Biochemistry and Biophysics Reports 39 (2024) 101770

Contents lists available at ScienceDirect

Biochemistry and Biophysics Reports

journal homepage: www.elsevier.com/locate/bbrep

Multiple types of nuclear localization signals in *Entamoeba histolytica*

Israel Canela-Pérez [a], Elisa Azuara-Liceaga [b], Patricia Cuéllar [b], Odila Saucedo-Cárdenas [c], Jesús Valdés [a,*]

[a] Departamento de Bioquímica, CINVESTAV-México, Av. IPN 2508 colonia San Pedro Zacatenco, GAM, CDMX, 07360, Mexico
[b] Posgrado en Ciencias Genómicas, Universidad Autónoma de la Ciudad de México, Mexico City, 03100, Mexico
[c] Departamento de Histología, Facultad de Medicina, Universidad Autónoma de Nuevo León, Monterrey, 67700, Mexico

A B S T R A C T

Entamoeba histolytica is a protozoan parasite that belongs to the Amoebozoa supergroup whose study related to the nucleocytoplasmic transport of proteins through the nucleus is poorly studied. In this work, we have performed *in silico* predictions of the potential nuclear localization signals (NLS) corresponding to the proteome of 8201 proteins from *Entamoeba histolytica* annotated in the AmoebaDB database. We have found the presence of monopartite nuclear localization signals (MNLSs), bipartite nuclear localization signals (BNLSs), and non-canonical monopartite NLSs with lengths exceeding 20 amino acid residues. Additionally, we detected a new type of NLS consisting of multiple juxtaposed bipartite NLSs (JNLSs) that have not been described in any eukaryotic organism. Also, we have generated consensus sequences for the nuclear import of proteins with the NLSs obtained. Docking experiments between EhImportin α and an MNLS, BNLS, and JNLS outlined the interacting residues between the Importin and cargo proteins, emphasizing their putative roles in nuclear import. By transfecting HA-tagged protein constructs, we assessed the nuclear localization of MNLS (U1A and U2AF1), JMNLS (U2AF2), and non-canonical NLS (N-terminus of Pol ll) *in vivo*. Our data provide the basis for understanding the nuclear transport process in *E. histolytica*.

1. Introduction

Entamoeba histolytica is a unicellular eukaryote belonging to the protist kingdom and forms part of the phylum Amoebozoa [1]. As a biological model for this project, we used the protozoan parasite *E. histolytica,* which is responsible for amoebic dysentery, or amebiasis, and is considered a global health problem. In nature, *E. histolytica* experiences a complex life cycle involving a vertebrate host, mainly transmitted by ingesting contaminated food or water. Ninety percent of infected individuals are asymptomatic, keeping the parasites in the colon's lumen, while the remaining 10 % are symptomatic [2]. Approximately 50 million people globally contract the infection, with over 100, 000 deaths due to amebiasis reported annually [3]. Only in the year 2022 in México, more than 162,000 cases of intestinal amebiasis and 428 cases of liver abscesses caused mainly by this protozoan parasite were reported in the national epidemiological magazine. This microorganism presents the typical characteristics of eukaryotic cells. However, it presents particularities, such as the mitosomes, whose function is to carry out the bioenergetic synthesis of the parasite through the sulfate activation pathway [4–7].

More than three decades have passed since the first studies of nuclear transport of proteins were carried out by Kalderon and co-workers (1984), in which they studied Ag-T large from SV-40 and described the first MNLS consisting of seven amino acid residues ($_{126}$PKKKRKV$_{132}$) sufficient to translocate the protein into the nucleoplasm of mammalian cells [8]. In 1988, Dingwall described the first BNLS of *Xenopus laevis* nucleoplasmin ($_{155}$KRPAATKKAGQAKKKK$_{170}$) [9], in which two Lysine/Arginine enriched motifs are separated by a spacer sequence not involved in the nuclear import process [10,11].

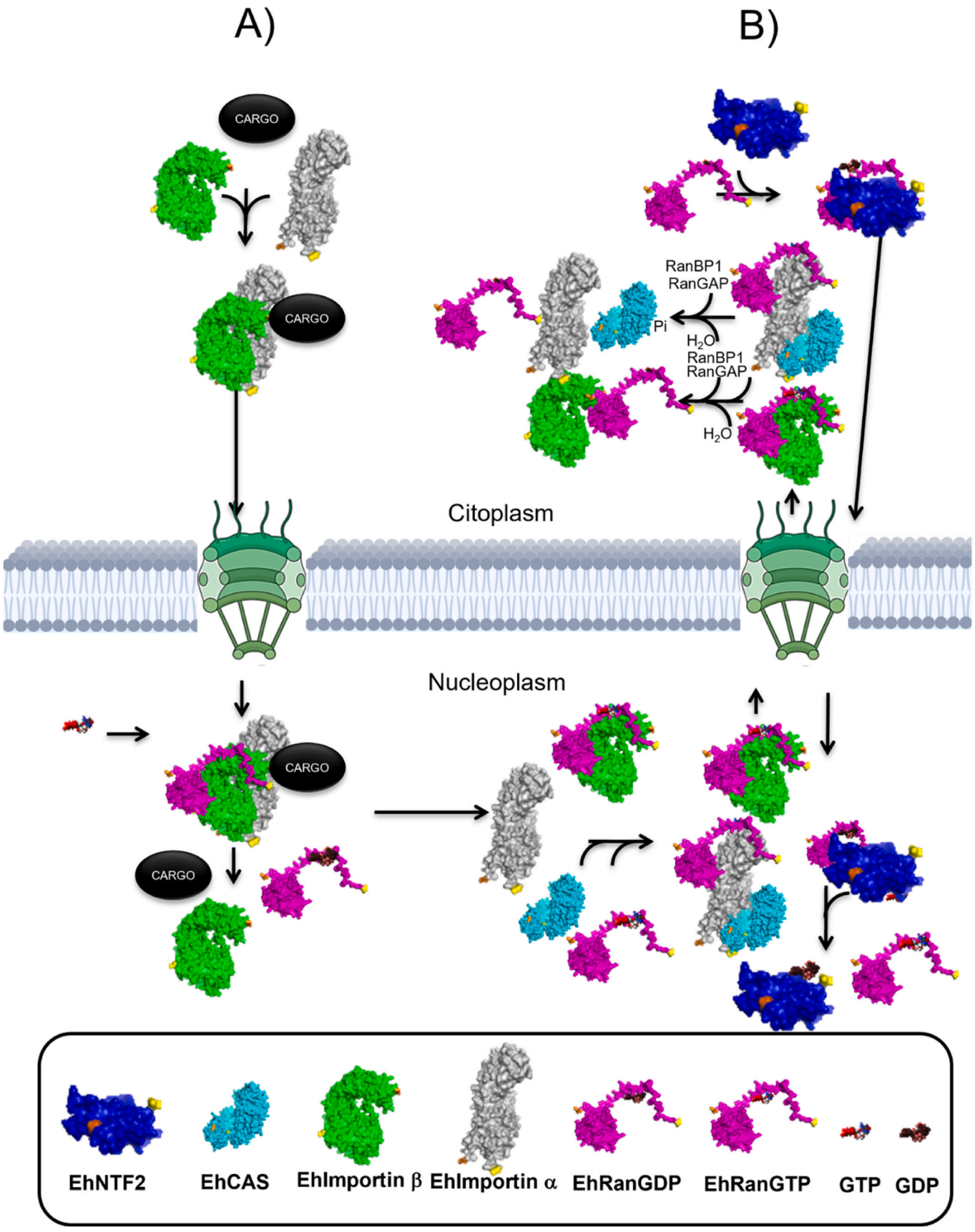
The classic nuclear transport pathway is a complex process that occurs through nuclear pore complexes (NPCs) involving proteins of the Karyopherin family, namely importins and exportins [12,13]. The translocation of proteins to the nucleoplasm mediated by Importin α and Importin β is the classical pathway of nuclear import of proteins [14]. The nuclear (cargo) proteins have protein motifs known as nuclear localization signals (NLS) to be transported from the cytoplasm. Many NLSs are recognized by the adaptor Importin α, forming the cargo/Importin α complex. Subsequently, the receptor protein Importin β binds to the IBB (Importin β binding domain) of Importin α, forming the cargo/Importin α/Importin β trimeric transport complex [15,16]. The trimeric complex is translocated to the nucleoplasm through the NPC (nuclear pore complex); this process is mediated through the interaction of Importin β with protein domains enriched with phenylalanine and glycine (FG) of NPC proteins [17]. Once the trimeric complex is in the nucleoplasm, a GTP binding protein called Ran-GTP (belonging to the RAS family of proteins) binds to Importin β, generating an allosteric

* Corresponding author.
E-mail address: jvaldes@cinvestav.mx (J. Valdés).

https://doi.org/10.1016/j.bbrep.2024.101770
Received 11 March 2024; Received in revised form 20 June 2024; Accepted 27 June 2024

Table 1
Accession numbers of the coding genes (and orthologs) for the main nuclear import and export transport factors in *E. histolytica* strain HM-1:IMSS, *E. dispar* strain SAW760, *E. invadens* strain IP1, *E. nuttalli* strain P19, *Trypanosoma cruzi* strain CLB, *T. brucei* strain TREU927, *S. cerevisiae,* and human.

Nuclear transport factor	E. histolytica	E. dispar	E. invadens	E. nuttalli	T. cruzi	T.brucei	S. cerevisiae	H. sapiens
Importin alpha	EHI_025350A EHI_179940A	EDI_165620A EDI_070520A EDI_256150A	EIN_065430-t26_1 EIN_018710-t26_1 EIN_087260-t26_1	ENU1_178740-t26_1 ENU1_126380-t26_1 ENU1_162570-t26_1	TcCLB.509965.110 TcCLB.509057.20	Tb927.6.2640	YNL189W	NM_002264.3
Importin beta	EHI_036520A	EDI_008990A	EIN_084990-t26_1	ENU1_099150-t26_1	TcCLB.504019.9 TcCLB.511529.9 TcCLB.511033.10 TcCLB.511031.49	Tb927.11.16340	YLR347C	NM_002265.5
Ran	EHI_148190A	EDI_335160A	EIN_026930-t26_1	ENU1_095110-t26_1	TcCLB.509455.80 TcCLB.503539.30	Tb927.3.1120	YLR293C	NM_001300797.1
NTF2	EHI_182470A	EDI_251900A	EIN_053790-t26_1	Absent	TcCLB.511407.10 TcCLB.509567.40	Tb927.8.4280	YER009W	NM_005796.3
CAS	EHI_164410A	EDI_085690A	EIN_152870-t26_1	ENU1_106980-t26_1	TcCLB.511725.150	Tb927.11.14340	YGR218W	NM_003400.4
RanGEF	EHI_049610A EHI_109960A	EDI_167680A EDI_063040A	EIN_030070-t26_1 EIN_023830-t26_1 EIN_403740-t26_1 EIN_408540-t26_1	ENU1_016460-t26_1 ENU1_187400-t26_1	TcCLB.509205.10	Tb927.11.11570	YLR310C	NM_001162383.2
RanGAP	EHI_124570A EHI_000430A EHI_004210A EHI_074060A EHI_079910A EHI_087740A EHI_100290A EHI_108750A EHI_137690A EHI_138500A EHI_187070A EHI_196550A	EDI_024010A EDI_009760A EDI_016520A EDI_026850A EDI_039650A EDI_065000A EDI_120050A EDI_127090A EDI_253040A EDI_275760A EDI_277700A EDI_339940A	EIN_051620-t26_1 EIN_055140-t26_1 EIN_130520-t26_1 EIN_135200-t26_1 EIN_151270-t26_1 EIN_154260-t26_1 EIN_172510-t26_1 EIN_176190-t26_1 EIN_229770-t26_1 EIN_244610-t26_1 EIN_250350-t26_1 EIN_281670-t26_1 EIN_328400-t26_1 EIN_370330-t26_1 EIN_372110-t26_1 EIN_411510-t26_1 EIN_428100-t26_1 EIN_492670-t26_1	ENU1_118310-t26_1 ENU1_049870-t26_1 ENU1_058350-t26_1 ENU1_078980-t26_1 ENU1_114980-t26_1 ENU1_131670-t26_1 ENU1_136870-t26_1 ENU1_162770-t26_1 ENU1_173900-t26_1 ENU1_175890-t26_1 ENU1_177550-t26_1 ENU1_191070-t26_1	TcCLB.506693.50	Tb927.3.2950	YMR235C	NM_002883.4
RanBP1	EHI_185290A EHI_082590A	EDI_320700A	EIN_141760-t26_1 EIN_141770-t26_1 EIN_176890-t26_1 EIN_204310-t26_1 EIN_204320-t26_1 EIN_449520-t26_1	ENU1_001580-t26_1	TcCLB.507099.30	Tb927.11.3380	YDR002W	NM_001278639.2



(caption on next page)

Fig. 1. Model for nuclear import of proteins of *E. histolytica*. Initially, a protein known as cargo contains a functional cNLS (MNLS or BNLS), identified by the nuclear transport factor Importin α (gray). Subsequently, a ternary import complex is formed when Importin β (green) recognizes the N-terminal IBB domain of Importin α. The ternary complex is transported through the nuclear pore complex from the cytoplasm to the nucleoplasm. The ternary complex is dissociated when Ran-GTP binds (purple) to Importin β, leaving Importin α and cargo in the nucleoplasm, whereas Importin β remains bound to Ran-GTP. B) Recycling of nuclear transport factors. The complex Importin β-RanGTP crosses the NPC from the nucleoplasm to the cytoplasm. The trimeric exporting complex Importin α-RanGTP-CAS can be returned to the cytoplasm. Both Importin α and Importin β are liberated once RanGTP is hydrolyzed. The exportin CAS (light blue) returns to the nucleoplasm by itself. The RanGDP produced by hydrolysis is recognized by the nuclear transport factor NTF2 (cobalt). Subsequently, RanGTP-NTF2 can cross to the nucleoplasm in which RanGTP is recharged with GTP by Ran, the chromatin-bound guanine nucleotide exchange factor (GEF). Individually in the cytoplasm, the hydrolysis of RanGTP is aided by RanGAP and RanBP1, which generates a higher gradient of RanGTP in the nucleoplasm. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

change and dissociating the trimeric complex [18]. Importin β remains bound to Ran-GTP to be exported to the cytoplasm. Thus, Ran-GTP fulfills a regulatory function of nuclear transport by maintaining a binding cycle with GTP and GDP, with a higher Ran GTP gradient in the nucleoplasm [19,20]. The binding of GTP to Ran is modulated by the guanine nucleotide exchange factor (RanGEF/RCC1) in the nucleus and Ran GTPase activating protein (Ran GAP) in the cytoplasm, both proteins distributed asymmetrically between the nucleus and cytoplasm, control the binding of Ran to GTP or GDP and consequently the regulation of nucleocytoplasmic transport [21–24].

On the other hand, Importin α is exported via the exportin CAS (Cellular apoptosis susceptibility protein)-RanGTP complex [16,25]. Once in the cytoplasm, the export complexes RanGTP/Importin α/CAS and RanGTP/Importin β are dissociated by GTP hydrolysis from RanGTP, generating RanGDP. The nuclear transport of RanGTP depends on NTF2 (nuclear transport factor 2), which has an affinity for the nucleoporin FG domains of the NPC [26,27]. Once in the nucleoplasm, NTF2-RanGDP is dissociated by RanGEF [21]. Notably, particular classes of proteins are not imported via this pathway; instead, their NLSs are recognized directly by Importin β homologs, which carry them into the nucleus in the same manner as Importin β carries Importin α. In mammals, Importin α is soluble and ubiquitous in the cell. In Trypanosomatids, recombinant versions of Importin α in *T. cruzi* have atypical nucleolar localization colocalizing with the nucleolar protein RPA31, a subunit of RNA Polymerase I [28,29]. Still, this adapter protein has not been characterized in Amoeba family members.

While nuclear transport has been analyzed in different eukaryotic models, in Amoebozoa, the classical nuclear transport pathway has yet to be studied. Despite this, in the AmoebaDB, we could identify the genes coding for the nuclear transport factors possibly participating in the canonical protein import and export pathways, namely, Importin α, Importin β, the Importin β homologs CAS and Ran, as well as NTF2, RanGAP, RanBP1, and RCC1 (Table 1). We propose a nuclear transport model with the aforementioned orthologous proteins of *E. histolytica* (Fig. 1).

