



Phage display identification of high-affinity ligands for human TSG101-UEV: A structural and thermodynamic study of PTAP recognition

Javier Murciano-Calles^a, Alejandro Rodríguez-Martínez^{a,c}, Andrés Palencia^{a,1}, Montserrat Andújar-Sánchez^b, Manuel Iglesias-Bexiga^{a,2}, Carles Corbi-Verge^{a,3}, Pedro Buzón^{a,4}, Javier Ruiz-Sanz^a, Jose C. Martínez^a, Horacio Pérez-Sánchez^c, Ana Cámara-Artigas^b, Irene Luque^{a,*}

^a Department of Physical Chemistry, Institute of Biotechnology and Unit of Excellence in Chemistry applied to Biomedicine and Environment, University of Granada, 18071 Granada, Spain

^b Department of Chemistry and Physics, Agrifood Campus of International Excellence (ceiA3) and CIAMBITAL, University of Almería, Carretera de Sacramento s/n 04120 Almería, Spain

^c Structural Bioinformatics and High-Performance Computing (BIO-HPC) Research Group, Universidad Católica de Murcia (UCAM), Guadalupe, Spain

ARTICLE INFO

Keywords:

Virus budding
Broad-spectrum antivirals
TSG101 UEV domain
Viral Late domains
Polyproline recognition domains

ABSTRACT

The ubiquitin E2 variant domain of TSG101 (TSG101-UEV) plays a pivotal role in protein sorting and virus budding by recognizing PTAP motifs within ubiquitinated proteins. Disrupting TSG101-UEV/PTAP interactions has emerged as a promising strategy for the development of novel host-oriented antivirals with a broad spectrum of action. Nonetheless, finding inhibitors with good properties as therapeutic agents remains a challenge since the key determinants of binding affinity and specificity are still poorly understood. Here we present a detailed thermodynamic, structural, and dynamic characterization viral PTAP Late domain recognition by TSG101-UEV, combining isothermal titration calorimetry, X-ray diffraction structural studies, molecular dynamics simulations, and computational analysis of intramolecular communication pathways. Our analysis highlights key contributions from conserved hydrophobic contacts and water-mediated hydrogen bonds at the PTAP binding interface. We have identified additional electrostatic hotspots adjacent to the core motif that modulate affinity. Using competitive phage display screening we have improved affinity by 1–2 orders of magnitude, producing novel peptides with low micromolar affinities that combine critical elements found in the best natural binders. Molecular dynamics simulations revealed that optimized peptides engage new pockets on the UEV domain surface. This study provides a comprehensive view of the molecular forces directing TSG101-UEV recognition of PTAP motifs, revealing that binding is governed by conserved structural elements yet tuneable through targeted optimization. These insights open new venues to design inhibitors targeting TSG101-dependent pathways with potential application as novel broad-spectrum antivirals.

Abbreviations: TSG101, Human Tumour-susceptibility Gene 101; ESCRT, Endosomal Sorting Complex Required for Transport; HRS, Hepatocyte Growth Factor-Regulated Tyrosine Kinase Substrate; ER, endoplasmic reticulum; UEV, Ubiquitin-conjugating enzyme E2 Variant; PRM, polyproline recognition modules; ITC, isothermal titration calorimetry; MD, molecular dynamics; RMSD, root-mean-square deviation.

* Corresponding author.

E-mail address: iluque@ugr.es (I. Luque).

¹ Current address: Institute for Advanced Biosciences, Structural Biology of Novel Targets in Human Diseases, Inserm U1209, CNRS UMR 5309, Université Grenoble-Alpes, 38000 Grenoble, France.

² Current address: Medical Department Menarini S. A., Carrer Alfons XII, 587, 08918 Badalona, Spain.

³ Current address: Recursion 336 Queen Street West Toronto - Ontario, Canada M5V2A2.

⁴ Current address: Department of Biochemistry, University of Zurich, 8057 Zurich, Switzerland.

<https://doi.org/10.1016/j.ijbiomac.2024.133233>

Received 26 March 2024; Received in revised form 6 June 2024; Accepted 15 June 2024

Available online 18 June 2024

0141-8130/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

1. Introduction

Human Tumour-susceptibility Gene 101 (TSG101) is a member of the class E family of vacuole protein-sorting proteins and an essential component of the Endosomal Sorting Complex Required for Transport (ESCRT) pathway, a conserved complex of about 30 proteins organized in four main complexes (ESCRT-0, -I, -II, and -III), which are sequentially recruited to drive membrane remodelling and fission in multiple cellular processes, including cytokinesis, autophagy, multivesicular body generation, extracellular vesicle biogenesis, as well as plasma, nuclear and *endo*-lysosomal membrane repair [1–3]. TSG101 controls ESCRT-I formation and orchestrates early events of membrane trafficking and scission through the interaction with a PTAP motif in HRS (Hepatocyte Growth Factor-Regulated Tyrosine Kinase Substrate) within the ESCRT-0 complex [4] and the recognition of ubiquitinated cargo, playing a key scaffolding role for the subsequent engagement of ESCRT-II and ESCRT-III.

Due to its functional versatility, the ESCRT machinery is sequestered by many intracellular pathogens for replication, assembly or egress [3,5,6] ESCRT pathogenic recruitment is typically mediated by small peptide sequences called viral Late (L-) domains found in the gag polypeptides of retrovirus and in the structural proteins of many families of enveloped RNA viruses in different numbers and combinations [7–9]. Three types of viral L-domains have been identified containing PT/SAP, YPX_nL/LxxLF, or PPxY conserved motifs that act as cellular adaptors engaging ESCRT functions through specific interactions with TSG101, ALIX, and NEDD4, respectively [10–12]. These interactions are essential to complete the last steps in the replication cycle of viruses that bud at the plasma membrane (filoviruses, arenaviruses, rhabdoviruses alphaviruses, and paramyxoviruses) as well as viruses budding at intracellular membranes, such as the endoplasmic reticulum (ER), the Golgi apparatus, or the ER-Golgi compartment (coronaviruses, flaviviruses, and bunyaviruses) [6,12]. Molecules disrupting ESCRT/L-domain interactions have been shown to block the egress of different viruses [13–17] and could be of interest as novel host-oriented antivirals with a wide spectrum of action and low susceptibility to the development of resistance.

PTAP Late domains recruit TSG101 through its UEV (Ubiquitin-conjugating enzyme E2 Variant) domain, mimicking TSG101/HRS interactions. TSG101-UEV shares the E2 ligase typical fold, although it contains an additional N-terminal α -helix, presents an extended β -hairpin (tongue) projecting from the main body of the domain [18], and lacks the two C-terminal helices, exposing a hydrophobic groove containing the PTAP binding site [18–20]. TSG101 binds ubiquitin with low affinity ($K_d = 500\text{--}800\ \mu\text{M}$) but is catalytically inactive due to the absence of the active-site cysteine, which is replaced by a tyrosine residue (Y₁₁₀) in the TSG101 sequence [21–23]. The ubiquitin binding site lies at a concave region at the lower half of the β -sheet, flanked by the $\beta_4\text{-}\alpha_3$ loop at the vestigial active site (lip), and the β -hairpin tongue [18,23]. The different location of the two interaction sites allows the simultaneous or even cooperative binding of ubiquitin and PTAP-containing proteins [23,24]. Additional sites for di-ubiquitin recognition have been recently described [24].

UEV domains belong to the family of polyproline recognition modules (PRM), which includes other families of protein-protein interaction domains (SH3, WW, GYF, and EVH1 domains), all characterized by the presence of one or more proline recognition pockets lined by highly conserved aromatic residues (xP pockets) [25]. Polyproline recognition by SH3 and WW domains has been widely studied and it is well established that these domains present structural and energetic features that complicate rational design and ligand optimization [26] including a) a shallow, relatively featureless and plastic binding sites; b) weak and promiscuous interactions with their natural ligands; c) binding energetics dominated by strong enthalpy/entropy compensation effects that hinder affinity optimization; d) complex thermodynamic profiles, inconsistent with the hydrophobic character of the binding interface,

that reveal significant contributions from water-mediated interactions and conformational effects [15,25,27–29]. The extent to which these features are shared by other PRM families, including UEV domains, remains to be established.

The search for therapeutic agents targeting TSG101-UEV/PTAP interactions requires developing high-affinity and specificity inhibitors with good pharmacological properties and, thus, a deep understanding of the molecular determinants of binding affinity and specificity. With this aim, we present here a comprehensive thermodynamic, structural, and dynamic study of the recognition of natural PTAP viral Late domains by TSG101-UEV combining isothermal titration calorimetry (ITC), X-ray crystallography, molecular dynamics (MD) simulations, and computational analysis of intramolecular communication pathways. Our results reveal that, as SH3 and WW domains, UEV interactions are governed by a complex interplay of different factors, raising questions about the feasibility of obtaining small drug-like ligands that could be of value as novel therapeutic agents: Is high binding affinity attainable for the binding of small monovalent ligands to TSG101-UEV? Can the molecular determinants of binding affinity be identified and ascribed to specific features within the ligand that could be integrated into small-molecule inhibitors? We have addressed these questions using a competitive phage display setup that has produced peptide sequences with low micromolar dissociation constants, improving binding affinity by 1–2 orders of magnitude compared to natural Late domain peptides thanks to optimized interactions at the C-terminal region of the ligand. This approach allowed us to overcome the complexity of PRM domains and identify highly localized elements crucial for high binding affinity, of great value to direct virtual screening and targeted optimization campaigns aimed at the development novel therapeutic agents targeting TSG101-UEV interactions.

2. Results and discussion

2.1. Thermodynamic analysis of the binding of Late domain peptides to TSG101-UEV

The binding energetics of a set of peptide ligands corresponding to the PTAP Late domains of different viruses were measured by ITC (Fig. 1 and Table 1). In all cases, Late domain peptides bind to a single site in TSG101-UEV with low to moderate dissociation constants, confirming that these sequences can recognize their host cellular target in the absence of other elements from the full-length protein. A dissociation constant of 71 μM was measured for the HIV-9 peptide, in good agreement with previous results [30], being very similar to the binding affinity reported for the full-length HIV-1 p6-gag protein [18]. Placing the PTAP motif at the C-terminus of the peptide (HTLV-ter) leads to a dramatic decrease in binding affinity, indicating that the residues immediately adjacent to the PTAP core motif are key for the interaction. This agrees with the relevant role previously proposed for the residue at position +1 in HIV-1 sequences [30]. Moreover, extending the peptide sequence by two additional residues at the C-terminus elicits very small effects ($\Delta\Delta G_{\text{HIV-11/HIV-9}} = -1\ \text{kJ}\cdot\text{mol}^{-1}$ and $\Delta\Delta G_{\text{Ebola-11/Ebola-9}} = -1.5\ \text{kJ}\cdot\text{mol}^{-1}$), suggesting that the determinants of binding affinity are highly localized around the PTAP motif. The presence of a PPxY Late domain motif in tandem with the PTAP core sequence (HTLV-ter or Ebola peptides) does not lead to improved binding affinities, in agreement with *in vivo* studies indicating the independent behaviour of different types of Late domains [31–33].

