Short linear motifs - Unexplored 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

* Corresponding author

E-mail: j.alvarado@imb-mainz.de (JAV)

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 20 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 27,832 motif matches in 296 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.

Author summary

Intracellular pathogens such as viruses, bacteria, and parasites infect the cells of their host and hijack their inner components to thrive and multiply. One way in which these pathogens interact with host proteins is through short linear motifs, small functional modules commonly

found in unstructured parts of proteins recognised by folded protein domains. *Toxoplasma gondii* is an intracellular parasite widely distributed in the human population and capable of infecting any warm-blooded animal. Given that unstructured protein regions are a common feature of Toxoplasma's proteins, we hypothesise that motifs play important roles during infection. Here, we collected representative examples from the literature to create a reference of the current state of motif usage during the Toxoplasma host cell invasion cycle. We developed a computational pipeline to predict thousands of motifs in Toxoplasma proteins. We then outline strategies for working with these predictions and show their application by experimentally testing proteins containing a motif involved in the innate immune response, successfully showing binding for two motifs to a human protein domain. 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). 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

84 their unbound state and often adopt an alpha-helix or beta-strand upon binding to partner
85 proteins (6). Because only a few mutations are needed to create a motif, they can evolve in
86 much shorter evolutionary timescales compared to folded protein regions (7). Thus,
87 pathogens have evolved to exploit these host motifs to infect their hosts, probably because
88 motifs are short and easily emerge in proteins (7), but also because of their many functional
89 roles in cell signalling. Many instances of host motif mimicry have been characterised for
90 viruses and bacteria (3,4,8–10). Although there have been some studies focusing on
91 intracellular eukaryotic parasites (11,12), the extent to which parasites use motifs is less
92 understood. *Toxoplasma gondii,* a member of the Apicomplexa phylum of unicellular
93 parasites, has one of the proteomes with higher levels of IDRs among Eukaryotes (13,14),
94 making it a good candidate to discover and characterise the role of motifs in its proteins during
95 parasitic infection.
96
97 *Toxoplasma* is a successful eukaryotic unicellular parasite, as it can infect any warm-blooded
98 animal and proliferate within any of its nucleated cells. It is estimated that a third of the human
99 population has been infected with Toxoplasma (15). Even though most of the infections in
100 humans are asymptomatic, immunocompromised individuals and susceptible populations can
101 develop different forms of Toxoplasmosis, from influenza-like symptoms to severe phenotypes
102 such as encephalitis, making *Toxoplasma* a global health burden (15,16). *Toxoplasma's*
103 definitive hosts are cats and all members of the feline family, where it reproduces sexually in
104 their gut to then disseminate in the environment. In all other hosts, such as all mammals and
105 birds, *Toxoplasma* reproduces asexually in different cell types to multiply its numbers and
106 disseminate along the food chain (17). Having multiple hosts means that *Toxoplasma* not only
107 has the molecular tools to infect different host cells but also to deal with diverse cell
108 programmes and immune responses, to thrive in each of its hosts. Motifs in Toxoplasma
109 proteins might play important roles in mediating the adaptation to the many different host
110 environments this parasite encounters.
111

112 Once consumed by its host, *T. gondii* cell invasion cycles depend on the sequential secretion
113 of proteins contained in its specialised infection organelles: the micronemes, rhoptries and
114 dense granules (Fig. 1A). Microneme proteins, or MICs, are targeted to the parasite cell
115 surface and are involved in host recognition and attachment, while rhoptry proteins, termed
116 RONs and ROPs, are secreted into the host cytoplasm and are involved in host invasion and
117 niche establishment (Fig. 1B). During the invasion process, a parasitophorous vacuole (PV)
118 is created from the invagination of the host cell membrane (18). From within this space,
119 Toxoplasma creates a suitable environment and continues to secrete proteins from
120 membraneless organelles called dense granules. These proteins, mostly known as GRAs, are
121 secreted either into the PV lumen, inserted into its membrane (PVM) or into the host cytoplasm
122 and even into the host nucleus, helping the parasite in modulating the host cell's regulatory
123 processes to proliferate (19). All these secreted proteins interact with host proteins throughout
124 the infection cycle, including host cell receptors, cytoskeletal proteins and transcription factors
125 (20,21). These host-parasite protein interactions are the subject of continuous study, serve as
126 a reference to study other Apicomplexans, and are promising targets for the development of
127 therapies. Thus, knowledge about motifs in *Toxoplasma* secreted proteins would increase our
128 mechanistic understanding of Toxoplasma infection strategies.
129
130 To identify motifs in *Toxoplasma gondii* proteins, we collected motif instances in *T. gondii* from
131 the literature, predicted occurrences of known motif types in its secreted proteins through the
132 development of an integrative discovery pipeline, and tested the binding of predicted TRAF6
133 binding motifs *in vitro,* which might have relevance in the regulation of the innate immune
134 response. The available motif predictions along with annotations will motivate further
135 functional studies in *Toxoplasma* and other Apicomplexan parasites to understand
136 mechanisms of infection and provide further examples of motif repurposing among
137 intracellular pathogens.
138
139

140 **Results**

141 **Curation of known motifs in *Toxoplasma gondii* secreted proteins**

142 We collected 21 experimentally validated motif instances from 11 Toxoplasma proteins from
143 the scientific literature that serve as a reference of the current motif knowledge in *Toxoplasma*
144 *gondii* (Table 1, Fig. 1). The Eukaryotic Linear Motif (ELM) resource is a reference point for
145 curated eukaryotic motifs from the scientific literature. Based on motif sequence patterns,
146 structures, binder information, and functional annotations, curators define motif types and
147 group motif types into broader motif classes (22). Two of the Toxoplasma motif instances were
148 previously annotated as substrate recognition sites in ELM (DOC). Eleven motif instances
149 were novel with respect to ELM content and were added to known motif classes of scaffolding
150 classes in ELM (LIG). The remaining eight novel motif instances are similar but could not able
151 to be directly annotated to ELM. Those correspond to one scaffolding class (LIG) and to three
152 subcellular localisation classes (TRG in ELM). We did not annotate motif instances in
153 Toxoplasma secreted proteins reported as motifs of cleavage sites (CLV in ELM), degradation
154 signals (DEG in ELM) or sites for post-translational modifications (MOD in ELM). As detailed
155 below, the curated 21 motif instances function throughout parasite invasion and establishment
156 (Fig. 1).