Characterizing nuclear transport signals is an almost unexplored field of molecular biology in the protozoan parasite *E. histolytica*. Nuclear proteins have been described in *E. histolytica*; however, the mechanism of nuclear import, specifically the nuclear localization signals, has yet to be critically described. Seminal studies addressing nuclear transport in members of the Amoebidae family are those by Feldherr and Akin, who conjugated colloidal gold particles to the BSA protein linked to the MNLS of the SV40 T antigen and with the BNLS of the nucleoplasmin of *X. laevis* (Fig. 2A and B) to prove its nuclear translocation in *Amoeba proteus* and *Chaos carolinensis*. They found that the MNLS has a higher import efficiency than the BNLS [30]. In a similar study, the *E. histolytica* enolase was fused to the NLS of the SV40 large T antigen (Fig. 2C), and increased levels of this protein were found in the nucleus compared to the cytoplasmic localization [31]. These studies did not address the presence or participation of nuclear transport factors mediating protein import and export.

A functional study addressed the nucleocytoplasmic transport of the *E. histolytica* F-Actin binding protein EhNCABP166 (EHI_093850) harboring 3 non-juxtaposed BNLSs and a nuclear export signal (NES) at its carboxyl terminus [32,33]. While the NES favored nuclear export, the authors found that EhNCABP166 is localized in the nucleoplasm and perinuclear regions, and the same was true when each of the EhNCABP166 NLS was fused to the Ac-d100 domain of the protein (Fig. 2D).

Recently, two monopartite-type NLSs have been identified in the major subunit of RNA Polymerase II towards its amino terminus [34]. HA-tagged constructs harboring full-length Pol II or the two NLSs fused to the C-terminal domain of Pol II were expressed in amoeba transformants. Both full-length and Pol II CTD localized to the nucleus. In addition, they allowed the recovery of promoter dsDNA fragments and nuclear noncoding RNAs of *E. histolytica* through chromatin and RNA immunoprecipitation experiments (Fig. 2E).

The amino acid sequence analysis identified the putative NLS of Ehp53, whose signal is at the carboxyl end and seems to be non-canonical [35]; furthermore, it is close to a leucine-rich NES that would be analogous to the first NES reported in viral proteins that are rich in leucine [36] (Fig. 2F). Also, a computer-predicted MNLS of 5 amino acid residues towards its carboxyl terminus has been reported for the GTP exchange factor regulatory protein RanGEF called EhRhoGAPnc [37]. In contrast, at its amino terminus, it contains an NES (Fig. 2G).

The publication of the *E. histolytica* genome made all the parasite's coding information available [38]. Using this database, we conducted a whole genome computational analysis to identify canonical and non-classical NLSs. The *E. histolytica* NLSs are complex, ranging from 4 to 698 residues. In the specific case of MNLSs, they range from 4 residues to atypical lengths of 41 residues. In the case of BNLSs, they range from 17 to 21 residues. Novel BNLSs identified here were those of 17 residues in which up to 10 are juxtaposed in their mid-part. Such NLSs have not been reported in any eukaryotic organism. In addition, we observed from 2 to 26 tandem juxtapositions of the 17-residues NLSs. We undertook the computer-assisted identification of the classical pathway of nuclear transporters of *E. histolytica*, and by identifying them, we hypothesized the potential presence of nuclear localization signals in the proteome of this parasite.

2. Materials and methods

2.1. Genome-wide prediction of NLSs in E. histolytica

We downloaded the protein sequences of the 8201 accession

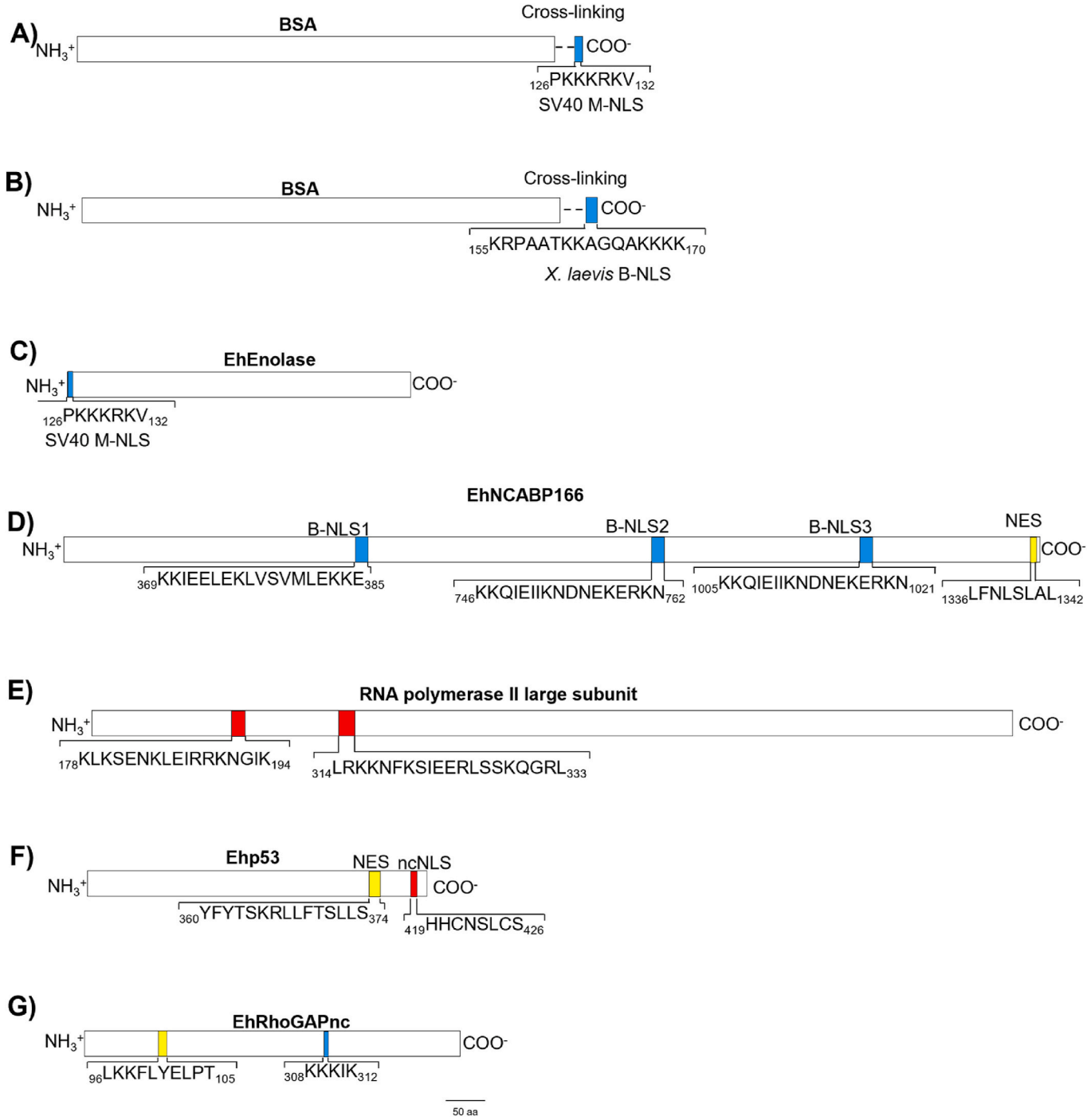


Fig. 2. Schematic representation (drawn to scale) of Amoeba nuclear proteins containing NLSs. **A-B**) Illustration of cases from the *Amoeba proteus* and *Chaos Carolinensis*; **C**) EhEnolase with an MNLS (for cNLSs, see Supplementary Fig. 1). The represented proteins have scale maps of their NLS elements. **D**) In the case of EhNCABP166 from *E. histolytica*, three motifs of basic residues, which potentially form a bipartite type of cNLS, are shown. **E**) Two non-canonical MNLSs of EhPol ll large subunit regions are shown towards its N-terminal. **F-G**) Ehp53 and EhRhoGAPnc are bioinformatic predictions of NLSs reported; however, they have not been tested experimentally. Blue and red boxes denote the positions of classical and non-classical NLS, respectively, while yellow boxes represent nuclear export signals. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Subcellular distribution of the 8201 *Entamoeba histolytica* proteins predicted with the MULocDeep software.

ORGANELLE	PROTEINS	%
NUCLEUS	3256	39.73
MITOCHONDRION	134	1.63
CYTOPLASM	2882	35.14
ER	646	7.87
SECRETET	345	4.2
MEMBRANE	787	9.59
LYSOSOME	10	0.12
GOLGI	139	1.69
PEROXISOME	2	0.03
TOTAL	8201	100

numbers in the database to avoid any bias. The protein sequences were analyzed in batches with the MULocDeep program [39] to obtain the proteins corresponding to each organelle. Subsequently, they were analyzed using the NLStradamus program [40] to obtain their putative nuclear localization signals.

The identified proteins with potential NLS were submitted to a second analysis with the PSORT II software to sort them between MNLSs and BNLSs [41]. Data was curated by individually analyzing each NLS, classifying them by type and length of amino acid residues. This led us to detect JNLSs in *E. histolytica* not previously reported visually. Finally, to predict the non-canonical NLS of RNA Polymerase II subunit RPB1 we used the software WolfPsort.

2.2. Alignment of sequences and obtaining frequency logos

Sequences of the same type and length were aligned with the CLUSTAL OMEGA program and subsequently entered into the Weblogo program to obtain the relative frequency logos (weblogo.berkeley.edu; last accessed: 06/30/2023) [42]. Only the set of NLSs of the same type with n>10 were considered to obtain the consensus sequences. The colors of the alignments corresponding to the JNLSs were made manually.

2.3. Modeling of proteins and their NLS motifs

Protein structures were obtained from the Alphafold.ebi.ac.uk database (last accessed: 06/30/2023), and their motifs were worked out using the PyMol program. One structure of each type was randomly chosen to exemplify the structures of the NLSs corresponding to each nuclear import consensus sequence. The accession numbers of the genes encoding the nuclear transport factors of *Saccharomyces cerevisiae* were obtained from the yeastgenome.org database, those of *Homo sapiens* from ncbi.nlm.nih.gov, those of *Trypanosoma cruzi* and *Trypanosoma brucei* from tritrypdb.org, while those of *E. histolytica*, *E. invadens*, *E. dispar* and *E. nuttalli* in the AmoebaDB.org database.

2.4. Docking

For the blind docking analysis, the three-dimensional structure of the JNLS of the EHI_001070 protein was initially predicted in the SWISS-MODEL [43], while the BNLS structures of the EHI_164340 protein and the MNLS of the EHI_142120 protein were predicted with RPBS Web Portal and the structure of EhImportin α (EHI_025350) was obtained with AlphaFold. Finally, the docking analysis between EhImportin α and the different NLSs was performed with the Global RAnge Molecular Matching (GRAMM) software [44]. The amino acid interactions of the docking were made with LIGPLOT [45].

2.5. Constructs, amoeba transfectants, and confocal microscopy

As previously described for U1A (EHI_050780) [46] and RNA Polymerase ll (EHI_121760) [34], the protein-coding regions of U2AF1 (EHI_192500), and U2AF2 (EHI_098300) genes were amplified by PCR using E. histolytica genomic DNA as a template and their respective primer pairs (U2AF33Nf taacagatctATGACAGAAACAACAAAAAAAGAAGAAAC, and U2AF33Nr taacagatctTTTGTTGTCATAATATCTTCTTCTTGAAG; U2AF84Nf taacggatccATGGCAGGAAGGTATGATAGGTCTCG, and U2AF 84Nr taacggatccTTCTTTTATTTCTTCTTCTGTTTTGAC). As previously described, amoeba transfectants were established and analyzed by confocal microscopy [34,46]. Ten to fourteen optical slices were obtained from the nuclei. Z-stacks were deconvoluted and 3D visualizations were generated using the IMARIS and MeshLab software as described [47].

3. Results

We analyzed the 8201 protein sequences annotated in the AmoebaDB database. We evaluated the subcellular distribution of the proteins with the MULocDeep software, obtaining 3256 putative nuclear proteins, 134 mitochondrial, 2282 cytoplasmic, 646 endoplasmic reticulum, 345 secretory, 787 membrane, 10 lysosomal, 139 from the Golgi apparatus and 2 from the peroxisome (Table 2).

To evaluate the presence of canonical and non-classical NLSs (cNLSs and ncNLSs, respectively), we used the NLStradamus software and, later, the PSORT II program. Of the 8201 proteins, 15.54 % have cNLSs (1275 proteins). Of the total number of proteins with NLS, only 34.82 % (444 proteins) exhibit a single NLS in their sequence, while the remaining 65.18 % (831 proteins) have multiple NLSs (Supplementary Fig. 2). These belong to the putative nuclear subgroup detected with the MULocDeep software.

From the 3256 proteins predicted as nuclear, we subtracted the 1275 proteins with cNLS to obtain 1981 (60.85 %) nuclear proteins that could potentially have ncNLSs (Supplementary Fig. 3). Of the 1275 proteins with NLS, 737 are hypothetical proteins (57.8 %). From the remaining 538 proteins, 165 (12.9 %) have a cytoplasmic and nucleocytoplasmic function, while 373 (29.3 %) proteins have a nuclear function (see Supplementary Table 1, which describes the functions of each protein identified with UniProt).

3.1. Consensus for nuclear import of proteins in E. histolytica

Regarding the study of the characterization of nuclear import, consensus signals of 4 residues in eukaryotic organisms have been extensively reviewed [48]. These include B_4, $B(B_2P)$ for *Trypanosoma cruzi* and *T. brucei* in which B is any basic residue (K, R or H) [49]; K (R/K)X(R/K) and KR(R/X)K for humans, adenovirus and polyomavirus [50,51]; K(K/R)X(K/R) and KRRR [52,53] or $(R/K)_4$ or $(R/K)_3(H/P)$ [41] for yeast; and KR(K/R)R or K(K/R)RK [54] for *Oryza sativa*. In *E. histolytica*, we identified the consensus of 4 residues K/R(K/RX)K/R, H $(K/R)_3$, $(K/R)_4$, and P$(K/R)_3$ (Fig. 3A–D, respectively); as well as the consensus of 5 and 6 residues K/R[$(K/R)X_2$]K/R and K/R[$(K/R)_2X_2$]K/R (Fig. 3E and F, respectively). In all formulas, the amino acids within parenthesis and brackets indicate that they can be interchangeable within (Tables 3 and 4).