All viral Late domain sequences bind TSG101-UEV with the thermodynamic signature previously described for other proline-rich recognition modules (large and negative binding enthalpies partially compensated by unfavourable entropic contributions (Fig. 1C). Nonetheless, large differences in binding enthalpies and heat capacities (up to 12 $\text{kJ}\cdot\text{mol}^{-1}$ and 1.4 $\text{kJ}\cdot(\text{K}\cdot\text{mol})^{-1}$ respectively) are observed between the different Late domains. The binding of the HIV-11 and Ebola-11 peptides is coupled to the uptake of 0.5 protons by protein/ligand

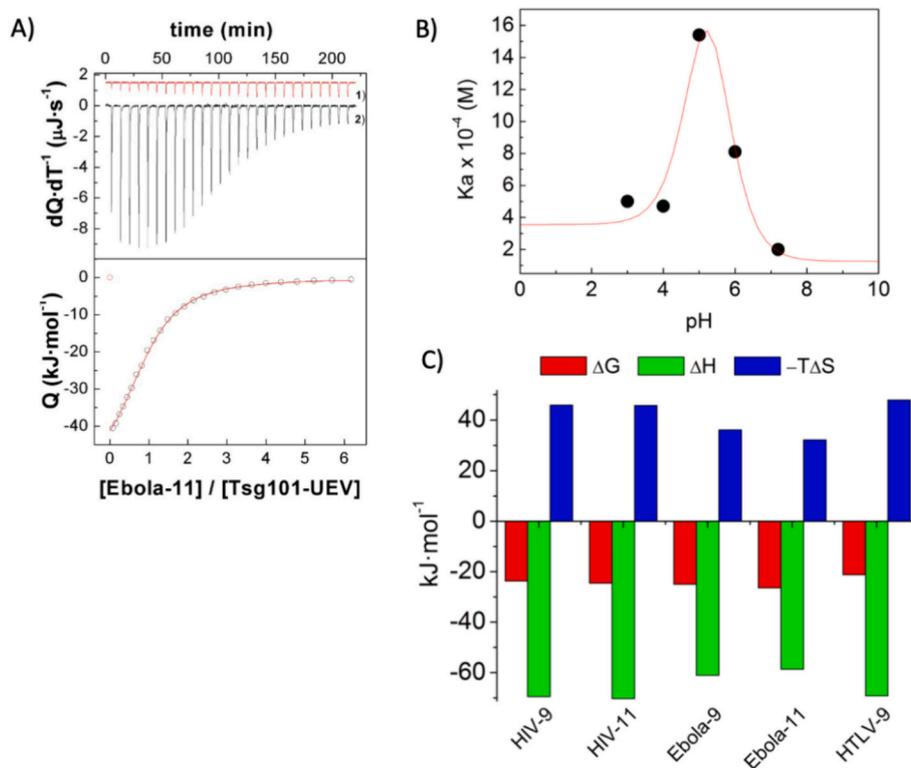


Fig. 1. Binding energetics of viral Late domains to TSG101-UEV. A) Calorimetric titration of the TSG101-UEV domain with the Ebola-11 peptide (ILPTAPPEYME) at 25 °C in 20 mM sodium phosphate, 5 mM β -mercaptoethanol pH 7.2. Upper panel: 1) Dilution experiment of Ebola-11 into the corresponding buffer under the same conditions and identical injection profile than the experiment. 2) Heat effects associated with the injection of Ebola-11 into the calorimetric cell containing the TSG101-UEV domain. The dilution curve has been displaced in the y-axis for representation purposes. Lower panel: Ligand concentration dependence of the heat released upon binding after normalization and correction for the heats of dilution. Symbols represent experimental data, and the continuous line corresponds to the best fitting to a model considering one set of binding sites. B) pH dependency of the Ebola-11 binding constant. The black points represent the experimental values and the red line correspond to the best fit to the equation described elsewhere [34]; C) Thermodynamic profile for the binding of Late-domain peptide to TSG101-UEV. Red bars represent the binding Gibbs energies at pH 7.2 and green and blue bars correspond, respectively, to the enthalpic and entropic contributions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

residues upon formation of the complex (Fig. 1B), associated to the ionization of Asp/Glu and His side chains (see Suppl. Information for a detailed analysis). Considering that the three His residues in TSG101-UEV (His₁₀₂ at the β 5- α 3 loop plus His₁₁₅ and His₁₁₉ at the α 3 helix) are >20 Å away from the binding site, these results suggest that binding of the PTAP peptide induces long-range changes in the environment of these residues leading to modulation of their pKa values.

After considering the ionization contributions, the binding of Late domain peptides to TSG101-UEV remains strongly exothermic, as typically found for other polyproline-recognition modules, characterized by markedly negative binding enthalpies that cannot be fully explained by direct contacts ligand/binding site contacts. The origin of this thermodynamic behaviour has been widely studied for SH3 and WW domains and attributed to the interplay of several factors complementing the direct interfacial interactions, *i.e.*, the establishment of extensive and conserved networks of water-mediated interactions at the binding interface [27,28,35] in combination with contributions arising from the modulation of the dynamic properties and conformational distribution of peptide and domain upon binding [29,36–38]. In the case of TSG101-UEV, the measured binding enthalpies differ strongly from the intrinsic values calculated from structure-based analysis of the interacting surfaces: $\Delta H_{\text{int}}(25\text{ °C}) = -35.8\text{ kJ}\cdot\text{mol}^{-1}$ and $\Delta C_p = 0.96\text{ kJ}\cdot(\text{K}\cdot\text{mol})^{-1}$ for HIV-9 and $\Delta H_{\text{int}}(25\text{ °C}) = -12.1\text{ kJ}\cdot\text{mol}^{-1}$, and $\Delta C_p = 1.11\text{ kJ}\cdot(\text{K}\cdot\text{mol})^{-1}$ for Ebola-9 [39,40]. These calculations, which typically predict binding enthalpies with an average error of $1.3\text{ kJ}\cdot\text{mol}^{-1}$ render theoretical values for TSG101-UEV that are $36\text{--}49\text{ kJ}\cdot\text{mol}^{-1}$ smaller than those determined experimentally. Such large discrepancies highlight the fact that direct PTAP-ligand/UEV interactions cannot solely account for the

absolute enthalpy and heat capacity values nor the large enthalpic differences observed between the ligands, suggesting that factors such as water-mediated interactions and energetic contributions arising from conformational effects are also determinant for PTAP ligand recognition by TSG101-UEV.

2.2. Structural characterization of TSG101-UEV/PTAP viral Late domain complexes

To further explore the origin of energetic variability between the different Late domain complexes, we determined the crystal structures of the complexes formed by TSG101-UEV with the Ebola and the HTLV Late domain peptides (Table 2). In all structures, the TSG101-UEV domain adopts very similar conformations, showing the typical UEV fold (Fig. S1A). The Late domain peptides in the complexes are well-ordered and have been fully modelled (Fig. S1B). The conformation and interaction pattern of the N-terminal region of the peptides and the PTAP core motif (positions -5 to 0) are very similar for the Ebola and HTLV Late domains, and highly conserved among the different TSG101-UEV complex structures available to date (Fig. 2).

As detailed in Tables S1-S4, the PTAP core motif shows a characteristic and highly conserved pattern of hydrophobic contacts and hydrogen bonds, in some cases mediated by partially buried water molecules found at conserved hydration sites (W1 to W6 in Figs. 2 and S2). Specifically: a) the first proline in the motif (P₃) inserts in a shallow hydrophobic pocket formed by P₇₁, T₅₈ and T₉₂; b) T₂ does not establish significant hydrophobic interactions with the domain but engages in three hydrogen bonds, including two direct interactions with the

Table 1
Binding energetics of viral Late domains to TSG101-UEV.

Ligand	pH	Buffer	T (°C)	ΔH_{app} (kJ·mol ⁻¹)	K_d (μ M)	$\Delta C_{p,app}$ (kJ·(K·mol) ⁻¹)	n_H	ΔH_{int}^a (kJ·mol ⁻¹)		
HIV-9 ^b	7.2	Phosphate	25	-69.5 ± 7.1	71.4 ± 6.2					
HIV-11 ^b	7.2	Phosphate	15	-49.4 ± 1.4	37.0 ± 2.8	-2.12 ± 0.49	0.5 ± 0.05	-72.0 ± 1.1		
			20	-64.5 ± 1.0	52.6 ± 1.7					
			25	-70.3 ± 5.1	50.0 ± 3.3					
	PIPES	25	-65.3 ± 1.8	83.3 ± 5.4						
		Imidazole	25	-53.6 ± 1.0	37.0 ± 2.0					
	6.0	Acetate	25	-73.1 ± 2.1	12.3 ± 0.6					
			5.0	-63.3 ± 1.6	6.5 ± 1.2					
			4.0	-58.9 ± 2.8	20.8 ± 2.0					
			3.0	-48.6 ± 1.0	20.0 ± 0.8					
	Ebola-9 ^c	7.2	Phosphate	25	-61.1 ± 2.0				41.7 ± 1.7	-0.93 ± 0.23
Ebola-11 ^c				7.2	Phosphate	20	-55.7 ± 0.5	19.6 ± 0.7		
	25	-58.6 ± 0.5	23.3 ± 0.6							
	30	-64.9 ± 0.5	33.3 ± 0.7							
PIPES	25	25	-56.8 ± 1.2	22.2 ± 0.7						
		Imidazole	25	-44.3 ± 2.3	22.2 ± 1.7					
		HTLV-9 ^d	7.2	Phosphate	20	-64.5 ± 1.9	196.1 ± 9.6	-0.69 ± 0.08		
25	-69.1 ± 0.7				188.7 ± 3.2					
30	-71.6 ± 1.9				303 ± 11.8					
HTLV-ter ^d	7.2	Phosphate	25	>1000						

^a These values were obtained considering the reported values for the ionization enthalpies of the different buffers: sodium phosphate ($\Delta H_{ion} = 3.6$ kJ·mol⁻¹), PIPES ($\Delta H_{ion} = 11.2$ kJ·mol⁻¹), imidazole ($\Delta H_{ion} = 36.64$ kJ·mol⁻¹), cacodilate ($\Delta H_{ion} = -3.0$ kJ·mol⁻¹), Acetate ($\Delta H_{ion} = -0.41$ kJ·mol⁻¹) and glycine ($\Delta H_{ion} = 4.0$ kJ·mol⁻¹).

^b Peptide sequences HIV-9 (PEP₃TAP₀PEE) and HIV-11 (PEP₃TAP₀PEESF) correspond to residues 453-460 and 453-462 respectively within the HIV-1-gag-p6 sequence (Uniprot: P04591).

^c Peptide sequences Ebola-9 (ILP₃TAP₀PEY) and Ebola-11 (ILP₃TAP₀PEYME) correspond to residues 5-13 and 5-15 respectively within the Ebola VP40 sequence (Uniprot: Q05128).

^d Peptide sequences HTLV-9 (YVEP₃TAP₀QVL) and HTLV_{ter} (PPPYVEP₃TAP₀) correspond to residues 118-127 and 124-133 respectively within the HTLV-1 p19-gag (Uniprot: P03345).

backbone atoms of N₆₉ and a water-mediated hydrogen bond with the carbonyl oxygen of T₆₇ (hydration site W1); c) the methyl group of A₁ is tightly packed against a small pocket formed by residues I₇₀, M₉₅, S₁₄₃ and V₁₄₁. In some complex structures (HIV-3obu and HRS-3obq), the A₁ backbone atoms are hydrogen bonded to the S₁₄₃ side chain, participating in a complex network of water-mediated hydrogen bonds implicating partially buried water molecules at hydration sites W2 and W3, the carbonyl oxygen of P₃, and the S₉₄ and T₉₂ sidechains, which are also part of the ubiquitin binding site; d) P₀ is the main anchor point for PTAP recognition. Its carbonyl oxygen forms a strong hydrogen bond with the backbone nitrogen of S₁₄₃, conserved in all Late domain complexes with an average distance of 2.8 Å, and contributing strongly to the binding energetics (up to 22 kJ·mol⁻¹ according to calculations with the software YASARA [41]). In addition, the P₀ sidechain is buried in a deep hydrophobic pocket formed by residues Y₆₃, Y₆₈, I₇₀, P₁₃₉ and F₁₄₂ and closely resembling the “xP pocket” of SH3 and WW domains [25], establishing strong contacts with the Y₆₈ side chain; e) Finally, Y₆₈, Y₆₃, N₆₆, and R₆₄ strongly coordinate a buried water molecule at hydration Site W5, which is highly conserved in most TSG101-UEV structures, including both free domains and complexes. Water molecules at this site do not interact directly with the peptide ligand but seem to stabilize Y₆₈ in an optimal conformation for binding. A similar situation is found for waters at site W6 that mediate the interaction between T₅₈ and N₆₉ in the protein, complementing the direct N₆₉/P₄ hydrogen bond.

Outside the PTAP motif peptide/domain interactions are scarce. A direct hydrogen bond between the carbonyl oxygen of residues at position -4 in the ligand and the N₆₉ side chain is consistently observed at the N-terminal region, which adopts very similar conformations in the different complexes. Conversely, residues C-terminal from the PTAP motif show high conformational variability, although some hydrophobic contacts between residues at position 2 in the ligand and F₁₄₂ and P₁₄₅ are observed.