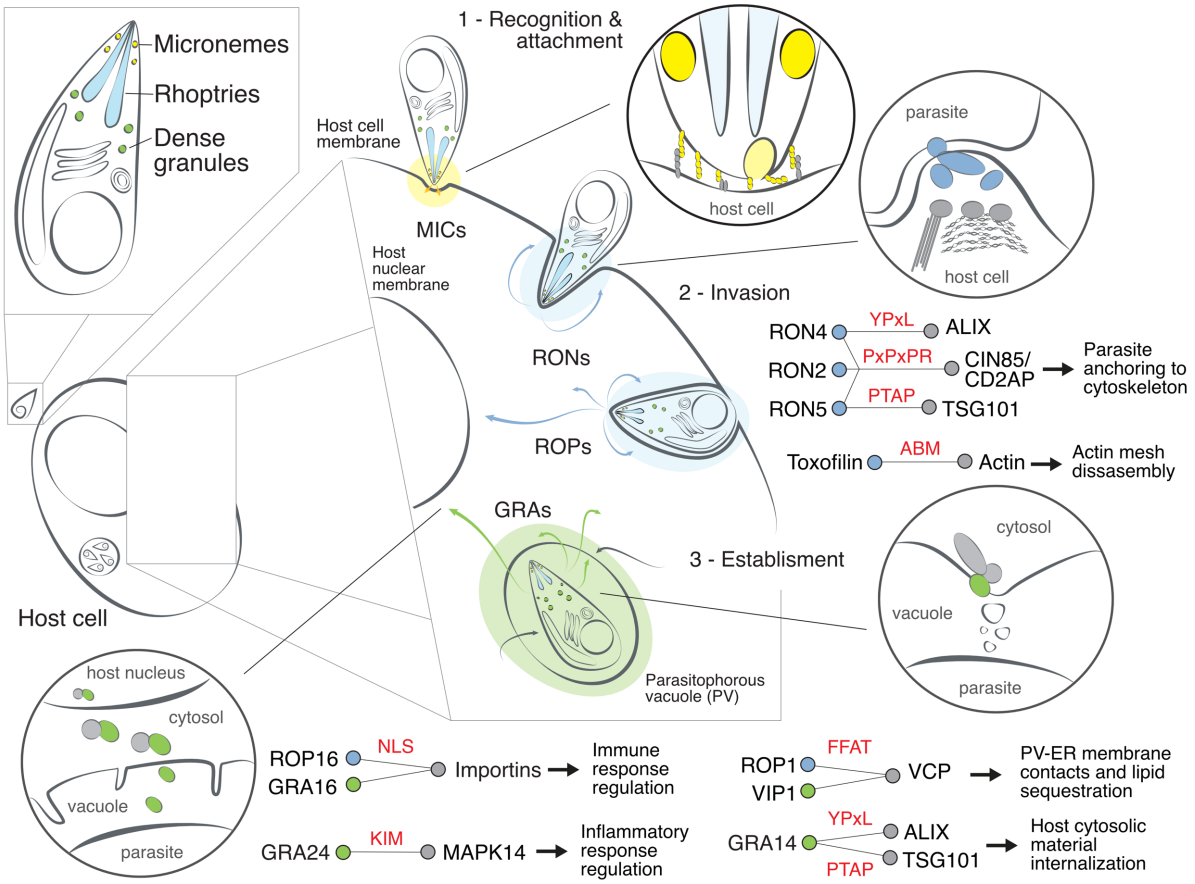
157

Protein name	ToxoDB (UniProt)	Motif name (ELM ID)	Host protein (Domain name)	Instances	Source
Rhoptry proteins					
RON2	TGME49_300100 (B6KV60)	PxPxPR (LIG_SH3_CIN85_PxpxPR_1)	CIN85 or CD2AP (SH3)	1	(23)
RON4	TGME49_229010 (S8F0B5)	PxPxPR (LIG_SH3_CIN85_PxpxPR_1)	CIN85 or CD2AP (SH3)	3	(23)
		LYPxL (LIG_LYPXL_S_1)	Alix (V-shaped)	2	(23)
RON5	TGME49_311470 (S8ERS3)	PxPxPR (LIG_SH3_CIN85_PxpxPR_1)	CIN85 or CD2AP (SH3)	1	(23)
		PTAP (LIG_PTAP_UEV_1)	Tsg101 (UEV)	2	(23)
ROP1	TGME49_309590 (A0A125YP48)	FFAT (TRG_ER_FFAT_1 like)	VAP (MSP)	1	(29)

ROP16	TGME49_262730 (Q3YJR5)	NLS (TRG_NLS_MonoExtN_4)	Exportin (Arm)	1	(36)
Toxofilin	TGME49_214080 (S8G4J8)	Actin-binding motif (LIG_Actin_WH2_1 like)	Actin	1	(27)
Dense granule proteins					
GRA14	TGME49_239740 (B5TVE8)	LYPxL (LIG_LYPXL_S_1)	Alix (V-shaped)	1	(31)
		PTAP (LIG_PTAP_UEV_1)	Tsg101 (UEV)	1	(31)
GRA16	TGME49_208830 (Q1JSN3)	Similar to PEXEL (TRG_Pf-PMV_PEXEL_1)	-	1	(35)
		NLS (TRG_NLS_MonoExtN_4)	Exportin (Arm)	2	(34)
GRA24	TGME49_230180 (A0A125YG37)	KIM (DOC_MAPK_GRA24_9)	MAPK (Kinase)	2	(32)
MYR1	TGME49_254470 (S8GT06)	Similar to PEXEL (TRG_Pf-PMV_PEXEL_1)	-	1	(35)
VIP1	TgME49_233220 (S8F2E2)	FFAT (TRG_ER_FFAT_1 like)	VAP (MSP)	1	(30)

158 **Table 1. Experimentally validated linear motif instances**. Known *Toxoplasma* motif instances in
159 rhoprty and dense granule proteins from the literature, and their interacting host protein domains.
160 Protein IDs correspond to the ToxoDB and UniProt accessions, while motif IDs correspond to the ELM
161 motif classes.
162
163 Rhoptry proteins RON2, RON4 and RON5 each contain PxpxPR motifs. During host cell
164 invasion, these motifs help the parasite connect its tight junction (TJ) complex to the host
165 cytoskeleton (23) (Fig. 1). The PxpxPR motifs bind to the SH3 domain of host adaptor proteins
166 CIN85 or CD2AP, which are involved in cytoskeletal rearrangements, vesicle trafficking and
167 endocytosis (24,25). Additionally, RON4 and RON5 contain instances of the motif types
168 LYPxL and PTAP. The LYPxL motif mediates the interaction of proteins with ALIX, while the
169 PTAP motif binds to the UEV domain of class E vacuolar sorting protein TSG101. ALIX and
170 TSG101 are both key components of the ESCRT membrane remodelling system, and during
171 *Toxoplasma* cell invasion, they are also recruited to the TJ for an efficient cell entry (23,26).
172 This process is also aided by parasite proteins such as Toxifilin, which binds to host actin
173 through an actin-binding motif (ABM) that is similar in sequence pattern to the annotated WH2-

174 binding motif in ELM (27). Toxofilin sequesters actin and promotes the disassembly of the
175 host cortical actin meshwork, facilitating an efficient internalisation (28).
176



177

178 **Figure 1. Motif use and hijacking during the *Toxoplasma* infection process.** *Toxoplasma gondii*
179 secretory organelles and infection cycle illustrating different host-pathogen interaction contexts, known
180 motif instances and their infection outcomes. Motif names are highlighted in red, proteins from
181 micronemes are highlighted in yellow, those from rhoptries in blue, and those from dense granules in
182 green.
183

184 Once *Toxoplasma* is established within the host cell, ROP1 and dense granule proteins, VIP1
185 and GRA14 are secreted to the parasitophorous vacuole membrane (PVM) (Fig. 1). ROP1
186 and VIP1 both contain FFAT motifs (similar in sequence motif pattern to the FFAT motif type
187 in ELM) that modulate the targeting of proteins to the endoplasmic reticulum via its interaction
188 with Vesicle Associated Membrane Protein (VAMP) Associated Proteins (VAPs). As ROP1

189 and VIP1 are inserted in the PVM, their interaction with VAP proteins leads to the association
190 of the PVM with the host ER membrane. Toxoplasma exploits this proximity to sequester the
191 lipids necessary for its vacuole and plasma membrane biogenesis (29,30). GRA14 harbours
192 both LYPxL and PTAP motifs in its C-terminal region. Since GRA14 is also inserted into the
193 PVM with its C-terminus facing the host cytosol, the motifs can recruit ALIX and TSG101,
194 which, in this case and differently from TJ proteins, help *Toxoplasma* sequester host cytosolic
195 material through the budding of its PVM (31).
196
197 Further, during establishment, more parasite proteins are secreted into the host cytosol and
198 may then be transported to the host nucleus. Within the host cytosol, secreted proteins interact
199 with host proteins via motifs. The intrinsically disordered protein GRA24 contains two
200 instances of a motif that binds to mitogen-activated protein kinase p38α (MAPK14). These
201 motifs allow GRA24 to bind to a MAPK14 dimer, forming a stable complex that leads to the
202 expression of proinflammatory factors, which in turn modulate parasite replication (32,33).
203 Parasite protein export is achieved by the presence of cell compartment localisation motifs,
204 which either use the parasite or host machinery to be secreted. In the first case, the correct
205 export of dense granule protein MYR1 to the PV and of GRA16 to the cytosol is dependent
206 on the cleavage of a short linear motif by a parasite aspartyl protease, in a similar way as the
207 Plasmodium PEXEL export motif (34,35). In the case of host nuclear targeting, GRA16, as
208 well as rhoptry protein kinase ROP16, are imported into the host nucleus via nuclear
209 localisation signals (NLS). ROP16 activates signal transducer and activator of transcription
210 (STAT) signalling pathways to inhibit host cell death and modulate the production of
211 interleukins (36,37). Within the nucleus, GRA16 regulates genes involved in cell cycle
212 progression (34). Overall, this set of curated motif instances exemplifies their role during
213 important stages of the Toxoplasma infection cycle. Nevertheless, the small number of known
214 motifs as well as their occurrence in a few proteins, highlight the need and potential to find
215 and characterise many more motif instances to facilitate further mechanistic studies of
216 Toxoplasma pathogenicity.

217

218 **Prediction of motif instances in *Toxoplasma* secreted proteins**

219 To identify novel motif instances that function in the Toxoplasma infection process via
220 interaction with host proteins, we focused on *Toxoplasma gondii* proteins from the strain ME49
221 that are secreted into host cells. To determine the subcellular localisation of secreted proteins,
222 we used the dataset from the hyperLOPIT experiment (38) and annotations from the ToxoDB,
223 a database from the VEuPathDB resource, the primary reference for the research of
224 eukaryotic pathogens (protists and fungi), their hosts and invertebrate vectors (39). The
225 secreted proteins were subsequently filtered for those with evidence of expression using mass
226 spectrometry datasets from the ToxoDB. To predict motif instances in these remaining 314
227 Toxoplasma proteins, we developed a computational pipeline based on sequence pattern
228 searches as well as structural and functional annotations of predicted motifs (Fig. 2). To find
229 putative motif instances, we used regular expressions defined for annotated motif classes in
230 the ELM database, which capture the preferences for certain amino acids at certain motif
231 positions. We scanned the selected proteins with 298 regular expressions corresponding to
232 motif classes with interacting domains that could be present either in Toxoplasma or in the
233 host using the taxonomic range annotations provided by the ELM resource. This motif search
234 resulted in a total of 68,066 motif matches in 309 proteins. Since motifs occur primarily in IDRs
235 and need to be accessible for binding to the partner proteins, we deployed structural
236 annotations to further restrict to more likely functional motif matches. To this end, we
237 evaluated the location of the motif match in IDRs using the software IUPRed3(40). In addition,
238 in cases where a full-length structural model of a given protein was available from the EBI
239 AlphaFold database (41), we evaluated motif accessibility through low pLDDT values and high
240 relative solvent accessibility (RSA). In cases where a structural model was not available, we
241 excluded motif matches that fell within annotated folded domain boundaries as annotated in
242 UniProt (42) (Fig. 2). Using this framework, we obtained a set of 24,291 motif matches within
243 295 proteins from micronemes, rhoptries and dense granules (Fig. 2, Fig. 3A) (Table S1).

