters actin and promotes the disassembly of the host
183 cortical actin meshwork, facilitating an efficient internalisation (29).
184
185 Once *Toxoplasma* is established within the host cell, ROP1 and dense granule
186 proteins, VIP1 and GRA14 are secreted to the parasitophorous vacuole membrane
187 (PVM) (Fig. 1). ROP1 and VIP1 both contain FFAT motifs (similar in sequence motif
188 pattern to the FFAT motif type in ELM) that modulate the targeting of proteins to the
189 endoplasmic reticulum via its interaction with Vesicle Associated Membrane Protein
190 (VAMP) Associated Proteins (VAPs). As ROP1 and VIP1 are inserted in the PVM,
191 their interaction with VAP proteins leads to the association of the PVM with the host
192 ER membrane. Toxoplasma exploits this proximity to sequester the lipids necessary
193 for its vacuole and plasma membrane biogenesis (30,31). GRA14 harbours both
194 LYPxL and PTAP motifs in its C-terminal region. Since GRA14 is also inserted into the
195 PVM with its C-terminus facing the host cytosol, the motifs can recruit ALIX and

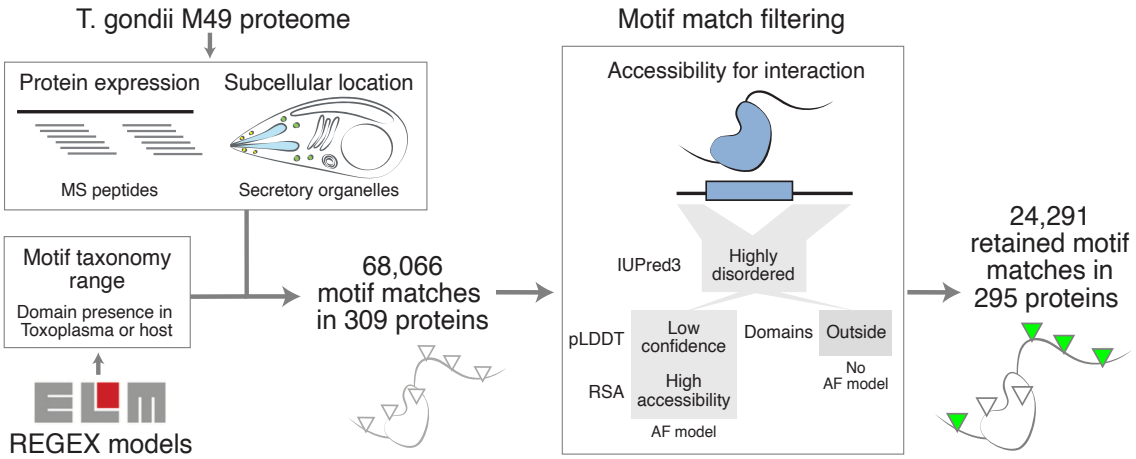
196 TSG101, which, in this case and differently from TJ proteins, help *Toxoplasma*
197 sequester host cytosolic material through the budding of its PVM (32).
198
199 Further, during establishment, more parasite proteins are secreted into the host
200 cytosol and may then be transported to the host nucleus. Within the host cytosol,
201 secreted proteins interact with host proteins via motifs. The intrinsically disordered
202 protein GRA24 contains two instances of a motif that binds to mitogen-activated
203 protein kinase p38α (MAPK14). These motifs allow GRA24 to bind to a MAPK14
204 dimer, forming a stable complex that leads to the expression of proinflammatory
205 factors, which in turn modulate parasite replication (33,34). Parasite protein export is
206 achieved by the presence of cell compartment localisation motifs, which either use the
207 parasite or host machinery to be secreted. In the first case, the correct export of dense
208 granule protein MYR1 to the PV and of GRA16 to the cytosol is dependent on the
209 cleavage of a short linear motif by a parasite aspartyl protease, in a similar way as the
210 Plasmodium PEXEL export motif (35,36). In the case of host nuclear targeting,
211 GRA16, as well as rhoptry protein kinase ROP16, are imported into the host nucleus
212 via nuclear localisation signals (NLS). ROP16 activates signal transducer and activator
213 of transcription (STAT) signalling pathways to inhibit host cell death and modulate the
214 production of interleukins (37,38). Within the nucleus, GRA16 regulates genes
215 involved in cell cycle progression (35). This set of curated motif instances provides a
216 non-exhaustive reference and exemplifies their role during important stages of the
217 Toxoplasma invasion cycle. Nevertheless, the small number of known motifs as well
218 as their occurrence in a few proteins, highlight the need and potential to find and
219 characterise many more motif instances to facilitate further mechanistic studies of
220 Toxoplasma infection strategies.

221

222 Prediction of motif instances in *Toxoplasma* secreted proteins

223 To identify novel motif instances that function in the Toxoplasma infection process via
224 interaction with host proteins, we focused on *Toxoplasma gondii* proteins from the
225 strain ME49 that are secreted into host cells. To determine the subcellular localisation
226 of secreted proteins, we used the dataset from the hyperLOPIT experiment (39) and
227 annotations from the ToxoDB, a database from the VEuPathDB resource, the
228 reference database for the research of eukaryotic pathogens (protists and fungi), their
229 hosts and invertebrate vectors (40). The secreted proteins were subsequently filtered
230 for those with evidence of expression using mass spectrometry datasets from the
231 ToxoDB. To predict motif instances in these 314 secreted proteins, we developed a
232 computational pipeline based on sequence pattern searches as well as structural and
233 functional annotations of predicted motifs (Fig. 2). To find putative motif instances, we
234 used regular expressions defined for annotated motif classes in the ELM database,
235 which capture the preferences for certain amino acids at certain motif positions. We
236 scanned the selected proteins with 298 regular expressions corresponding to motif
237 classes with interacting domains that could be present either in Toxoplasma or in the
238 host using the taxonomic range annotations provided by the ELM resource. This motif
239 search resulted in a total of 68,066 motif matches in 309 proteins. Since motifs occur
240 primarily in IDRs and need to be accessible for binding to the partner proteins, we
241 deployed structural annotations to further restrict to more likely functional motif
242 matches. To this end, we evaluated the location of the motif match in IDRs using the
243 software IUPRed3(41). In addition, in cases where a full-length structural model of a
244 given protein was available from the EBI AlphaFold database (42), we evaluated motif
245 accessibility through low pLDDT values and high relative solvent accessibility (RSA).

246 In cases where a structural model was not available, we excluded motif matches that
247 fell within annotated folded domain boundaries as annotated in UniProt (43) (Fig. 2).
248 Using this framework, we obtained a set of 24,291 motif matches within 295 proteins
249 from micronemes, rhoptries and dense granules (Fig. 2, Fig. 3A) (Table S1).
250



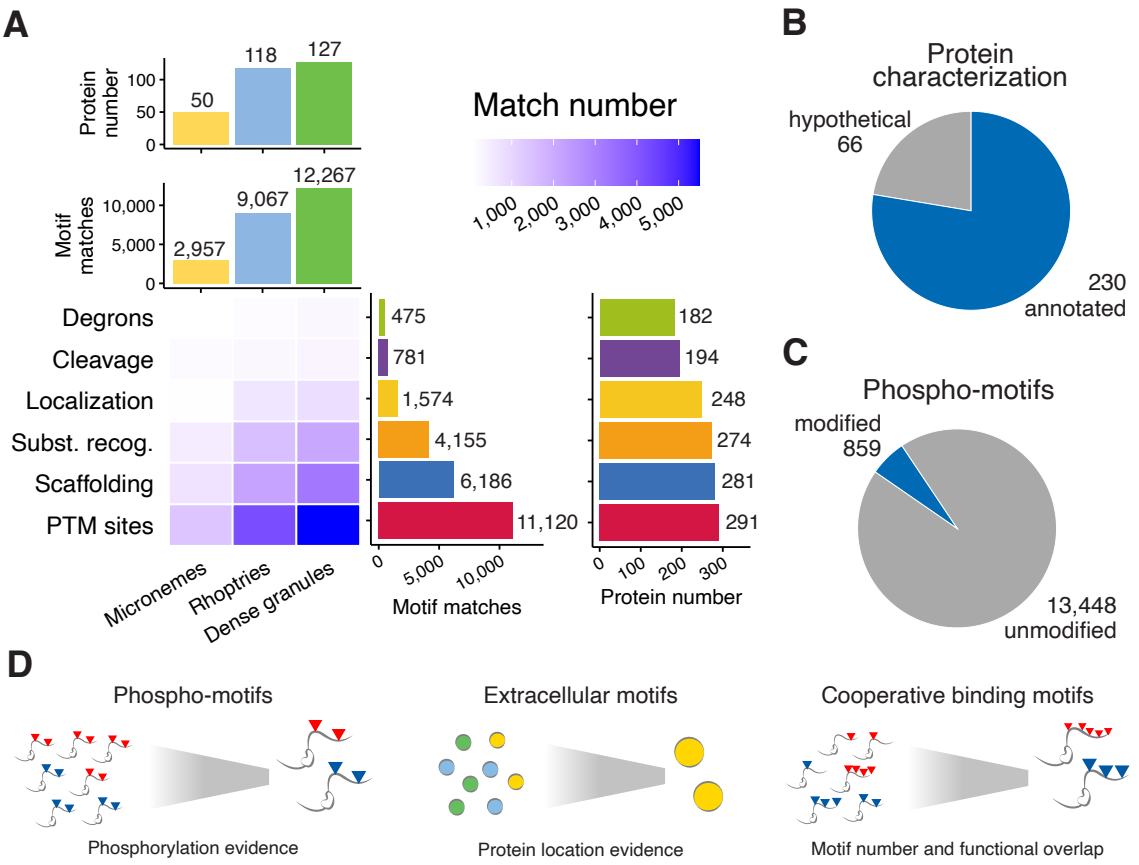
251

252 **Figure 2 - Motif prediction pipeline.** Diagram outlining the protein and motif model selection
253 for motif prediction, as well as motif accessibility filtering criteria.
254

255 Our computational pipeline captured 16 of the 21 known motif instances within our
256 retained motif matches. Two PTAP motifs in RON5 were missed because they have
257 low protein IDR scores. The FFAT motifs in ROP1 and VIP1, and the actin-binding
258 motif in Toxofilin were not captured because their sequences did not match any ELM
259 regular expression. These last two examples highlight that current annotated motif
260 types in the ELM resource do not fully capture the diversity of motifs in *Toxoplasma*
261 proteins. Our curated motif instances could thus be used to reevaluate and extend
262 ELM motif type annotations. When further analysing the curated motif instances here,
263 we found a previously unreported PxpxPR motif in the N-terminal region of RON4 as
264 well as two SxIP motifs. SxIP motifs modulate the binding to microtubule plus-end

265 tracking proteins (+TIPs), which have an important role in chromosome segregation,
266 cell signalling and organelle trafficking (36). SxIP motifs are already known to be
267 exploited by the Apicomplexan parasite *Theileria* (13). RON8, which, together with
268 RON2, RON4, and RON5, forms part of the TJ (22), also contained matches of the
269 SxIP and LYPxL motifs. The presence of extra motif matches related to host
270 cytoskeleton in TJ proteins indicates that the current model of motif-hijacking during
271 this stage of host cell invasion is still incomplete and will benefit from our motif
272 predictions. Furthermore, the identification of 4,822 strong motif matches in our
273 dataset from 66 proteins annotated as "hypothetical" in ToxoDB and with expression
274 evidence (Fig. 3B) represents a potential venue to explore, assign and understand
275 their function.