In summary, the structural analysis suggests that the recognition of PTAP-containing peptides by TSG101-UEV follows the paradigm previously described for other proline-recognition modules (*i.e.* SH3 and

WW domains) [27,28,35], in which extensive networks of water-mediated hydrogen bonds complement direct ligand-domain interactions. Late domain recognition is mainly driven by the P₃TAP₀ core sequence. The second proline in the motif (P₀) binds at the canonical xP proline binding pocket and is the main anchor point, contributing 40–60 % of the total contact energy and 25–55 % of the energy associated to direct ligand/protein hydrogen bonds (Tables S1 and S4).

2.3. Molecular dynamics studies of TSG101-UEV/PTAP viral Late domain complexes

The crystal structures do not reveal any remarkable sequence-specific interactions that could account for the reduced binding affinity found for the HTLV-1 Late domain compared to the Ebola and HIV-1 peptides. To gain further insight into the origin of these energetic differences and confirm the relevance of water-mediated interactions, 100 ns molecular dynamics trajectories were calculated using the different TSG101-UEV X-ray structures as starting points.

In all cases, the backbone root-mean-square (RMSD) deviation for the TSG101-UEV domain rapidly reached a plateau without abrupt transitions in the energy profile (Fig. S3), indicating that the UEV domain is stable throughout the simulations. The RMSF profile of the TSG101-UEV domain is similar for all complexes, showing two regions of higher flexibility corresponding to the “lip” and “tongue” regions conforming the ubiquitin binding site (Fig. S4). Interestingly, binding of the PTAP peptides seems to enhance the mobility of these regions, suggesting a long-range cooperative connection between the PTAP and ubiquitin binding sites. All ligands show a similar RMSF pattern, with P₀, A₁, and T₂ being the main anchor points.

MD simulations reproduced the interaction pattern observed in the X-ray crystal structures, showing highly homogenous peptide-ligand interactions for all complexes within the PTAP core-motif region (Fig. 3A), dominated by packing interactions at the A₁ (I₇₀, M₉₅, S₁₄₃ and V₁₄₁) and P₀ (S₁₄₃, Y₆₃, Y₆₈, I₇₀, P₁₃₉ y F₁₄₂) pockets and complemented by strong direct hydrogen bonds with Y₆₉ and S₁₄₃. The

Table 2
Data collection and refinement statistics.

	TSG101UEV/ HTLV	TSG101UEV/ Ebola	TSG101UEV- P321
PDB code	4ZNY	4EJE	4YC1
Wavelength (Å)	0.97	0.98	0.98
Resolution range	19.96–2.40 (2.49–2.40)	20.00–2.20 (2.32–2.20)	19.60–2.00 (2.06–2.00)
Space group	C222 ₁	I4	P321
Unit cell (Å ³)	78.62	105.54	169.74
	119.74	105.54	169.74
	41.64	75.30	39.71
	90 90 90	90 90 90	90 90 120
Total reflections	37,420 (4376)	171,665 (19410)	381,505 (29916)
Unique reflections	8020 (839)	21,003 (3046)	44,232 (3606)
Multiplicity	4.7 (5.3)	8.2 (6.4)	8.6 (8.3)
Completeness (%)	99.80 (100)	99.7 (100)	99.7 (100)
Mean I/sigma(I)	34.6 (9.7)	19.7 (4.7)	22.6 (4.5)
Wilson B-factor	54.96	41.71	31.97
R-merge	0.012 (0.12)	0.059 (0.404)	0.019 (0.170)
Reflections used in refinement	8020(750)	20,948 (2059)	44,065 (4389)
Reflections used for R-free	416 (39)	2134 (227)	2130 (227)
R-work	0.24 (0.40)	0.20 (0.32)	0.17 (0.24)
R-freeCou	0.2459 (0.3662)	0.24 (0.38)	0.20 (0.31)
Non-hydrogen atoms	1168	2530	3713
Macromolecules	1158	2445	3464
Ligands	5	5	25
Solvent	5	80	224
Protein residues	152	303	440
RMS (bonds)	0.003	0.010	0.013
RMS (angles)	0.87	1.35	1.38
Ramachandran favored (%)	97.97	96.27	98.39
Ramachandran allowed (%)	2.03	3.39	1.61
Ramachandran outliers (%)	0.00	0.34	0.00
Rotamer outliers (%)	0.00	5.96	0.26
Clashscore	1.76	9.11	2.01
Average B-factor	71.03	27.76	48.85
Macromolecules	71.10	27.63	48.97
Ligands	67.00	79.65	59.18
Solvent	57.56	28.52	45.77
Number of TLS groups	9	20	24

Statistics for the highest-resolution shell are shown in parenthesis.

pattern of water-mediated interactions derived from the crystallographic structures is also reproduced in all trajectories (see Table 3 for average values over all complexes and Table S5 for a more detailed analysis). Long-lived water molecules (average residence times >10 ns) were found to mediate ligand-domain interactions in hydration sites W1-W4 with high frequency (between 60 and 90 % of the simulation time on average) and small dispersion between complexes. Also, structural water molecules at sites W5 and W6 mediate hydrogen bonds between TSG101-UEV residues in the free domain and complex simulations, with occupancies over 90 %. Water molecules at these sites are tightly coordinated, showing high residence times that are increased upon ligand binding (from 8 to 21 ns on average for R₆₄-w-Y₆₈ at site W5 and from 25 to 61 ns for N₆₉-w-T₅₈ at site W6, reaching maximum values of 44 ns and 135 ns, respectively).

The MD simulations unveil differences in the interaction pattern of residues outside the PTAP core motif that were not apparent in the crystal structures. The strongest and most discriminating interactions implicate glutamate residues at positions +2 and +3 at the C-terminal end of the HIV-1 and Ebola peptides, which seem to be responsible for the stronger binding of these Late domains (K_d ~ 20–50 μM). Throughout the MD trajectories, these residues remain near positively charged residues (R₆₄, K₉₈, and R₁₄₄) with high frequency (Fig. 3A). Specifically, E₊₂ and R₁₄₄ remain in close contact throughout the HIV-1

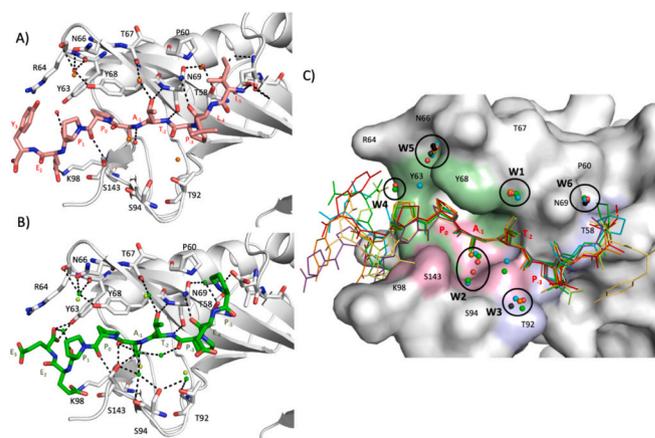


Fig. 2. Water-mediated interactions observed in the high-resolution structures of the Ebola (A) and HTLV (B) Late domain complexes with TSG101-UEV. Partially buried interfacial water molecules (ASA < 30 Å²) are shown as red spheres and water-mediated hydrogen bonds (polar contacts identified by PyMOL) are shown as discontinuous black lines. (C) Conserved hydration sites (black circles) observed in the crystal structures of the TSG101-UEV domain (grey surface). The P₀, A₁, and P₃ binding pockets have been highlighted in green, pink, and purple respectively. Ligand residues (lines) and interfacial water-molecules (spheres) have been coloured according to the different structures (Ebola-4eje-A: red; Ebola-4eje-B: orange; HTLV_4zny: yellow; HIV1-3obu: dark green; HIV1mutant-3obx: light green; Hrs-3obq: cyan; HEV-7nlc: purple; free domain structures 4yc1, 3obs and 2f0r: black). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and Ebola simulations (100 % and 50 % of the simulation time, respectively), establishing strong electrostatic interactions and long-lived hydrogen bonds. Such interactions are not apparent from the Ebola and HIV-1 crystal structures (Fig. 3B), where these glutamate side chains adopt alternative conformations resembling those observed for the uncharged side chains present in the weaker binding HTLV, HEV, and HRS peptides (K_d ~ 200–300 μM). Interestingly, the P₂/A substitution within the HIV-1 Late domain, which produces a 50 % reduction in binding affinity, is also associated to weakened E₊₂ and E₊₃ electrostatic interactions.

In summary, we find that some Late domain sequences achieve improved binding affinities with respect to TSG101-UEV cellular partners even in the absence of additional elements from the viral partners. These short Late domain sequences contain, thus, the most binding determinants, which are highly localized around the PTAP core motif. Despite the conservation of the PTAP interaction pattern for the different complexes, our results show that optimized interactions implicating C-terminal residues immediately adjacent to the core motif (positions +2 y +3) seem to be key for tight binding to TSG101-UEV. The structural and MD studies reveal the importance of several interfacial waters at highly conserved hydration sites, including water molecules mediating ligand/protein interactions and long-lived structural waters present in the free TSG101-UEV domain that become more tightly coordinated upon ligand binding. These interfacial water molecules are expected to have a significant impact on the binding energetics, as reflected in the largely exothermic enthalpies, and might also be important for enhancing enthalpy/entropy compensation effects [27,35].

2.4. Phage display study of TSG101-UEV binding preferences for affinity optimization

To further elucidate the binding preferences of TSG101-UEV and explore the possibility of generating high-affinity ligands, a competitive phage display screening approach was devised, following a strategy that

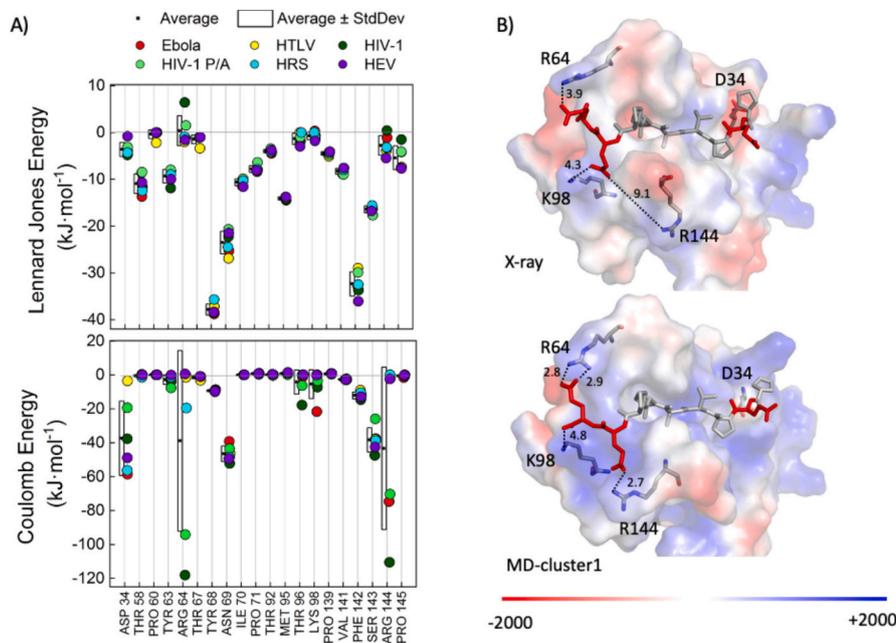


Fig. 3. Predominant interactions in MD trajectories for the different TSG101-UEV complexes. A) Trajectory averages of the packing (upper panel) and electrostatic (lower panel) contributions to the binding energy for the most relevant residues in the TSG101-UEV domain calculated by Gromacs ver2018.2 suite [42]. For each complex, the energetic contributions for each residue averaged throughout the MD trajectory are shown as coloured circles: Ebola-4eje-A (red), HTLV-4zny (yellow), HIV-1-3obu (dark green), HIV1mutant-3obx (light green), Hrs-3obq (cyan), HEV-7nlc (purple). Black squares represent the residue contributions averaged over all trajectories and open bars the corresponding standard deviation. B) Surface representation of the TSG101-UEV domain coloured according to the electrostatic potential calculated by PyMOL-APBS using the crystal structure of the 3obu-HIV-1 complex (upper panel) and a representative structure of the most populated cluster derived from the HIV-1 MD trajectory (lower panel). Ligand residues are shown as grey or red (negatively charged side chains) sticks. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
MD analysis of water-mediated interactions.