244

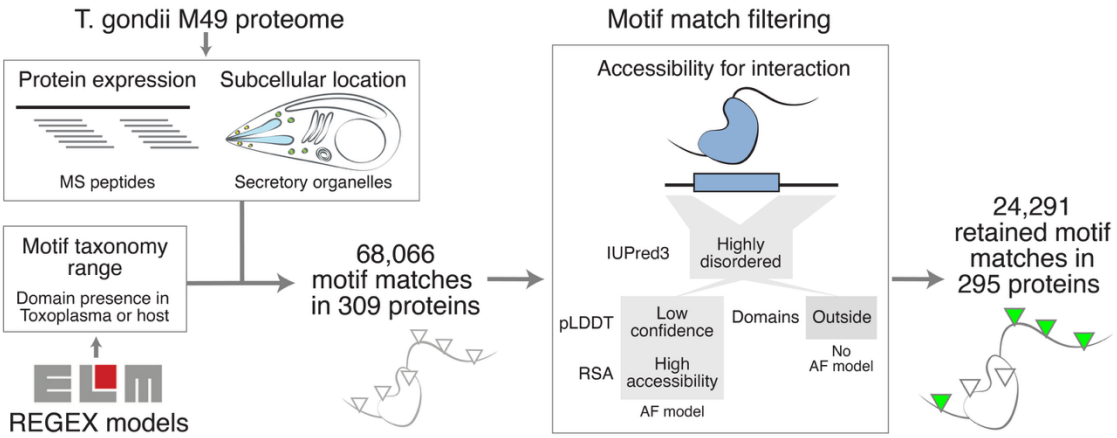
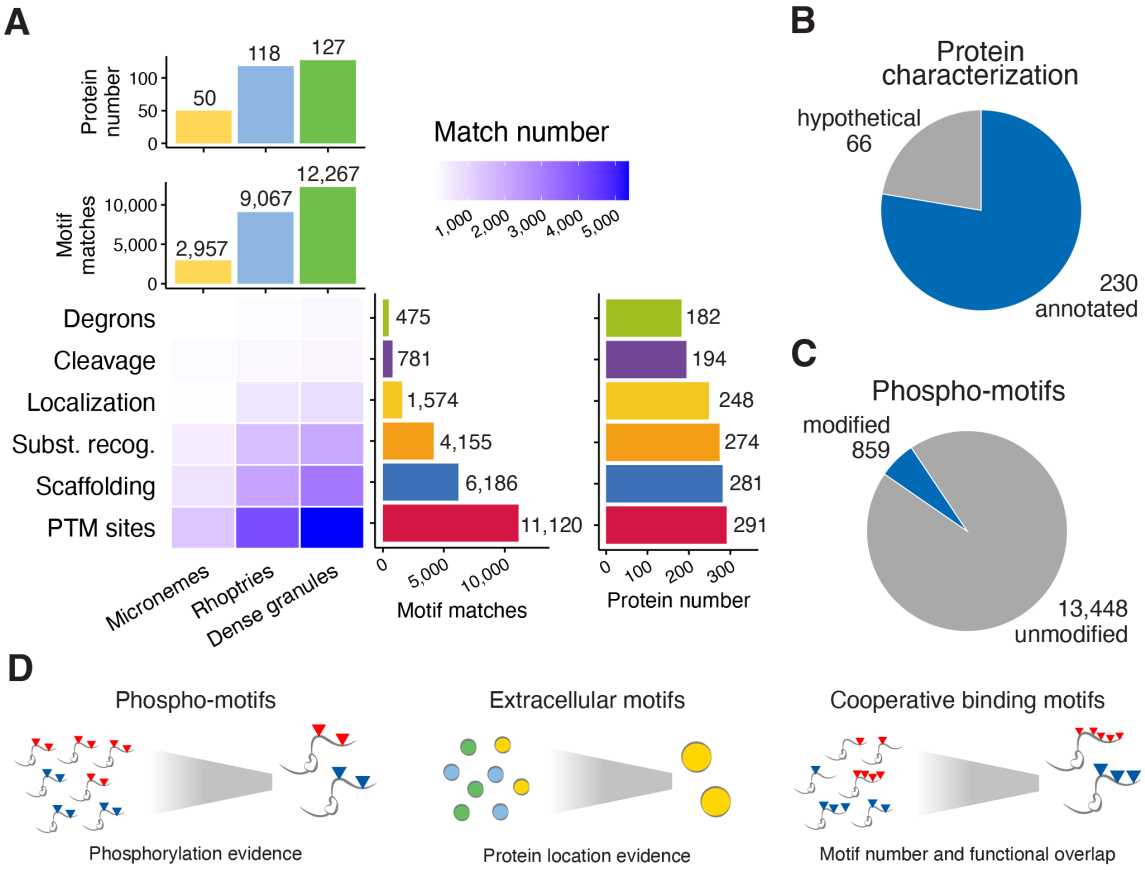


Figure 2 - Motif prediction pipeline. Diagram outlining the protein and motif model selection for motif prediction, as well as motif accessibility filtering criteria.

Our computational pipeline captured 16 of the 21 known motif instances within our retained motif matches. Two PTAP motifs in RON5 were missed because they have low protein IDR scores. The FFAT motifs in ROP1 and VIP1, and the actin-binding motif in Toxofilin were not captured because their sequences did not match any ELM regular expression. These last three examples highlight that current annotated motif types in the ELM resource do not fully capture the diversity of motifs in *Toxoplasma* proteins. Our curated motif instances could thus be used to reevaluate and extend ELM motif type annotations. When further analysing the proteins with curated motif instances here, we found a previously unreported PxpxPR motif in the N-terminal region of RON4 as well as two SxIP motifs. SxIP motifs modulate the binding to microtubule plus-end tracking proteins (+TIPs), which have an important role in chromosome segregation, cell signalling and organelle trafficking (36). SxIP motifs are already known to be exploited by the Apicomplexan parasite *Theileria* (12). RON8, which, together with RON2, RON4, and RON5, forms part of the TJ (22), also contained matches of the SxIP and LYPxL motifs. The presence of extra motif matches related to host cytoskeleton in TJ proteins indicates that the current model of motif-hijacking during this stage of host cell invasion is still incomplete and will benefit from our motif predictions. Furthermore, the identification of 4,822 motif matches in our dataset from 66 proteins annotated as

266 “hypothetical” in ToxoDB and with expression evidence (Fig. 3B) represents a potential venue
267 to explore, assign and understand their function.
268



269

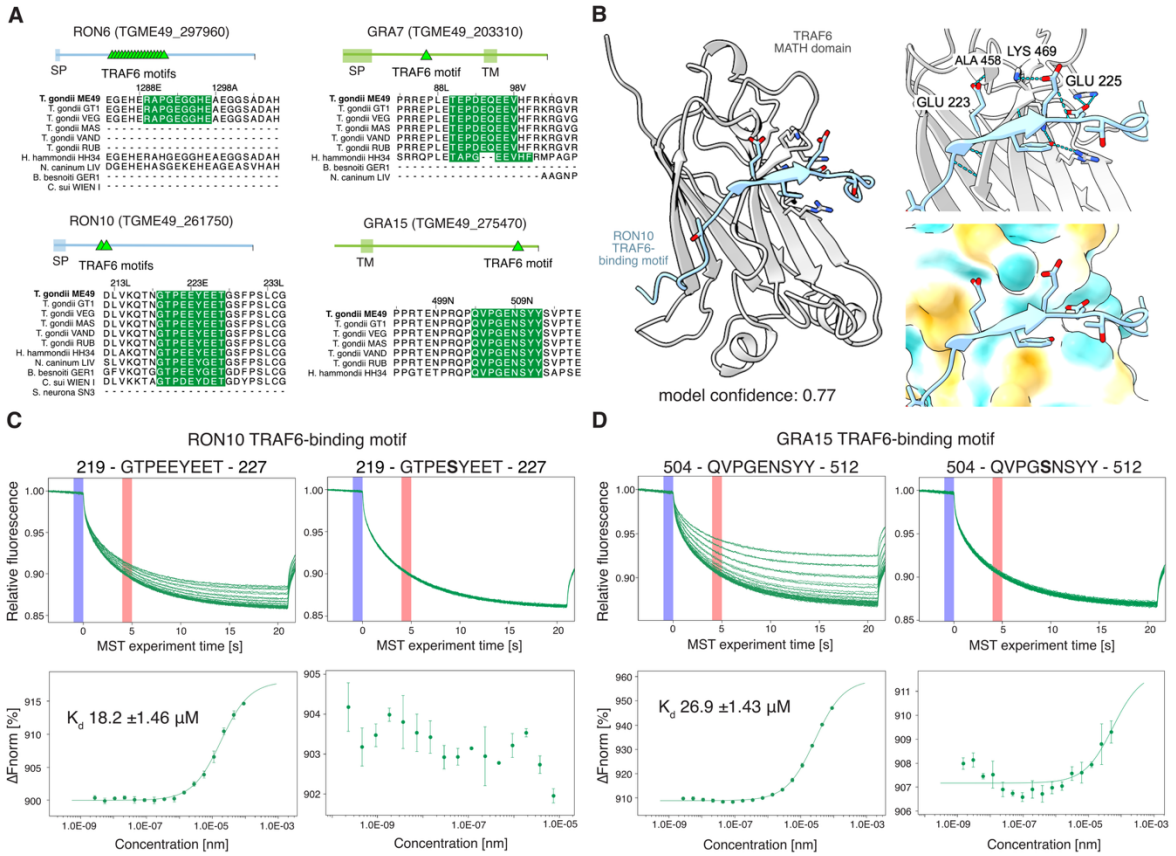
270 **Figure 3 - Matches in secreted proteins. A**. Number of filtered motif matches from secreted proteins
271 across different motif categories. **B**. Split of proteins with filtered matches according to their annotation.
272 **C**. Split of filtered phospho-motifs according to whether they have phosphorylation evidence. **D**.
273 Additional motif filtering strategies according to motif type functions and properties.
274