Seven-residue MNLSs consensuses have been reported for

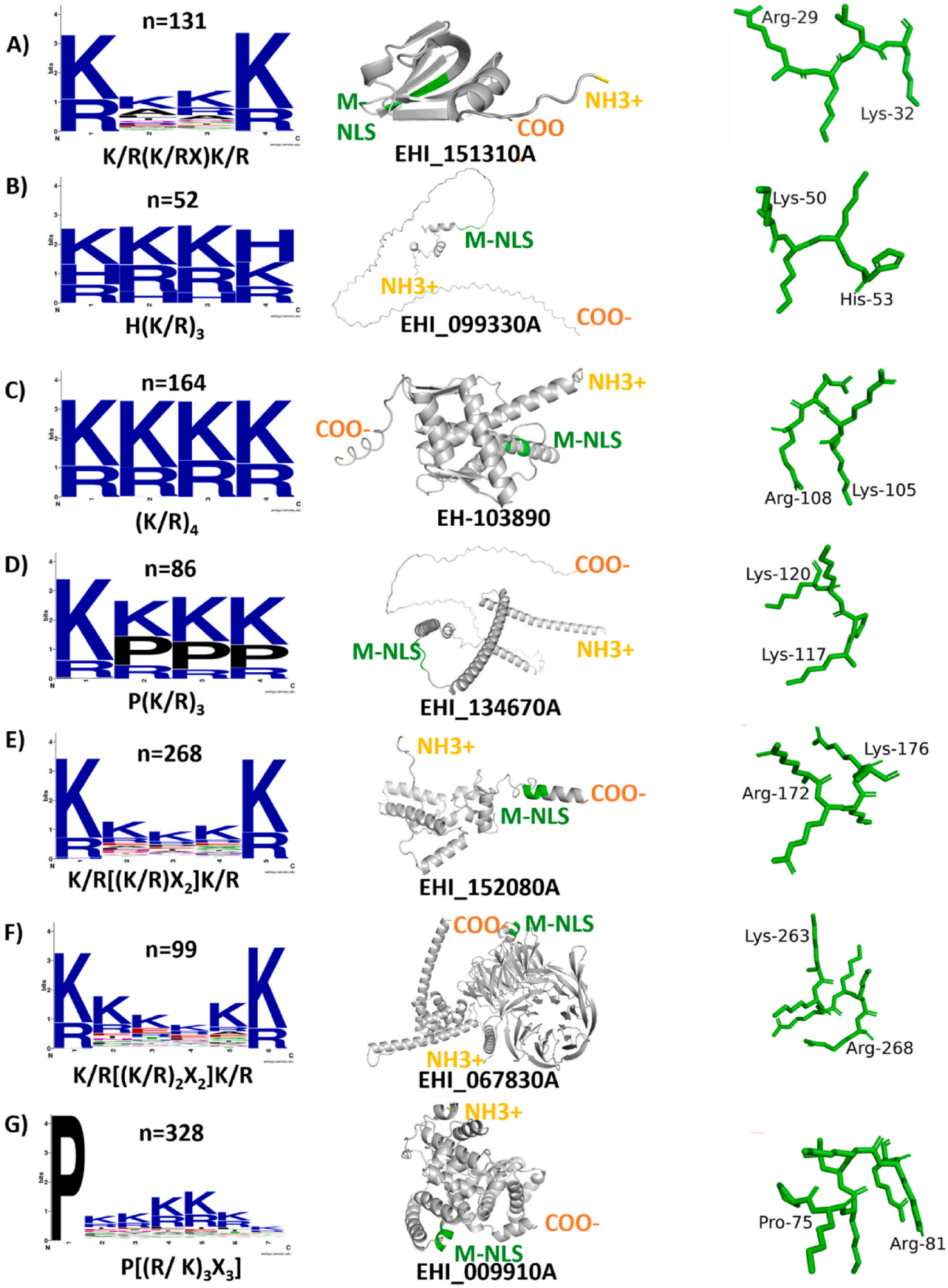


Fig. 3. Frequency Logos of MNLSs length classes from *E. histolytica.* The consensus sequence shows an example of each case's structure (in green). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

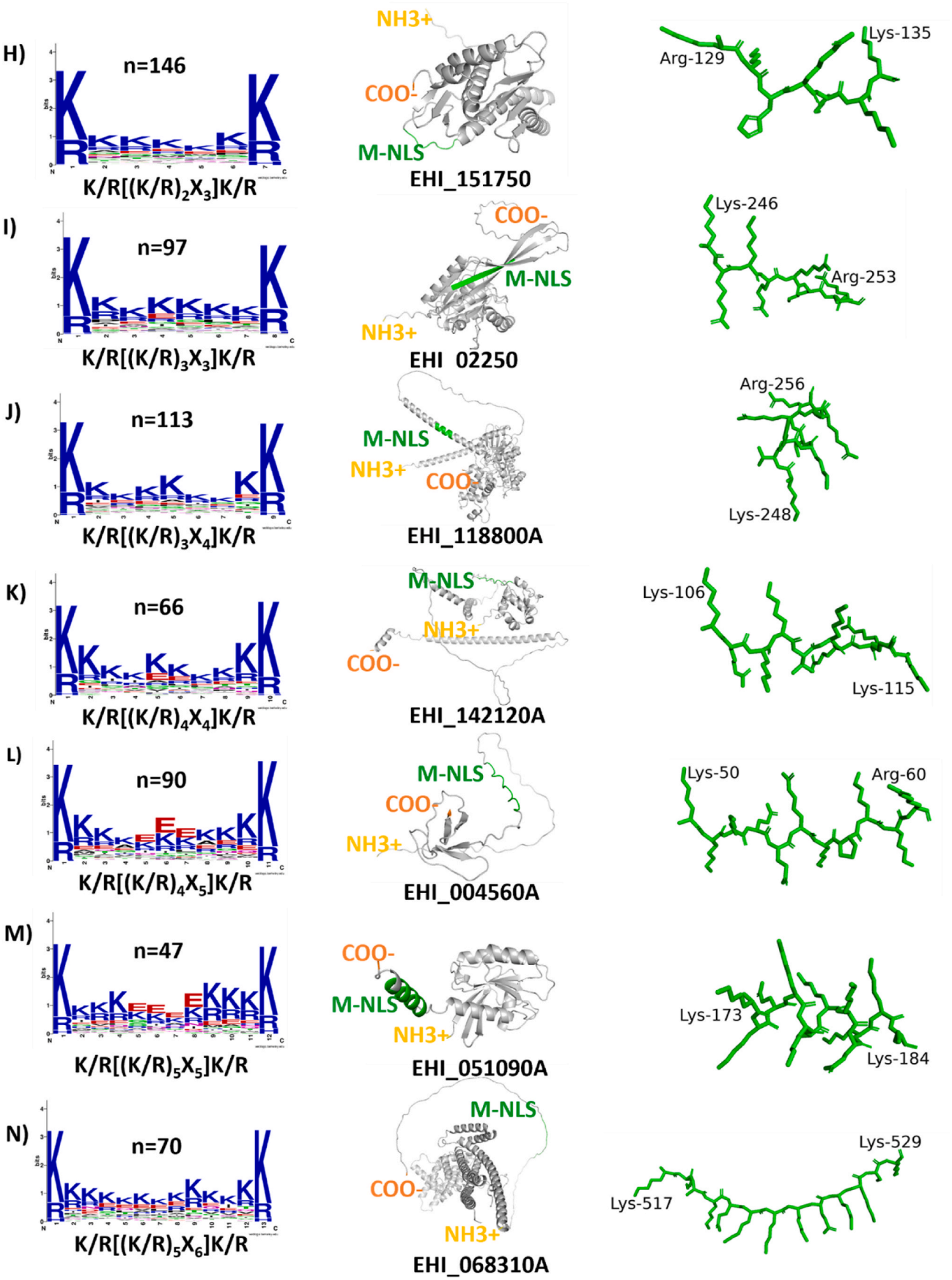


Fig. 3. (*continued*).

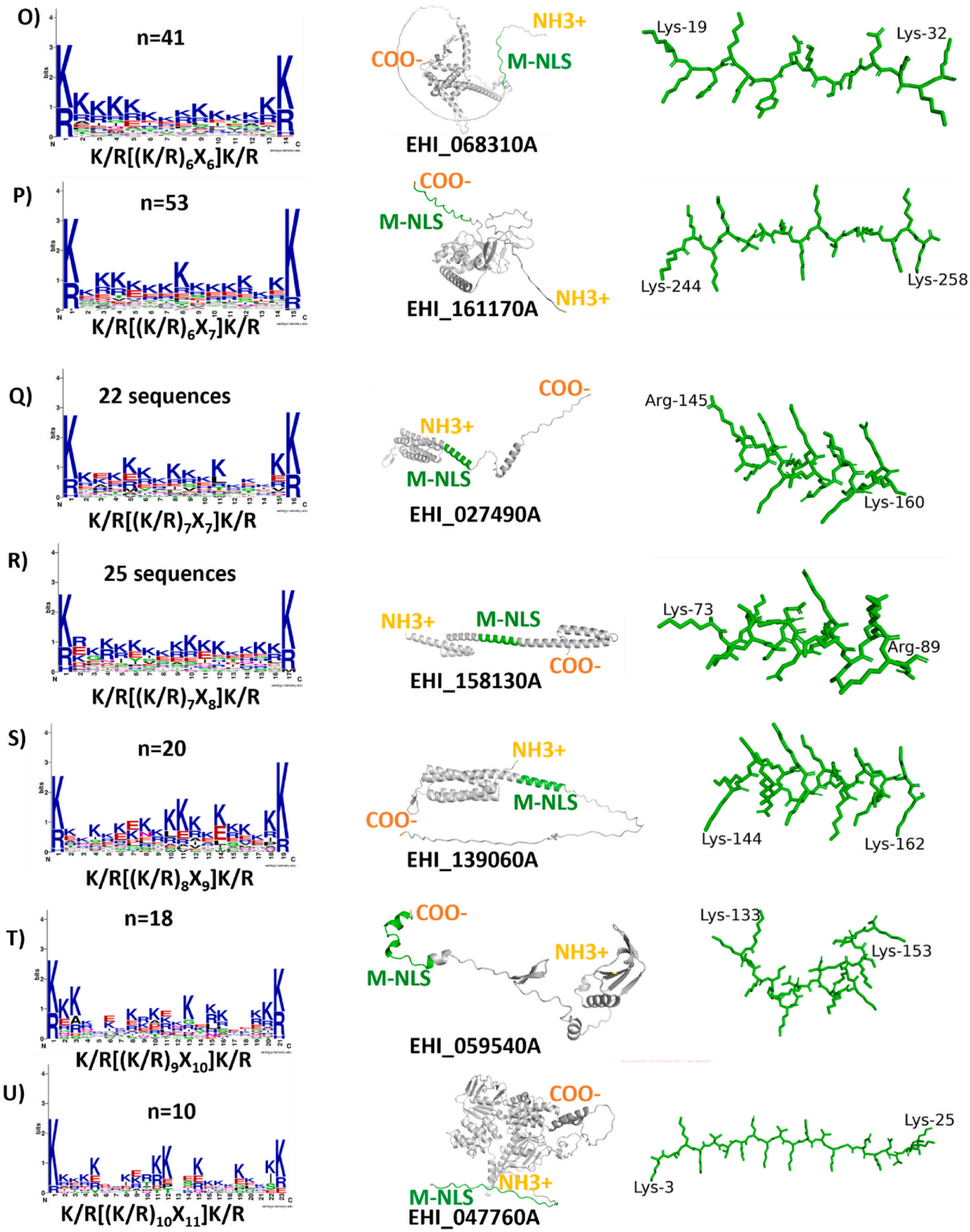


Fig. 3. (*continued*).

Table 3
Total canonical NLSs of each type and length of the entire *E. histolytica* genome.

Residues	Consensus sequence	Sequences
4	K/R(K/RX)K/R	131
4	H$(K/R)_3$	52
4	4(K/R)	164
4	P$(K/R)_3$	86
5	K/R[$(K/R)X_2$]K/R	268
6	K/R[$(K/R)_2X_2$]K/R	99
7	P[$(R/K)_3X_3$]	328
7	K/R[$(K/R)_2X_3$]K/R	146
8	K/R[$(K/R)_3X_3$]K/R	97
9	K/R[$(K/R)_3X_4$]K/R	113
10	K/R[$(K/R)_4X_4$]K/R	66
11	K/R[$(K/R)_4X_5$]K/R	90
12	K/R[$(K/R)_5X_5$]K/R	47
13	K/R[$(K/R)_5X_6$]K/R	70
14	K/R[$(K/R)_6X_6$]K/R	41
15	K/R[$(K/R)_6X_7$]K/R	53
16	K/R[$(K/R)_7X_7$]K/R	22
17	K/R[$(K/R)_7X_8$]K/R	25
18	K/R[$(K/R)_8X_8$]K/R	5
19	K/R[$(K/R)_8X_9$]K/R	20
20	NC	3
21	K/R[$(K/R)_9X_{10}$]K/R	18
22	K/R[$(K/R)_{10}X_{10}$]K/R	8
23	K/R[$(K/R)_{10}X_{11}$]K/R	10
24	K/R[$(K/R)_{11}X_{11}$]K/R	5
25	K/R[$(K/R)_{11}X_{12}$]K/R	4
26	NC	2
27	NC	1
31	NC	2
33	NC	2
41	NC	1
	TOTAL	1979
17	$(K/R)_2X_{10}[(R/K)_3X_2]$	469
18	$(K/R)_2X_{10}[(R/K)_3X_2]$K/R	95
18	$(K/R)_3X_{10}[(R/K)_3X_2]$	26
19	$(K/R)_3X_{10}[(R/K)_4X_2]$	10
19	(R/K)X$(K/R)_2X_{10}[(R/K)_3X_2]$	77
19	$(K/R)_2X_{10}[(R/K)_4X_3]$	88
20	$(K/R)_2X_{10}[(R/K)_4X_4]$	70
20	(K/R)$X_2(K/R)_2X_{10}[(R/K)_3X_2]$	39
20	(K/R)X$(K/R)_2X_{10}[(R/K)_3X_2]$	18
21	$(K/R)_2X_{10}[(R/K)_5X_4]$	35
	TOTAL	927

trypanosomatids P$[(R/K)_3X_3]$ and for yeast PXX$[(R/K)_3X]$, respectively [49]. We detected 2 consensus sequences of 7 residues. One identical to that of trypanosomatids (Fig. 3G) and one exclusive to Entamoeba K/R $[(K/R)_2X_3]$K/R, in which X can be found from the second to the fifth position (Fig. 3H).

We found that *E. histolytica* contains multiple atypical consensus MNLSs of different length composition: 8 residues, K/R$[(K/R)_3X_3]$K/R; 9 residues, K/R$[(K/R)_3X_4]$K/R; 10 residues, K/R$[(K/R)_4X_4]$K/R; 11 residues, K/R$[(K/R)_4X_5]$K/R; 12 residues, K/R$[(K/R)_5X_5]$K/R; 13 residues; K/R$[(K/R)_5X_6]$K/R, 14 residues; K/R$[(K/R)_6X_6]$K/R, 15 residues; K/R$[(K/R)_6X_7]$K/R, 16 residues; K/R$[(K/R)_7X_7]$K/R, 17 residues; K/R $[(K/R)_7X_8]$K/R, 18 residues; K/R$[(K/R)_8X_8]$K/R, 19 residues; K/R$[(K/R)_8X_9]$K/R, 21 residues; K/R$[(K/R)_9X_{10}]$K/R, 22 residues; K/R$[(K/R)_{10}X_{10}]$K/R, 23 residues; K/R$[(K/R)_{10}X_{11}]$K/R, 24 residues; K/R$[(K/R)_{11}X_{11}]$K/R and 25 residues; K/R$[(K/R)_{11}X_{12}]$K/R (Fig. 3I-U, respectively). Additionally, we found proteins with MNLSs with 20, 26, 27, 31, 33, and 41 residues. However, obtaining a consensus sequence or relative frequency logos of n < 10 sequences was impossible (Supplementary Fig. 4).

Regarding the consensuses belonging to BNLSs, there are those of yeast and trypanosomes whose sequence $(K/R)_2X_{10}[(R/K)_3X_2]$ is identical for both groups [49]. We found that amoeba also has this same bipartite nuclear import sequence (Fig. 4A). However, we obtained 9 more consensuses for this type of signals that are mentioned below: two consensus sequences of 18 residues; $(K/R)_2X_{10}[(R/K)_3X_2]$K/R and $(K/R)_3X_{10}[(R/K)_3X_2]$ (Fig. 4 B and C, respectively), three consensus sequences of 19 residues; $(K/R)_3X_{10}[(R/K)_4X_2]$, (R/K)X $(K/R)_2X_{10}[(R/K)_3X_2]$ and $(K/R)_2X_{10}[(R/K)_4X_3]$ (Fig. 4 D-F respectively), three sequences of 20 residues; $(K/R)_2X_{10}[(R/K)_4X_4]$, (K/R) $X_2(K/R)_2X_{10}[(R/K)_3X_2]$ and (K/R)X$(K/R)_2X_{10}[(R/K)_3X_2]$ (Fig. 4 G-I respectively), we finally obtained a sequence of 21 residues; $(K/R)_2X_{10}[(R/K)_5X_4]$ (Fig. 4J).