Hydration site	Interaction	Average complexes		Average free domain	
		Occupancy ^a (% sim. time)	<time> ^b (ns)	Occupancy ^a (% sim. time)	<time> ^b (ns)
Site 1	N69-T67	73 ± 5	10 ± 1	72 ± 4	10 ± 3
	L-N69	90 ± 5	9.4 ± 0.8	–	–
	L-T67	52 ± 4	3.1 ± 0.4	–	–
Site 2	S94-S143	80 ± 1	15.0 ± 0.8	74.6 ± 0.3	8 ± 1
	L-S94	87 ± 4	9.3 ± 0.4	–	–
	L-S143	90 ± 1	13 ± 2	–	–
Site 3	T92-S94	84.0 ± 0.8	13 ± 7	80 ± 3	9.3 ± 0.9
Site 4	K98-Y63	71 ± 2	14 ± 3	75 ± 3	10 ± 3
	L-K98	57 ± 15	13 ± 3	–	–
	L-Y63	60 ± 12	15 ± 5	–	–
Site 5	Y63-R64	93 ± 1	10 ± 1	96.2 ± 0.6	9 ± 3
	Y63-Y68	46 ± 8	8 ± 1	52 ± 22	5 ± 1
	Y63-N66	15 ± 3	3.7 ± 0.4	28 ± 3	2.3 ± 0.3
Site 6	R64-Y68	74 ± 5	21 ± 6	49 ± 19	8 ± 3
	R64-N66	38 ± 4	4.3 ± 0.8	46 ± 7	5.3 ± 0.9
	Y68-N66	87 ± 3	9 ± 2	57 ± 22	9 ± 2
	N69-T58	86 ± 2	61 ± 14	90.3 ± 0.3	25 ± 2
	L-N69	89 ± 5	9.4 ± 0.8	–	–
	L-T58	37 ± 7	12 ± 3	–	–

^a Occupancy: Average values for the % of simulation time that water mediated interactions implicating binding site residues and PTAP ligands.

^b <time>: Average values for the residence times of interfacial water molecules calculated over the 100 ns of the MD trajectory.

previously had allowed us to produce nanomolar binders for the WW3 domain of NEDD4 [15]. In this way, monovalent phage display screening at pIII was performed using a randomized library with a fixed P(T/S)AP core motif centred in the sequence (pIII-x₅P(T/S)APx₅). To favour the identification of high-affinity sequences, additional rounds of selection were carried out using the tightest binding sequence from the first round (PD1) as a competing ligand. Analysis of the best sequences (Fig. 4A-B) revealed a very strong preference for a proline residue at position +1, consistent with previous reports suggesting a P(T/S)AP₀P core motif for TSG101-UEV binding [30]. In agreement with our thermodynamic and structural results, the N-terminal section of the peptide was found to be highly variable, although bulky and highly constrained residues were frequently found at positions -4 and -5. Conversely, the C-terminal region showed well-defined preferences at positions +2 to +4: Trp residues were found with very high frequency at position +2,

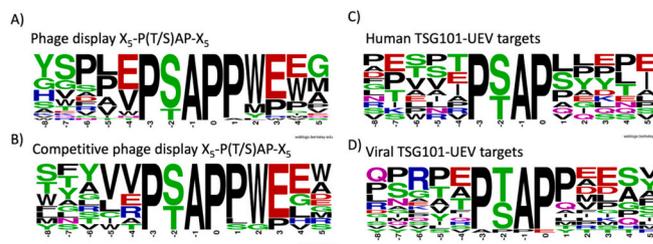


Fig. 4. Web-logo (<https://weblogo.berkeley.edu>) showing the frequency of occurrence of the different amino acids for TSG101-UEV ligands A) Best 30 sequences derived from the randomized pIII-x₅P(T/S)APx₅ library; B) best 12 sequences derived from the competitive selection round; C) naturally occurring sequences from human proteins reported to bind TSG101-UEV D) naturally occurring sequences from viral proteins reported to bind TSG101-UEV. TSG101-UEV interacting sequences from natural human and viral partners were extracted from the ELM database (<http://elm.eu.org/>).

followed by negatively charged residues (mostly Glu) at position +3 and, less frequently, at position +4, supporting the hypothesis that electrostatic interactions at the C-terminus are important for high affinity.

The binding energetics of four representative peptide sequences were determined by ITC. These included two sequences (PD1 and PD2) corresponding to the most frequent patterns from round 1 and two sequences (PD3 and PD4) designed to incorporate the most salient features observed in the rounds of competitive selection (Fig. 5 and Table S6). As expected, we found that the PD peptides present binding affinities 1 to 2 orders of magnitude higher than the natural Late-domain sequences. The best binder (PD4) combines elements found in the strongest natural binders: a) bulky hydrophobic residues at positions -4 and -5 plus an aromatic residue at position +2 (Ebola) and b) proline at +1 and glutamate at +3 (HIV-1). Interestingly, the improvement in binding affinity arises from optimized enthalpic ($\Delta\Delta H_{PD4-Ebola} = -4.5 \text{ kJ}\cdot\text{mol}^{-1}$) and entropic ($-\Delta\Delta S_{PD4-Ebola} = -2.3 \text{ kJ}\cdot\text{mol}^{-1}$) contributions, something difficult to achieve by rational design. We find that substituting the core-motif Ser₁ by Thr₁ (PD4_T) or removing the S₈Y₉T₆ N-terminal residues (PD5) have negligible energetic effects, confirming that residues beyond position -5 are not relevant for binding. Taking PD5 as a reference, C-terminal extensions were incorporated to test the impact of including an additional -E₄W₅ moiety, frequently found within the phage display sequences (peptides PD5_E and PD5_EW) and reminiscent of the HIV-1 Late domain. The introduction of a second glutamate in PD5_E produced a large increase in binding enthalpy ($\Delta\Delta H_{PD5-PD5_E} = -7.6 \text{ kJ}\cdot\text{mol}^{-1}$), reaching values very similar to the HIV-1 peptide. Nonetheless, this improvement is fully overcome by more unfavourable entropic contributions, leading finally to a reduced binding affinity. Such strong enthalpy/entropy compensation effects, possibly reflecting the complexity of binding and the strong influence of water-mediated interactions and conformational reorganization [27,29,35], have been systematically observed for SH3 and WW domains and are, indeed, one of the major roadblocks for rational ligand optimization.

To gain further insight into the origin of the improved binding affinities, structural models of the TSG101-UEV domain in complex with the different phage display ligands were generated using the crystal structure of the Ebola complex (4EJE) as a template. Using the modelled structures as starting points, 100 ns MD trajectories were calculated for the different complexes. Phage display peptides maintain the conformation and the conserved pattern of interactions found for the natural ligands at the N-terminal and PTAP regions. Nonetheless, significant

differences are observed in the C-terminal region. As illustrated in Fig. 6A the tryptophan sidechain at position +2 in the PD peptides, strongly selected in all phage display experiments, is inserted in a new binding pocket formed by residues K₉₈, T₉₆, F₁₄₂, and P₁₄₅, where it establishes highly optimized interactions, including better packing and stronger hydrogen bonds. These new interactions seem to outweigh the loss of the E₊₂/R₆₄-R₁₄₄ coulombic interactions characteristic of the best natural binders (Fig. 6B-C). In this way, the hydrophobic contact energies calculated by YASARA for residues at position +2, ranging from 3 to 9 $\text{kJ}\cdot\text{mol}^{-1}$ for the natural Late domains (Table S4), rise to 17 $\text{kJ}\cdot\text{mol}^{-1}$ for PD5. Also, even though the geometry and frequency of all hydrogen bonds are highly conserved between PD and natural ligands, the hydrogen bond between E₊₂:N/S₁₄₃:O, observed with relatively low frequency (occupancies of 20 % for Ebola and 40 % HIV-1) for the best natural Late domains coupled to the E₊₂/R₆₄-R₁₄₄ coulombic interactions, is strengthened by the presence of W₊₂, with occupancies up to 66 % for the PD ligands (Table S5).

In summary, even though we have shown that the recognition of PTAP-containing sequences by TSG101-UEV shares the complexity, low affinity, and promiscuity inherent to polyproline recognition by SH3 or WW domains, the phage display approach used here has allowed us to overcome these challenges, achieving up to 2 orders of magnitude improvements in binding affinity. The fact that most binding energy is highly localized around a relatively small PTAP₀PW sequence is a valuable opportunity for drug design. Our results suggest that identifying small molecule ligands of TSG101-UEV and improving their binding affinity through targeted optimization is feasible, potentially leading to drug-like inhibitors of TSG101-UEV/PTAP interactions of therapeutic interest as novel broad-spectrum antivirals.

2.5. Computational study of the impact of PTAP ligand binding on TSG101-UEV intramolecular communication networks

Functional dynamics and energetic effects associated with the modulation of the conformational distribution of proteins and ligands upon binding have been reported to contribute significantly to the binding energetics of polyproline recognition domains [29,36–38]. To explore to which extent the engagement of a new pocket by W₊₂, largely responsible for the high-affinity binding of the phage display peptides, modulates the conformational properties of the TSG101-UEV domain, the intramolecular cooperative networks of the Ebola and PD5 complexes with respect to the free UEV domain were analysed. Aiming at generating a robust consensus description, three algorithms based on different principles were used (see Methods for details): pPerturb (statistical mechanics analysis of thermodynamic perturbations) [43], ProteinLens (graph-based calculation of bond-to-bond propensities) [44], and WebPSN (protein structure network and elastic network model/normal mode analysis, ENM-NMA) [45]. These web-based algorithms were applied to the high-resolution structures of the free TSG101-UEV domain (2FOR) and the Ebola complex (4EJE) as well as to the representative structure of the most populated cluster from the PD5 MD trajectory.

A consensus of the most relevant residues for allosteric coupling within the TSG101-UEV domain (allosteric hotspots) was generated to a robust description that reflects the underlying energetic coupling within the protein, minimizing eventual artifacts associated with each approach (see supplemental information for a detailed description). For the free TSG101-UEV domain, all algorithms consistently identified a cluster of highly energetically coupled residues within the protein core, composed mostly of aromatic residues at the $\alpha 1$ - $\alpha 2$ (Y₁₅, Y₁₇) and $\beta 3$ - $\beta 4$ (W₇₅, D₇₈, Y₈₀, P₈₁, Y₈₂) loops as well as the $\alpha 3$ (W₁₁₃, W₁₁₇) and $\alpha 4$ (L₁₂₄, F₁₃₅) helices (Fig. 7A). As shown in Fig. 8A, WebPSN identified a network of strong and highly recurrent links centred at Y₈₀ ($\beta 3$ - $\beta 4$ loop) connecting the PTAP and ubiquitin binding region (see supplementary information for details).