276



277

278 **Figure 3 - Matches in secreted proteins. A**. Number of filtered motif matches from secreted
279 proteins across different motif categories. **B**. Split of proteins with filtered matches according
280 to their annotation. **C**. Split of filtered phospho-motifs according to whether they have
281 phosphorylation evidence. **D**. Additional motif filtering strategies according to motif type
282 functions and properties.
283
284 On average, our pipeline obtained 82 retained motifs per secreted protein: 59 in
285 micronemes, 77 in rhoptries and 97 in dense granules (Fig. 3A). While proteins,
286 especially those with longer IDRs such as dense granule proteins (Fig. S1A), are
287 expected to have multiple motif occurrences, the relatively high number of predicted
288 motif instances still points to some of those being likely non-functional. Most of the
289 motif matches come from motif types that denote sites for PTMs such as
290 phosphorylation (Fig. 3A). This most likely results from very degenerate regular
291 expressions available for motif types from the PTM class, such as …[ST]...[ST]
292 defining sites phosphorylated by GSK kinase. In regular expressions, a "." corresponds
293 to any amino acid possible at this position and amino acids in brackets denote possible
294 amino acids at this position in the motif. More than 2,000 motif matches alone were
295 identified with this GSK kinase phosphorylation motif pattern. Phosphoproteomics data
296 can be used to provide evidence for actual modification of these predicted motifs (45).
297 Using available phosphosite datasets from ToxoDB, we found evidence of
298 phosphorylation for ~6% of predicted phosphorylation sites (Fig. 3C-D), assigning
299 higher confidence to these predicted motifs (Table S1). We generally observe that for
300 most motif classes, motif matches were identified only for a fraction of annotated motif
301 types in ELM. For example, 38% of motif matches from the targeting class (TRG)
302 correspond to diArg motifs, 35% of motif matches from the degradation class (DEG)

303 correspond to SPOP degrons (see below), and 18% of scaffolding motifs bind SH3
304 domains (Fig. S1B).
305
306 To further select for more likely functional motif matches, one can take into account
307 functional aspects of the different motif types and the cellular context of secreted
308 proteins (Fig. 3D). For example, as microneme proteins will be present in extracellular
309 environments, motifs known to interact with domains from extracellular proteins would
310 be expected to have more likely functional relevance. In our dataset (Table S1), we
311 found 401 motif matches from 12 motif extracellular types in 48 microneme proteins
312 that could be relevant during the recognition and attachment stages. Integrin binding
313 is a common strategy for pathogens to attach to the host cell membrane (46–48). The
314 most common integrin-binding motif is the RGD motif, which can bind to eight of the
315 different alpha and beta integrin heterodimers (49). Within our dataset, we found three
316 motif matches of the RGD motif in three microneme proteins: MIC17C, Toxolysin 4
317 (TLN4) and chitinase-like protein CLP1. As some microneme proteins might be
318 localised to the host cytosol, e.g. during host cell lysis and egress (50), we did not filter
319 out matches from other cytosolic motif types in microneme proteins. Degradation
320 signals, or degrons, are motifs recognised by E3 ubiquitin ligases and mark proteins
321 for proteasomal degradation (51). Degrons might be relevant motif candidates for
322 Toxoplasma infection, as other pathogens are known to modify protein levels via
323 degrons to achieve outcomes such as the stabilisation of host components necessary
324 for their survival (2) or modulating the activity of their proteins once inside the host
325 (52). We found 475 degrons in 182 secreted proteins: 73 in micronemes, 157 in
326 rhoptries and 245 in dense granule proteins. Speckle-type POZ protein (SPOP), an
327 E3 ubiquitin ligase, recognises its substrates via an ST-rich degron (53). We found

155 SPOP degron matches in 80 secreted proteins. Since multivalency is a known mechanism by which SPOP substrates enhance their affinity to SPOP (53), one could prioritise the 34 secreted proteins containing two or more SPOP motif matches in our dataset. Rhoptry protein toxolysin 1 (TLN1), a metalloprotease with a role in rhoptry maturation (54), but potentially detrimental for host establishment once secreted into the host cytosol, has 10 identical SPOP degron matches, which could be used to modulate TLN1 levels once within the host cytoplasm. Overall, these examples illustrate how our dataset can be further explored to prioritise motif candidates for experimental characterisation.

Experimental validation of predicted TRAF6-binding motifs

For the experimental validation of predicted motifs, we focused on motifs that bind the tumour necrosis factor (TNF) receptor-associated factor 6 (TRAF6) protein. TRAF6 binds to proteins containing a P.E..[DE(Φ)] motif through its C-terminal MATH domain (Φ denotes hydrophobic residues). TRAF6 plays important roles in immunity and development by contributing to the formation of various signalling protein complexes. During the innate immune response, TRAF6 is activated by interleukin 1 receptor (IL-1R), linking this receptor to the canonical activation of the transcription factor NF-κB (55). Through the addition of K63-linked polyubiquitin chains, TRAF6 primes proteins for complex assembly and not for proteasomal degradation. The TRAF6-binding motif is present in proteins such as IL-1R, Tumour necrosis factor protein CD40, and Mitochondrial antiviral-signalling protein (MAVS) (56). In the context of infection, herpesviruses have been shown to use the TRAF6 motif-containing protein Tio to stimulate NF-κB activation (57).

353 Among our predicted motif instances, we found 111 motif matches in 68 rhoptry and
354 dense granule proteins. Given that the TRAF6-binding motif is known to appear
355 multiple times within proteins (58), we hypothesised that the motif repetitions could
356 interact with the TRAF6 trimer in a multivalent way. Following this criterion, we
357 selected rhoptry neck protein RON6, which harbours 17 matches of the TRAF6-
358 binding motif, most of them being identical repeats, and RON10, which harbours two
359 different TRAF6-binding motif matches (Fig. 4A). RON10 is known to interact with
360 RON9, but RON10 and RON9 do not have clear roles during infection (59). TRAF6
361 has been shown previously to interact with the dense granule proteins GRA7 and
362 GRA15, which themselves play a role in modulating the innate immune response
363 during *Toxoplasma* infection (60–63). However, the actual interaction interfaces of
364 GRA7 and GRA15 with TRAF6 have not been described (22), making the TRAF6 motif
365 matches in GRA7 and GRA15 interesting candidates for further experimental
366 characterisation (Fig. 4A).
367
368 We proceeded to determine the dissociation constants of the TRAF6-binding motifs in
369 RON6, RON10, GRA7 and GRA15 to the TRAF6 MATH domain using the microscale
370 thermophoresis (MST) binding assay. To this end, we expressed and purified the
371 human TRAF6 MATH domain and ordered peptides containing the sequence of the
372 following motif matches. In the case of RON6, we selected the most frequent version
373 of its motif repeats. For RON10, we selected the second motif match that contained
374 multiple negative charges, which might contribute to binding (Fig. 4B). GRA7 and
375 GRA15 only have one TRAF6-binding motif match each, which was selected for
376 experimental testing. In addition, we ordered these wildtype peptide sequences with a
377 Glutamate to Serine mutation at position 3 in the motifs supposed to weaken or

abrogate binding to the MATH domain because the conserved glutamate residue forms a hydrogen bond with backbone atoms of residue A458 in TRAF6 (Fig. 4B) (Table S2). Our experimental results confirm that the peptides containing the motifs from RON10 and GRA15 can bind to the TRAF6 MATH domain, and these interactions were lost when testing the mutant peptide versions. The dissociation constants for the RON10 (Kd 18.2 ±1.46 μM) and GRA15 peptides (Kd 26.9 ±1.43 μM) are as expected, for functional motif interactions, in the lower micromolar range (Fig. 4C-D, Data S1, Table S2). These results represent the first evidence that RON10 can bind TRAF6, pointing to a role of RON10 in the regulation of the innate immune response.