275 On average, our pipeline obtained 82 retained motifs per secreted protein: 59 in micronemes,
276 77 in rhoptries and 97 in dense granules (Fig. 3A). While proteins, especially those with longer
277 IDRs such as dense granule proteins (Fig. S1A), are expected to have multiple motif
278 occurrences, the relatively high number of predicted motif instances still points to some of
279 those being likely non-functional. Most of the motif matches come from motif types that denote
280 sites for PTMs such as phosphorylation (Fig. 3A). This most likely results from very

281 degenerate regular expressions available for motif types from the PTM class, such as
282 …[ST]...[ST] defining sites phosphorylated by GSK kinase. In regular expressions, a "."
283 corresponds to any amino acid possible at this position and amino acids in brackets denote
284 possible amino acids at this position in the motif. More than 2,000 motif matches alone were
285 identified with this GSK kinase phosphorylation motif pattern. Phosphoproteomics data can
286 be used to provide evidence for actual modification of these predicted motifs (44). Using
287 available phosphosite datasets from ToxoDB, we found evidence of phosphorylation for ~6%
288 of predicted phosphorylation sites (Fig. 3C-D), assigning higher confidence to these predicted
289 motifs (Table S1). We generally observe that for most motif classes, motif matches were
290 identified only for a fraction of annotated motif types in ELM. For example, 38% of motif
291 matches from the targeting class (TRG) correspond to diArg motifs, 35% of motif matches
292 from the degradation class (DEG) correspond to SPOP degrons (see below), and 18% of
293 scaffolding motifs bind SH3 domains (Fig. S1B).
294
295 To further select for more likely functional motif matches, one can consider functional aspects
296 of the different motif types and the cellular context of secreted proteins (Fig. 3D). For example,
297 as microneme proteins will be present in extracellular environments, motifs known to interact
298 with domains from extracellular proteins would be expected to have more likely functional
299 relevance. In our dataset (Table S1), we found 401 motif matches from 12 motif extracellular
300 types in 48 microneme proteins that could be relevant during the recognition and attachment
301 stages. Integrin binding is a common strategy for pathogens to attach to the host cell
302 membrane (45–47). The most common integrin-binding motif is the RGD motif, which can
303 bind to eight of the different alpha and beta integrin heterodimers (48). Within our dataset, we
304 found three motif matches of the RGD motif in three microneme proteins: MIC17C, Toxolysin
305 4 (TLN4) and chitinase-like protein CLP1. As some microneme proteins might be localised to
306 the host cytosol, e.g. during host cell lysis and egression (49), we did not filter out matches
307 from other cytosolic motif types in microneme proteins. Degradation signals, or degrons, are
308 motifs recognised by E3 ubiquitin ligases and mark proteins for proteasomal degradation (50).

309 Degrons might be relevant motif candidates for Toxoplasma infection, as other pathogens are
310 known to modify protein levels via degrons to achieve outcomes such as the stabilisation of
311 host components necessary for their survival (3) or modulating the activity of their proteins
312 once inside the host (51). We found 475 degrons in 182 secreted proteins: 73 in micronemes,
313 157 in rhoptries and 245 in dense granule proteins. Speckle-type POZ protein (SPOP), an E3
314 ubiquitin ligase, recognises its substrates via an ST-rich degron (52). We found 155 SPOP
315 degron matches in 80 secreted proteins. Since multivalency is a known mechanism by which
316 SPOP substrates enhance their affinity to SPOP (52), one could prioritise the 34 secreted
317 proteins containing two or more SPOP motif matches in our dataset. Rhoptry protein toxolysin
318 1 (TLN1), a metalloprotease with a role in rhoptry maturation (53), but potentially detrimental
319 for host establishment once secreted into the host cytosol, has 10 identical SPOP degron
320 matches, which could be used to modulate TLN1 levels once within the host cytoplasm. These
321 examples illustrate how our dataset can be further explored to prioritise motif candidates for
322 experimental characterisation.
323
324 **Experimental validation of predicted TRAF6-binding motifs**
325 For the experimental validation of predicted motifs, we selected motifs that bind the tumour
326 necrosis factor (TNF) receptor-associated factor 6 (TRAF6) protein. TRAF6 binds to proteins
327 containing a P.E..[DE(Φ)] motif through its C-terminal MATH domain (Φ denotes hydrophobic
328 residues). TRAF6 plays important roles in immunity and development by contributing to the
329 formation of various signalling protein complexes. During the innate immune response,
330 TRAF6 is activated by interleukin 1 receptor (IL-1R), linking this receptor to the canonical
331 activation of the transcription factor NF-κB (54). Through the addition of K63-linked
332 polyubiquitin chains, TRAF6 primes proteins for complex assembly and not for proteasomal
333 degradation. The TRAF6-binding motif is present in proteins such as IL-1R, Tumour necrosis
334 factor protein CD40, and Mitochondrial antiviral-signalling protein (MAVS) (55). In the context
335 of infection, herpesviruses have been shown to use the TRAF6 motif-containing protein Tio
336 to stimulate NF-κB activation (56).

Among our predicted motif instances, we found 111 motif matches in 68 rhoptry and dense granule proteins. Given that the TRAF6-binding motif is known to appear multiple times within proteins (57), we hypothesised that the motif repetitions could interact with the TRAF6 trimer in a multivalent way. Following this criterion, we selected rhoptry neck protein RON6, which harbours 17 matches of the TRAF6-binding motif, most of them being identical repeats, and RON10, which harbours two different TRAF6-binding motif matches (Fig. 4A). RON10 is known to interact with RON9, but RON10 and RON9 do not have clear roles during infection (58). TRAF6 has been shown previously to interact with the dense granule proteins GRA7 and GRA15, which themselves play a role in modulating the innate immune response during *Toxoplasma* infection (59–62). However, the actual interaction interfaces of GRA7 and GRA15 with TRAF6 have not been described (21), making the TRAF6 motif matches in GRA7 and GRA15 interesting candidates for further experimental characterisation (Fig. 4A).



351 **Figure 4 - TRAF6 motif binding assays. A.** Protein diagrams of RON6, RON10, GRA7 and GRA15
352 indicating the position of their TRAF6-binding motifs with accompanying multiple sequence alignment
353 showing their presence in different Toxoplasma strains and Coccidian species. **B.** AlphaFold model of
354 TRAF6 MATH domain with TRAF6-binding motif from RON10. Right panels display the potential
355 hydrogen bond network at the interaction interface (top) and secondary structure adopted by the
356 TRAF6-binding motif (bottom). **C.** RON10 TRAF6 motif testing. MST traces (top graph) and dose-
357 response curves (bottom panel) of TRAF6 MATH domain binding to RON10 TRAF6 motif containing
358 peptide (left) and the E>S mutant (right). In the MST traces, the cold region is set at 0 s (blue) and the
359 hot detection region at 5 s (red). **D**. Same as in A, but for the GRA15 TRAF6 motif and the E>S motif
360 mutant.
361
362 We proceeded to determine the dissociation constants of the TRAF6-binding motifs in RON6,
363 RON10, GRA7 and GRA15 to the TRAF6 MATH domain using the microscale thermophoresis
364 (MST) binding assay. To this end, we expressed and purified the human TRAF6 MATH
365 domain and ordered peptides containing the sequence of the following motif matches. In the
366 case of RON6, we selected the most frequent version of its motif repeats. For RON10, we
367 selected the second motif match that contained multiple negative charges, which might
368 contribute to binding (Fig. 4B). GRA7 and GRA15 only have one TRAF6-binding motif match
369 each, which was selected for experimental testing. In addition, we ordered these wildtype
370 peptide sequences with a Glutamate to Serine mutation at position 3 in the motifs supposed
371 to weaken or abrogate binding to the MATH domain because the conserved glutamate residue
372 forms a hydrogen bond with backbone atoms of residue A458 in TRAF6 (Fig. 4B) (Table S2).
373 Our experimental results confirm that the peptides containing the motifs from RON10 and
374 GRA15 can bind to the TRAF6 MATH domain, and these interactions were lost when testing
375 the mutant peptide versions. The dissociation constants for the RON10 (Kd 18.2 ±1.46 μM)
376 and GRA15 peptides (Kd 26.9 ±1.43 μM) are as expected, for functional motif interactions, in
377 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.