The binding of the Ebola ligand modulates this cooperative network,

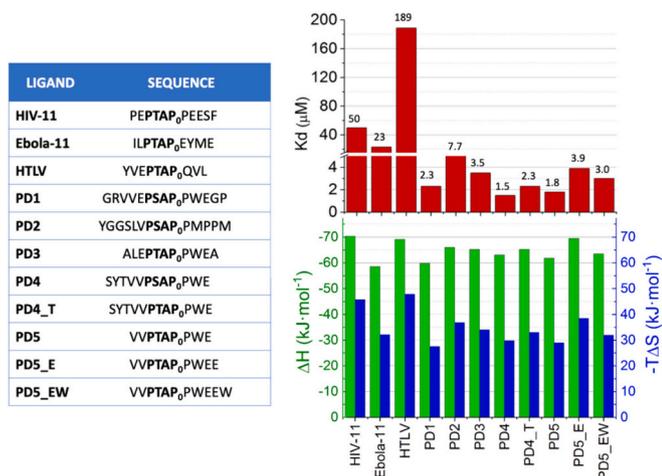


Fig. 5. Binding energetics of phage display peptides to TSG101-UEV. Red bars represent the dissociation constants measured at pH 7.2 and 25 °C. The enthalpic and entropic contributions to the binding affinity are shown as green and blue bars, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

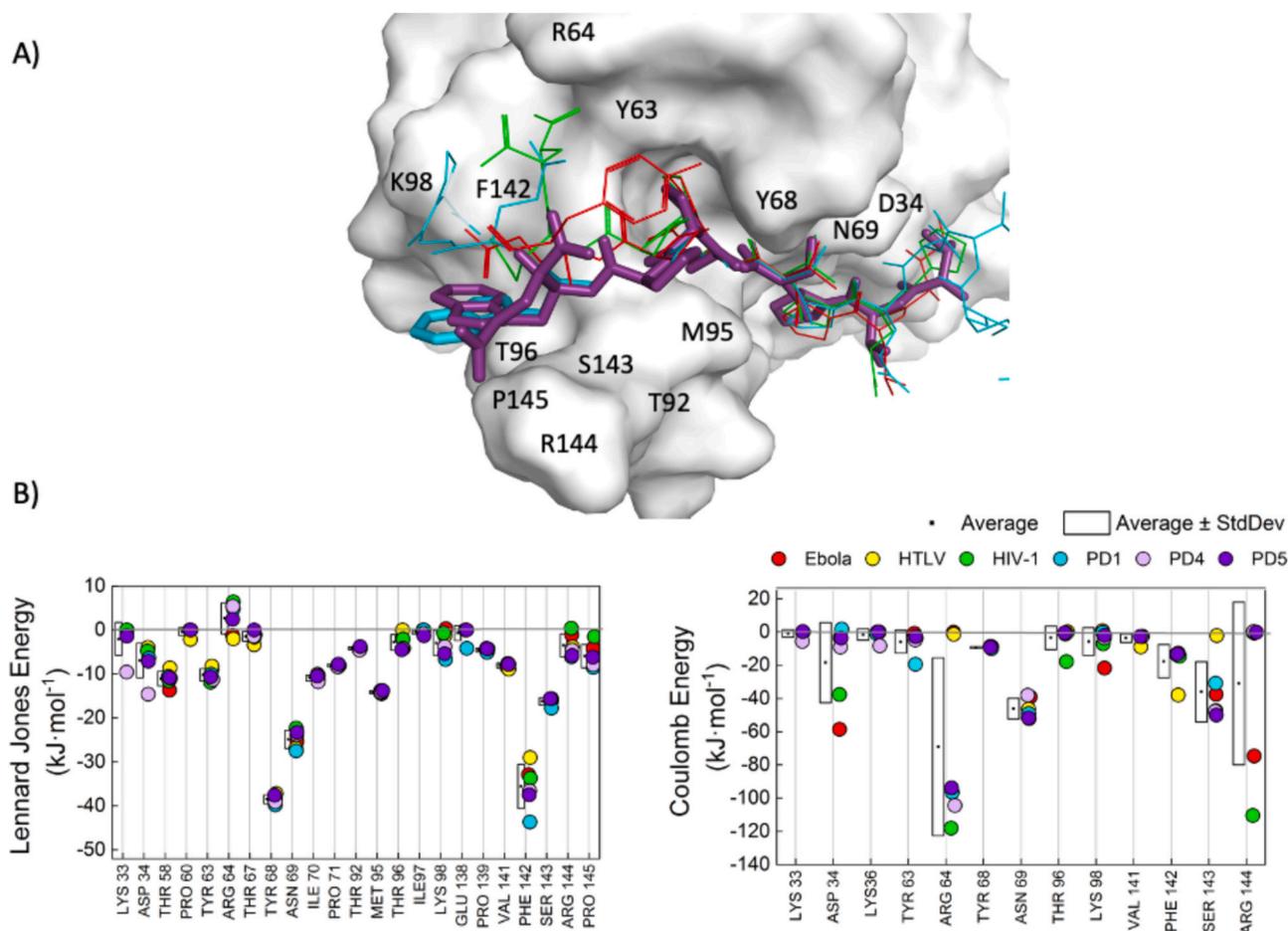


Fig. 6. A) Binding mode of high affinity phage display PD1 and PD5 peptides. Superposed are the crystal structures of the HIV-1 (3OBU) and Ebola (4EJE-A) ligands in complex with TSG101-UEV and the most populated clusters derived from the MD simulations for high affinity ligands PD1 and PD5 (40 % and 50 % of the simulation time, respectively). The structure of the TSG101-UEV domain in the PD5 complex is shown as a grey surface. Ligand residues are represented as coloured sticks (Ebola: red, HIV-1: green, PD1: cyan, and PD5: purple). PD5 residues and key W_{+2} from PD1 are shown as sticks. The most relevant interacting residues within TSG101-UEV have been labelled. B) Energetic contributions of TSG101-UEV residues to the binding of PD1 and PD5 in comparison with HIV-1 and Ebola Late domain sequences. Shown are the trajectory averages of the packing (left panel) and electrostatic (right panel) contributions to the binding energy for the most relevant residues in the TSG101-UEV domain calculated by Gromacs ver2018.2 suite [42]. For each complex, the energetic contributions for each residue averaged throughout the MD trajectory are shown as coloured circles: Ebola (red), HTLV (yellow), HIV1 (green), PD1 (cyan), PD4 (violet), and PD5 (purple). Black squares represent the residue contributions averaged over all trajectories and open bars the corresponding standard deviation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

increasing the number of allosteric hotspots, the interconnection between residues, and the coupling intensity (Tables S7 and S8). In this way, additional residues along the central $\beta 2$ and $\beta 3$ strands are engaged, strengthening the linkage between the PTAP binding site region and the Y_{80} allosteric cluster while weakening the coupling with the ubiquitin binding site (Fig. 8B).

Binding of PD5 further increase the number of allosteric hotspots (Fig. 8C and Table S9), including now residues W_2 , P_0 , and V_{-5} from the ligand. The optimized interactions between the W_{+2} side chain and the TSG101-UEV C-terminal residues (F₁₄₂ through P₁₄₅) and shift the most relevant communication pathway to the C-terminal half of the domain, enhancing the communication between the PTAP and ubiquitin binding sites (see supplementary information for details).

In summary, our results consistently indicate that the binding of PTAP-ligands modulates the conformational and dynamic properties of the TSG101-UEV domain. MD simulations reveal that Late domain binding enhances the mobility of the lip and tongue region, defining the canonical E2 binding site for monoubiquitin, in agreement with recent NMR studies reporting dynamic coupling between these two functional sites [24]. The measured pH dependency of the binding energetics, suggesting changes in the pKa of histidine residues >20 Å away from the

PTAP binding site, further sustains this idea. Two of the three histidine residues in the domain (H_{102} and H_{119}) are identified among the most relevant residues for allosteric coupling in bound and free TSG101-UEV, in some cases ranging within the top 10 % (Tables S7-S10). Moreover, H_{119} and H_{115} are close to the vestigial active site (Y_{110}), within the $\alpha 3$ helix, which also contains two of the strongest allosteric hotspots (Y_{113} and W_{117}). The relevance of these two histidine residues is modulated by the presence of the ligand (reduced in the Ebola complex and enhanced in the PD5 complex). Interestingly, the $\alpha 3$ helix contributes little to intramolecular communication in the free domain but is engaged in the most relevant communication pathways by the presence of the Ebola Late domain and, most significantly, by the PD5 high-affinity ligand (Fig. 8). These observations agree with recent reports showing that the removal of H_{115} side chain (H_{115A} mutant) impairs HIV-1-gag binding as efficiently as binding site (M_{95A}) or core motif (P_{-3}/L_3TAP_0) mutations. Similar effects are observed for other alanine substitutions (Y_{82A} , Y_{110A} , and W_{117A}) in the vicinity of the vestigial active site [24].

Overall, this analysis provides a rationale for the measured pH dependency of the Ebola binding energetics suggesting the ligand-induced modulation of the pKa values of distal His residues upon binding and confirms the allosteric coupling between the PTAP and ubiquitin

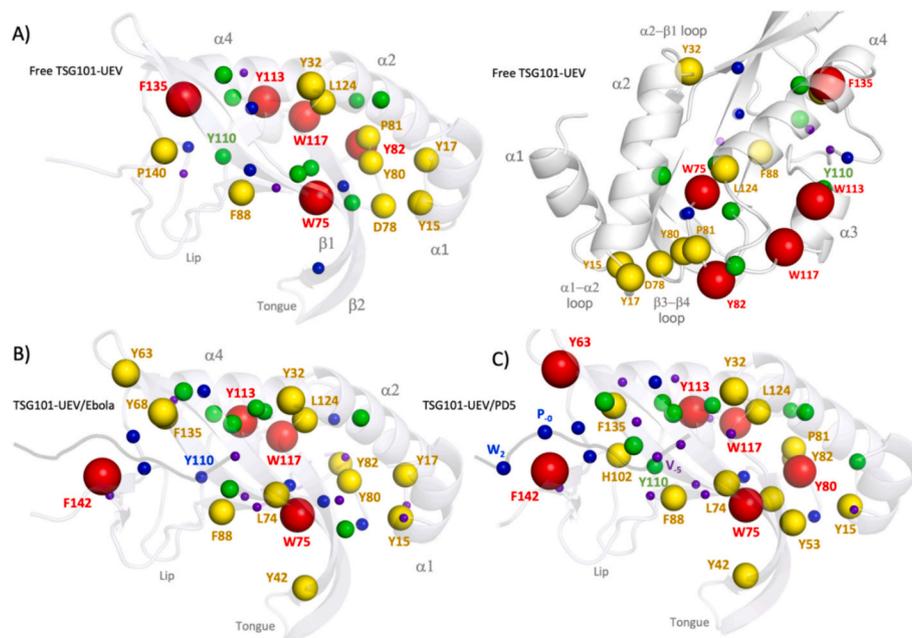


Fig. 7. Consensus selection of allosteric hotspots for A) the free TSG101-UEV domain in two different orientations; B) the TSG101-UEV/Ebola complex, and C) the TSG101-UEV/PD5 complex. Shown is a cartoon representation of the TSG101-UEV domain structure (white) and peptide ligands (grey). Residues ranking within the top 10 % most relevant sites for intra-molecular cooperative interactions are shown as spheres. Red spheres: top 10 % in the three algorithms; yellow: top 10 % in two algorithms; green: top 10 % in one algorithm and high score in at least another [>0.9 ProteinLens, >400 pPerturb, hub (>4 links) in Web-PSN]; blue: top 10 % in one algorithm and medium score in at least another [>0.7 ProteinLens, >400 pPerturb, hub (>4 links) in Web-PSN]; purple: top 10 % in one algorithm and low score in another. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

binding sites in TSG101-UEV, providing mechanistic insight into the intramolecular communication pathways and how they are modulated by the binding of different ligands. Elucidating the details of allosteric communication between the PTAP and mono- and di-ubiquitin binding sites is of enormous functional interest to fully understand the role of TSG101 in trafficking, MVB formation, cytokinesis, or viral budding, as well as its interplay with other ESCRT elements. In this sense, the high-affinity PD peptides, establishing strong interactions with new regions in the UEV domain and modulating the intramolecular communication pathways, could be interesting tools to probe intramolecular cooperativity within the UEV domain and its relevance for TSG101 functions.

3. Conclusions

Our integrated approach, which combines structural, biophysical, and computational analysis, has provided comprehensive insights into the molecular determinants underlying the recognition of PTAP motifs by the ubiquitin E2 variant domain of TSG101. Binding energy is focused on a few anchor residues within the motif itself, dominated by the P₀ proline, which provides most of the affinity but limited selectivity. Additional interactions outside this motif, many of them water-mediated, contribute to binding specificity and energetic optimization.