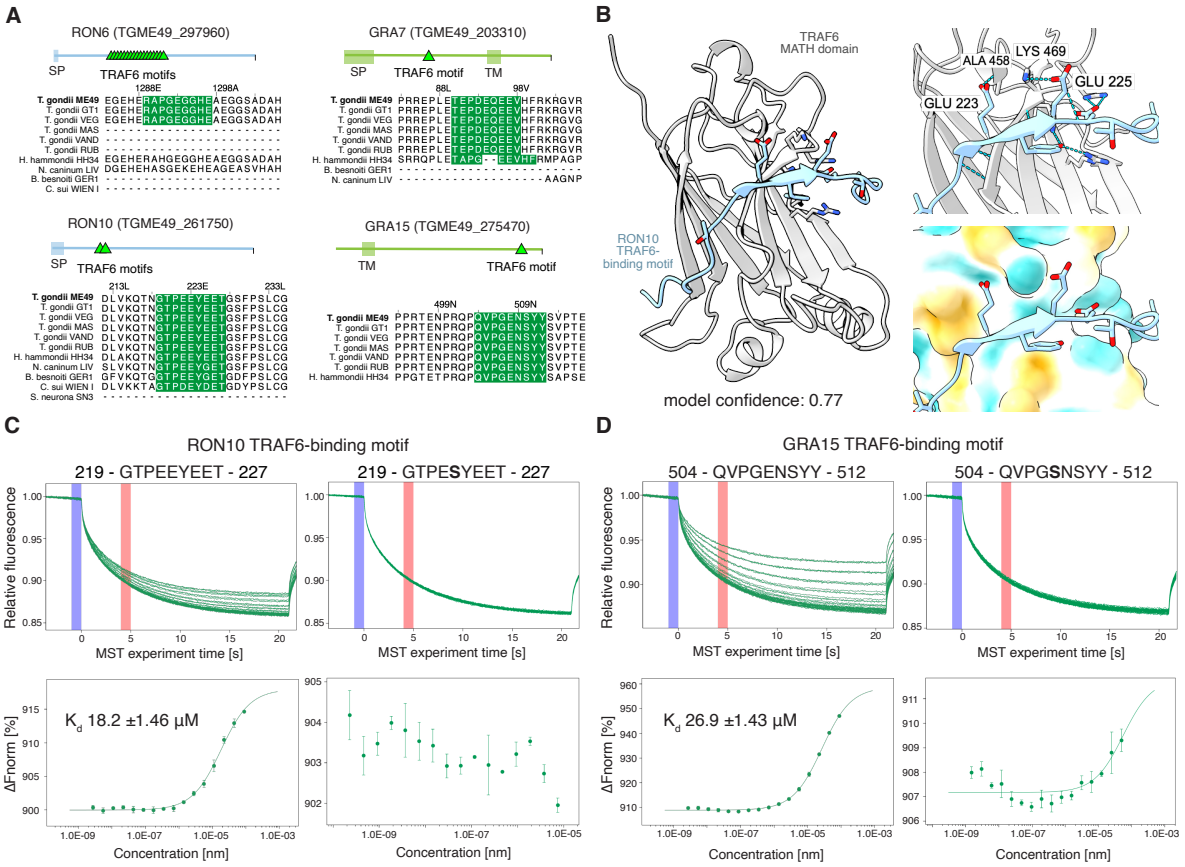


Figure 4 - TRAF6 motif binding assays. A. Protein diagrams of RON6, RON10, GRA7 and GRA15 indicating the position of their TRAF6-binding motifs with accompanying multiple sequence alignment showing their presence in different Toxoplasma strains and Coccidian

species. **B.** AlphaFold model of TRAF6 MATH domain with TRAF6-binding motif from RON10. Right panels display the potential hydrogen bond network at the interaction interface (top) and secondary structure adopted by the TRAF6-binding motif (bottom). **C.** RON10 TRAF6 motif testing. MST traces (top graph) and dose-response curves (bottom panel) of TRAF6 MATH domain binding to RON10 TRAF6 motif containing peptide (left) and the E>S mutant (right). In the MST traces, the cold region is set at 0 s (blue) and the hot detection region at 5 s (red). **D**. Same as in A, but for the GRA15 TRAF6 motif and the E>S motif mutant.

Discussion

It has been previously speculated that *Toxoplasma* induces NF-kB activation and macrophage recruitment to limit tissue invasion, assuring host survival and promoting development into the chronic bradyzoite form (50). The experimentally confirmed TRAF6-binding motifs in RON10 and GRA15 might support this process by connecting TRAF6-dependent IL-1 signalling with NF-kB activation. However, proving the functional relevance of these motifs would require further experimentation. Even though the RON6 peptide did not show binding to the TRAF6 MATH domain, it might still display binding in cells via multivalent interactions involving its 17 putative TRAF6-binding motif matches and TRAF6 trimers.

Our filtered dataset of 24,291 motif matches within 295 Toxoplasma proteins still likely contains many non-functional motif matches. More sophisticated computational approaches for match filtering would require machine learning on larger sets of validated motif instances, which, at the moment, are not available for *Toxoplasma* and other parasites. Conservation of motif matches is often used as an additional line of evidence for the prioritisation of likely functional motif matches. Unfortunately, Toxoplasma secreted proteins tend to be more novel, less conserved overall and

418 display longer regions of disorder (Fig. S1A) (59). This makes it difficult to build good
419 sequence alignments to compute common motif conservation metrics like in (65).
420 Nevertheless, we included the assessment of the presence of our motif matches in
421 different *Toxoplasma* strains and the few closely related parasite species as
422 supplementary conservation metrics (Table S1), which could be useful when working
423 with different strains or in understanding motif variation and evolution.
424
425 This study represents the first comprehensive survey of short linear motifs in
426 *Toxoplasma gondii* or any other Apicomplexan parasite. There have been other efforts
427 to predict and validate short linear motifs in pathogenic proteins (32,66). However,
428 they focus on only a subset of proteins or specific motifs, or other pathogens such as
429 viruses (61). Identification of novel motif instances in parasites such as Toxoplasma
430 also enables refinement of current motif type annotations, as highlighted by some
431 examples described in this study.
432
433 Our systematic approach for predicting instances of known motif types in Toxoplasma
434 secreted proteins indicates that there might be a substantial number of motifs in
435 *Toxoplasma* to be characterised, also in non-secreted proteins, e.g. motifs that
436 participate in endogenous processes for parasite motility, cytoskeletal
437 rearrangements, and parasite development. Thus, this dataset represents a valuable
438 resource for hypothesis generation and design of downstream experiments to study
439 molecular mechanisms of *Toxoplasma gondii* infection and biology. Structural models
440 of predicted domain-motif interactions could further be used to design disruptive
441 mutations to validated predicted modes of binding (4). Identified and experimentally
442 validated pathogen-host motif-domain interactions can form the basis for the

443 development of therapeutic strategies aimed at perturbing these interactions and
444 thereby interfering with infection processes (4,10). Motif discovery in pathogens
445 remains a much unexplored area of research, but holds great promise for improved
446 understanding of infection mechanisms and the better design of effective treatments
447 for parasite infection.
448

449 ## Acknowledgements

450 We are thankful to members of the Gibson team for discussions at various stages of
451 this work. We also express our gratitude to the curators and contributors of the ELM
452 and the ToxoDB resources. This project was partly funded by the Deutsche
453 Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID
454 449991970 granted to K.Lu.
455

456 ## Author contributions

457 J.A.V., T.G. conceptualised the study, J.A.V. produced code and performed
458 computational analyses, K.R., K.La., A.B. performed experiments, J.A.V. wrote the
459 manuscript. J.A.V., T.G., K.R., K.La., K.Lu. edited the manuscript, T.G. and K.Lu.
460 provided funding.
461

462 ## Data and Code Availability

463 The motif discovery pipeline, processing and analysis code is made available on
464 GitHub.(https://github.com/JesnsAV/Toxo_Motifs)
465
466
467

468 ## Materials and Methods

469 Retrieval of sequence, structure and motif data

470 *Toxoplasma* strain proteomes, as well as those of five closely related organisms from
471 the Sarcocystidae clade, were retrieved from the ToxoDB resource (40). The fasta
472 files of the annotated proteomes were collected for the following organisms:
473 *Toxoplasma gondii* GT1 (Type I), *T. gondii* MAS and *T. gondii* ME49 (Type II and
474 reference strain), *T. gondii* RUB and *T. gondii* VEG (Type III), *T. gondii* VAND,
475 *Hammondia hammondii, Neospora caninum, Besnoitia besnoitii, Cystoisospora suis*
476 Wien I and *Sarcosystis neurona* SN3. Using the ToxoDB search strategies, we
477 retrieved mass spectrometry proteomic peptide mapping data, phosphosites, and
478 cellular location predictions from the HyperLOPIT experiment (39) for the proteins of
479 *T. gondii* ME49. All available AlphaFold database (AFDB) (67) predicted structures for
480 *Toxoplasma gondii* ME49 were retrieved from the Google Cloud Console public-
481 datasets-deepmind-alphafold-v4, taxonomy id: 50877 (02.05.2025). The latest dataset
482 of motif classes was downloaded from the ELM resource (30.04.2025) (23). The file
483 contained each motif code name, its regular expression (REGEX) and supplementary
484 information. We did manual annotation of the motif classes to determine if their
485 taxonomic range included vertebrates and *Toxoplasma*, i.e. if the domain was
486 annotated in ELM to be present in the respective organism for interaction with the
487 motif, ultimately excluding motif classes of organisms such as plants, insects and
488 fungi. We also added a manual annotation on whether the motif is functional in an
489 extracellular context and if it requires a phosphorylation or modification to interact with
490 its domain partner. We retrieved available T. gondii domain mappings from UniProt
491 (29.07.2025) (43), which came from Pfam, PROSITE and SMART.

492

493 Motif prediction and structural assessments

494 We determined the presence of motif matches in the *T. gondii* ME49 by doing a
495 baseline REGEX search. To assess the accessibility of these motif predictions, two
496 approaches were selected: one that takes into account the intrinsic disorder of the
497 sequence and another that evaluates the accessibility of the motif from predicted
498 AlphaFold DB (AFDB) structures. The IUPred3 software (41) was used to calculate
499 the disorder probability of the full proteome of *T. gondii* ME49. Once a REGEX match
500 was determined, the IUPred3 score from all motif residues was retrieved and
501 averaged. Complementarily, using available AFDB structures, we took the pLDDT
502 scores from all motif residues and averaged them (68). We calculated residue
503 accessibility using DSSP (Dictionary of Secondary Structure of Proteins) software (69)
504 on each AFDB structure. The individual residue scores were then standardised
505 according to (70) to obtain relative surface accessibility scores (RSA). We then took
506 the RSA scores of all motif residues and averaged them. We determined if a motif
507 match overlapped with known domain boundaries from UniProt by using positional
508 information.

509

510 Motif organism presence assessment

511 A reciprocal best hit search was carried out using BLAST to construct sets of
512 orthologous sequence groups among *Toxoplasma* strains and related species. These
513 sequence groups contained the sequences with the lowest e-value from each
514 organism's proteome, so alignments contained a standard number of sequences and
515 were easily comparable. Multiple sequence alignments were produced for each
516 sequence group using ClustalO with default parameters (71). Standard conservation
517 measurements (65) were not possible as our alignments contained only a few species.