There are no reported consensuses regarding the JNLS. To our knowledge, we identified these motifs for the first time. The consensuses obtained for NLSs made up of 2 JNLSs are, according to their length listed: 24 residues, $(K/R)_2X_5[(R/K)_2(K/R)X_2]X_7[(R/K)_3X_2]$; 25 residues, $(K/R)_2X_6[(R/K)_2(K/R)X_2]$ $X_7[(R/K)_3X_2]$; 26 residues, $(K/R)_2X_7[(R/K)_2(K/R)X_2]$ $X_7[(R/K)_3X_2]$; 29 residues, $(K/R)_2X_{10}[(R/K)_2(K/R)X_2]$ $X_7[(R/K)_3X_2]$; 30 residues, $(K/R)_2X_{10}[(R/K)_2(K/R)X_2]$ $X_8[(R/K)_3X_2]$; 31 residues, $(K/R)_2X_{10}[(R/K)_2(K/R)X_2]$ $X_9[(R/K)_3X_2]$; 32 residues $(K/R)_2X_{10}[(K/R)X_2(R/K)_2]$ $X_{10}[(R/K)_3X_2]$ and 33 residues, $(K/R)_2X_{10}[(K/R)_2X_2(R/K)](R/K)X_{10}[(R/K)_3X_2]$ (Fig. 5A–H, respectively). We could not generate consensus sequences for motifs of 23 and 28 residues due to their scarce (n < 10) representation (Supplementary Fig. 5). We also detected NLSs made up of 3 BNLSs up to 26 residues, and as before, we could not generate consensus sequences (Table 4 and Supplementary Figs. 6 and 7).

We performed blind dockings between EhImportin α (3873 atoms) with one MNLS (95 Atoms), one BNLS (159 atoms), and one JNLS (263 atoms) and observed that the three NLSs interact with the central part of the Armadillo domain of EhImportin α, while the Importin β binding domain (IBB domain) remains without any interaction (Fig. 6A). In the case of the EhImportin α and MNLS interaction, a ΔG of −170 was obtained; main interactions occurred between Arg 7 of MNLS with Arg29, Asp225 and His18 of EhImportin α (Fig. 6B). In the case of EhImportin α and BNLS we obtained a ΔG of −265, where Lys18 of BNLS interacts with Glu208 and Met133 of EhImportin α, while Lys1 of BNLS interacts with Asp294 of EhImportin α (Fig. 6C). The interaction between EhImportin α with JNLS, with a ΔG of −282, showed that Lys18, Lys17 and Glu1 of JNLS interact with Leu251, Glu254 and Asn209 of Importin α, respectively (Fig. 6D). It should be noted that the number of matches for each docking was 30,000, while the Grid box size was 64 (Fig. 6).

We observed that proteins U1A, U2AF1, U2AF2, and the N-terminus of RNA Polymerase ll localized to the nucleus of *E. histolytica* (Fig. 7). While U1A contains a MNLS of 4 amino acid residues 117RKPK120 (Fig. 8), U2AF1 has a MNLS of 14 residues 191KPRRRGRNRERRRR204 (Fig. 9), U2AF2 contains a JNLS of 31 residues 54RREYSRNEDREDRHRRVPEEERYNRSIRRRA84 (Fig. 10), whereas the N-terminus of Pol ll contains 2 non-canonical NLSs, one of 17 residues 178KLKSENKLEIRRKNGIK194 and the second of 20 residues 314LRKKNFKSIEERLSSKQGRL333 (Fig. 11), respectively. The negative controls (empty vector and the initial 40 amino acids of U2AF2) were devoid of HA signals (Supplementary Fig. 8).

Table 4
Total non-canonical NLSs of each type and length of the entire *E. histolytica* genome. NC (No consensus sequence), JNLS (Juxtaposed NLS).

Residues	Consensus sequence	JNLSs of 2 bipartite Sequences
23	NC	7
24	$(K/R)_2X_5[(R/K)_2(K/R)X_2]\ X_7[(R/K)_3X_2]$	10
25	$(K/R)_2X_6[(R/K)_2(K/R)X_2]\ X_7[(R/K)_3X_2]$	11
26	$(K/R)_2X_7[(R/K)_2(K/R)X_2]\ X_7[(R/K)_3X_2]$	10
27	NC	3
28	NC	7
29	$(K/R)_2X_{10}[(R/K)_2(K/R)X_2]\ X_7[(R/K)_3X_2]$	33
30	$(K/R)_2X_{10}[(R/K)_2(K/R)X_2]\ X_8[(R/K)_3X_2]$	26
31	$(K/R)_2X_{10}[(R/K)_2(K/R)X_2]\ X_9[(R/K)_3X_2]$	31
32	$(K/R)_2X_{10}[(K/R)X_2(R/K)_2]\ X_{10}[(R/K)_3X_2]$	21
33	$(K/R)_2X_{10}[(K/R)_2X_2(R/K)](R/K)X_{10}[(R/K)_3X_2]$	13
	TOTAL	172

Residues	Consensus sequence	JNLSs of 3 bipartite Sequences
35	NC	3
36	NC	3
37	NC	2
38	NC	1
39	NC	1
40	NC	5
41	NC	4
42	NC	3
43	NC	3
44	NC	3
45	NC	5
46	NC	4
47	NC	2
48	NC	1
51	NC	1
	TOTAL	41

Residues	Consensus sequence	JNLSs of 4 bipartite Sequences
41	NC	1
43	NC	2
45	NC	2
48	NC	1
53	NC	1
55	NC	2
60	NC	1
61	NC	1
	TOTAL	11

Residues	Consensus sequence	JNLSs of 5 bipartite Sequences
45	NC	1
47	NC	1
54	NC	1
58	NC	1
67	NC	1
68	NC	1
	TOTAL	6

Residues	Consensus sequence	JNLSs of 6 bipartite Sequences
79	NC	2
82	NC	2
	TOTAL	4
Residues	Consensus sequence	JNLSs of 7 bipartite Sequences
94	NC	1
106	NC	1
	TOTAL	2
Residues	Consensus sequence	JNLSs of 8 bipartite Sequences
78	NC	1
Residues	Consensus sequence	JNLSs of 12 bipartite Sequences
94	NC	1
124	NC	1
	TOTAL	2
Residues	Consensus sequence	JNLS of 26 bipartite Sequences
346	NC	1
Residues	Consensus sequence	Non canonical NLS
686	NC	1
698	NC	1
	TOTAL	2

4. Discussion

Although nuclear transport has been studied for more than 3 decades with initial studies, studies have yet to be performed on nuclear transport in protozoa. Initially, the cNLS of *X. laevis* and SV40 virus were tested with exogenous proteins in members of the phylum Amoebozoa [30].

A higher nuclear concentration compared to the cytoplasmic concentration of exogenous proteins such as BSA and EhEnolase fused to the NLS of the large T antigen of SV40 transfected in Chaos carolinensis, Amoeba proteus and E. histolytica and the conservation of MNLSs and BNLSs from our analyzes suggest that the Importin α/β pathway in the phylum Amoebozoa is functional [30,31]. Furthermore, the presence of nuclear transport factors in *E. histolytica* (Fig. 6) reinforces the hypothesis that nuclear transport mediated by the canonical nuclear protein import pathway is present from an early stage of eukaryotic evolution, as observed in phylogenetic analyses of Importin α conservation and Importin β in yeast and mammalian species members of the supergroup Opisthokonta [28,49,55]. The fact that we identified the EhCAS Exportin (CCR1), Ran, RanBP1, RandGAP, RanGEF, and NTF2 annotated in the *E. histolytica* genome indicates that the nuclear protein export process is present being supported by the fact that the protein EhNCABP166 contains a functional NES rich in Leucines and the Ehp53 protein that has anNES of the same type and EhRhoGAPnc as well [33,35,37] however they need to be tested experimentally. The function of the paralogous nucleotide exchange factors RhoGEF and RanGEF is to stimulate the exchange of GDP to GTP on a signaling GTPase. Specifically, RanGEF exchanges GDP to GTP in the nuclear transport of proteins to export different nucleocytoplasmic proteins [56–58]. In AmoebaDB, we have identified RanGEF, which supports the hypothesis of GTP-dependent nuclear export and its subsequent hydrolysis to GDP in the cytoplasm in E. histolytica. We have chosen RanGEF for the model of the nuclear transport pathway in Fig. 1 because it has been proved in other species to be the main GTP exchange factor in this pathway in eukaryotes. Our search for nuclear transport factors allowed us to generate a model of nuclear import and export of proteins for *E. histolytica* in Fig. 1, in which, as a novel contribution, we incorporated the factor NTF2 and the GTP hydrolysis proteins RanBP1, Ran GAP, and RanGEF since they were not taken into account in the models suggested by Gwairgi and Ghildyal, we have additionally seen their conservation in lower and higher eukaryotes (see Table 2). Regarding the identification of EhRan, it suggests a dependence on GTP for disintegrating the trimeric nuclear import complex of proteins once they cross the NPC [30,55]. However, these RhoGAP variants from E. histolytica present domains such as plekstrin homology (PH), calmodulin-binding (IQ), LIM-type zinc finger (LIM), DBL homology subfamily GEF, suggesting other functions from eukaryotes [58].

We found that 39.73 % of the proteins of *E. histolytica* can potentially be nuclear, while 15.55 % (1275 proteins) of the proteome contain cNLSs, resulting in 39.15 % of nuclear proteins with cNLS. As a reference, the nuclear proteins from *Trypanosoma cruzi* contain 41 % of their nuclear proteins with cNLS. *Trypanosoma brucei* has 68 % and yeast cells present an intermediate situation with 57 % of their nuclear proteins with possible cNLS [49]. This similarity in the distribution ranges of proteins with cNLS between *T. cruzi* and *E. histolytica* may be because many nuclear proteins could enter through non-canonical pathways [49,59,60]. As we have seen, EhRNA Polymerase II, whose nuclear localization has been demonstrated with an ncNLS, could be imported by a mechanism other than the Importin α/β pathways, such as the snurportin pathway or by a piggyback process upon binding with another or other proteins that do contain a cNLS. In addition to the cNLS, several non-classical NLSs have been identified, such as the PY-NLS sequences composed of the R/H/KX(2–5)PY consensus [61–63]. NLS M9 functions as an exchange motif between the nucleus and cytoplasm, is composed of 38 residues (YNDFGNYNNQSSNFGPMKGGNFGGRSSGPY), and is found in the hnRNP A1 protein [64]. The beta-like import receptor binding domain (NLS BIB) corresponds to a 42-residue signal enriched in arginine and glycine (VHSHKKKKIRTSPTFRRPKTLRLRRQPKYRRKSAPRRNK) found in the ribosomal protein rpL23 [65]. The RS domain corresponds to a signal rich in arginine and serine; in some proteins, it

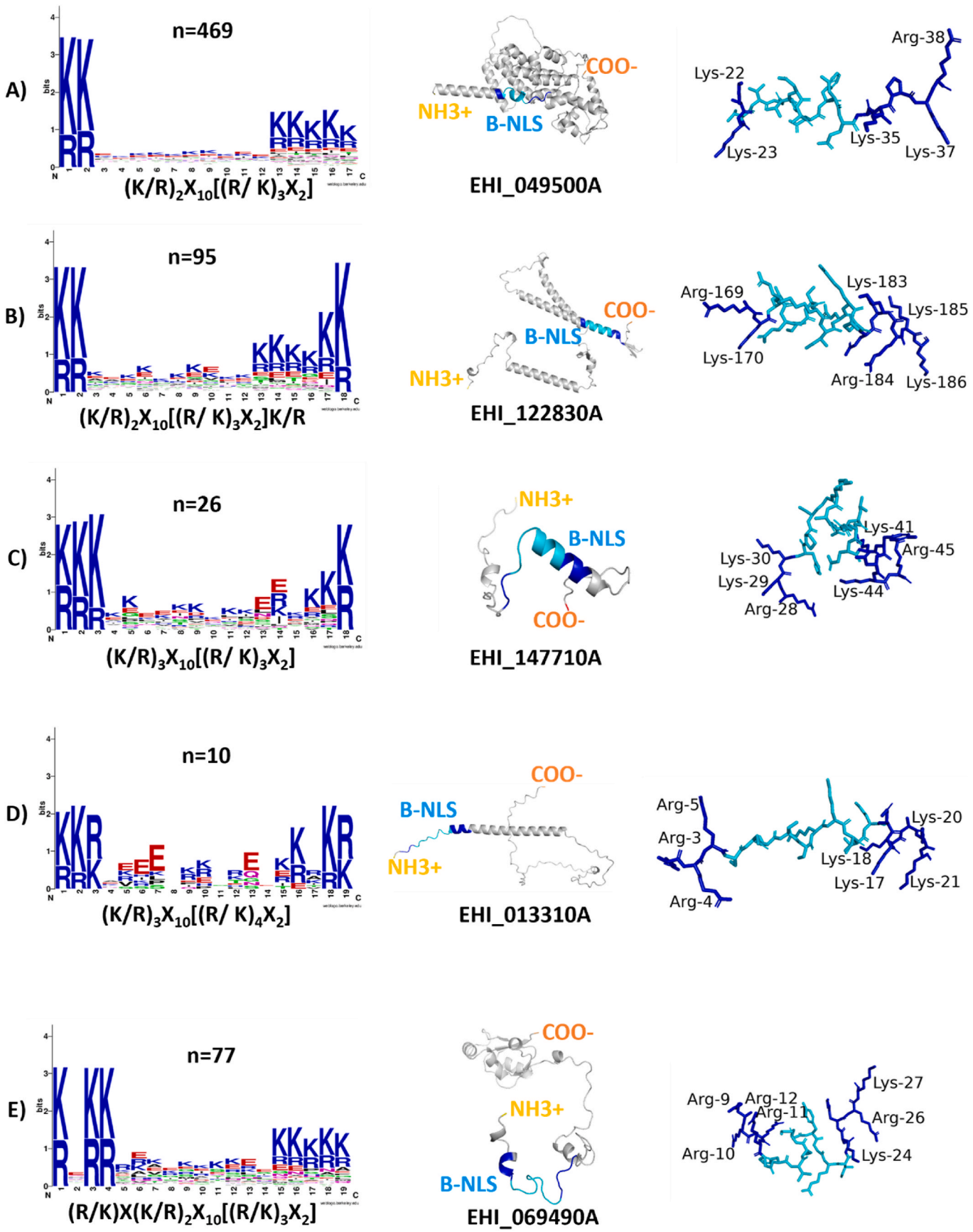


Fig. 4. Logos of NNLSs length classes with frequencies obtained from *E. histolytica* proteome. It is shown the consensus sequence with an individual structure for each case. Structures of B-NLSs are shown in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

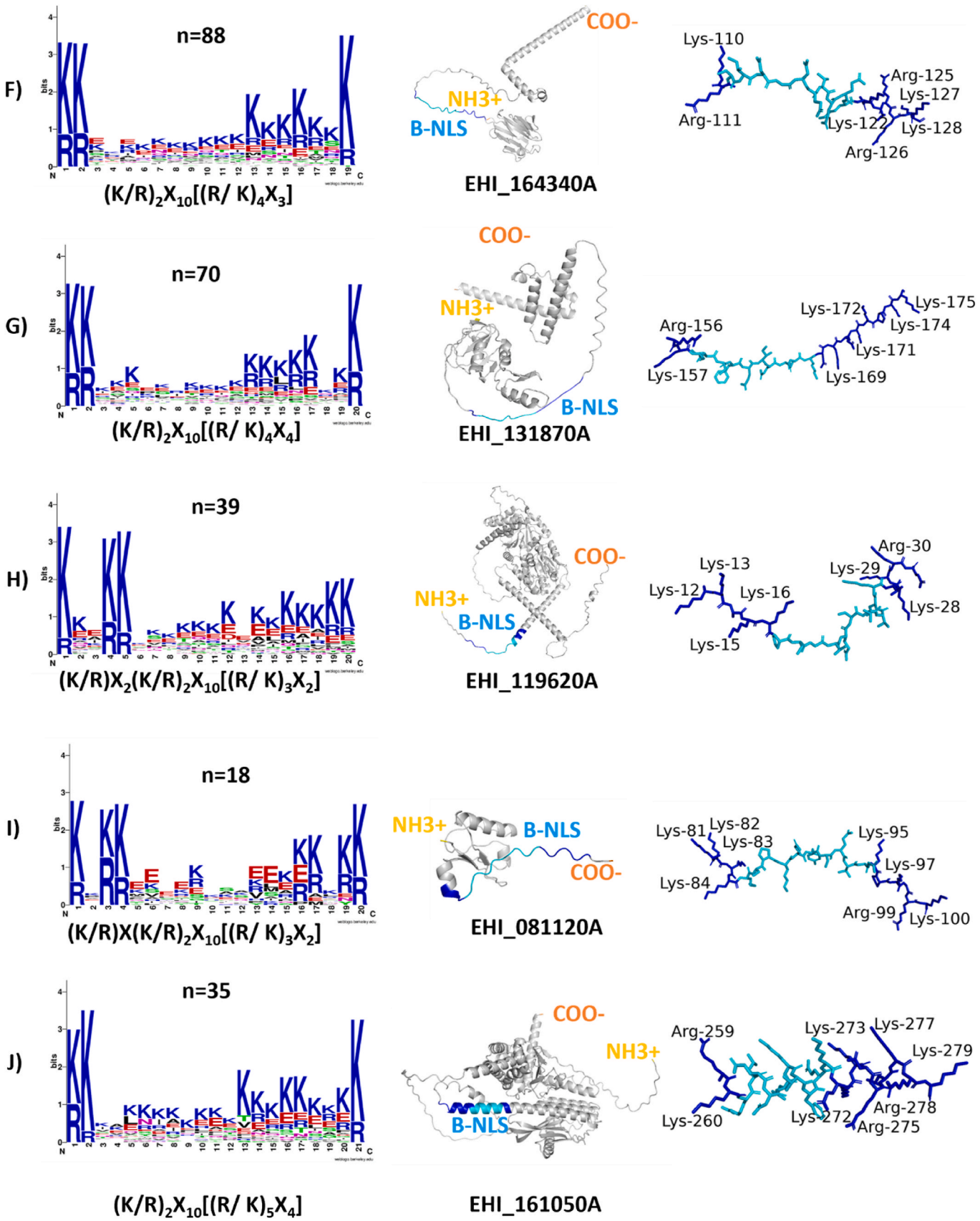


Fig. 4. (*continued*).