Despite the conserved architecture of the binding site, we have demonstrated that binding affinity to TSG101-UEV can be tuned through sequence optimization at hotspot positions flanking the core PTAP motif. Through massive phage display screening of randomized peptide libraries, we have optimized binding affinity, obtaining 1–2 order of magnitude improvements in the dissociation constants compared to the HRS cellular partner. We found that binding can be improved through anchor residues C-terminal from the core motif that establish optimized van der Waals at new proximal pockets and enhance electrostatic complementarity. These results are of high value to direct virtual screening and targeted optimization campaigns aimed at developing therapeutic agents targeting TSG101-UEV interactions.

This study provides a strategic blueprint for understanding and

modulating the recognition of PTAP-containing short linear motifs, that mediate key protein-protein interactions and constitute important pathogenicity factors in viruses and other disease-related pathways. Our findings may direct future efforts to develop inhibitors or molecular probes targeting TSG101-dependent sorting and budding processes through virtual screening or structure-based design focused on disrupting key energetic hotspots.

Finally, the results presented here also provide valuable mechanistic insights into the intramolecular communication pathways within TSG101-UEV and how they are modulated by ligand binding. Our high-affinity PD peptides are, in this respect, interesting tools to study intramolecular cooperativity within TSG101-UEV and the relevance of allosteric communication between the PTAP and mono- and di-ubiquitin binding sites for the regulation of TSG101 function.

4. Materials and methods

4.1. Protein expression and purification

For the structural and thermodynamic study, TSG101-UEV (residues 1–145 in uniprot Q99816) was expressed and purified as described before [20]. For the phage display studies, the plasmid encoding a His-tag-GST-TSG101-UEV construct, generously provided by Prof. Sachdev Sidhu (University of Toronto), was transformed into *Escherichia coli* BL21/DE3 cells. Cells were grown at 37 °C to an OD of 0.8. IPTG was added at 0.2 mM and expression was allowed overnight at room temperature. Cells were harvested, resuspended in 20 mM Na-Pi, 500 mM NaCl, and 20 mM imidazole, pH 7.4, broken with a French press, and ultracentrifuged at 30,000 rpm for 30 min. The supernatant was passed through a Ni-NTA column, with two steps of washes with an increasing imidazole concentration (20 mM and 50 mM), eluted at 500 mM of imidazole. The identity and purity of the protein was checked by SDS-PAGE and matrix-assisted laser desorption ionization time of flight experiments carried out at the Center of Scientific Instrumentation of the University of Granada. The purified protein (>95 % pure) was

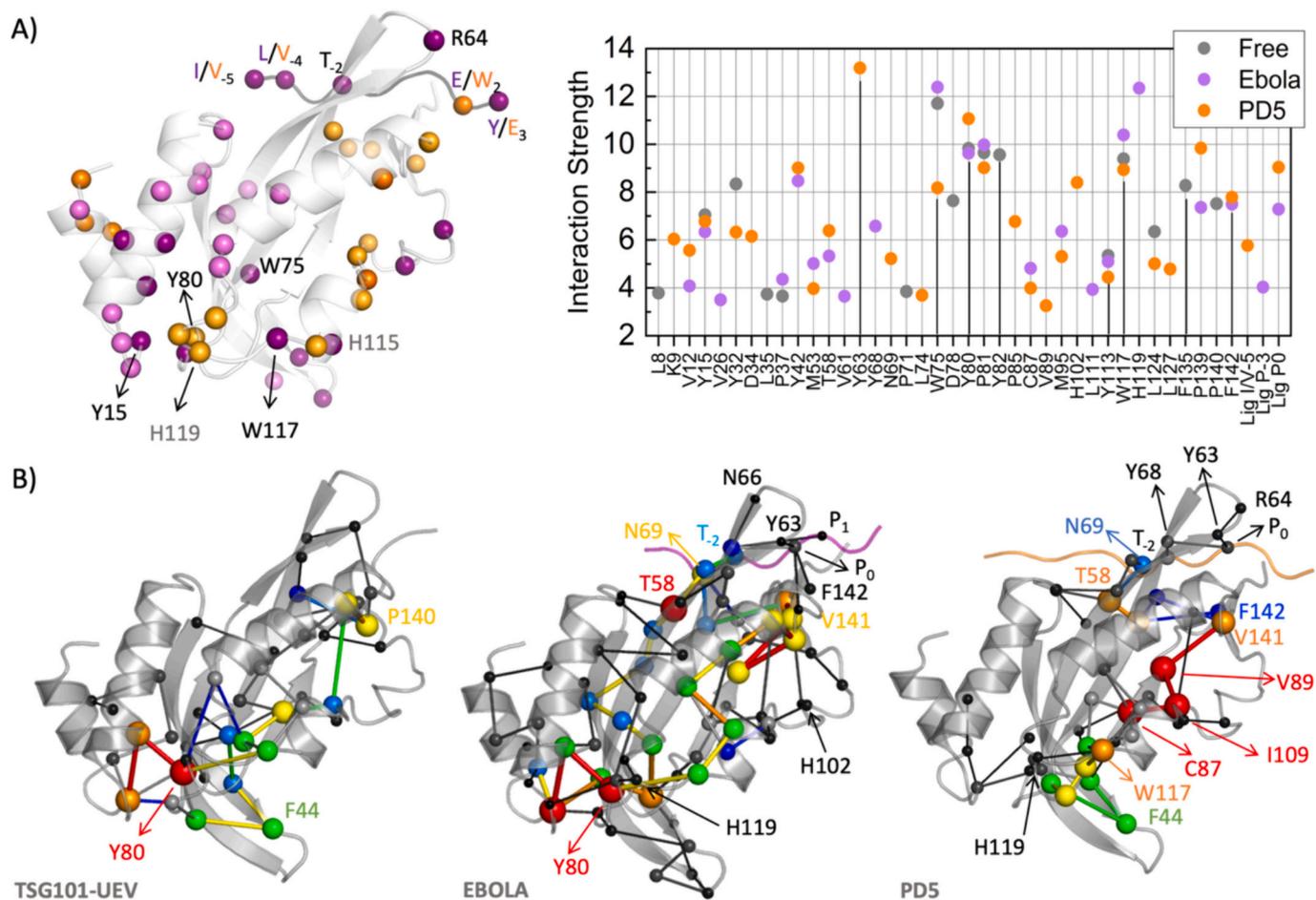


Fig. 8. Effect of ligand binding on allosteric communication within the TSG101-UEV domain. A) Differences on coupling intensity between the Ebola and PD5 complexes. The left panel shows residues with differences in coupling intensity above 20 % of the maximum interaction strengths. Residues interacting stronger in the Ebola complex are shown as purple spheres (pPerturb $\Delta\text{SumdQ}_{\text{Ebola-PD5}} > 40$: dark purple; ProteinLens $\Delta\text{Quantile Score} > 0.2$: light purple) and residues with stronger coupling in the PD5 complex are shown as orange spheres (pPerturb $\Delta\text{SumdQ}_{\text{Ebola-PD5}} < 40$: dark orange; ProteinLens $\Delta\text{Quantile Score} < 0.2$: light orange). The right panel summarized the interaction strengths of the WebPSN hubs for the free TSG101-UEV domain (grey symbols), the Ebola complex (purple symbols) and the PD5 complex (orange symbols). Black lines highlight the consensus allosteric hotspots selected by the Pperturb, ProteinLens and WebPSN algorithms. B) Global allosteric metapaths (see [Materials and methods](#) for details) calculated by WebPSN for the free TSG101-UEV domain (left panel), the Ebola complex (central panel), and the PD5 complex (right panel). The structure of the TSG101-UEV domain, Ebola ligand and PD5 ligands are shown as grey, purple, and orange cartoons, respectively. Nodes and hubs (nodes with >4 links) are shown as spheres and links are represented as lines connecting two nodes. Nodes and links are sized and coloured according to their relative recurrence (red: 90–100; orange: 80–90; yellow: 70–80; green: 60–70; light blue: 50–60; dark blue: 40–50; grey: 30–40; black 25–30). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

extensively dialyzed in glycine 50 mM pH 3 buffer for flash freezing with liquid nitrogen. For phage display experiments the protein was dialyzed in PBS and concentration was measured using an extinction coefficient of $44,600 \text{ M}^{-1} \cdot \text{cm}^{-1}$ at 280 nm.

4.2. Isothermal titration calorimetry

ITC experiments were performed with high-precision ITC-200 or VP-ITC calorimeters (Microcal Inc., Northampton, Massachusetts) as described before [29]. In most cases, the TSG101-UEV domain solution (40–60 μM) was titrated with peptide ligands (2 mM) dissolved in the dialysis buffer. For the natural ligands, showing lower affinity, a profile of injection volumes from 4 to 21 μL was devised to better define the titration curve. All titrations were performed in 20 mM buffer, 5 mM β -mercaptoethanol at several pH values and temperatures. The heat evolved after each peptide injection was obtained from the integral of the calorimetric signal. The heat associated with the binding process was calculated as the difference between the heat of reaction and the corresponding heat of dilution, obtained from independent titrations of the peptides into the buffer. The resulting binding isotherms were analysed

by non-linear least-squares fittings to a model corresponding to a single set of identical sites, as described before [35]. A set of additional titrations at different pH values and using buffers with different ionization enthalpies was performed to assess the impact of coupled ionization equilibria titration experiments on the binding energetics [34].

4.3. Protein crystallization and data collection

The TSG101-UEV domain and its complexes with the Late domains from the HTLV and Ebola virus (HTLV and Ebola-9 peptides) were crystallized using the vapor diffusion technique with a sitting drop set-up. Briefly, droplets were prepared with 1.5 μL of the protein (7 $\text{mg} \cdot \text{mL}^{-1}$ in 10 mM Tris buffer pH 8.0) and 1.5 μL of reservoir solution. The mixture was vapor equilibrated against 100 μL of reservoir solution. The best crystal of the free TSG101-UEV domain diffracted to 2.0 \AA resolution and belonged to the trigonal space group P321, with unit-cell parameters $a = b = 169.74$, $c = 39.71$, $\alpha = \beta = 90.00$ and $\gamma = 120.00$. These crystals were obtained using as precipitant solution 25 % PEG4000, 0.2 M ammonium sulfate, 0.1 M sodium acetate at pH 4.6. To crystallize the complexes TSG101-UEV at 8 mg/mL was mixed with the

corresponding peptide solutions at a 1:2 protein:peptide molar ratio. The TSG101-UEV/HTLV complex crystallized in the C2221 space group, with unit-cell parameters $a = 78.62$, $b = 119.74$, $c = 41.64$, $\alpha = \beta = \gamma = 90.00$ and diffracted to 2.4 \AA resolution. The best crystals grew in 30 % PEG4000, 0.1 M ammonium sulfate, 0.1 M MES at pH 6.0. The TSG101-UEV/Ebola9 complex was crystallized in 0.1 M ammonium sulfate, 25 % PEG 4 K, 0.1 M Hepes at pH 7.0. These crystals belong to the I4 space group with unit-cell parameters $a = b = 105.54$, $c = 75.30$, $\alpha = \beta = \gamma = 90.00$ and diffracted to 2.0 \AA .

During data collection, crystals were placed in a cold nitrogen stream maintained at 110 K. X-ray diffraction of the TSG101/Ebola complex, TSG101-HTLV complex, and the free TSG101-UEV domain were collected at the beamlines BM16, ID29-2, and ID23-1 of the European Synchrotron Radiation Facility (ESRF), respectively. Data were indexed and processed with the software XDS [46] in the autoPROC toolbox (Vonrhein et al., 2011). Data were scaled using the package Aimless [47] from the CCP4 suite [48]. Statistics are shown in Table 2.

4.4. Structure refinement and model building

The structures of the TSG101-UEV free domain and HTLV and Ebola complexes were obtained using the PHENIX suite [49]. Molecular-replacement phasing using the AutoMR feature of PHENIX [50] was performed with the coordinates of the TSG101-UEV domain (2FOR [20]). The final model was obtained after several cycles of manual building performed using the resulting A-weighted 2FO-Fc and FO-Fc electron-density maps in COOT [51,52]. Water molecules were modelled automatically using phenix.refine in PHENIX [50] and manually inspected in difference electron-density maps in COOT. In the final rounds of refinement, some molecules belonging to the precipitant solution were modelled. As the resolution is $>1.7 \text{ \AA}$, individual isotropic B-factor and TLS refinement were applied [48]. The final models were validated using MolProbity and PDB-REDO [53,54]. Structure solution and refinement statistics are shown in Table 2. The atomic coordinates of all structures have been deposited in the Protein Data Bank (PDB) (see Table 2).