518 Thus, the conservation of the motifs was approximated by carrying out a presence
519 assessment. A motif was considered present in another organism if a match of the
520 same class was present in its sequence within a window of 15 amino acids in the
521 alignment around the start site of the motif match of the reference *Toxoplasma* strain
522 ME49. And then the presence proportion of the total number of both strains and
523 species was calculated and associated with each motif match, resulting in three
524 different scores: the proportion of strains, species and overall organisms.
525
526 Data enrichment and filtering
527 The data from the ELM REGEX motifs matches, their conservation, IUPred3, pLDDT,
528 and accessibility scores were combined into a single table together with their
529 respective protein information on subcellular location, annotations, domain overlap,
530 and modifications. This was carried out using their unique ToxoDB identifiers and the
531 motif name, match number and residue positions. Data filtering to define a set of motif
532 matches was carried out with three different strategies. Firstly, we retained motifs from
533 proteins with at least one peptide mapped to them from retrieved mass spectrometry
534 data, to make sure proteins were expressed. We then made use of the ELM taxonomy
535 range by ensuring that the taxonomic range of the motif instance, if the motif has been
536 reported in a certain group of organisms, was valid for the *Toxoplasma* matches. Only
537 motif classes that apply to Vertebrates and Apicomplexans were kept, mainly
538 excluding those characterised in plants, insects or fungi. The classes valid in
539 Vertebrates were kept if the motifs were located in secreted proteins, as they would
540 be in contact with the vertebrate host (human, a mammal or birds) cell components.
541 The classes only valid in extracellular contexts were kept if the motifs were located in
542 microneme proteins, as among secreted proteins, they would be the ones in this

location. We applied structural filters in two steps; only motif matches with IUPred3 scores above 0.4 were considered, following disordered region filters in (72). Secondly, for motifs in proteins with available AFDB structures, we retained the ones with a pLDDT<=80 (68) and an average RSA>= 0.25 (73). For motifs without AFDB structures, we retained only the ones outside of known domain boundaries.

AlphaFold modelling

The TRAF6 MATH domain and the RON10 TRAF6-binding motif were selected for structural modelling. We used a local installation of AlphaFold Multimer version 2.3.219 (74). We manually defined domain boundaries to include missing secondary structure with high pLDDT values. We extended the motif sequence by including 5 residues at each sequence end. Structural model visualisations were produced using ChimeraX-1.10 (75).

The local installation of AlphaFold Multimer version 2.3.2 was run using the following parameters and following Alphafold GitHub instructions (https://github.com/deepmind/alphafold#running-alphafold):

```
--model_preset=multimer \
--db_preset=full_dbs \
--max_template_date=2020-05-14 \
--num_multimer_predictions_per_model=1 \
--use_gpu_relax=True \
--
bfd_database_path=/mnt/storage/alphafold/v232/bfd/bfd_metaclust_clu_complete_id
30_c90_final_seq.sorted_opt \
```

```
--
mgnify_database_path=/mnt/storage/alphafold/v232/mgnify/mgy_clusters_2022_05.f
a \
--obsolete_pdbs_path=/mnt/storage/alphafold/v232/pdb_mmcif/obsolete.dat \
--
pdb_seqres_database_path=/mnt/storage/alphafold/v232/pdb_seqres/pdb_seqres.tx
t \
--template_mmcif_dir=/mnt/storage/alphafold/v232/pdb_mmcif/mmcif_files \
--uniprot_database_path=/mnt/storage/alphafold/v232/uniprot/uniprot.fasta \
--uniref90_database_path=/mnt/storage/alphafold/v232/uniref90/uniref90.fasta \
--uniref30_database_path=/mnt/storage/alphafold/v232/uniref30/UniRef30_2021_03
\
--use_precomputed_msas=True
```

Protein Expression and Purification

A gene encoding the human TRAF6 C-domain (residues 346-504, UniProt ID Q9Y4K3) was synthesised (GeneArt, Thermo Fisher Scientific) and subcloned into the pCoofy4 plasmid (76) containing an N-terminal His6-MBP-3C fusion tag. For the human TRAF6 C-domain (InterPro: IPR008974) gene, the native DNA sequence was used. The resulting bacterial expression construct was freshly transformed into *E. coli* BL21(DE3) codon + RIL cells (Stratagene). Precultures were grown overnight at 37°C in LB medium supplemented with 30 µg/ml Kanamycin and 35 µg/ml Chloramphenicol and used to inoculate the large-scale expression cultures. 10 ml preculture was added to 1 litre of TB-FB medium supplemented with 30 µg/ml Kanamycin and 35 µg/ml Chloramphenicol. The cultures were grown at 37°C until the OD600 reached ~0.8,

593 after which the temperature was reduced to 18°C and expression of the hTRAF6 C-
594 domain was induced with 0.5 mM IPTG. After overnight expression at 18°C, the
595 cultures were harvested by centrifugation (30 min, 5000 g, 4°C), and the pellets were
596 flash-frozen in liquid nitrogen and stored at -80°C until the start of the protein
597 purification.
598
599 The cells were lysed by 5 passages through a microfluidiser device, followed by
600 centrifugation (30 min, 140.000 g, 4°C). The cleared lysate was loaded onto a 5 ml
601 Protino Ni-NTA column (Macherey-Nagel) pre-equilibrated with 50 mM Tris-HCl, pH
602 8.0, 500 mM NaCl and 20 mM imidazole. After loading, the Ni-NTA column was
603 washed with equilibration buffer until the A280 nm signal returned to baseline and
604 eluted in a 60 ml gradient going from 0 to 100% of Ni-NTA elution buffer (50 mM Tris-
605 HCl, pH 8.0, 500 mM NaCl, 350 mM imidazole). The Ni-NTA elution fractions were
606 analysed by SDS-PAGE, and the fractions containing the His6-MBP-3C-hTRAF6 C-
607 domain fusion protein were pooled. 2 mg of His6-tagged HRV 3C protease was added,
608 and the sample was dialysed overnight at 4°C to IEX buffer (25 mM HEPES, pH 7.5,
609 100 mM NaCl, 20 mM imidazole). The next day, the dialysed sample was loaded onto
610 a 5 ml Protino Ni-NTA column (Macherey-Nagel) coupled to a 5 ml HiTrap Q HP anion
611 exchange column (Cytiva), both equilibrated with IEX buffer. The untagged hTRAF6
612 C-domain protein was present in the flow-through of the coupled columns, which was
613 concentrated to 5 ml and injected into a HiLoad 16/600 Superdex 75 pg size exclusion
614 chromatography column (Cytiva) pre-equilibrated with 25 mM HEPES pH 7.5, 150 mM
615 NaCl and 1 mM DTT. The elution fractions containing pure hTRAF6 C-domain protein
616 were pooled and concentrated to 2.1 mg/ml (111 μM). The final hTRAF6 C-domain
617 protein samples were aliquoted, flash-frozen in liquid nitrogen and stored at -80°C until

618 further usage. The identity of the hTRAF6 C-domain protein was verified by mass
619 spectrometry (EMBL Proteomics Core Facility), and the oligomerisation state
620 (monomeric) and absence of aggregates were confirmed by size exclusion
621 chromatography column coupled to multi-angle light scattering (SEC-MALS). The
622 stability of the hTRAF6 C-domain protein in the storage buffer was assessed by nano-
623 Differential Scanning Fluorimetry (nano-DSF; Tm ~ 46.8 °C).
624
625 Binding assays
626 We selected 4 motifs from 4 different *Toxoplasma* proteins with different AF model
627 scores for carrying out binding assays as described in Results, together with the
628 human MAVS TRAF6 motif, being selected as a control (56). Peptides of 15 amino
629 acids containing the motifs were defined, as well as versions containing E>S point
630 mutations at motif position +3. The peptides were labelled with 5-Carboxyfluorescein
631 (5-Fluo) at the C-terminus by Biosyntan GmbH (https://www.biosyntan.de) to carry out
632 Microscale thermophoresis (MST). The lyophilised peptides were resuspended in
633 25mM HEPES, pH 7.5 and 150mM NaCl to a concentration of 2mM, and the pH was
634 adjusted to 7.5 when necessary. MST measurements were performed using a
635 Monolith NT.115 (NanoTemper Technology). Two-fold serial dilutions of the hTRAF6
636 protein (99 µM) were mixed in a 1:1 ratio with 100nM peptides. Titration measurements
637 were taken for all *Toxoplasma* peptides, and the human MAVS control, with two-fold
638 serial dilutions of the hTRAF6 protein (99 µM) mixed in a 9:1 ratio with 500nM. All
639 measurements were performed in triplicate and carried out at 25°C with an LED
640 excitation power of 20% and a medium MST power. Finally, the data were analysed
641 assuming a 1:1 binding model and graphed using the MO.Affinity Software
642 (NanoTemper Technology).

643 ## References

644 1. Omotade TO, Roy CR. Manipulation of Host Cell Organelles by Intracellular Pathogens.
645 Cossart P, Roy CR, Sansonetti P, editors. Microbiol Spectr. 2019 Apr 12;7(2):7.2.37.

646 2. Davey NE, Travé G, Gibson TJ. How viruses hijack cell regulation. Trends Biochem Sci.
647 2011 Mar;36(3):159–69.

648 3. Sámano-Sánchez H, Gibson TJ. Mimicry of Short Linear Motifs by Bacterial Pathogens: A
649 Drugging Opportunity. Trends Biochem Sci. 2020 Jun;45(6):526–44.

650 4. Baptista D, Gomez-Lucas L, Jänes J, Krogan NJ, Amorim MJ, Ivarsson Y, et al. AlphaFold
651 models of host-pathogen interactions elucidate the prevalence and structural modes of
652 molecular mimicry [Internet]. Bioinformatics; 2025 [cited 2025 Aug 8]. Available from:
653 http://biorxiv.org/lookup/doi/10.1101/2025.06.04.657796

654 5. Davey NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, et al. Attributes of
655 short linear motifs. Mol Biosyst. 2012 Jan;8(1):268–81.

656 6. Van Roey K, Uyar B, Weatheritt RJ, Dinkel H, Seiler M, Budd A, et al. Short linear motifs:
657 ubiquitous and functionally diverse protein interaction modules directing cell regulation.
658 Chem Rev. 2014 Jul 9;114(13):6733–78.

659 7. Davey NE, Cyert MS, Moses AM. Short linear motifs – ex nihilo evolution of protein
660 regulation. Cell Commun Signal. 2015 Dec;13(1):43.

661 8. Mészáros B, Sámano-Sánchez H, Alvarado-Valverde J, Čalyševa J, Martínez-Pérez E,
662 Alves R, et al. Short linear motif candidates in the cell entry system used by SARS-CoV-
663 2 and their potential therapeutic implications. Sci Signal. 2021 Jan 12;14(665):eabd0334.

664 9. Mihalič F, Simonetti L, Giudice G, Sander MR, Lindqvist R, Peters MBA, et al. Large-scale
665 phage-based screening reveals extensive pan-viral mimicry of host short linear motifs. Nat
666 Commun. 2023 Apr 26;14(1):2409.