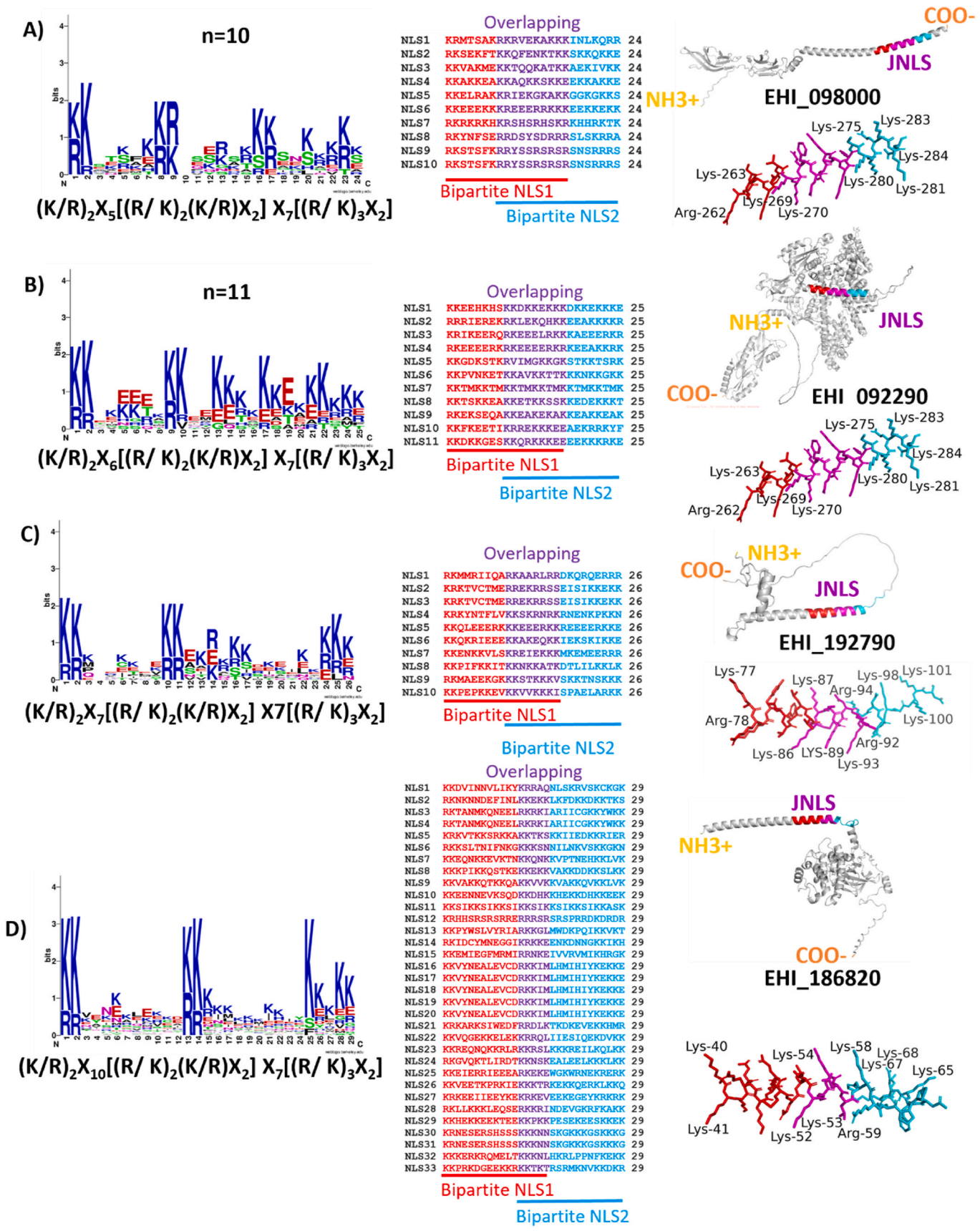


Fig. 5. Alignments and frequency logos of NLSs comprised 2 JNLSs and were separated by length classes obtained from the *E. histolytica* proteome. The consensus sequence with an individual structure for each case is shown. J-NLS structures are shown in red, purple, and blue, and juxtaposed positions in purple. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

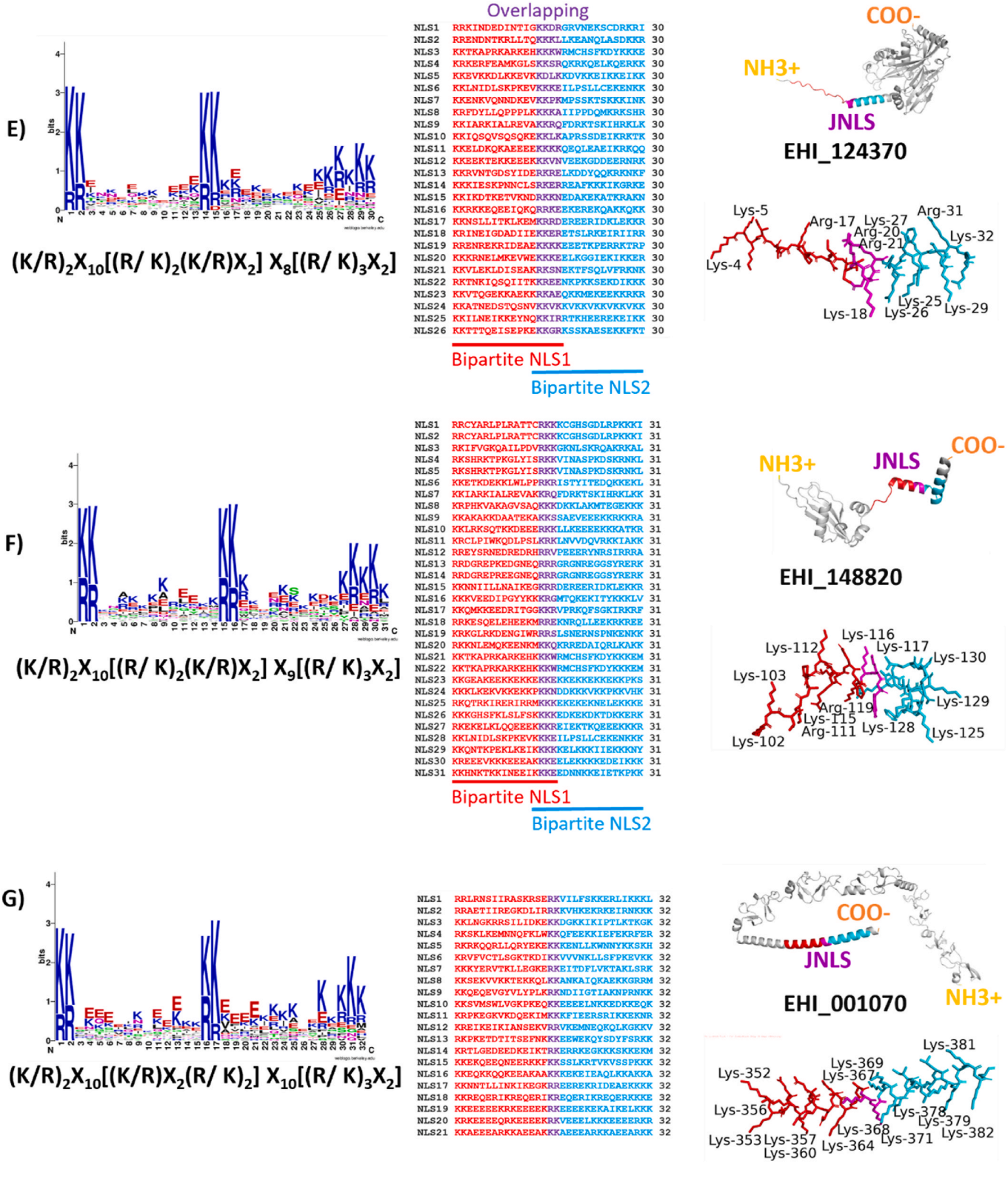


Fig. 5. (*continued*).

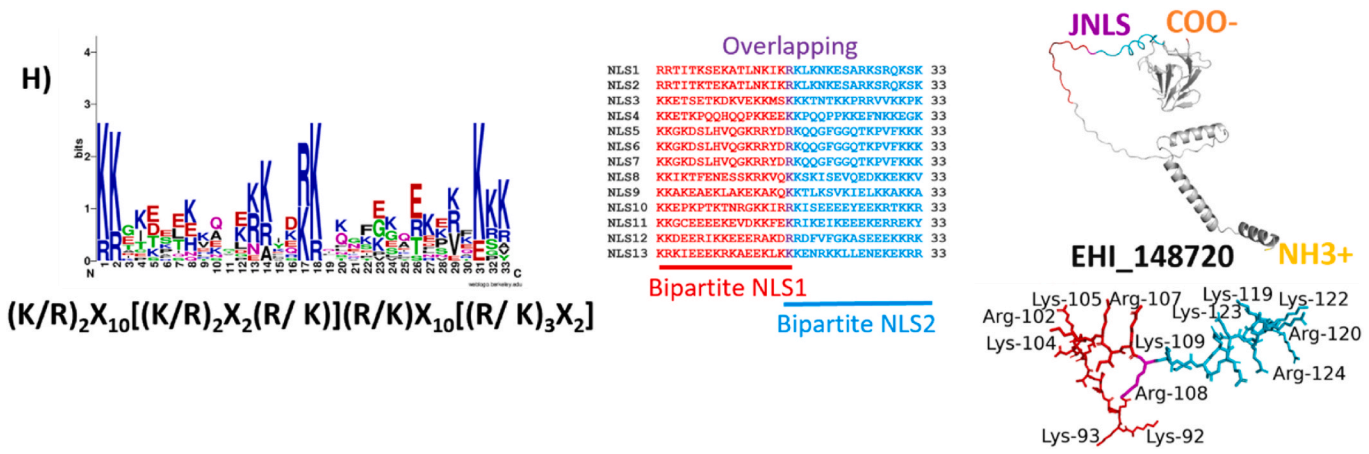


Fig. 5. (*continued*).

functions as an import signal and in others as a nuclear export signal [66]. Additionally, the non-canonical bipartite NLS is conserved in GATA factors from Aspergillus, yeast, and mammals, indicating an ancient origin [67]. Some proteins are transported to the nucleus directly by Importin β, without the participation of Importin α. The proteins that have been described are the parathyroid protein PTHrP [68], the nuclear transcription factor SCREBP2, which is essential for cholesterol metabolism [69], and the zinc finger protein C2H2 [70].

When analyzing the distribution of the type of cNLS in *E. histolytica,* we obtained 62.88 % MNLSs and 29.44 % BNLSs; this majority inclination towards MNLSs has been reported in other parasitic protozoan species such as *T. brucei* and *T. cruzi* where it is evident that the fraction of proteins that have MNLS is the majority when compared to those that have BNLSs.

In the case of the MNLSs in *T. brucei,* they are 62.6 % and in *T. cruzi* 59.8 %, while the BNLSs in *T. brucei* are 12.1 % and in *T. cruzi* it is 21.3 % [49]. This inclination of favored selective pressure toward MNLS has also been reported in *S. cerevisiae*, where 25.8 % are BNLSs and 30.9 % MNLSs [71].

With these and previous [30] data, it can be hypothesized that the initial Importin α receptor in protozoan parasites is more adapted to the recognition of MNLSs than bipartite ones or possibly there has been a coevolution process between the Importin α receptor with the MNLSs since, as shown in Table 3 *E. histolytica* has 25 consensus sequences for MNLSs against 10 consensus sequences for BNLSs [28]. Although this NTF has not been studied in *E. histolytica*, we can mention that Importin α from yeast and mammals recognizes MNLSs structurally in both regions corresponding to the significant and minor pockets [72,73]. Structurally, our models of the NLSs are mainly in *trans*-spatial configuration, which is most likely necessary for their recognition by EhImportin α.

Protein analyses with cNLS show a fraction of multiple NLSs in their sequence, with 65.18 % of the 1275 proteins with NLSs (Supplementary Figure 2) (MNLS and BNLS). This has been reported in *T. brucei,* where it is 25.3 %, and in *T. cruzi,* it is 19.9 %. Multiple NLSs in a single protein sequence have been reported in *Aspergillus nidulans* with the AreA factor that activates the transcription of nitrogen metabolism genes, and may contain multiple NLSs and serve as a mechanism for regulating nucleocytoplasmic localization [67,74].

The JNLSs in *E. histolytica* that correspond to 7.68 % could serve as signals that regulate the expression and nuclear localization of proteins with this type of signal since a regulation mechanism for nuclear concentration has been reported. The enzyme 5-lipoxygenase (5-LO) with 3 NLSs that, when individually fused to the green fluorescent protein (GFP), caused quantitatively and statistically less import than the SV40 NLS. However, when the 3 NLSs are combined, nuclear import to the SV40 NLS occurs [75].

Additionally, it has been shown that the Nuclear factor erythroid 2-related factor 2 (Nrf2) from Mus musculus mediates the transcriptional response of cells to oxidative stress and is translocated into the nucleus through two NLSs [76]. Another pathway consists of the participation of Snurportin, which interacts with Importin α instead of β import ribonucleoproteins, proteins that make up the spliceosomal U snRNPs [77].

However, we cannot rule out the potential plasticity of EhImportin α for the recognition of BNLSs since, in addition to having this type of signal, we detected a large number of proteins with multiple BNLSs, especially with non-canonical J-NLSs up to now as a novel contribution in this study.

It has been shown that the larger pocket can bind to at least 5 basic amino acid residues (K and R), while the smaller pocket can bind to two basic amino acid residues [73]. With this information, we can hypothesize that the EhImportin α major and minor pockets could recognize MNLSs similarly. Our docking assays indicate that the central region of the armadillo EhImportin α domain interacts with the MNLS, BNLS, and JNLS in a similar way as occurs in *S. cerevisiae* [78], suggesting an evolutionary conservation in the mechanism and plasticity of recognition of this nuclear transport factor.