4.5. Phage display

Peptide-phage selections were performed using a constrained library with a fixed P(T/S)AP motif flanked by five random amino acids at N- and C-termini (x_5 -P(T/S)AP- x_5). The library was fused to the gene-3 major coat protein of M13 phage and contained $>10^{10}$ unique members. Binding selections were performed in 96-well Maxisorp plates using 6 wells (for the first round) or 3 wells (for the rest of the rounds) previously coated overnight at $4 \text{ }^\circ\text{C}$ with GST-TSG101-UEV at $10 \text{ }\mu\text{g/mL}$. Selections were performed through five rounds as described [55] and enrichment was followed by phage titration after each round showing a plateau at the third round of selection. Hence, individual clones from rounds 3 and 4 were subjected to phage ELISA to discriminate specific binding using as control BSA, as previously described [56]. The thirty individual clones that raised the highest absorbance values in the ELISA were sequenced.

A second set of peptide-phage selections was performed including a competition step. The same x_5 -P(T/S)AP- x_5 library was freshly generated and a cycle of two rounds of enrichment was performed. For the third round, before adding the phage pool GST-TSG101-UEV was incubated for 30 min with PD1 at saturating concentrations ($100 \text{ }\mu\text{M}$). Up to round 5, three cycles were performed with this competition step, again following the enrichment by phage titration. The maximum enrichment was found at round 5, and 96 individual clones from that round were analysed by ELISA twice. The first ELISA was regularly performed and the second, using the same clones, was done including the competition step with PD1 as described for the phage selections. The twelve clones with their highest ELISA ratios were sequenced and aligned for analysis to decipher a consensus sequence.

4.6. Construction of structural models for TSG101-UEV in complex with the phage display peptides

The crystal structure of the TSG101-UEV/Ebola Late domain complex (4EJE) was used as a starting point for model construction. First, all ligand atoms and extraneous molecules were removed from the structure. The resulting protein pdb file was preprocessed with Maestro (Schrödinger Release 2022-1: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2021. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2021. Maestro), with the Protein Preparation Wizard and System builder tools, employing the OPLS3e forcefield [57]. Next, each peptide sequence was inputted into Maestro to generate a corresponding 3D structure for peptide modelling. Bond orders were assigned, and the LigPrep task was executed using the OPLS3e forcefield to incorporate appropriate ionization and partial charges. Finally, the generated peptide structures were meticulously aligned with the natural peptide of the original pdb structure to secure an initial position close to the binding site. This rigorous model-building process ensures the accuracy of our computational predictions and the reliability of our findings.

4.7. Molecular dynamics simulations

Molecular dynamics simulations were conducted with the Gromacs ver2018.2 suite [42]. Topology files for the TSG101-UEV complex and each peptide were generated using the gmxdump tool for the target protein and acpype (<https://doi.org/10.1186/1756-0500-5-367>) for the complex topology. Prior to the molecular dynamics (MD) simulation several preparatory steps were executed. These included box generation, solvation procedures, a 20 ps minimization step for system optimization, a 400 ps nVT equilibration step for system stabilization at constant volume and temperature, and five 400 ps NPT equilibration steps for system equilibration at constant number of particles, pressure, and temperature. The MD simulation was then carried out for 100 ns to investigate the system dynamics.

Upon completion of the simulation, various analyses were conducted to examine the interactions between proteins and peptides and the potential conformational changes observed in each case. We employed the gmxdump and gmxdump tools (<https://pubs.acs.org/doi/10.1021/ct7000045>) to determine the nonbonded interactions (Coulomb and Lennard-Jones forces), as well as hydrogen bonding, for every residue within the protein. Cluster studies were performed for each complex with gmxdump (Gmx Cluster-GROMACS 2018. GROMACS User Guide—Gmx Cluster: <http://manual.gromacs.org/documentation/2018/onlinehelp/gmx-cluster.html> (accessed on 1 May 2020), providing a detailed analysis of the five most prominent conformations. The water occupancy calculation was carried out using Bridge2 software [58] to assess the occupancy and endurance time of each interaction of interest.

4.8. Structure-based analysis of intramolecular interaction networks

All calculations were performed using the crystal structure of the free TSG101-UEV domain (2FOR), the crystal structure of the TSG101-UEV/Ebola complex (4EJE, chains AC), and a representative structure of the most populated cluster (accounting for 50 % of the simulation time) obtained from the MD trajectory for PD5 ligand.

The extent to which each residue in the TSG101-UEV domain is energetically coupled to its neighbours (up to a distance of $12\text{--}15 \text{ \AA}$) was calculated using the pPerturb algorithm [43], which provides a statistical thermodynamic calculation of the extent of propagation of a mutational perturbation (side chain truncation to alanine/glycine) into the protein structure using empirical relations derived from large-scale mutational analysis, and focusing primarily on van der Waals interactions (electrostatic and conformational effects are not considered). The interaction-network profiles for the free TSG101-UEV and the Ebola and PD5 complexes were calculated perturbing all residues in the UEV

domain. The extent to which the perturbation percolates into the interaction network (coupling distance, d_c) and the overall perturbation magnitude (Sum ΔQ) were obtained for the different structures. Residues with Sum ΔQ values >450 were considered as allosteric hotspots.

ProteinLens [44] calculates bond-to-bond propensities from an energy-weighted atomistic graph description of the protein structure, which are presented as statistical quantile scores that account for distance bias and allow the identification of those residues showing a statistically significant higher propensity. In the case of the complex structures both chains (protein and ligand) were considered for graph construction. Two sets of bond-to-bond propensities were calculated using the default settings: one considering residues in the ligand as the perturbation source and another perturbing TSG101-UEV residues conforming the PTAP binding site (D34, Y63, Y68, N66, I70, T92, M95, V141, F142, R144). Residues with quantile scores >0.9 were considered as allosteric hotspots.

WebPSN [45] generates information about structural communication by combining a graph-based description of the high-resolution structure with information on system dynamics (cross-correlation of atomic motions) derived by ENM-NMA. This algorithm provides a detailed description of the communication networks, providing information on nodes, links, hubs (nodes connected by at least four links) as well as communication pathways (metapaths), providing a global picture of the structural communication within the protein. In the case of the complex structures both chains (protein and ligand) were considered for the calculations. Residues identified as hubs and showing interaction forces higher than 3 were selected as allosteric hotspots.

Funding

This work was supported by the MCIN/AEI/10.13039/501100011033/and FEDER Una manera de hacer Europa [grant numbers BIO2016-78746-C2-1-R, PID2020-112895RB-I00]; the FEDER/Junta de Andalucía-Consejería de Economía y Conocimiento [grant number CV20-19149]; and the Consejería de Universidad, Investigación e Innovación and FEDER Andalusian Program 2021-2027 [grant number C-EXP-295-UGR23].

CRedit authorship contribution statement

Javier Murciano-Calles: Writing – review & editing, Resources, Methodology, Investigation. **Alejandro Rodríguez-Martínez:** Writing – review & editing, Investigation. **Andrés Palencia:** Writing – review & editing, Investigation. **Montserrat Andújar-Sánchez:** Writing – review & editing, Investigation. **Manuel Iglesias-Bexiga:** Writing – review & editing, Investigation. **Carles Corbi-Verge:** Investigation. **Pedro Buzón:** Investigation. **Javier Ruiz-Sanz:** Writing – review & editing, Methodology, Investigation. **Jose C. Martínez:** Writing – review & editing, Resources, Methodology. **Horacio Pérez-Sánchez:** Writing – review & editing, Validation, Software, Resources, Methodology, Formal analysis. **Ana Cámara-Artigas:** Writing – review & editing, Validation, Resources, Methodology, Formal analysis. **Irene Luque:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no conflicts of interest with the contents of this article.

Data availability statement

The X-ray structures and diffraction data reported in this paper are deposited in the Protein Data Bank under the accession codes 4ZNY, 4EJE, 4YC1.

Acknowledgements

Supercomputing resources in this work have been supported by the Plataforma Andaluza de Bioinformática of the University of Málaga, by the supercomputing infrastructure of the NLHPC (ECM-02, Powered@NLHPC), and by the Extremadura Research Centre for Advanced Technologies (CETA–CIEMAT), funded by the European Regional Development Fund (ERDF). CETA–CIEMAT belongs to CIEMAT and the Government of Spain. We thank the support of the C.I.C. of the University of Granada. J.M.C. acknowledges a reincorporation research contract from the University of Granada.

Appendix A. Supplementary data

1.- Details on the assessment of the impact of coupled ionization equilibria on the binding energetics of viral Late domains to TSG101-UEV. 2.- Details on the Computational study of the impact of PTAP ligand binding on TSG101-UEV intramolecular communication networks. Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijbiomac.2024.133233>.

References

- [1] M. Vietri, M. Radulovic, H. Stenmark, The many functions of ESCRTs, *Nat. Rev. Mol. Cell Biol.* 21 (1) (2020) 25–42.
- [2] L. Lemus, V. Goder, Membrane trafficking: ESCRTs act here, there, and everywhere, *Curr. Biol.* 32 (6) (2022) R292–R294.
- [3] A. Calistri, A. Reale, G. Palu, C. Parolin, Why cells and viruses cannot survive without an ESCRT, *Cells* 10 (3) (2021).
- [4] Q. Lu, L.W. Hope, M. Brasch, C. Reinhard, S.N. Cohen, TSG101 interaction with HRS mediates endosomal trafficking and receptor down-regulation, *Proc. Natl. Acad. Sci. U. S. A.* 100 (13) (2003) 7626–7631.
- [5] B. Meng, A.M.L. Lever, The interplay between ESCRT and viral factors in the enveloped virus life cycle, *Viruses* 13 (2) (2021).
- [6] Y. Rivera-Cuevas, V.B. Carruthers, The multifaceted interactions between pathogens and host ESCRT machinery, *PLoS Pathog.* 19 (5) (2023) e1011344.
- [7] I. Ahmed, Z. Akram, H.M.N. Iqbal, A.L. Munn, The regulation of endosomal sorting complex required for transport and accessory proteins in multivesicular body sorting and enveloped viral budding - an overview, *Int. J. Biol. Macromol.* 127 (2019) 1–11.
- [8] R.M. Ferraiuolo, K.C. Manthey, M.J. Stanton, A.A. Triplett, K.U. Wagner, The multi-faceted roles of the tumor susceptibility gene 101 (TSG101) in normal development and disease, *Cancers (Basel)* 12 (2) (2020).
- [9] K.M. Rose, When in need of an ESCRT: the nature of virus assembly sites suggests mechanistic parallels between nuclear virus egress and retroviral budding, *Viruses* 13 (6) (2021).
- [10] E.O. Freed, Viral late domains, *J. Virol.* 76 (10) (2002) 4679–4687.
- [11] P.D. Bieniasz, Late budding domains and host proteins in enveloped virus release, *Virology* 344 (1) (2006) 55–63.
- [12] L. Welker, J.C. Paillart, S. Bernacchi, Importance of viral late domains in budding and release of enveloped RNA viruses, *Viruses-Basel* 13 (8) (2021).
- [13] Z.Y. Han, J.H. Lu, Y.L. Liu, B. Davis, M.S. Lee, M.A. Olson, G. Ruthel, B. D. Freedman, M.J. Schnell, J.E. Wrobel, A.B. Reitz, R.N. Harty, Small-molecule probes targeting the viral PPxY-host Nedd4 interface block egress of a broad range of RNA viruses, *J. Virol.* 88 (13) (2014) 7294–7306.
- [14] H.M. Loughran, Z. Han, J.E. Wrobel, S.E. Decker, G. Ruthel, B.D. Freedman, R. N. Harty, A.B. Reitz, Quinoxaline-based inhibitors of Ebola and Marburg VP40 egress, *Bioorg. Med. Chem. Lett.* 26 (15) (2016) 3429–3435.
- [15] F. Castillo, C. Corbi-Verge, J. Murciano-Calles, A.M. Candel, Z. Han, M. Iglesias-Bexiga, J. Ruiz-Sanz, P.M. Kim, R.N. Harty, J.C. Martinez, I. Luque, Phage display identification of nanomolar ligands for human NEDD4-WW3: energetic and dynamic implications for the development of broad-spectrum antivirals, *Int. J. Biol. Macromol.* 207 (2022) 308–323.
- [16] K.R. Lennard, R.M. Gardner, C. Doigneaux, F. Castillo, A. Tavassoli, Development of a cyclic peptide inhibitor of the p6/UEV protein–protein interaction, *ACS Chem. Biol.* 14 (9) (2019) 1874–1878.
- [17] A. Tavassoli, Q. Lu, J. Gam, H. Pan, S.J. Benkovic, S.N. Cohen, Inhibition of HIV budding by a genetically selected cyclic peptide targeting the Gag-TSG101 interaction, *ACS Chem. Biol.* 3 (12) (2008) 757–764.
- [18] O. Pornillos, S.L. Alam, R.L. Rich, D.G. Myszka, D.R. Davis, W.I. Sundquist, Structure and functional interactions of the Tsg101 UEV domain, *EMBO J.* 21 (10) (2002) 2397–2406.
- [19] O. Pornillos, S.L. Alam, D.R. Davis, W.I. Sundquist, Structure of the Tsg101 UEV domain in complex with the PTAP motif of the HIV-1 p6 protein, *Nat. Struct. Biol.* 9 (11) (2002) 812–817.
- [20] A. Palencia, J.C. Martinez, P.L. Mateo, I. Luque, A. Cámara-Artigas, Structure of human TSG101 UEV domain, *Acta Crystallogr. D Biol. Crystallogr.* 62 (Pt 4) (2006) 458–464.