667 10. Simonetti L, Nilsson J, McInerney G, Ivarsson Y, Davey NE. SLiM-binding pockets: an
668 attractive target for broad-spectrum antivirals. Trends Biochem Sci. 2023 May;48(5):420–
669 7.

670 11. Giménez A, Del Giudice MG, López PV, Guaimas F, Sámano-Sánchez H, Gibson TJ, et
671 al. *Brucella* NpeA is a secreted Type IV effector containing an N-WASP-binding short
672 linear motif that promotes niche formation. Roy CR, editor. mBio. 2024 Jul
673 17;15(7):e00726-24.

674 12. Boddey JA, Moritz RL, Simpson RJ, Cowman AF. Role of the *Plasmodium* Export Element
675 in Trafficking Parasite Proteins to the Infected Erythrocyte. Traffic. 2009 Mar;10(3):285–
676 99.

677 13. Huber S, Karagenc T, Ritler D, Rottenberg S, Woods K. Identification and characterisation
678 of a *Theileria annulata* proline-rich microtubule and SH3 domain-interacting protein
679 (TaMISHIP) that forms a complex with CLASP1, EB1, and CD2AP at the schizont surface.
680 Cell Microbiol. 2018 Jul;20(7):e12838.

681 14. Mohan A, Sullivan Jr WJ, Radivojac P, Dunker AK, Uversky VN. Intrinsic disorder in
682 pathogenic and non-pathogenic microbes: discovering and analyzing the unfoldomes of
683 early-branching eukaryotes. Mol Biosyst. 2008;4(4):328.

684 15. Pancsa R, Tompa P. Structural Disorder in Eukaryotes. Buday L, editor. PLoS ONE. 2012
685 Apr 5;7(4):e34687.

686 16. De Barros RAM, Torrecilhas AC, Marciano MAM, Mazuz ML, Pereira-Chioccola VL, Fux
687 B. Toxoplasmosis in Human and Animals Around the World. Diagnosis and Perspectives
688 in the One Health Approach. Acta Trop. 2022 Jul;231:106432.

689 17. Milne G, Webster JP, Walker M. Toxoplasma gondii: An Underestimated Threat? Trends
690 Parasitol. 2020 Dec;36(12):959–69.

691 18. Dubey JP. Toxoplasmosis of Animals and Humans [Internet]. 3rd ed. Boca Raton: CRC
692 Press; 2021 [cited 2025 Aug 8]. Available from:
693 https://www.taylorfrancis.com/books/9781003199373

694 19. Boothroyd JC, Dubremetz JF. Kiss and spit: the dual roles of Toxoplasma rhoptries. Nat
695 Rev Microbiol. 2008 Jan;6(1):79–88.

696 20. Rastogi S, Cygan AM, Boothroyd JC. Translocation of effector proteins into host cells by
697 Toxoplasma gondii. Curr Opin Microbiol. 2019 Dec;52:130–8.

698 21. Kemp LE, Yamamoto M, Soldati-Favre D. Subversion of host cellular functions by the
699 apicomplexan parasites. FEMS Microbiol Rev. 2013 Jul;37(4):607–31.

700 22. Hakimi MA, Olias P, Sibley LD. *Toxoplasma* Effectors Targeting Host Signaling and
701 Transcription. Clin Microbiol Rev. 2017 Jul;30(3):615–45.

702 23. Kumar M, Michael S, Alvarado-Valverde J, Zeke A, Lazar T, Glavina J, et al. ELM-the
703 Eukaryotic Linear Motif resource-2024 update. Nucleic Acids Res. 2024 Jan
704 5;52(D1):D442–55.

705 24. Guérin A, Corrales RM, Parker ML, Lamarque MH, Jacot D, El Hajj H, et al. Efficient
706 invasion by Toxoplasma depends on the subversion of host protein networks. Nat
707 Microbiol. 2017 Aug 28;2(10):1358–66.

708 25. Havrylov S, Jolanta Redowicz M, Buchman VL. Emerging Roles of Ruk/CIN85 in Vesicle-
709 Mediated Transport, Adhesion, Migration and Malignancy. Traffic. 2010 Mar
710 17;11(6):721–31.

711 26. Jozic D, Cárdenes N, Deribe YL, Moncalián G, Hoeller D, Groemping Y, et al. Cbl
712 promotes clustering of endocytic adaptor proteins. Nat Struct Mol Biol. 2005
713 Nov;12(11):972–9.

714 27. Henne WM, Buchkovich NJ, Emr SD. The ESCRT Pathway. Dev Cell. 2011 Jul;21(1):77–
715 91.

716 28. Lee SH, Hayes DB, Rebowski G, Tardieux I, Dominguez R. Toxofilin from *Toxoplasma*
717 *gondii* forms a ternary complex with an antiparallel actin dimer. Proc Natl Acad Sci. 2007
718 Oct 9;104(41):16122–7.

719 29. Delorme-Walker V, Abrivard M, Lagal V, Anderson K, Perazzi A, Gonzalez V, et al.
720 Toxofilin upregulates the host cortical actin cytoskeleton dynamics facilitating *Toxoplasma*
721 invasion. J Cell Sci. 2012 Jan 1;jcs.103648.

722 30. Mehra C, Alvarado Valverde J, Matias AMN, Torelli F, Medeiros TC, Straub J, et al.
723 Toxoplasma effector TgROP1 establishes membrane contact sites with the endoplasmic
724 reticulum during infection. Nat Microbiol. 2025 Nov 25;10(12):3331–45.

725 31. Romano JD, Buh R, Grudda T, Box JR, Beltran J, Khan SM, et al. Toxoplasma gondii
726 VIP1 mediates parasitophorous vacuole–host endoplasmic reticulum interactions to
727 facilitate parasite development. Nat Microbiol. 2025 Oct 10;10(12):3315–30.

728 32. Rivera-Cuevas Y, Mayoral J, Di Cristina M, Lawrence ALE, Olafsson EB, Patel RK, et al.
729 Toxoplasma gondii exploits the host ESCRT machinery for parasite uptake of host
730 cytosolic proteins. Olias P, editor. PLOS Pathog. 2021 Dec 13;17(12):e1010138.

731 33. Braun L, Brenier-Pinchart MP, Yogavel M, Curt-Varesano A, Curt-Bertini RL, Hussain T,
732 et al. A *Toxoplasma* dense granule protein, GRA24, modulates the early immune
733 response to infection by promoting a direct and sustained host p38 MAPK activation. J
734 Exp Med. 2013 Sep 23;210(10):2071–86.

735 34. Pellegrini E, Palencia A, Braun L, Kapp U, Bougdour A, Belrhali H, et al. Structural Basis
736 for the Subversion of MAP Kinase Signaling by an Intrinsically Disordered Parasite
737 Secreted Agonist. Structure. 2017 Jan;25(1):16–26.

738 35. Bougdour A, Durandau E, Brenier-Pinchart MP, Ortet P, Barakat M, Kieffer S, et al. Host
739 Cell Subversion by Toxoplasma GRA16, an Exported Dense Granule Protein that Targets
740 the Host Cell Nucleus and Alters Gene Expression. Cell Host Microbe. 2013
741 Apr;13(4):489–500.

742 36. Coffey MJ, Sleebs BE, Uboldi AD, Garnham A, Franco M, Marino ND, et al. An aspartyl
743 protease defines a novel pathway for export of Toxoplasma proteins into the host cell.
744 eLife. 2015 Nov 18;4:e10809.

745 37. Saeij JPJ, Coller S, Boyle JP, Jerome ME, White MW, Boothroyd JC. Toxoplasma co-opts
746 host gene expression by injection of a polymorphic kinase homologue. Nature. 2007
747 Jan;445(7125):324–7.

748 38. Chang S, Shan X, Li X, Fan W, Zhang SQ, Zhang J, et al. *Toxoplasma gondii* Rhoptry
749 Protein ROP16 Mediates Partially SH-SY5Y Cells Apoptosis and Cell Cycle Arrest by
750 Directing Ser15/37 Phosphorylation of p53. Int J Biol Sci. 2015;11(10):1215–25.

751 39. Barylyuk K, Koreny L, Ke H, Butterworth S, Crook OM, Lassadi I, et al. A Comprehensive
752 Subcellular Atlas of the Toxoplasma Proteome via hyperLOPIT Provides Spatial Context
753 for Protein Functions. Cell Host Microbe. 2020 Nov;28(5):752-766.e9.

754 40. Alvarez-Jarreta J, Amos B, Aurrecoechea C, Bah S, Barba M, Barreto A, et al.
755 VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center in
756 2023. Nucleic Acids Res. 2024 Jan 5;52(D1):D808–16.

757 41. Erdős G, Pajkos M, Dosztányi Z. IUPred3: prediction of protein disorder enhanced with
758 unambiguous experimental annotation and visualization of evolutionary conservation.
759 Nucleic Acids Res. 2021 Jul 2;49(W1):W297–303.

760 42. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate
761 protein structure prediction with AlphaFold. Nature. 2021 Aug;596(7873):583–9.

762 43. The UniProt Consortium, Bateman A, Martin MJ, Orchard S, Magrane M, Ahmad S, et al.
763 UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 2023 Jan
764 6;51(D1):D523–31.

765 44. Honnappa S, Gouveia SM, Weisbrich A, Damberger FF, Bhavesh NS, Jawhari H, et al.
766 An EB1-Binding Motif Acts as a Microtubule Tip Localization Signal. Cell. 2009
767 Jul;138(2):366–76.

768 45. Sámano-Sánchez H, Gibson TJ, Chemes LB. Using Linear Motif Database Resources to
769 Identify SH2 Domain Binders. In: Carlomagno T, Köhn M, editors. SH2 Domains [Internet].
770 New York, NY: Springer US; 2023 [cited 2025 Oct 29]. p. 153–97. (Methods in Molecular
771 Biology; vol. 2705). Available from: https://link.springer.com/10.1007/978-1-0716-3393-
772 9_9

773 46. Berinstein A, Roivainen M, Hovi T, Mason PW, Baxt B. Antibodies to the vitronectin
774 receptor (integrin alpha V beta 3) inhibit binding and infection of foot-and-mouth disease
775 virus to cultured cells. J Virol. 1995 Apr;69(4):2664–6.

776 47. Kwok T, Zabler D, Urman S, Rohde M, Hartig R, Wessler S, et al. Helicobacter exploits
777 integrin for type IV secretion and kinase activation. Nature. 2007 Oct;449(7164):862–6.