Regarding the conservation of basic amino acids Arginine and Lysine in the NLSs of the supergroups (Ophystokonta, Excabata, and Amoebozoa), it indicates a selective pressure in the structural conservation of the recognition regions of nuclear transport factors, specifically with the initial path α/β adapter; Importin α in the conserved glutamic acid and tryptophan residues of the major and minor pockets that recognize the basic residues K, R and in very few cases H [62,79]. However, it is necessary to experimentally test and characterize different NLSs, such as those we have predicted in *E. histolytica*. We can propose that the nuclear localization of U1A, U2AF2, U2AF1, and RNA Polymerase II is because they contain NLSs rich in basic amino acids, mainly K and R. However, it will be crucial to dissect and mutagenize these NLSs to ensure their participation in the nuclear translocation of these proteins.

Our data strengthen the hypothesis that the nuclear import of proteins mediated by highly conserved nuclear transport factors arose in the early unicellular eukaryotes.

The NLSs seem to have arisen at an early stage in the evolution of eukaryotes as they are present in protozoans such as Trypanosomatidae and, in this case, in the supergroup, Amoebozoa with high conservation in their sequences, enabling coevolution between NLSs and NTF in eukaryotic organisms and most likely in the last common ancestor of eukaryotes.

CRediT authorship contribution statement

Israel Canela-Pérez: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal

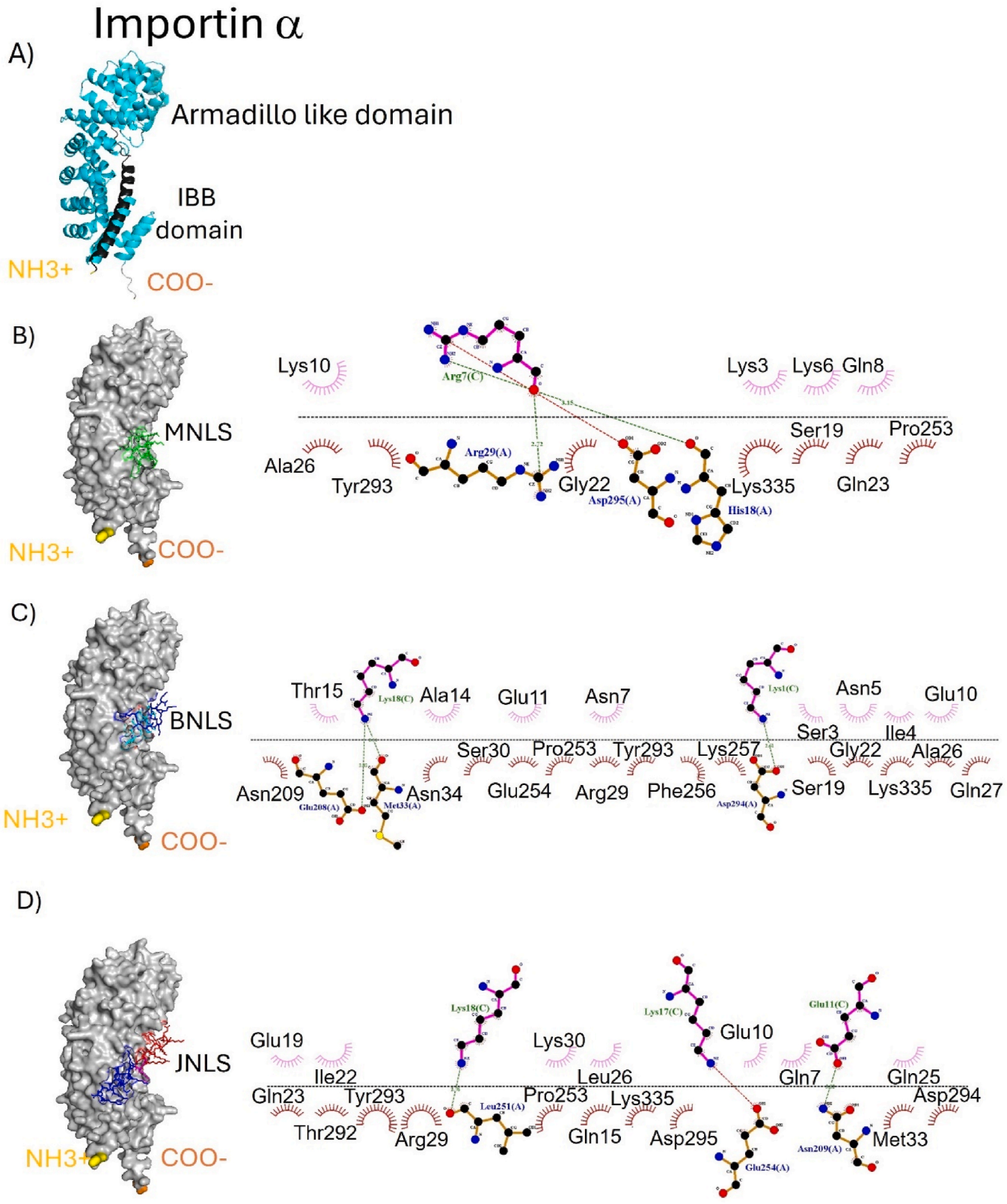


Fig. 6. Blind docking experiments between EhImportin α and NLSs. The EhImportin α structure shows the armadillo domain (blue) and the Importin β binding domain (IBB, in black) appears in (A). Docking interactions with an MNLS (B), a BNLS (C), and a JNLS (D). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

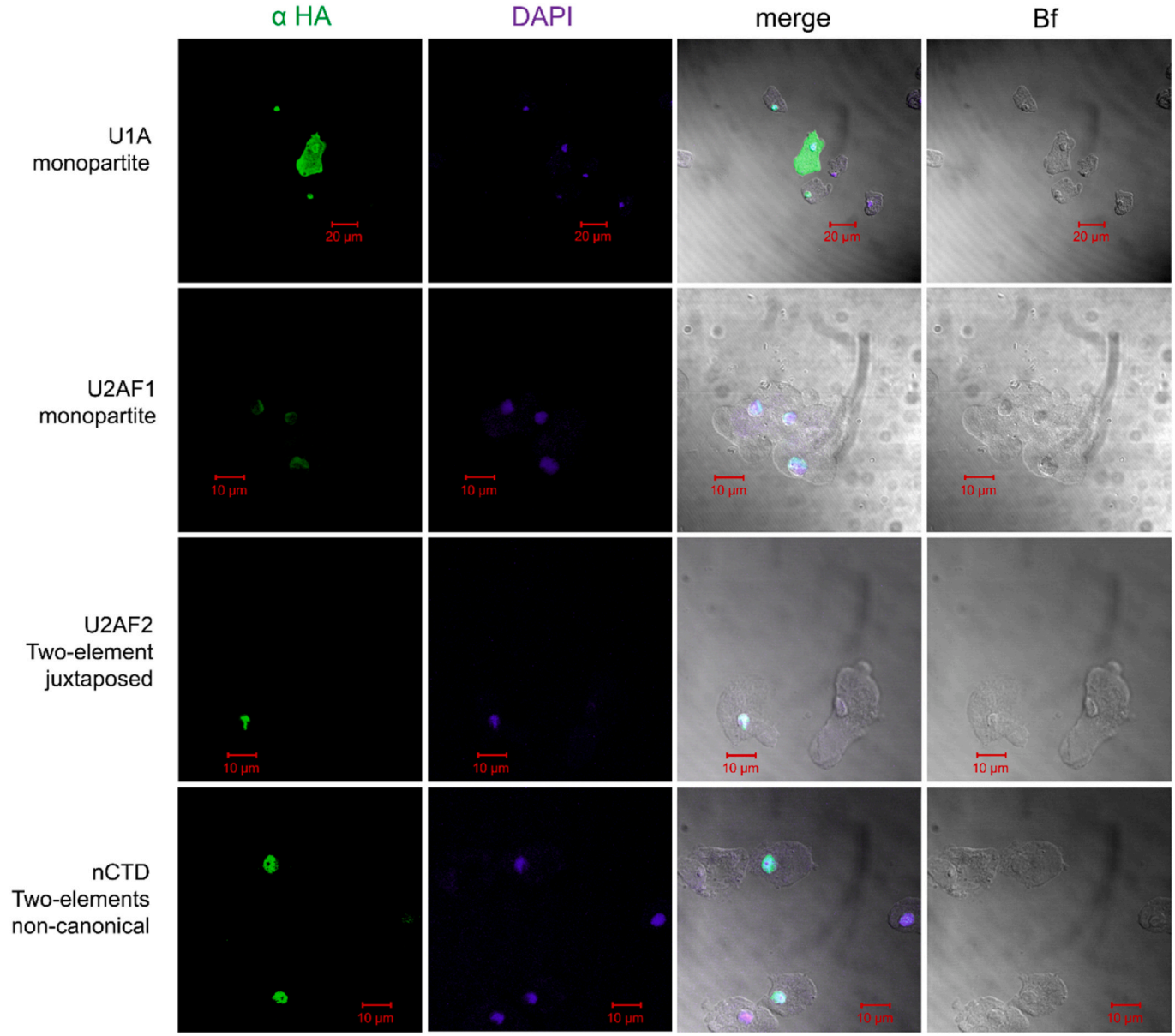


Fig. 7. Nuclear localization of HA-tagged U1A, U2AF1, U2AF2, and the N-terminus of RNA Polymerase ll. Green signals correspond to anti-HA Alexa antibodies and blue signals correspond to DAPI-contrasted nuclear DNA. Merged and bright-field (Bf) channels are also shown. The U1A sample shows an artifactual broken cell where the protein permeated to the cytoplasm. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

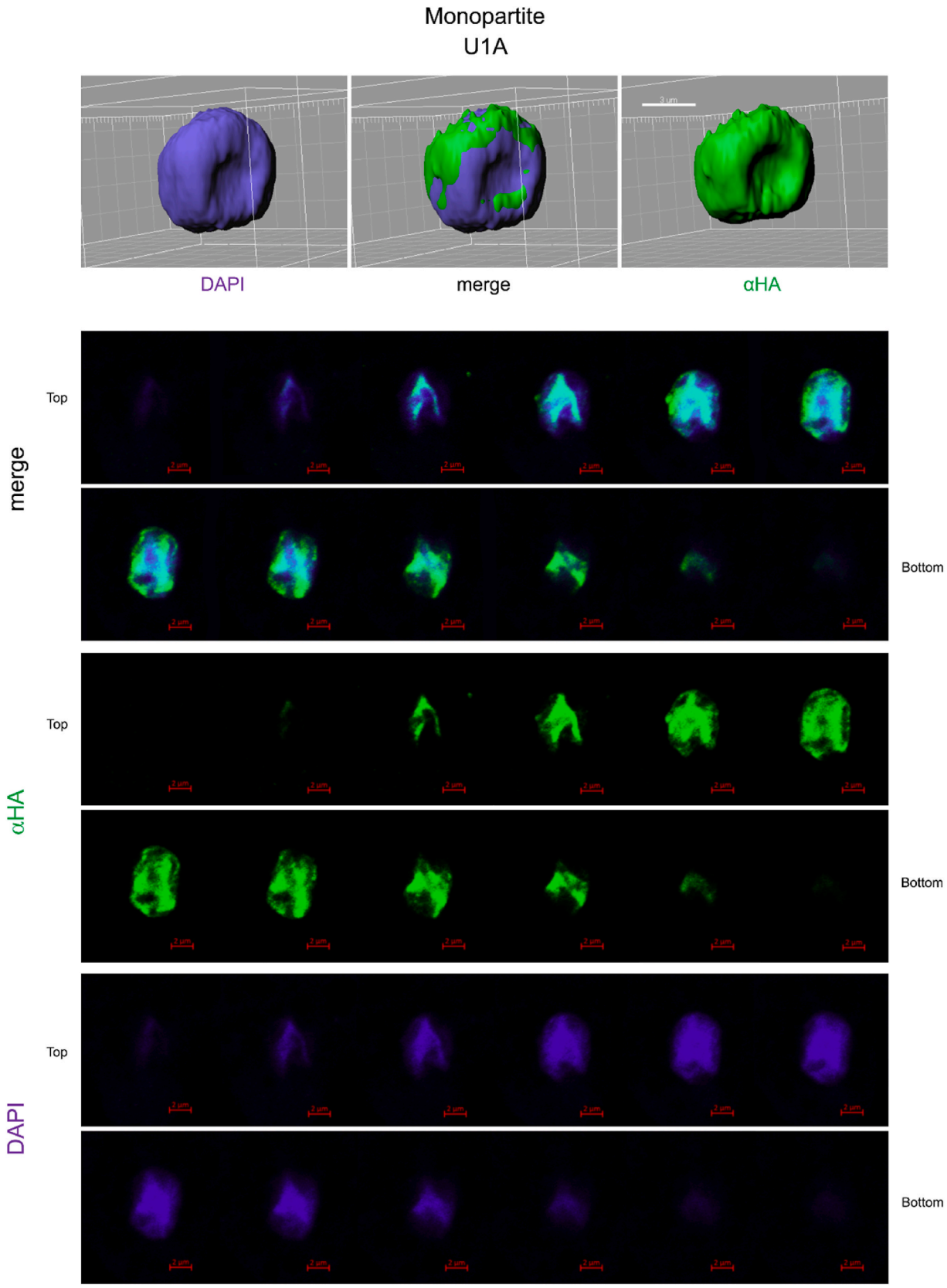


Fig. 8. Nuclear localization and 3D visualization of U1A harboring an MNLS. The lower panels show the sequential optical slices, ordered from top to bottom. Green signals: anti-HA Alexa antibodies; blue signals: DAPI-contrasted nuclear DNA. A merged channel is also shown. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

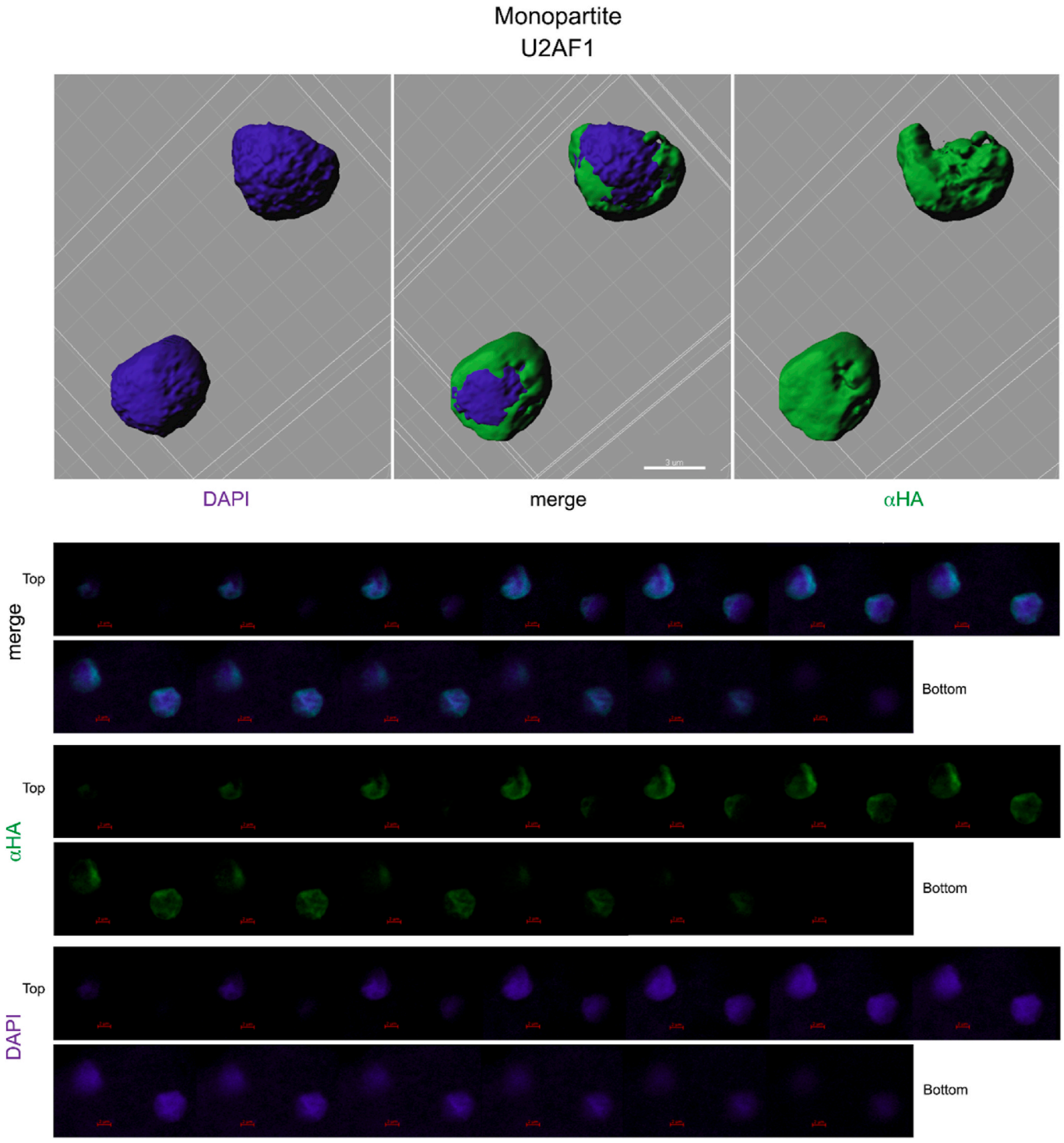


Fig. 9. Nuclear localization and 3D visualization of U2AF1 harboring an MNLS. The lower panels show the sequential optical slices ordered from top to bottom. Green signals: anti-HA Alexa antibodies; blue signals: DAPI-contrasted nuclear DNA. A merged channel is also shown. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