- [21] J.E. Garrus, U.K. von Schwedler, O.W. Pornillos, S.G. Morham, K.H. Zavitz, H. E. Wang, D.A. Wettstein, K.M. Stray, M. Cote, R.L. Rich, Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding, *Cell* 107 (1) (2001) 55–65.
- [22] D.J. Katzmann, M. Babst, S.D. Emr, Ubiquitin-dependent sorting into the multivesicular body pathway requires the function of a conserved endosomal protein sorting complex, ESCRT-I, *Cell* 106 (2) (2001) 145–155.
- [23] W.I. Sundquist, H.L. Schubert, B.N. Kelly, G.C. Hill, J.M. Holton, C.P. Hill, Ubiquitin recognition by the human TSG101 protein, *Mol. Cell* 13 (6) (2004) 783–789.
- [24] M. Strickland, S. Watanabe, S.M. Bonn, C.M. Camara, M.R. Starich, D. Fushman, C. A. Carter, N. Tjandra, Tsg101/ESCRT-I recruitment regulated by the dual binding modes of K63-linked diubiquitin, *Structure* 30 (2) (2022) 289–299 e6.
- [25] L.J. Ball, R. Kühne, J. Schneider-Mergener, H. Oshkinat, Recognition of proline-rich motifs by protein-protein-interaction domains, *Angew. Chem. Int. Ed. Eng.* 44 (19) (2005) 2852–2869.
- [26] J.C. Martinez, F. Castillo, J. Ruiz-Sanz, J. Murciano-Calles, A. Camara-Artigas, I. Luque, Understanding binding affinity and specificity of modular protein domains: a focus in ligand design for the polyproline-binding families, *Adv. Protein Chem. Struct. Biol.* 130 (2022) 161–188.
- [27] A. Palencia, A. Camara-Artigas, M.T. Pisabarro, J.C. Martinez, I. Luque, Role of interfacial water molecules in proline-rich ligand recognition by the Src homology 3 domain of Abl, *J. Biol. Chem.* 285 (4) (2010) 2823–2833.
- [28] J.M. Martin-García, J. Ruiz-Sanz, I. Luque, Interfacial water molecules in SH3 inter- actions: a revised paradigm for polyproline recognition, *Biochem. J.* 442 (2) (2012) 443–451.
- [29] M. Iglesias-Bexiga, A. Palencia, C. Corbi-Verge, P. Martin-Malpartida, F.J. Blanco, M.J. Macias, E.S. Cobos, I. Luque, Binding site plasticity in viral PPxY late domain recognition by the third WW domain of human NEDD4, *Sci. Rep.* 9 (1) (2019) 15076.
- [30] Y.J. Im, L. Kuo, X. Ren, P.V. Burgos, X.Z. Zhao, F. Liu, T.R. Burke Jr., J. S. Bonifacino, E.O. Freed, J.H. Hurley, Crystallographic and functional analysis of the ESCRT-I/HIV-1 gag PTAP interaction, *Structure* 18 (11) (2010) 1536–1547.
- [31] F. Bouamr, J.A. Melillo, M.Q. Wang, K. Nagashima, M. de Los Santos, A. Rein, S. P. Goff, PPPYEPTAP motif is the late domain of human T-cell leukemia virus type 1 gag and mediates its functional interaction with cellular proteins Nedd4 and Tsg101, *J. Virol.* 77 (22) (2003) 11882–11895.
- [32] A. Sakurai, J. Yasuda, H. Takano, Y. Tanaka, M. Hatakeyama, H. Shida, Regulation of human T-cell leukemia virus type 1 (HTLV-1) budding by ubiquitin ligase Nedd4, *Microbes Infect.* 6 (2) (2004) 150–156.
- [33] T. Irie, J.M. Licata, J.P. McGettigan, M.J. Schnell, R.N. Harty, Budding of PPxY-containing rhabdoviruses is not dependent on host proteins TGS101 and VPS4A, *J. Virol.* 78 (6) (2004) 2657–2665.
- [34] A. Velazquez-Campoy, I. Luque, M.J. Todd, M. Milutinovich, Y. Kiso, E. Freire, Thermodynamic dissection of the binding energetics of KNI-272, a potent HIV-1 protease inhibitor, *Protein Sci.* 9 (9) (2000) 1801–1809.
- [35] A. Palencia, E.S. Cobos, P.L. Mateo, J.C. Martinez, I. Luque, Thermodynamic dissection of the binding energetics of proline-rich peptides to the Abl-SH3 domain: implications for rational ligand design, *J. Mol. Biol.* 336 (2) (2004) 527–537.
- [36] J.C. Ferreón, V.J. Hilsner, Thermodynamics of binding to SH3 domains: the energetic impact of polyproline II (PII) helix formation, *Biochemistry* 43 (24) (2004) 7787–7797.
- [37] C. Wang, N.H. Pawley, L.K. Nicholson, The role of backbone motions in ligand binding to the c-Src SH3 domain, *J. Mol. Biol.* 313 (4) (2001) 873–887.
- [38] S. Arold, R. O'Brien, P. Franken, M.P. Strub, F. Hoh, C. Dumas, J.E. Ladbury, RT loop flexibility enhances the specificity of Src family SH3 domains for HIV-1 Nef, *Biochemistry* 37 (42) (1998) 14683–14691.
- [39] I. Luque, E. Freire, Structure-based prediction of binding affinities and molecular design of peptide ligands, *Methods Enzymol.* 295 (1998) 100–127.
- [40] I. Luque, E. Freire, Structural parameterization of the binding enthalpy of small ligands, *Proteins* 49 (2) (2002) 181–190.
- [41] E. Krieger, G. Vriend, YASARA view - molecular graphics for all devices - from smartphones to workstations, *Bioinformatics* 30 (20) (2014) 2981–2982.
- [42] C. Kutzner, S. Pall, M. Fechner, A. Esztermann, B.L. de Groot, H. Grubmüller, More bang for your buck: improved use of GPU nodes for GROMACS 2018, *J. Comput. Chem.* 40 (27) (2019) 2418–2431.
- [43] S. Gopi, D. Devanshu, N. Rajasekaran, S. Anantkrishnan, A.N. Naganathan, pPerturb: a server for predicting long-distance energetic couplings and mutation-induced stability changes in proteins via perturbations, *ACS Omega* 5 (2) (2020) 1142–1146.
- [44] S.F. Mersmann, L. Stromich, F.J. Song, N. Wu, F. Vianello, M. Barahona, S. N. Yaliraki, ProteinLens: a web-based application for the analysis of allosteric signalling on atomistic graphs of biomolecules, *Nucleic Acids Res.* 49 (W1) (2021) W551–W558.
- [45] A. Felline, M. Seeber, F. Fanelli, webPSN v2.0: a webserver to infer fingerprints of structural communication in biomacromolecules, *Nucleic Acids Res.* 48 (W1) (2020) W94–W103.
- [46] W. Kabsch, Xds, *Acta Crystallogr. D Biol. Crystallogr.* 66 (Pt 2) (2010) 125–132.
- [47] P.R. Evans, An introduction to data reduction: space-group determination, scaling and intensity statistics, *Acta Crystallogr. D Biol. Crystallogr.* 67 (Pt 4) (2011) 282–292.
- [48] M.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, R. M. Keegan, E.B. Krissinel, A.G. Leslie, A. McCoy, S.J. McNicholas, G.N. Murshudov, N.S. Pannu, E.A. Potterton, H.R. Powell, R.J. Read, A. Vagin, K.S. Wilson, Overview of the CCP4 suite and current developments, *Acta Crystallogr. D Biol. Crystallogr.* 67 (Pt 4) (2011) 235–242.
- [49] P.D. Adams, P.V. Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, J. J. Headd, L.W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N. W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T. C. Terwilliger, P.H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr. D Biol. Crystallogr.* 66 (Pt 2) (2010) 213–221.
- [50] P.V. Afonine, R.W. Grosse-Kunstleve, N. Echols, J.J. Headd, N.W. Moriarty, M. Mustyakimov, T.C. Terwilliger, A. Urzhumtsev, P.H. Zwart, P.D. Adams, Towards auto- mated crystallographic structure refinement with phenix.refine, *Acta Crystallogr. D Biol. Crystallogr.* 68 (Pt 4) (2012) 352–367.
- [51] P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics, *Acta Crystallogr. D Biol. Crystallogr.* 60 (Pt 12 Pt 1) (2004) 2126–2132.
- [52] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, *Acta Crystallogr. D Biol. Crystallogr.* 66 (Pt 4) (2010) 486–501.
- [53] C.J. Williams, J.J. Headd, N.W. Moriarty, M.G. Prisant, L.L. Videau, L.N. Deis, V. Verma, D.A. Keedy, B.J. Hintze, V.B. Chen, S. Jain, S.M. Lewis, W. B. Arendall 3rd, J. Snoeyink, P.D. Adams, S.C. Lovell, J.S. Richardson, D. C. Richardson, MolProbity: more and better reference data for improved all-atom structure validation, *Protein Sci.* 27 (1) (2018) 293–315.
- [54] R.P. Joosten, F. Long, G.N. Murshudov, A. Perrakis, The PDB_REDO server for macromolecular structure model optimization, *IUCrJ* 1 (Pt 4) (2014) 213–220.
- [55] R. Tonikian, Y. Zhang, C. Boone, S.S. Sidhu, Identifying specificity profiles for peptide recognition modules from phage-displayed peptide libraries, *Nat. Protoc.* 2 (6) (2007) 1368–1386.
- [56] K. Deshayes, M.L. Schaffer, N.J. Skelton, G.R. Nakamura, S. Kadkhodayan, S. S. Sidhu, Rapid identification of small binding motifs with high-throughput phage display: discovery of peptidic antagonists of IGF-1 function, *Chem. Biol.* 9 (4) (2002) 495–505.
- [57] K. Roos, C. Wu, W. Damm, M. Reboul, J.M. Stevenson, C. Lu, M.K. Dahlgren, S. Mondal, W. Chen, L. Wang, R. Abel, R.A. Friesner, E.D. Harder, OPLS3e: extending force field coverage for drug-like small molecules, *J. Chem. Theory Comput.* 15 (3) (2019) 1863–1874.
- [58] M. Siemers, A.N. Bondar, Interactive interface for graph-based analyses of dynamic H-bond networks: application to spike protein S, *J. Chem. Inf. Model.* 61 (6) (2021) 2298–3014.