778 48. Zimmermann L, Peterhans E, Frey J. RGD Motif of Lipoprotein T, Involved in Adhesion of
779 *Mycoplasma conjunctivae* to Lamb Synovial Tissue Cells. J Bacteriol. 2010 Jul
780 15;192(14):3773–9.

781 49. D'Souza SE, Ginsberg MH, Plow EF. Arginyl-glycyl-aspartic acid (RGD): a cell adhesion
782 motif. Trends Biochem Sci. 1991 Jan;16:246–50.

783 50. Liu Q, Li FC, Zhou CX, Zhu XQ. Research advances in interactions related to Toxoplasma
784 gondii microneme proteins. Exp Parasitol. 2017 May;176:89–98.

785 51. Mészáros B, Kumar M, Gibson TJ, Uyar B, Dosztányi Z. Degrons in cancer. Sci Signal.

786 2017 Mar 14;10(470):eaak9982.

787 52. Schnupf P, Portnoy DA, Decatur AL. Phosphorylation, ubiquitination and degradation of
788 listeriolysin O in mammalian cells: role of the PEST-like sequence. Cell Microbiol. 2006
789 Feb;8(2):353–64.

790 53. Pierce WK, Grace CR, Lee J, Nourse A, Marzahn MR, Watson ER, et al. Multiple Weak
791 Linear Motifs Enhance Recruitment and Processivity in SPOP-Mediated Substrate
792 Ubiquitination. J Mol Biol. 2016 Mar;428(6):1256–71.

793 54. Hajagos BE, Turetzky JM, Peng ED, Cheng SJ, Ryan CM, Souda P, et al. Molecular
794 dissection of novel trafficking and processing of the Toxoplasma gondii rhoptry
795 metalloprotease toxolysin-1. Traffic Cph Den. 2012 Feb;13(2):292–304.

796 55. Takeda K, Akira S. TLR signaling pathways. Semin Immunol. 2004 Feb;16(1):3–9.

797 56. Shi Z, Zhang Z, Zhang Z, Wang Y, Li C, Wang X, et al. Structural Insights into
798 Mitochondrial Antiviral Signaling Protein (MAVS)-Tumor Necrosis Factor Receptor-
799 associated Factor 6 (TRAF6) Signaling. J Biol Chem. 2015 Oct;290(44):26811–20.

800 57. Heinemann S, Biesinger B, Fleckenstein B, Albrecht JC. NFκB Signaling Is Induced by
801 the Oncoprotein Tio through Direct Interaction with TRAF6. J Biol Chem. 2006
802 Mar;281(13):8565–72.

803 58. Ye H, Arron JR, Lamothe B, Cirilli M, Kobayashi T, Shevde NK, et al. Distinct molecular
804 mechanism for initiating TRAF6 signalling. Nature. 2002 Jul;418(6896):443–7.

805 59. Lamarque MH, Papoin J, Finizio AL, Lentini G, Pfaff AW, Candolfi E, et al. Identification
806 of a New Rhoptry Neck Complex RON9/RON10 in the Apicomplexa Parasite Toxoplasma
807 gondii. Moreno SN, editor. PLoS ONE. 2012 Mar 12;7(3):e32457.

808 60. Alaganan A, Fentress SJ, Tang K, Wang Q, Sibley LD. *Toxoplasma* GRA7 effector

809 increases turnover of immunity-related GTPases and contributes to acute virulence in the
810 mouse. Proc Natl Acad Sci. 2014 Jan 21;111(3):1126–31.

811 61. Hunter CA, Sibley LD. Modulation of innate immunity by Toxoplasma gondii virulence
812 effectors. Nat Rev Microbiol. 2012 Nov;10(11):766–78.

813 62. Ihara F, Fereig RM, Himori Y, Kameyama K, Umeda K, Tanaka S, et al. Toxoplasma gondii
814 Dense Granule Proteins 7, 14, and 15 Are Involved in Modification and Control of the
815 Immune Response Mediated via NF-κB Pathway. Front Immunol. 2020 Jul 31;11:1709.

816 63. Yang CS, Yuk JM, Lee YH, Jo EK. Toxoplasma gondii GRA7-Induced TRAF6 Activation
817 Contributes to Host Protective Immunity. Adams JH, editor. Infect Immun. 2016
818 Jan;84(1):339–50.

819 64. Wilson BA, Foy SG, Neme R, Masel J. Young genes are highly disordered as predicted
820 by the preadaptation hypothesis of de novo gene birth. Nat Ecol Evol. 2017 Apr
821 24;1(6):0146.

822 65. Davey NE, Cowan JL, Shields DC, Gibson TJ, Coldwell MJ, Edwards RJ. SLiMPrints:
823 conservation-based discovery of functional motif fingerprints in intrinsically disordered
824 protein regions. Nucleic Acids Res. 2012 Nov;40(21):10628–41.

825 66. Idrees S, Paudel KR, Banik M, Suwal N, Thapa R, Bashyal S. Predicting Motif-Mediated
826 Interactions Based on Viral Genomic Composition. Int J Mol Sci. 2025 Apr 13;26(8):3674.

827 67. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold
828 Protein Structure Database: massively expanding the structural coverage of protein-
829 sequence space with high-accuracy models. Nucleic Acids Res. 2022 Jan
830 7;50(D1):D439–44.

831 68. Wilson CJ, Choy WY, Karttunen M. AlphaFold2: A Role for Disordered Protein/Region
832 Prediction? Int J Mol Sci. 2022 Apr 21;23(9):4591.

833 69. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of
834 hydrogen-bonded and geometrical features. Biopolymers. 1983 Dec;22(12):2577–637.

835 70. Tien MZ, Meyer AG, Sydykova DK, Spielman SJ, Wilke CO. Maximum Allowed Solvent
836 Accessibilites of Residues in Proteins. Porollo A, editor. PLoS ONE. 2013 Nov
837 21;8(11):e80635.

838 71. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation
839 of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol.
840 2011 Jan;7(1):539.

841 72. Davey NE, Seo M, Yadav VK, Jeon J, Nim S, Krystkowiak I, et al. Discovery of short linear
842 motif-mediated interactions through phage display of intrinsically disordered regions of the
843 human proteome. FEBS J. 2017 Feb;284(3):485–98.

844 73. Levy ED. A simple definition of structural regions in proteins and its use in analyzing
845 interface evolution. J Mol Biol. 2010 Nov 5;403(4):660–70.

846 74. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex
847 prediction with AlphaFold-Multimer [Internet]. Bioinformatics; 2021 [cited 2025 Jun 5].
848 Available from: http://biorxiv.org/lookup/doi/10.1101/2021.10.04.463034

849 75. Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, et al. UCSF
850 CHIMERAX : Tools for structure building and analysis. Protein Sci. 2023 Nov;32(11):e4792.

851 76. Scholz J, Besir H, Strasser C, Suppmann S. A new method to customize protein
852 expression vectors for fast, efficient and background free parallel cloning. BMC
853 Biotechnol. 2013 Dec;13(1):12.

854

855

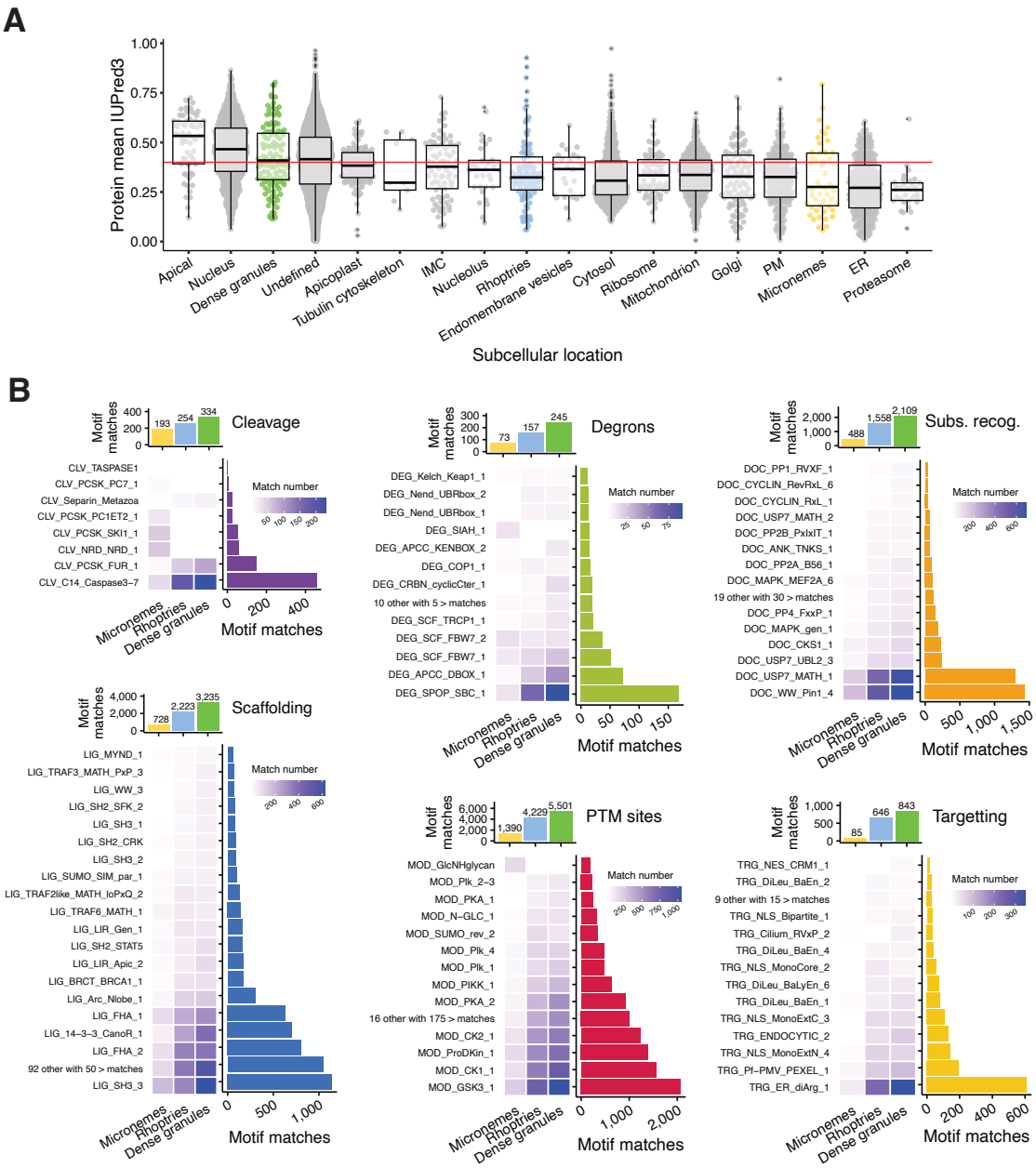
856

Supplementary information

Table S1 – Motif predictions in microneme, rhoptry and dense granule proteins

Table S2 – Selected TRAF6 motif-containing peptides and binding assay results

Data S1 – MST binding extended results for the selected TRAF6 motif-containing peptides



Supplementary Figure 1 - Protein disorder in different locations and matches in secreted proteins. A. IUPred3 mean score for proteins of the different LOPIT subcellular locations. **B**. Motif matches of the most frequent motif classes by motif.