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 display longer IDRs (Fig. S1A) (59). This makes it difficult to build good sequence alignments to compute common motif conservation metrics like in (64). Nevertheless, we included the assessment of the presence of our motif matches in different *Toxoplasma* strains and the few closely related parasite species as supplementary conservation metrics (Table S1), which could be useful when working with different strains or in understanding motif variation and evolution.

404 This study represents the first comprehensive survey of short linear motifs in *Toxoplasma*
405 *gondii* or any other Apicomplexan parasite. There have been other efforts to predict and
406 validate short linear motifs in pathogenic proteins (31,65). However, they focus on only a
407 subset of proteins or specific motifs, or other pathogens such as viruses (61). Identification of
408 novel motif instances in parasites such as Toxoplasma also enables refinement of current
409 motif type annotations, as highlighted by some examples described in this study.
410
411 Our systematic approach for predicting instances of known motif types in Toxoplasma
412 secreted proteins indicates that there might be a substantial number of motifs in *Toxoplasma*
413 to be characterised, also in non-secreted proteins, e.g. motifs that participate in endogenous
414 processes for parasite motility, cytoskeletal rearrangements, and parasite development. Thus,
415 this dataset represents a valuable resource for hypothesis generation and design of
416 downstream experiments to study molecular mechanisms of *Toxoplasma gondii* infection and
417 biology. Structural models of predicted domain-motif interactions could further be used to
418 design disruptive mutations to validated predicted modes of binding (2). Identified and
419 experimentally validated pathogen-host motif-domain interactions can form the basis for the
420 development of therapeutic strategies aimed at perturbing these interactions and thereby
421 interfering with infection processes (4,10). Motif discovery in pathogens remains a much
422 unexplored area of research but holds great promise for improved understanding of infection
423 mechanisms and the better design of effective treatments for parasite infection.
424

425 **Acknowledgements**
426 We are thankful to members of the Gibson team for discussions at various stages of this work.
427 We also express our gratitude to the curators and contributors of the ELM and the ToxoDB
428 resources.
429
430

Author contributions

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

Data and Code Availability

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

Materials and Methods

Retrieval of sequence, structure and motif data

Toxoplasma strain proteomes, as well as those of five closely related organisms from the Sarcocystidae clade, were retrieved from the ToxoDB resource (39). The fasta files of the annotated proteomes were collected for the following organisms: *Toxoplasma gondii* GT1 (Type I), *T. gondii* MAS and *T. gondii* ME49 (Type II and reference strain), *T. gondii* RUB and *T. gondii* VEG (Type III), *T. gondii* VAND, *Hammondia hammondii, Neospora caninum, Besnoitia besnoitii, Cystoisospora suis* Wien I and *Sarcosystis neurona* SN3. Using the ToxoDB search strategies, we retrieved mass spectrometry proteomic peptide mapping data, phosphosites, and cellular location predictions from the HyperLOPIT experiment (38) for the proteins of *T. gondii* ME49. All available AlphaFold database (AFDB) (66) predicted structures for *Toxoplasma gondii* ME49 were retrieved from the Google Cloud Console public-datasets-deepmind-alphafold-v4, taxonomy id: 50877 (02.05.2025). The latest dataset of motif classes was downloaded from the ELM resource (30.04.2025) (22). The file contained each motif code name, its regular expression (REGEX) and supplementary information. We did manual annotation of the motif classes to determine if their taxonomic range included vertebrates and *Toxoplasma*, i.e. if the domain was annotated in ELM to be present in the respective organism for interaction with the motif, ultimately excluding motif classes of organisms such as plants,

458 insects and fungi. We also added a manual annotation on whether the motif is functional in
459 an extracellular context and if it requires a phosphorylation or modification to interact with its
460 domain partner. We retrieved available T. gondii domain mappings from UniProt (29.07.2025)
461 (42), which came from Pfam, PROSITE and SMART.
462
463 Motif prediction and structural assessments
464 We determined the presence of motif matches in the *T. gondii* ME49 by doing a baseline
465 REGEX search. To assess the accessibility of these motif predictions, two approaches were
466 selected: one that takes into account the intrinsic disorder of the sequence and another that
467 evaluates the accessibility of the motif from predicted AlphaFold DB (AFDB) structures. The
468 IUPred3 software (40) was used to calculate the disorder probability of the full proteome of *T.*
469 *gondii* ME49. Once a REGEX match was determined, the IUPred3 score from all motif
470 residues was retrieved and averaged. Complementarily, using available AFDB structures, we
471 took the pLDDT scores from all motif residues and averaged them (67). We calculated residue
472 accessibility using DSSP (Dictionary of Secondary Structure of Proteins) software (68) on
473 each AFDB structure. The individual residue scores were then standardised according to (69)
474 to obtain relative surface accessibility scores (RSA). We then took the RSA scores of all motif
475 residues and averaged them. We determined if a motif match overlapped with known domain
476 boundaries from UniProt by using positional information.
477
478 Motif organism presence assessment
479 A reciprocal best hit search was carried out using BLAST to construct sets of orthologous
480 sequence groups among *Toxoplasma* strains and related species. These sequence groups
481 contained the sequences with the lowest e-value from each organism's proteome, so
482 alignments contained a standard number of sequences and were easily comparable. Multiple
483 sequence alignments were produced for each sequence group using ClustalO with default
484 parameters (70). Standard conservation measurements (64) were not possible as our
485 alignments contained only a few species. Thus, the conservation of the motifs was

486 approximated by carrying out a presence assessment. A motif was considered present in
487 another organism if a match of the same class was present in its sequence within a window
488 of 15 amino acids in the alignment around the start site of the motif match of the reference
489 *Toxoplasma* strain ME49. And then the presence proportion of the total number of both strains
490 and species was calculated and associated with each motif match, resulting in three different
491 scores: the proportion of strains, species and overall organisms.
492
493 Data enrichment and filtering
494 The data from the ELM REGEX motifs matches, their conservation, IUPred3, pLDDT, and
495 accessibility scores were combined into a single table together with their respective protein
496 information on subcellular location, annotations, domain overlap, and modifications. This was
497 carried out using their unique ToxoDB identifiers and the motif name, match number and
498 residue positions. Data filtering to define a set of motif matches was carried out with three
499 different strategies. Firstly, we retained motifs from proteins with at least one peptide mapped
500 to them from retrieved mass spectrometry data, to make sure proteins were expressed. We
501 then made use of the ELM taxonomy range by ensuring that the taxonomic range of the motif
502 instance, if the motif has been reported in a certain group of organisms, was valid for the
503 *Toxoplasma* matches. Only motif classes that apply to Vertebrates and Apicomplexans were
504 kept, mainly excluding those characterised in plants, insects or fungi. The classes valid in
505 Vertebrates were kept if the motifs were located in secreted proteins, as they would be in
506 contact with the vertebrate host (human, a mammal or birds) cell components. The classes
507 only valid in extracellular contexts were kept if the motifs were located in microneme proteins,
508 as among secreted proteins, they would be the ones in this location. We applied structural
509 filters in two steps; only motif matches with IUPred3 scores above 0.4 were considered,
510 following disordered region filters in (71). Secondly, for motifs in proteins with available AFDB
511 structures, we retained the ones with a pLDDT<=80 (67) and an average RSA>= 0.25 (72).
512 For motifs without AFDB structures, we retained only the ones outside of known domain
513 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 (73). 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 (74).