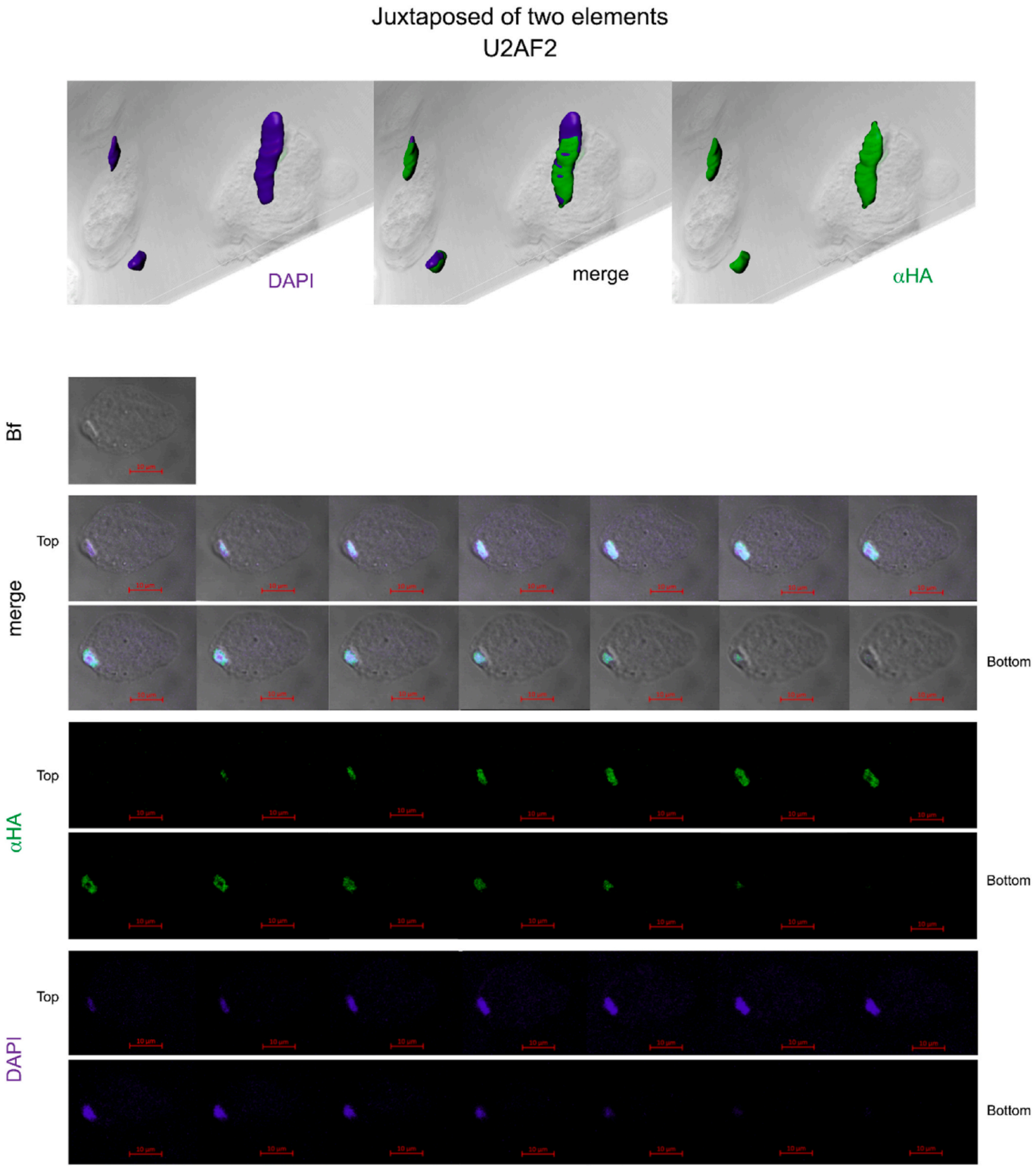


Fig. 10. Nuclear localization and 3D visualization of U2AF2 harboring a JNLS. The lower panels show the sequential optical slices ordered from top to bottom. Green signals: anti-HA Alexa antibodies; blue signals: DAPI-contrasted nuclear DNA. Merged and bright-field (Bf) channels are also shown. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

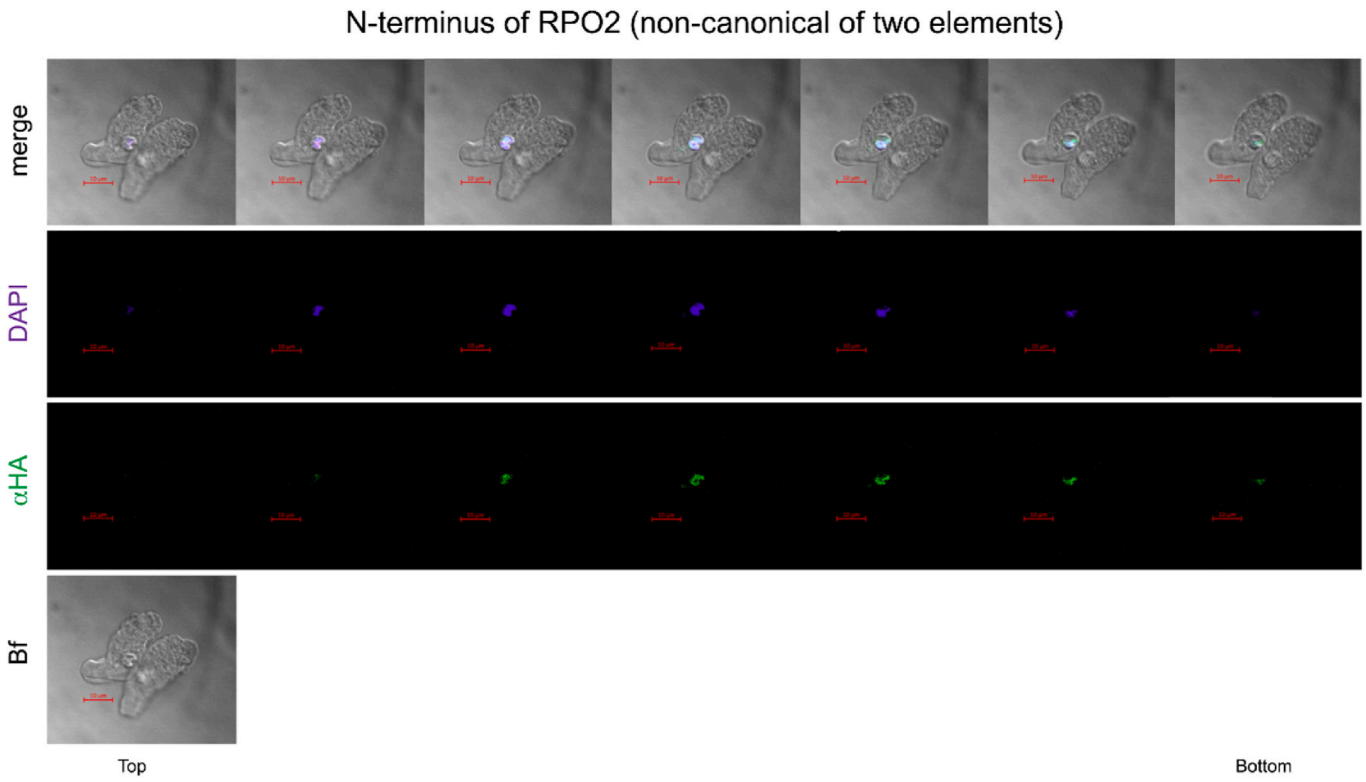


Fig. 11. Nuclear localization of the N-terminus of RNA Polymerase ll harboring a JNLS. The lower panels show the sequential optical slices ordered from top to bottom. Green signals: anti-HA Alexa antibodies; blue signals: DAPI-contrasted nuclear DNA; BF: bright field. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

analysis, Data curation, Conceptualization. **Elisa Azuara-Liceaga:** Writing – review & editing, Resources, Investigation, Funding acquisition, Data curation, Conceptualization. **Patricia Cuéllar:** Writing – review & editing, Investigation, Conceptualization. **Odila Saucedo-Cárdenas:** Writing – review & editing, Investigation, Formal analysis, Data curation. **Jesús Valdés:** Writing – review & editing, Writing – original draft, Project administration, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jesus Valdes reports financial support and equipment, drugs, or supplies were provided by CONAHCYT México grant CF-2019-194163. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbrep.2024.101770.

References

[1] Y.I. Tekle, F. Wang, F.C. Wood, O.R. Anderson, A. Smirnov, New insights on the evolutionary relationships between the major lineages of Amoebozoa, Sci. Rep. 12 (2022) 11173, https://doi.org/10.1038/s41598-022-15372-7.
[2] S. Begum, H. Gorman, A. Chadha, K. Chadee, Entamoeba histolytica, Trends Parasitol. 37 (2021) 676–677, https://doi.org/10.1016/j.pt.2021.01.001.
[3] H. Zulfiqar, G. Mathew, S. Horrall, Amebiasis, StatPearls, Treasure Island (FL), 2023.
[4] R. Haque, C.D. Huston, M. Hughes, E. Houpt, W.A. Petri Jr., Amebiasis, N. Engl. J. Med. 348 (2003) 1565–1573, https://doi.org/10.1056/NEJMra022710.
[5] S.L. Stanley Jr., Amoebiasis, Lancet 361 (2003) 1025–1034, https://doi.org/10.1016/S0140-6736(03)12830-9.
[6] B.S. Pritt, C.G. Clark, Amebiasis, Mayo Clin. Proc. 83 (2008) 1154–1159, https://doi.org/10.4065/83.10.1154, quiz 1159-1160.
[7] F. Mi-ichi, M. Abu Yousuf, K. Nakada-Tsukui, T. Nozaki, Mitosomes in Entamoeba histolytica contain a sulfate activation pathway, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 21731–21736, https://doi.org/10.1073/pnas.0907106106.
[8] D. Kalderon, B.L. Roberts, W.D. Richardson, A.E. Smith, A short amino acid sequence able to specify nuclear location, Cell 39 (1984) 499–509, https://doi.org/10.1016/0092-8674(84)90457-4.
[9] C. Dingwall, J. Robbins, S.M. Dilworth, B. Roberts, W.D. Richardson, The nucleoplasmin nuclear location sequence is larger and more complex than that of SV-40 large T antigen, J. Cell Biol. 107 (1988) 841–849, https://doi.org/10.1083/jcb.107.3.841.
[10] L. Xu, J. Massague, Nucleocytoplasmic shuttling of signal transducers, Nat. Rev. Mol. Cell Biol. 5 (2004) 209–219, https://doi.org/10.1038/nrm1331.
[11] D. Xu, A. Farmer, Y.M. Chook, Recognition of nuclear targeting signals by Karyopherin-beta proteins, Curr. Opin. Struct. Biol. 20 (2010) 782–790, https://doi.org/10.1016/j.sbi.2010.09.008.
[12] C. Strambio-De-Castillia, M. Niepel, M.P. Rout, The nuclear pore complex: bridging nuclear transport and gene regulation, Nat. Rev. Mol. Cell Biol. 11 (2010) 490–501, https://doi.org/10.1038/nrm2928.
[13] S.J. Kim, J. Fernandez-Martinez, I. Nudelman, Y. Shi, W. Zhang, B. Raveh, T. Herricks, B.D. Slaughter, J.A. Hogan, P. Upla, I.E. Chemmama, R. Pellarin, I. Echeverria, M. Shivaraju, A.S. Chaudhury, J. Wang, R. Williams, J.R. Unruh, C. H. Greenberg, E.Y. Jacobs, Z. Yu, M.J. de la Cruz, R. Mironska, D.L. Stokes, J. D. Aitchison, M.F. Jarrold, J.L. Gerton, S.J. Ludtke, C.W. Akey, B.T. Chait, A. Sali, M.P. Rout, Integrative structure and functional anatomy of a nuclear pore complex, Nature 555 (2018) 475–482, https://doi.org/10.1038/nature26003.
[14] B. Cautain, R. Hill, N. de Pedro, W. Link, Components and regulation of nuclear transport processes, FEBS J. 282 (2015) 445–462, https://doi.org/10.1111/febs.13163.
[15] M. Stewart, D. Rhodes, Switching affinities in nuclear trafficking, Nat. Struct. Biol. 6 (1999) 301–304, https://doi.org/10.1038/7529.
[16] M. Stewart, Molecular mechanism of the nuclear protein import cycle, Nat. Rev. Mol. Cell Biol. 8 (2007) 195–208, https://doi.org/10.1038/nrm2114.
[17] J. Ma, A. Goryaynov, W. Yang, Super-resolution 3D tomography of interactions and competition in the nuclear pore complex, Nat. Struct. Mol. Biol. 23 (2016) 239–247, https://doi.org/10.1038/nsmb.3174.
[18] I.R. Vetter, C. Nowak, T. Nishimoto, J. Kuhlmann, A. Wittinghofer, Structure of a Ran-binding domain complexed with Ran bound to a GTP analogue: implications for nuclear transport, Nature 398 (1999) 39–46, https://doi.org/10.1038/17969.