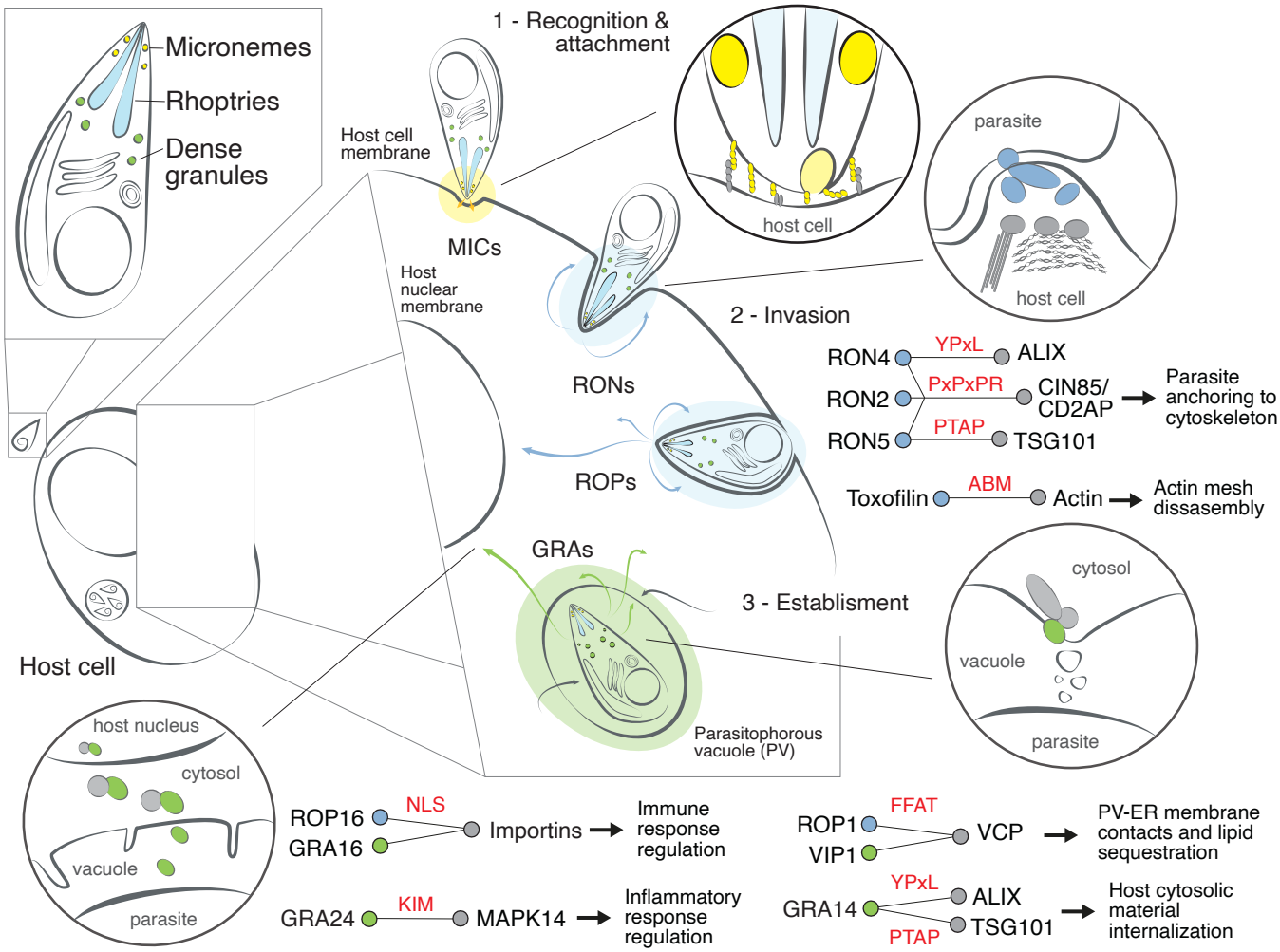


Figure 1. Motif use and hijacking during the *Toxoplasma* infection process. *Toxoplasma gondii* secretory organelles and infection cycle illustrating different host-pathogen interaction contexts, known motif instances and their infection outcomes. Motif names are highlighted in red, proteins from micronemes are highlighted in yellow, those from rhoptries in blue, and those from dense granules in green.

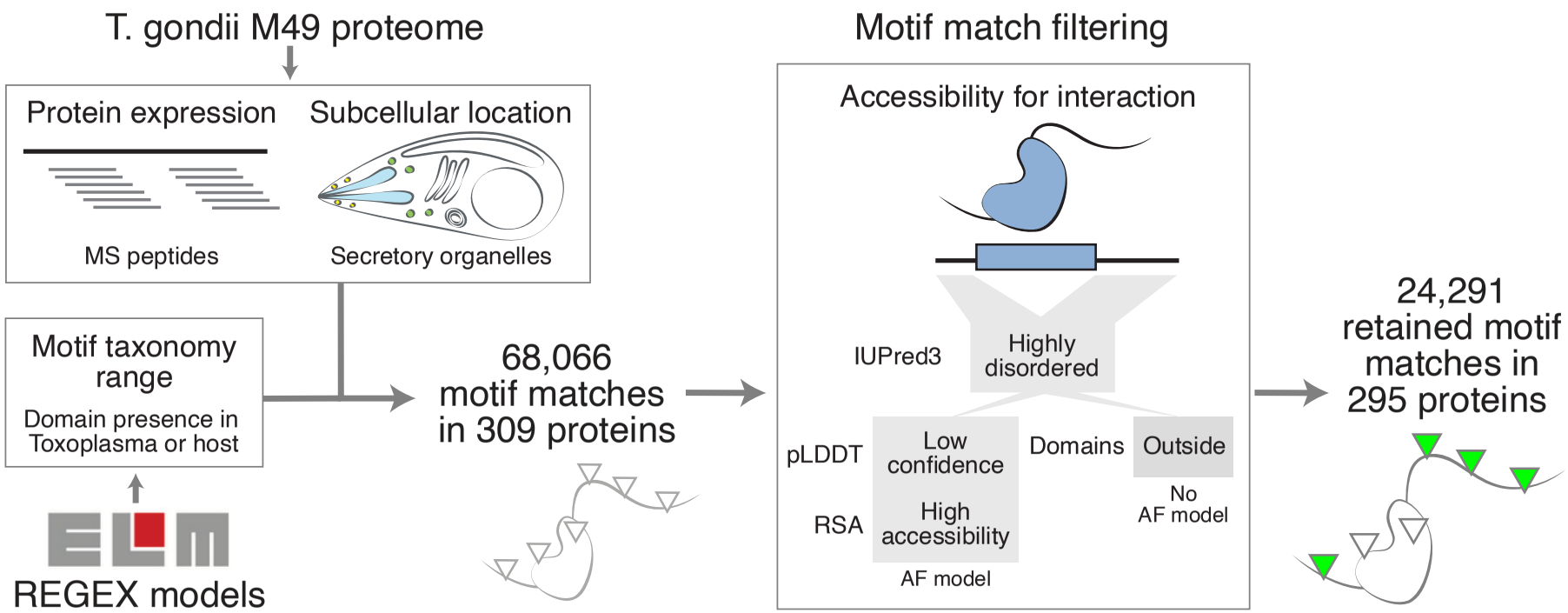


Figure 2 - Motif prediction pipeline. Diagram outlining the protein and motif model selection for motif prediction, as well as motif accessibility filtering criteria.