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_id30_c90
_final_seq.sorted_opt \
--mgnify_database_path=/mnt/storage/alphafold/v232/mgnify/mgy_clusters_2022_05.fa \
--obsolete_pdbs_path=/mnt/storage/alphafold/v232/pdb_mmcif/obsolete.dat \
--pdb_seqres_database_path=/mnt/storage/alphafold/v232/pdb_seqres/pdb_seqres.txt \
--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
```

542 Protein Expression and Purification

543 A gene encoding the human TRAF6 C-domain (residues 346-504, UniProt ID Q9Y4K3) was
544 synthesised (GeneArt, Thermo Fisher Scientific) and subcloned into the pCoofy4 plasmid (75)
545 containing an N-terminal His6-MBP-3C fusion tag. For the human TRAF6 C-domain (InterPro:
546 IPR008974) gene, the native DNA sequence was used. The resulting bacterial expression
547 construct was freshly transformed into *E. coli* BL21(DE3) codon + RIL cells (Stratagene).
548 Precultures were grown overnight at 37°C in LB medium supplemented with 30 µg/ml
549 Kanamycin and 35 µg/ml Chloramphenicol and used to inoculate the large-scale expression
550 cultures. 10 ml preculture was added to 1 litre of TB-FB medium supplemented with 30 µg/ml
551 Kanamycin and 35 µg/ml Chloramphenicol. The cultures were grown at 37°C until the OD600
552 reached ~0.8, after which the temperature was reduced to 18°C and expression of the
553 hTRAF6 C-domain was induced with 0.5 mM IPTG. After overnight expression at 18°C, the
554 cultures were harvested by centrifugation (30 min, 5000 g, 4°C), and the pellets were flash-
555 frozen in liquid nitrogen and stored at -80°C until the start of the protein purification.

556

557 The cells were lysed by 5 passages through a microfluidiser device, followed by centrifugation
558 (30 min, 140.000 g, 4°C). The cleared lysate was loaded onto a 5 ml Protino Ni-NTA column
559 (Macherey-Nagel) pre-equilibrated with 50 mM Tris-HCl, pH 8.0, 500 mM NaCl and 20 mM
560 imidazole. After loading, the Ni-NTA column was washed with equilibration buffer until the
561 A280 nm signal returned to baseline and eluted in a 60 ml gradient going from 0 to 100% of
562 Ni-NTA elution buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 350 mM imidazole). The Ni-
563 NTA elution fractions were analysed by SDS-PAGE, and the fractions containing the His6-
564 MBP-3C-hTRAF6 C-domain fusion protein were pooled. 2 mg of His6-tagged HRV 3C
565 protease was added, and the sample was dialysed overnight at 4°C to IEX buffer (25 mM
566 HEPES, pH 7.5, 100 mM NaCl, 20 mM imidazole). The next day, the dialysed sample was
567 loaded onto a 5 ml Protino Ni-NTA column (Macherey-Nagel) coupled to a 5 ml HiTrap Q HP
568 anion exchange column (Cytiva), both equilibrated with IEX buffer. The untagged hTRAF6 C-
569 domain protein was present in the flow-through of the coupled columns, which was

570 concentrated to 5 ml and injected into a HiLoad 16/600 Superdex 75 pg size exclusion
571 chromatography column (Cytiva) pre-equilibrated with 25 mM HEPES pH 7.5, 150 mM NaCl
572 and 1 mM DTT. The elution fractions containing pure hTRAF6 C-domain protein were pooled
573 and concentrated to 2.1 mg/ml (111 μM). The final hTRAF6 C-domain protein samples were
574 aliquoted, flash-frozen in liquid nitrogen and stored at -80°C until further usage. The identity
575 of the hTRAF6 C-domain protein was verified by mass spectrometry (EMBL Proteomics Core
576 Facility), and the oligomerisation state (monomeric) and absence of aggregates were
577 confirmed by size exclusion chromatography column coupled to multi-angle light scattering
578 (SEC-MALS). The stability of the hTRAF6 C-domain protein in the storage buffer was
579 assessed by nano-Differential Scanning Fluorimetry (nano-DSF; Tm ~ 46.8 °C).
580
581 Binding assays
582 We selected 4 motifs from 4 different *Toxoplasma* proteins with different AF model scores for
583 carrying out binding assays as described in Results, together with the human MAVS TRAF6
584 motif, being selected as a control (55). Peptides of 15 amino acids containing the motifs were
585 defined, as well as versions containing E>S point mutations at motif position +3. The peptides
586 were labelled with 5-Carboxyfluorescein (5-Fluo) at the C-terminus by Biosyntan GmbH
587 (https://www.biosyntan.de) to carry out Microscale thermophoresis (MST). The lyophilised
588 peptides were resuspended in 25mM HEPES, pH 7.5 and 150mM NaCl to a concentration of
589 2mM, and the pH was adjusted to 7.5 when necessary. MST measurements were performed
590 using a Monolith NT.115 (NanoTemper Technology). Two-fold serial dilutions of the hTRAF6
591 protein (99 μM) were mixed in a 1:1 ratio with 100nM peptides. Titration measurements were
592 taken for all *Toxoplasma* peptides, and the human MAVS control, with two-fold serial dilutions
593 of the hTRAF6 protein (99 μM) mixed in a 9:1 ratio with 500nM. All measurements were
594 performed in triplicate and carried out at 25°C with an LED excitation power of 20% and a
595 medium MST power. Finally, the data were analysed assuming a 1:1 binding model and
596 graphed using the MO.Affinity Software (NanoTemper Technology).
597

598 **References**

599 1. Kellermann M, Scharte F, Hensel M. Manipulation of Host Cell Organelles by Intracellular
600 Pathogens. Int J Mol Sci. 2021 June 17;22(12):6484.
601 2. Baptista D, Gomez-Lucas L, Jänes J, Krogan NJ, Amorim MJ, Ivarsson Y, et al. AlphaFold
602 models of host-pathogen interactions elucidate the prevalence and structural modes of
603 molecular mimicry [Internet]. Bioinformatics; 2025 [cited 2025 Aug 8]. Available from:
604 http://biorxiv.org/lookup/doi/10.1101/2025.06.04.657796
605 3. Davey NE, Travé G, Gibson TJ. How viruses hijack cell regulation. Trends Biochem Sci.
606 2011 Mar;36(3):159–69.
607 4. Sámano-Sánchez H, Gibson TJ. Mimicry of Short Linear Motifs by Bacterial Pathogens:
608 A Drugging Opportunity. Trends Biochem Sci. 2020 June;45(6):526–44.
609 5. Davey NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, et al. Attributes of
610 short linear motifs. Mol Biosyst. 2012 Jan;8(1):268–81.
611 6. Van Roey K, Uyar B, Weatheritt RJ, Dinkel H, Seiler M, Budd A, et al. Short linear motifs:
612 ubiquitous and functionally diverse protein interaction modules directing cell regulation.
613 Chem Rev. 2014 July 9;114(13):6733–78.
614 7. Davey NE, Cyert MS, Moses AM. Short linear motifs – ex nihilo evolution of protein
615 regulation. Cell Commun Signal. 2015 Dec;13(1):43.
616 8. Mészáros B, Sámano-Sánchez H, Alvarado-Valverde J, Čalyševa J, Martínez-Pérez E,
617 Alves R, et al. Short linear motif candidates in the cell entry system used by SARS-CoV-
618 2 and their potential therapeutic implications. Sci Signal. 2021 Jan 12;14(665):eabd0334.
619 9. Mihalič F, Simonetti L, Giudice G, Sander MR, Lindqvist R, Peters MBA, et al. Large-scale