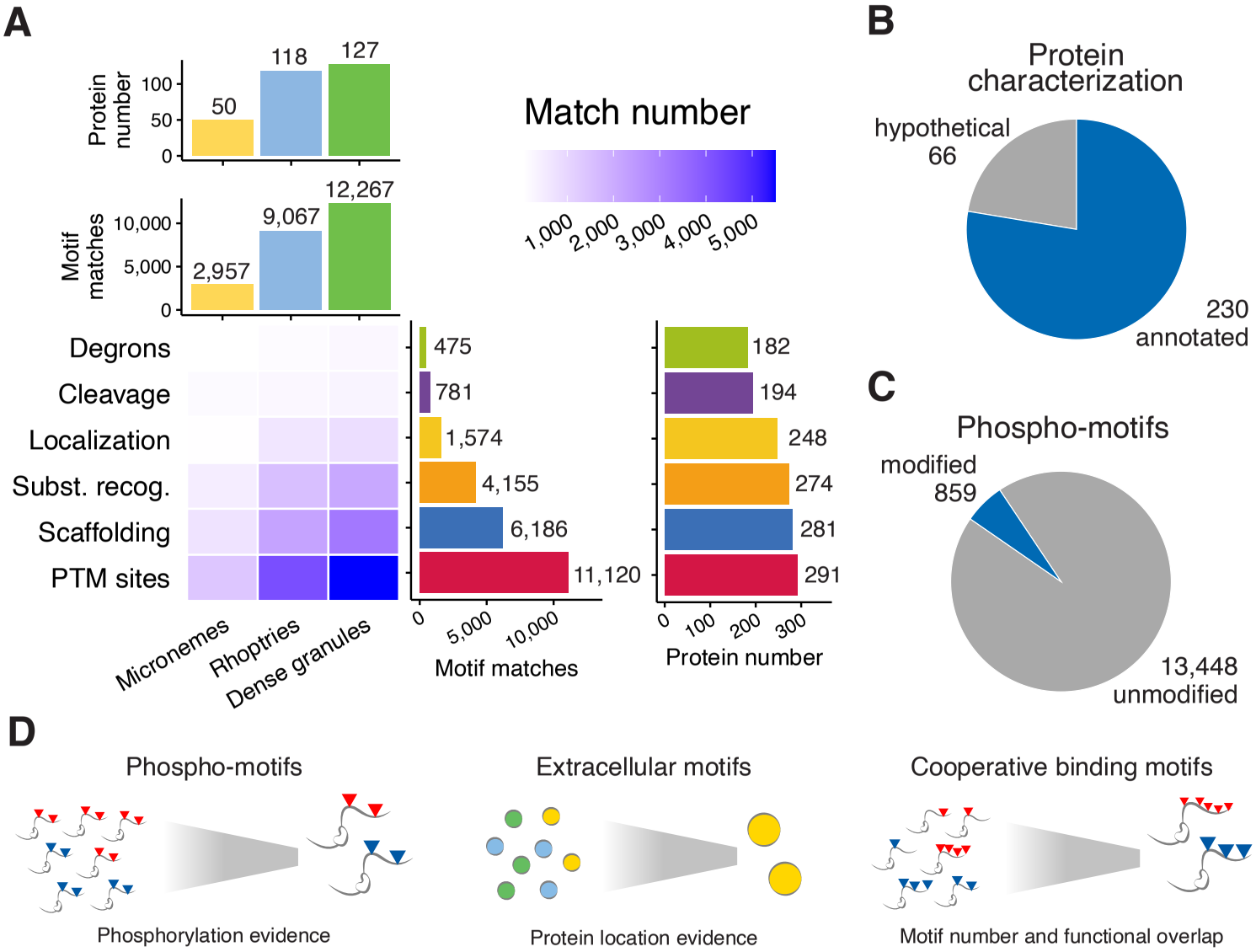


Figure 3 - Matches in secreted proteins. A. Number of filtered motif matches from secreted proteins across different motif categories. **B**. Split of proteins with filtered matches according to their annotation. **C**. Split of filtered phospho-motifs according to whether they have phosphorylation evidence. **D**. Additional motif filtering strategies according to motif type functions and properties.

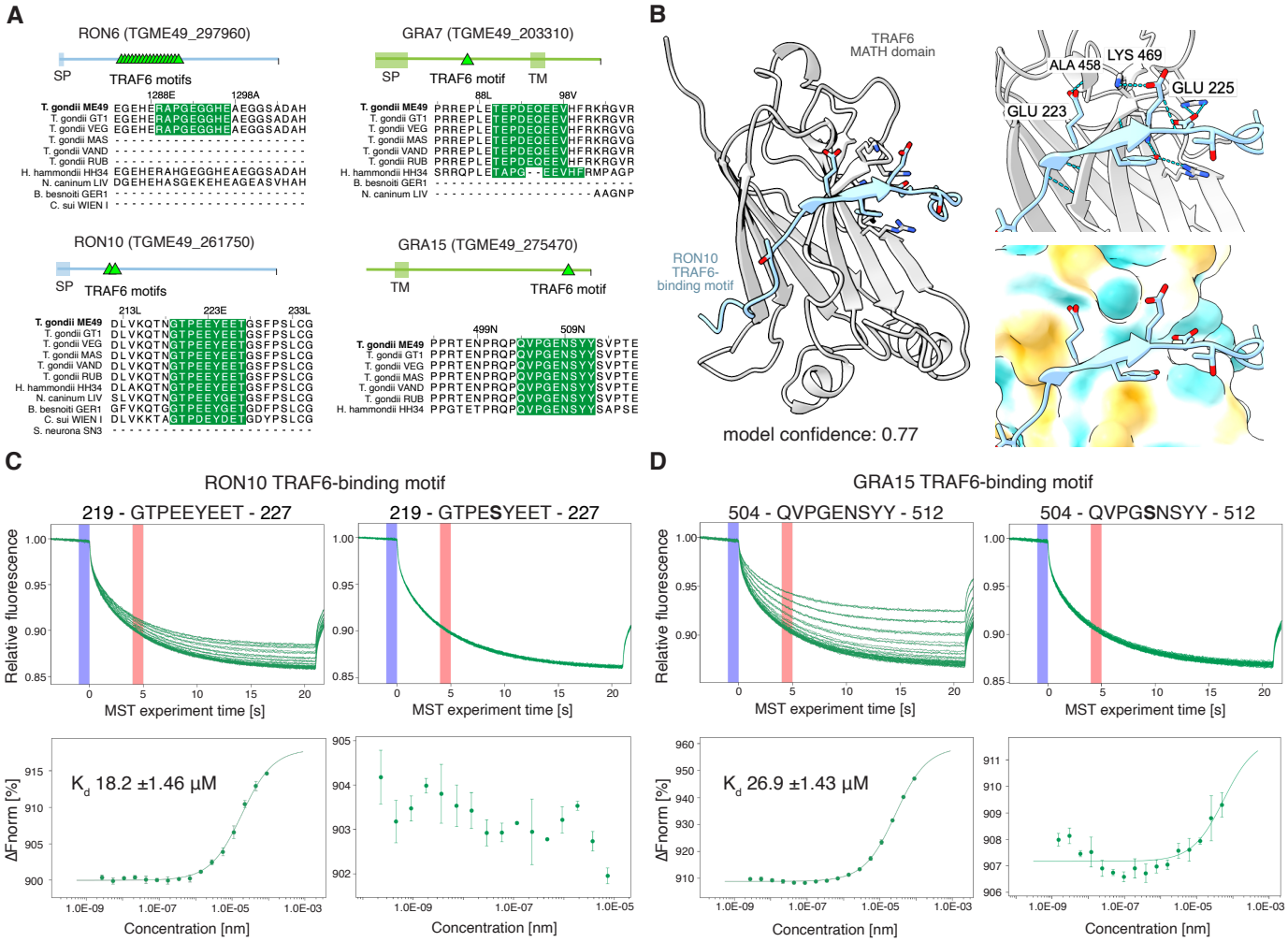
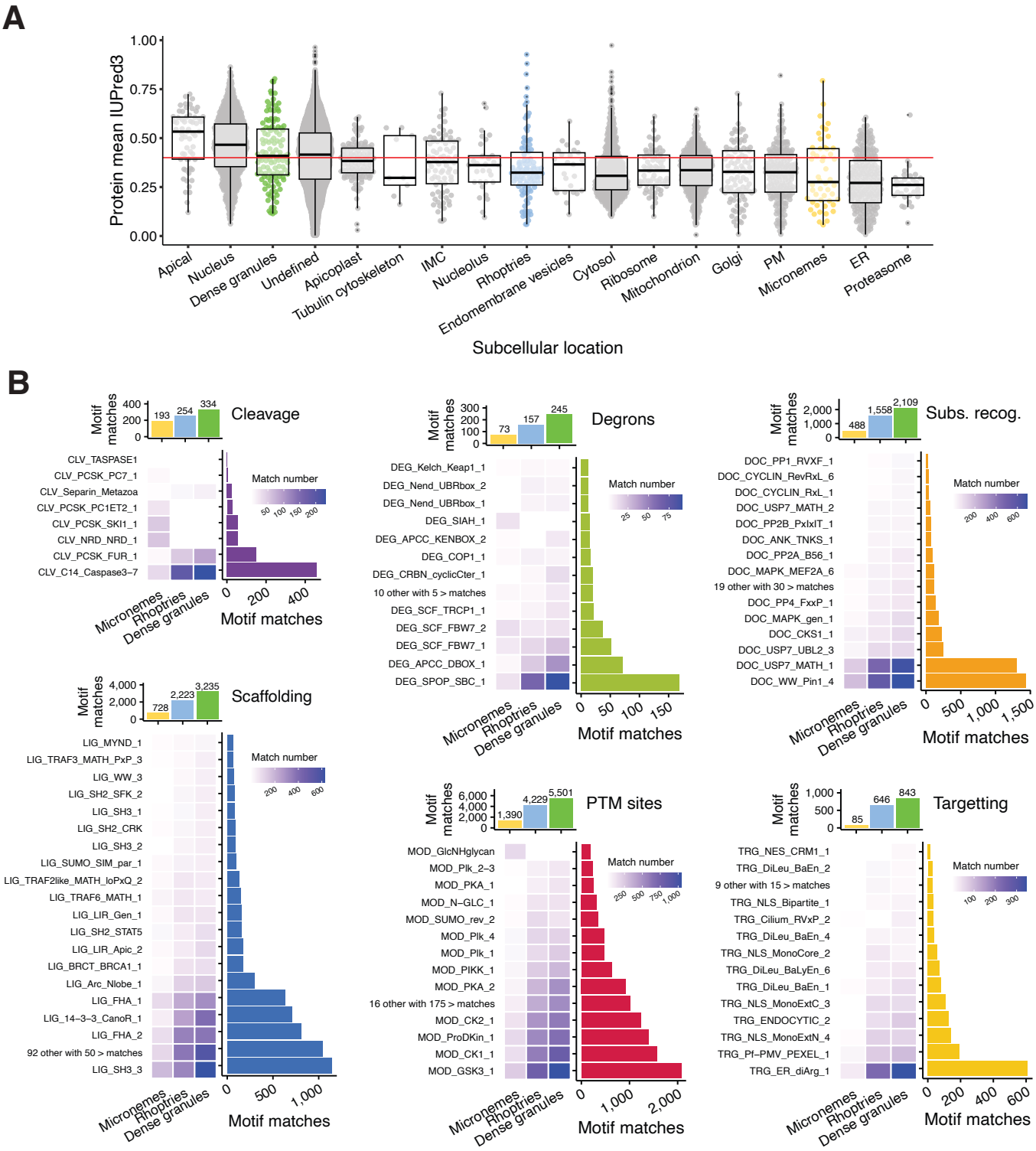


Figure 4 - TRAF6 motif binding assays. A. Protein diagrams of RON6, RON10, GRA7 and GRA15 indicating the position of their TRAF6-binding motifs with accompanying multiple sequence alignment showing their presence in different Toxoplasma strains and Coccidian species. **B.** AlphaFold model of TRAF6 MATH domain with TRAF6-binding motif from RON10. Right panels display the potential hydrogen bond network at the interaction interface (top) and secondary structure adopted by the TRAF6-binding motif (bottom). **C.** RON10 TRAF6 motif testing. MST traces (top graph) and dose-response curves (bottom panel) of TRAF6 MATH domain binding to RON10 TRAF6 motif containing peptide (left) and the E>S mutant (right). In the MST traces, the cold region is set at 0 s (blue) and the hot detection region at 5 s (red). **D**. Same as in A, but for the GRA15 TRAF6 motif and the E>S motif mutant.



Supplementary Figure 1 - Protein disorder in different locations and matches in secreted proteins. A. IUPred3 mean score for proteins of the different LOPIT subcellular locations. **B**. Motif matches of the most frequent motif classes by motif.