620 phage-based screening reveals extensive pan-viral mimicry of host short linear motifs.
621 Nat Commun. 2023 Apr 26;14(1):2409.
622 10. Simonetti L, Nilsson J, McInerney G, Ivarsson Y, Davey NE. SLiM-binding pockets: an
623 attractive target for broad-spectrum antivirals. Trends Biochem Sci. 2023 May;48(5):420–
624 7.

625 11. Boddey JA, Moritz RL, Simpson RJ, Cowman AF. Role of the *Plasmodium* Export Element
626 in Trafficking Parasite Proteins to the Infected Erythrocyte. Traffic. 2009 Mar;10(3):285–
627 99.
628 12. Huber S, Karagenc T, Ritler D, Rottenberg S, Woods K. Identification and characterisation
629 of a *Theileria annulata* proline-rich microtubule and SH3 domain-interacting protein
630 (TaMISHIP) that forms a complex with CLASP1, EB1, and CD2AP at the schizont surface.
631 Cell Microbiol. 2018 July;20(7):e12838.
632 13. Mohan A, Sullivan Jr WJ, Radivojac P, Dunker AK, Uversky VN. Intrinsic disorder in
633 pathogenic and non-pathogenic microbes: discovering and analyzing the unfoldomes of
634 early-branching eukaryotes. Mol Biosyst. 2008;4(4):328.
635 14. Pancsa R, Tompa P. Structural Disorder in Eukaryotes. Buday L, editor. PLoS ONE. 2012
636 Apr 5;7(4):e34687.
637 15. De Barros RAM, Torrecilhas AC, Marciano MAM, Mazuz ML, Pereira-Chioccola VL, Fux
638 B. Toxoplasmosis in Human and Animals Around the World. Diagnosis and Perspectives
639 in the One Health Approach. Acta Trop. 2022 July;231:106432.
640 16. Milne G, Webster JP, Walker M. Toxoplasma gondii: An Underestimated Threat? Trends
641 Parasitol. 2020 Dec;36(12):959–69.
642 17. Dubey JP. Toxoplasmosis of Animals and Humans [Internet]. 3rd ed. Boca Raton: CRC
643 Press; 2021 [cited 2025 Aug 8]. Available from:
644 https://www.taylorfrancis.com/books/9781003199373
645 18. Boothroyd JC, Dubremetz JF. Kiss and spit: the dual roles of Toxoplasma rhoptries. Nat
646 Rev Microbiol. 2008 Jan;6(1):79–88.
647 19. Rastogi S, Cygan AM, Boothroyd JC. Translocation of effector proteins into host cells by
648 Toxoplasma gondii. Curr Opin Microbiol. 2019 Dec;52:130–8.
649 20. Kemp LE, Yamamoto M, Soldati-Favre D. Subversion of host cellular functions by the
650 apicomplexan parasites. FEMS Microbiol Rev. 2013 July;37(4):607–31.
651 21. Hakimi MA, Olias P, Sibley LD. *Toxoplasma* Effectors Targeting Host Signaling and
652 Transcription. Clin Microbiol Rev. 2017 July;30(3):615–45.

653 22. Kumar M, Michael S, Alvarado-Valverde J, Zeke A, Lazar T, Glavina J, et al. ELM-the
654 Eukaryotic Linear Motif resource-2024 update. Nucleic Acids Res. 2024 Jan
655 5;52(D1):D442–55.
656 23. Guérin A, Corrales RM, Parker ML, Lamarque MH, Jacot D, El Hajj H, et al. Efficient
657 invasion by Toxoplasma depends on the subversion of host protein networks. Nat
658 Microbiol. 2017 Aug 28;2(10):1358–66.
659 24. Havrylov S, Jolanta Redowicz M, Buchman VL. Emerging Roles of Ruk/CIN85 in Vesicle-
660 Mediated Transport, Adhesion, Migration and Malignancy. Traffic. 2010 Mar
661 17;11(6):721–31.
662 25. Jozic D, Cárdenes N, Deribe YL, Moncalián G, Hoeller D, Groemping Y, et al. Cbl
663 promotes clustering of endocytic adaptor proteins. Nat Struct Mol Biol. 2005
664 Nov;12(11):972–9.
665 26. Henne WM, Buchkovich NJ, Emr SD. The ESCRT Pathway. Dev Cell. 2011
666 July;21(1):77–91.
667 27. Lee SH, Hayes DB, Rebowski G, Tardieux I, Dominguez R. Toxofilin from *Toxoplasma*
668 *gondii* forms a ternary complex with an antiparallel actin dimer. Proc Natl Acad Sci. 2007
669 Oct 9;104(41):16122–7.
670 28. Delorme-Walker V, Abrivard M, Lagal V, Anderson K, Perazzi A, Gonzalez V, et al.
671 Toxofilin upregulates the host cortical actin cytoskeleton dynamics facilitating *Toxoplasma*
672 invasion. J Cell Sci. 2012 Jan 1;jcs.103648.
673 29. Mehra C, Alvarado Valverde J, Matias AMN, Torelli F, Medeiros TC, Straub J, et al.
674 Toxoplasma effector TgROP1 establishes membrane contact sites with the endoplasmic
675 reticulum during infection. Nat Microbiol. 2025 Nov 25;10(12):3331–45.
676 30. Romano JD, Buh R, Grudda T, Box JR, Beltran J, Khan SM, et al. Toxoplasma gondii
677 VIP1 mediates parasitophorous vacuole–host endoplasmic reticulum interactions to
678 facilitate parasite development. Nat Microbiol. 2025 Oct 10;10(12):3315–30.
679 31. Rivera-Cuevas Y, Mayoral J, Di Cristina M, Lawrence ALE, Olafsson EB, Patel RK, et al.
680 Toxoplasma gondii exploits the host ESCRT machinery for parasite uptake of host

681 cytosolic proteins. Olias P, editor. PLOS Pathog. 2021 Dec 13;17(12):e1010138.
682 32. Braun L, Brenier-Pinchart MP, Yogavel M, Curt-Varesano A, Curt-Bertini RL, Hussain T,
683 et al. A *Toxoplasma* dense granule protein, GRA24, modulates the early immune
684 response to infection by promoting a direct and sustained host p38 MAPK activation. J
685 Exp Med. 2013 Sept 23;210(10):2071–86.
686 33. Pellegrini E, Palencia A, Braun L, Kapp U, Bougdour A, Belrhali H, et al. Structural Basis
687 for the Subversion of MAP Kinase Signaling by an Intrinsically Disordered Parasite
688 Secreted Agonist. Structure. 2017 Jan;25(1):16–26.
689 34. Bougdour A, Durandau E, Brenier-Pinchart MP, Ortet P, Barakat M, Kieffer S, et al. Host
690 Cell Subversion by Toxoplasma GRA16, an Exported Dense Granule Protein that Targets
691 the Host Cell Nucleus and Alters Gene Expression. Cell Host Microbe. 2013
692 Apr;13(4):489–500.
693 35. Coffey MJ, Sleebs BE, Uboldi AD, Garnham A, Franco M, Marino ND, et al. An aspartyl
694 protease defines a novel pathway for export of Toxoplasma proteins into the host cell.
695 eLife. 2015 Nov 18;4:e10809.
696 36. Saeij JPJ, Coller S, Boyle JP, Jerome ME, White MW, Boothroyd JC. Toxoplasma co-opts
697 host gene expression by injection of a polymorphic kinase homologue. Nature. 2007
698 Jan;445(7125):324–7.
699 37. Chang S, Shan X, Li X, Fan W, Zhang SQ, Zhang J, et al. *Toxoplasma gondii* Rhoptry
700 Protein ROP16 Mediates Partially SH-SY5Y Cells Apoptosis and Cell Cycle Arrest by
701 Directing Ser15/37 Phosphorylation of p53. Int J Biol Sci. 2015;11(10):1215–25.
702 38. Barylyuk K, Koreny L, Ke H, Butterworth S, Crook OM, Lassadi I, et al. A Comprehensive
703 Subcellular Atlas of the Toxoplasma Proteome via hyperLOPIT Provides Spatial Context
704 for Protein Functions. Cell Host Microbe. 2020 Nov;28(5):752-766.e9.
705 39. Alvarez-Jarreta J, Amos B, Aurrecoechea C, Bah S, Barba M, Barreto A, et al.
706 VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center in
707 2023. Nucleic Acids Res. 2024 Jan 5;52(D1):D808–16.
708 40. Erdős G, Pajkos M, Dosztányi Z. IUPred3: prediction of protein disorder enhanced with

709 unambiguous experimental annotation and visualization of evolutionary conservation.
710 Nucleic Acids Res. 2021 July 2;49(W1):W297–303.
711 41. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate
712 protein structure prediction with AlphaFold. Nature. 2021 Aug;596(7873):583–9.
713 42. The UniProt Consortium, Bateman A, Martin MJ, Orchard S, Magrane M, Ahmad S, et al.
714 UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 2023 Jan
715 6;51(D1):D523–31.
716 43. Honnappa S, Gouveia SM, Weisbrich A, Damberger FF, Bhavesh NS, Jawhari H, et al.
717 An EB1-Binding Motif Acts as a Microtubule Tip Localization Signal. Cell. 2009
718 July;138(2):366–76.
719 44. Sámano-Sánchez H, Gibson TJ, Chemes LB. Using Linear Motif Database Resources to
720 Identify SH2 Domain Binders. In: Carlomagno T, Köhn M, editors. SH2 Domains [Internet].
721 New York, NY: Springer US; 2023 [cited 2025 Oct 29]. p. 153–97. (Methods in Molecular
722 Biology; vol. 2705). Available from: https://link.springer.com/10.1007/978-1-0716-3393-
723 9_9
724 45. Berinstein A, Roivainen M, Hovi T, Mason PW, Baxt B. Antibodies to the vitronectin
725 receptor (integrin alpha V beta 3) inhibit binding and infection of foot-and-mouth disease
726 virus to cultured cells. J Virol. 1995 Apr;69(4):2664–6.
727 46. Kwok T, Zabler D, Urman S, Rohde M, Hartig R, Wessler S, et al. Helicobacter exploits
728 integrin for type IV secretion and kinase activation. Nature. 2007 Oct;449(7164):862–6.
729 47. Zimmermann L, Peterhans E, Frey J. RGD Motif of Lipoprotein T, Involved in Adhesion of
730 *Mycoplasma conjunctivae* to Lamb Synovial Tissue Cells. J Bacteriol. 2010 July
731 15;192(14):3773–9.
732 48. D'Souza SE, Ginsberg MH, Plow EF. Arginyl-glycyl-aspartic acid (RGD): a cell adhesion
733 motif. Trends Biochem Sci. 1991 Jan;16:246–50.
734 49. Liu Q, Li FC, Zhou CX, Zhu XQ. Research advances in interactions related to Toxoplasma
735 gondii microneme proteins. Exp Parasitol. 2017 May;176:89–98.
736 50. Mészáros B, Kumar M, Gibson TJ, Uyar B, Dosztányi Z. Degrons in cancer. Sci Signal.

737 2017 Mar 14;10(470):eaak9982.
738 51. Schnupf P, Portnoy DA, Decatur AL. Phosphorylation, ubiquitination and degradation of
739 listeriolysin O in mammalian cells: role of the PEST-like sequence. Cell Microbiol. 2006
740 Feb;8(2):353–64.
741 52. Pierce WK, Grace CR, Lee J, Nourse A, Marzahn MR, Watson ER, et al. Multiple Weak
742 Linear Motifs Enhance Recruitment and Processivity in SPOP-Mediated Substrate
743 Ubiquitination. J Mol Biol. 2016 Mar;428(6):1256–71.
744 53. Hajagos BE, Turetzky JM, Peng ED, Cheng SJ, Ryan CM, Souda P, et al. Molecular
745 dissection of novel trafficking and processing of the Toxoplasma gondii rhoptry
746 metalloprotease toxolysin-1. Traffic Cph Den. 2012 Feb;13(2):292–304.
747 54. Takeda K, Akira S. TLR signaling pathways. Semin Immunol. 2004 Feb;16(1):3–9.
748 55. Shi Z, Zhang Z, Zhang Z, Wang Y, Li C, Wang X, et al. Structural Insights into
749 Mitochondrial Antiviral Signaling Protein (MAVS)-Tumor Necrosis Factor Receptor-
750 associated Factor 6 (TRAF6) Signaling. J Biol Chem. 2015 Oct;290(44):26811–20.
751 56. Heinemann S, Biesinger B, Fleckenstein B, Albrecht JC. NFκB Signaling Is Induced by
752 the Oncoprotein Tio through Direct Interaction with TRAF6. J Biol Chem. 2006
753 Mar;281(13):8565–72.
754 57. Ye H, Arron JR, Lamothe B, Cirilli M, Kobayashi T, Shevde NK, et al. Distinct molecular
755 mechanism for initiating TRAF6 signalling. Nature. 2002 July;418(6896):443–7.
756 58. Lamarque MH, Papoin J, Finizio AL, Lentini G, Pfaff AW, Candolfi E, et al. Identification
757 of a New Rhoptry Neck Complex RON9/RON10 in the Apicomplexa Parasite Toxoplasma
758 gondii. Moreno SN, editor. PLoS ONE. 2012 Mar 12;7(3):e32457.
759 59. Alaganan A, Fentress SJ, Tang K, Wang Q, Sibley LD. *Toxoplasma* GRA7 effector
760 increases turnover of immunity-related GTPases and contributes to acute virulence in the
761 mouse. Proc Natl Acad Sci. 2014 Jan 21;111(3):1126–31.
762 60. Hunter CA, Sibley LD. Modulation of innate immunity by Toxoplasma gondii virulence
763 effectors. Nat Rev Microbiol. 2012 Nov;10(11):766–78.
764 61. Ihara F, Fereig RM, Himori Y, Kameyama K, Umeda K, Tanaka S, et al. Toxoplasma

765 gondii Dense Granule Proteins 7, 14, and 15 Are Involved in Modification and Control of
766 the Immune Response Mediated via NF-κB Pathway. Front Immunol. 2020 July
767 31;11:1709.
768 62. Yang CS, Yuk JM, Lee YH, Jo EK. Toxoplasma gondii GRA7-Induced TRAF6 Activation
769 Contributes to Host Protective Immunity. Adams JH, editor. Infect Immun. 2016
770 Jan;84(1):339–50.
771 63. Wilson BA, Foy SG, Neme R, Masel J. Young genes are highly disordered as predicted
772 by the preadaptation hypothesis of de novo gene birth. Nat Ecol Evol. 2017 Apr
773 24;1(6):0146.
774 64. Davey NE, Cowan JL, Shields DC, Gibson TJ, Coldwell MJ, Edwards RJ. SLiMPrints:
775 conservation-based discovery of functional motif fingerprints in intrinsically disordered
776 protein regions. Nucleic Acids Res. 2012 Nov;40(21):10628–41.
777 65. Idrees S, Paudel KR, Banik M, Suwal N, Thapa R, Bashyal S. Predicting Motif-Mediated
778 Interactions Based on Viral Genomic Composition. Int J Mol Sci. 2025 Apr 13;26(8):3674.
779 66. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold
780 Protein Structure Database: massively expanding the structural coverage of protein-
781 sequence space with high-accuracy models. Nucleic Acids Res. 2022 Jan
782 7;50(D1):D439–44.
783 67. Wilson CJ, Choy WY, Karttunen M. AlphaFold2: A Role for Disordered Protein/Region
784 Prediction? Int J Mol Sci. 2022 Apr 21;23(9):4591.
785 68. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of
786 hydrogen-bonded and geometrical features. Biopolymers. 1983 Dec;22(12):2577–637.
787 69. Tien MZ, Meyer AG, Sydykova DK, Spielman SJ, Wilke CO. Maximum Allowed Solvent
788 Accessibilites of Residues in Proteins. Porollo A, editor. PLoS ONE. 2013 Nov
789 21;8(11):e80635.
790 70. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation
791 of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol.
792 2011 Jan;7(1):539.

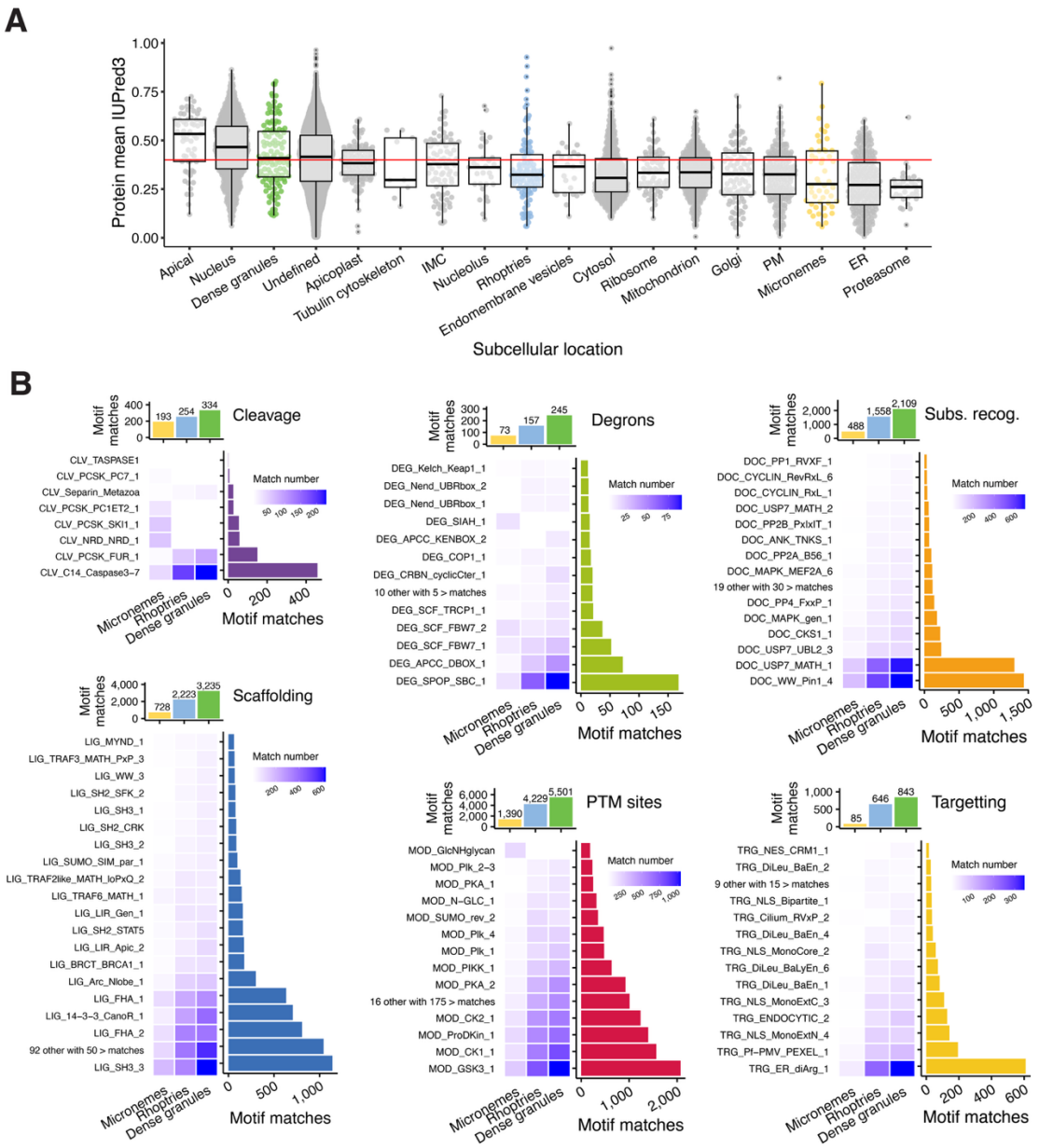
71. Davey NE, Seo M, Yadav VK, Jeon J, Nim S, Krystkowiak I, et al. Discovery of short linear motif-mediated interactions through phage display of intrinsically disordered regions of the human proteome. FEBS J. 2017 Feb;284(3):485–98.
72. Levy ED. A simple definition of structural regions in proteins and its use in analyzing interface evolution. J Mol Biol. 2010 Nov 5;403(4):660–70.
73. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer [Internet]. Bioinformatics; 2021 [cited 2025 June 5]. Available from: http://biorxiv.org/lookup/doi/10.1101/2021.10.04.463034
74. Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, et al. UCSF CHIMERAX : Tools for structure building and analysis. Protein Sci. 2023 Nov;32(11):e4792.
75. Scholz J, Besir H, Strasser C, Suppmann S. A new method to customize protein expression vectors for fast, efficient and background free parallel cloning. BMC Biotechnol. 2013 Dec;13(1):12.

Supportive 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 ELM accession.