[19] D. Gorlich, N. Pante, U. Kutay, U. Aebi, F.R. Bischoff, Identification of different roles for RanGDP and RanGTP in nuclear protein import, EMBO J. 15 (1996) 5584–5594.
[20] A. Hoelz, G. Blobel, Cell biology: popping out of the nucleus, Nature 432 (2004) 815–816, https://doi.org/10.1038/432815a.
[21] F.R. Bischoff, H. Ponstingl, Catalysis of guanine nucleotide exchange on Ran by the mitotic regulator RCC1, Nature 354 (1991) 80–82, https://doi.org/10.1038/354080a0.
[22] F.R. Bischoff, C. Klebe, J. Kretschmer, A. Wittinghofer, H. Ponstingl, RanGAP1 induces GTPase activity of nuclear Ras-related Ran, Proc. Natl. Acad. Sci. U. S. A. 91 (1994) 2587–2591, https://doi.org/10.1073/pnas.91.7.2587.
[23] J. Moroianu, G. Blobel, A. Radu, Nuclear protein import: Ran-GTP dissociates the karyopherin alphabeta heterodimer by displacing alpha from an overlapping binding site on beta, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 7059–7062, https://doi.org/10.1073/pnas.93.14.7059.
[24] M.C. King, C.P. Lusk, G. Blobel, Karyopherin-mediated import of integral inner nuclear membrane proteins, Nature 442 (2006) 1003–1007, https://doi.org/10.1038/nature05075.
[25] U. Kutay, F.R. Bischoff, S. Kostka, R. Kraft, D. Gorlich, Export of importin alpha from the nucleus is mediated by a specific nuclear transport factor, Cell 90 (1997) 1061–1071, https://doi.org/10.1016/s0092-8674(00)80372-4.
[26] M. Stewart, Insights into the molecular mechanism of nuclear trafficking using nuclear transport factor 2 (NTF2), Cell Struct. Funct. 25 (2000) 217–225, https://doi.org/10.1247/csf.25.217.
[27] T. Jovanovic-Talisman, J. Tetenbaum-Novatt, A.S. McKenney, A. Zilman, R. Peters, M.P. Rout, B.T. Chait, Artificial nanopores that mimic the transport selectivity of the nuclear pore complex, Nature 457 (2009) 1023–1027, https://doi.org/10.1038/nature07600.
[28] I. Canela-Perez, I. Lopez-Villasenor, A.M. Cevallos, R. Hernandez, Trypanosoma cruzi Importin alpha: ability to bind to a functional classical nuclear localization signal of the bipartite type, Parasitol. Res. 119 (2020) 3899–3907, https://doi.org/10.1007/s00436-020-06885-z.
[29] I. Canela-Perez, I. Lopez-Villasenor, A.M. Cevallos, R. Hernandez, Nuclear distribution of the Trypanosoma cruzi RNA Pol I subunit RPA31 during growth and metacyclogenesis, and characterization of its nuclear localization signal, Parasitol. Res. 117 (2018) 911–918, https://doi.org/10.1007/s00436-018-5747-4.
[30] C.M. Feldherr, D. Akin, Signal-mediated nuclear transport in the amoeba, J. Cell Sci. 112 (Pt 12) (1999) 2043–2048, https://doi.org/10.1242/jcs.112.12.2043.
[31] A. Tovy, R. Siman Tov, R. Gaentzsch, M. Helm, S. Ankri, A new nuclear function of the Entamoeba histolytica glycolytic enzyme enolase: the metabolic regulation of cytosine-5 methyltransferase 2 (Dnmt2) activity, PLoS Pathog. 6 (2010) e1000775, https://doi.org/10.1371/journal.ppat.1000775.
[32] A.D. Campos-Parra, N.A. Hernandez-Cuevas, R. Hernandez-Rivas, M. Vargas, EhNCABP166: a nucleocytoplasmic actin-binding protein from Entamoeba histolytica, Mol. Biochem. Parasitol. 172 (2010) 19–30, https://doi.org/10.1016/j.molbiopara.2010.03.010.
[33] R. Uribe, J. Almaraz Barrera Mde, M. Robles-Flores, G. Mendoza Hernandez, A. Gonzalez-Robles, R. Hernandez-Rivas, N. Guillen, M. Vargas, A functional study of nucleocytoplasmic transport signals of the EhNCABP166 protein from Entamoeba histolytica, Parasitology 139 (2012) 1697–1710, https://doi.org/10.1017/S0031182012001199.
[34] J.A. Garcia-Lerena, G. Gonzalez-Blanco, O. Saucedo-Cardenas, J. Valdes, Promoter-bound full-length intronic circular RNAs-RNA Polymerase II complexes regulate gene expression in the human parasite Entamoeba histolytica, Noncoding RNA 8 (2022), https://doi.org/10.3390/ncrna8010012.
[35] L. Mendoza, E. Orozco, M.A. Rodriguez, G. Garcia-Rivera, T. Sanchez, E. Garcia, P. Gariglio, Ehp53, an Entamoeba histolytica protein, ancestor of the mammalian tumour suppressor p53, Microbiology (Read.) 149 (2003) 885–893, https://doi.org/10.1099/mic.0.25892-0.
[36] M. Fukuda, S. Asano, T. Nakamura, M. Adachi, M. Yoshida, M. Yanagida, E. Nishida, CRM1 is responsible for intracellular transport mediated by the nuclear export signal, Nature 390 (1997) 308–311, https://doi.org/10.1038/36894.
[37] A. Hernandez-Flores, M.J. Almaraz-Barrera, D. Lozano-Amado, J. Correa-Basurto, A. Rojo-Dominguez, E. Luna-Rivera, M. Schnoor, N. Guillen, R. Hernandez-Rivas, M. Vargas, A new nucleocytoplasmic RhoGAP protein contributes to control the pathogenicity of Entamoeba histolytica by regulating EhRacC and EhRacD activity, Cell Microbiol. 18 (2016) 1653–1672, https://doi.org/10.1111/cmi.12603.
[38] B. Loftus, I. Anderson, R. Davies, U.C. Alsmark, J. Samuelson, P. Amedeo, P. Roncaglia, M. Berriman, R.P. Hirt, B.J. Mann, T. Nozaki, B. Suh, M. Pop, M. Duchene, J. Ackers, E. Tannich, M. Leippe, M. Hofer, I. Bruchhaus, U. Willhoeft, A. Bhattacharya, T. Chillingworth, C. Churcher, Z. Hance, B. Harris, D. Harris, K. Jagels, S. Moule, K. Mungall, D. Ormond, R. Squares, S. Whitehead, M.A. Quail, E. Rabbinowitsch, H. Norbertczak, C. Price, Z. Wang, N. Guillen, C. Gilchrist, S. E. Stroup, S. Bhattacharya, A. Lohia, P.G. Foster, T. Sicheritz-Ponten, C. Weber, U. Singh, C. Mukherjee, N.M. El-Sayed, W.A. Petri Jr., C.G. Clark, T.M. Embley, B. Barrell, C.M. Fraser, N. Hall, The genome of the protist parasite Entamoeba histolytica, Nature 433 (2005) 865–868, https://doi.org/10.1038/nature03291.
[39] Y. Jiang, D. Wang, Y. Yao, H. Eubel, P. Kunzler, I.M. Moller, D. Xu, MULocDeep: a deep-learning framework for protein subcellular and suborganellar localization prediction with residue-level interpretation, Comput. Struct. Biotechnol. J. 19 (2021) 4825–4839, https://doi.org/10.1016/j.csbj.2021.08.027.
[40] A.N. Nguyen Ba, A. Pogoutse, N. Provart, A.M. Moses, NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction, BMC Bioinf. 10 (2009) 202, https://doi.org/10.1186/1471-2105-10-202.
[41] K. Nakai, P. Horton, PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization, Trends Biochem. Sci. 24 (1999) 34–36, https://doi.org/10.1016/s0968-0004(98)01336-x.
[42] G.E. Crooks, G. Hon, J.M. Chandonia, S.E. Brenner, WebLogo: a sequence logo generator, Genome Res. 14 (2004) 1188–1190, https://doi.org/10.1101/gr.849004.
[43] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T.A.P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, SWISS-MODEL: homology modelling of protein structures and complexes, Nucleic Acids Res. 46 (2018) W296–W303, https://doi.org/10.1093/nar/gky427.
[44] A. Singh, M.M. Copeland, P.J. Kundrotas, I.A. Vakser, GRAMM Web server for protein docking, Methods Mol. Biol. 2714 (2024) 101–112, https://doi.org/10.1007/978-1-0716-3441-7_5.
[45] A.C. Wallace, R.A. Laskowski, J.M. Thornton, LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions, Protein Eng. 8 (1995) 127–134, https://doi.org/10.1093/protein/8.2.127.
[46] J. Valdes, T. Nozaki, E. Sato, Y. Chiba, K. Nakada-Tsukui, N. Villegas-Sepulveda, R. Winkler, E. Azuara-Liceaga, M.S. Mendoza-Figueroa, N. Watanabe, H.J. Santos, Y. Saito-Nakano, J.M. Galindo-Rosales, Proteomic analysis of Entamoeba histolytica in vivo assembled pre-mRNA splicing complexes, J. Proteonomics 111 (2014) 30–45, https://doi.org/10.1016/j.jprot.2014.07.027.
[47] Z.H. Abu Bakar, J.P. Bellier, W.Z. Wan Ngah, D. Yanagisawa, K.I. Mukaisho, I. Tooyama, Optimization of 3D immunofluorescence analysis and visualization using IMARIS and MeshLab, Cells 12 (2023), https://doi.org/10.3390/cells12020218.
[48] M. Marfori, A. Mynott, J.J. Ellis, A.M. Mehdi, N.F. Saunders, P.M. Curmi, J. K. Forwood, M. Boden, B. Kobe, Molecular basis for specificity of nuclear import and prediction of nuclear localization, Biochim. Biophys. Acta 1813 (2011) 1562–1577, https://doi.org/10.1016/j.bbamcr.2010.10.013.
[49] I. Canela-Perez, I. Lopez-Villasenor, L. Mendoza, A.M. Cevallos, R. Hernandez, Nuclear localization signals in trypanosomal proteins, Mol. Biochem. Parasitol. 229 (2019) 15–23, https://doi.org/10.1016/j.molbiopara.2019.02.003.
[50] D. Chelsky, R. Ralph, G. Jonak, Sequence requirements for synthetic peptide-mediated translocation to the nucleus, Mol. Cell Biol. 9 (1989) 2487–2492, https://doi.org/10.1128/mcb.9.6.2487-2492.1989.
[51] M.R. Fontes, T. Teh, D. Jans, R.I. Brinkworth, B. Kobe, Structural basis for the specificity of bipartite nuclear localization sequence binding by importin-alpha, J. Biol. Chem. 278 (2003) 27981–27987, https://doi.org/10.1074/jbc.M303275200.
[52] M.R. Hodel, A.H. Corbett, A.E. Hodel, Dissection of a nuclear localization signal, J. Biol. Chem. 276 (2001) 1317–1325, https://doi.org/10.1074/jbc.M008522200.
[53] S. Kosugi, M. Hasebe, T. Entani, S. Takayama, M. Tomita, H. Yanagawa, Design of peptide inhibitors for the importin alpha/beta nuclear import pathway by activity-based profiling, Chem. Biol. 15 (2008) 940–949, https://doi.org/10.1016/j.chembiol.2008.07.019.
[54] S. Kosugi, M. Hasebe, N. Matsumura, H. Takashima, E. Miyamoto-Sato, M. Tomita, H. Yanagawa, Six classes of nuclear localization signals specific to different binding grooves of importin alpha, J. Biol. Chem. 284 (2009) 478–485, https://doi.org/10.1074/jbc.M807017200.
[55] M.A. Gwairgi, R. Ghildyal, Nuclear transport in Entamoeba histolytica: knowledge gap and therapeutic potential, Parasitology 145 (2018) 1378–1387, https://doi.org/10.1017/S0031182018000252.
[56] K. Nakada-Tsukui, H. Okada, B.N. Mitra, T. Nozaki, Phosphatidylinositol-phosphates mediate cytoskeletal reorganization during phagocytosis via a unique modular protein consisting of RhoGEF/DH and FYVE domains in the parasitic protozoon Entamoeba histolytica, Cell Microbiol. 11 (2009) 1471–1491, https://doi.org/10.1111/j.1462-5822.2009.01341.x.
[57] D.E. Bosch, D.P. Siderovski, G protein signaling in the parasite Entamoeba histolytica, Exp. Mol. Med. 45 (2013) e15, https://doi.org/10.1038/emm.2013.30.
[58] C.C. Hon, K. Nakada-Tsukui, T. Nozaki, N. Guillen, Dissecting the actin cytoskeleton of Entamoeba histolytica from a genomic perspective, anaerobic parasitic Protozoa. Genomics and Molecular Biology, 2010, pp. 81–118.
[59] C. dos Santos Junior Ade, D.E. Kalume, R. Camargo, D.P. Gomez-Mendoza, J. R. Correa, S. Charneau, M.V. de Sousa, B.D. de Lima, C.A. Ricart, Unveiling the trypanosoma cruzi nuclear proteome, PLoS One 10 (2015) e0138667, https://doi.org/10.1371/journal.pone.0138667.
[60] C. Goos, M. Dejung, C.J. Janzen, F. Butter, S. Kramer, The nuclear proteome of Trypanosoma brucei, PLoS One 12 (2017) e0181884, https://doi.org/10.1371/journal.pone.0181884.
[61] B.J. Lee, A.E. Cansizoglu, K.E. Suel, T.H. Louis, Z. Zhang, Y.M. Chook, Rules for nuclear localization sequence recognition by karyopherin beta 2, Cell 126 (2006) 543–558, https://doi.org/10.1016/j.cell.2006.05.049.
[62] Y.M. Chook, K.E. Suel, Nuclear import by karyopherin-betas: recognition and inhibition, Biochim. Biophys. Acta 1813 (2011) 1593–1606, https://doi.org/10.1016/j.bbamcr.2010.10.014.
[63] A. Lange, R.E. Mills, S.E. Devine, A.H. Corbett, A PY-NLS nuclear targeting signal is required for nuclear localization and function of the Saccharomyces cerevisiae mRNA-binding protein Hrp1, J. Biol. Chem. 283 (2008) 12926–12934, https://doi.org/10.1074/jbc.M800898200.
[64] H. Siomi, G. Dreyfuss, A nuclear localization domain in the hnRNP A1 protein, J. Cell Biol. 129 (1995) 551–560, https://doi.org/10.1083/jcb.129.3.551.
[65] S. Jakel, D. Gorlich, Importin beta, transportin, RanBP5 and RanBP7 mediate nuclear import of ribosomal proteins in mammalian cells, EMBO J. 17 (1998) 4491–4502, https://doi.org/10.1093/emboj/17.15.4491.

[66] D. Cazalla, J. Zhu, L. Manche, E. Huber, A.R. Krainer, J.F. Caceres, Nuclear export and retention signals in the RS domain of SR proteins, Mol. Cell Biol. 22 (2002) 6871–6882, https://doi.org/10.1128/MCB.22.19.6871-6882.2002.
[67] C.C. Hunter, K.S. Siebert, D.J. Downes, K.H. Wong, S.D. Kreutzberger, J.A. Fraser, D.F. Clarke, M.J. Hynes, M.A. Davis, R.B. Todd, Multiple nuclear localization signals mediate nuclear localization of the GATA transcription factor AreA, Eukaryot. Cell 13 (2014) 527–538, https://doi.org/10.1128/EC.00040-14.
[68] G. Cingolani, J. Bednenko, M.T. Gillespie, L. Gerace, Molecular basis for the recognition of a nonclassical nuclear localization signal by importin beta, Mol. Cell 10 (2002) 1345–1353, https://doi.org/10.1016/s1097-2765(02)00727-x.
[69] S.J. Lee, T. Sekimoto, E. Yamashita, E. Nagoshi, A. Nakagawa, N. Imamoto, M. Yoshimura, H. Sakai, K.T. Chong, T. Tsukihara, Y. Yoneda, The structure of importin-beta bound to SREBP-2: nuclear import of a transcription factor, Science 302 (2003) 1571–1575, https://doi.org/10.1126/science.1088372.
[70] S. Choi, E. Yamashita, N. Yasuhara, J. Song, S.Y. Son, Y.H. Won, H.R. Hong, Y. S. Shin, T. Sekimoto, I.Y. Park, Y. Yoneda, S.J. Lee, Structural basis for the selective nuclear import of the C2H2 zinc-finger protein Snail by importin beta, Acta Crystallogr. D Biol. Crystallogr. 70 (2014) 1050–1060, https://doi.org/10.1107/S1399004714000972.
[71] A. Lange, R.E. Mills, C.J. Lange, M. Stewart, S.E. Devine, A.H. Corbett, Classical nuclear localization signals: definition, function, and interaction with importin alpha, J. Biol. Chem. 282 (2007) 5101–5105, https://doi.org/10.1074/jbc.R600026200.
[72] E. Conti, J. Kuriyan, Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha, Structure 8 (2000) 329–338, https://doi.org/10.1016/s0969-2126(00)00107-6.
[73] E. Conti, M. Uy, L. Leighton, G. Blobel, J. Kuriyan, Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha, Cell 94 (1998) 193–204, https://doi.org/10.1016/s0092-8674(00)81419-1.
[74] L. Fischer-Fantuzzi, C. Vesco, Cell-dependent efficiency of reiterated nuclear signals in a mutant simian virus 40 oncoprotein targeted to the nucleus, Mol. Cell Biol. 8 (1988) 5495–5503, https://doi.org/10.1128/mcb.8.12.5495-5503.1988.
[75] M. Luo, C.W. Pang, A.E. Gerken, T.G. Brock, Multiple nuclear localization sequences allow modulation of 5-lipoxygenase nuclear import, Traffic 5 (2004) 847–854, https://doi.org/10.1111/j.1600-0854.2004.00227.x.
[76] M. Theodore, Y. Kawai, J. Yang, Y. Kleshchenko, S.P. Reddy, F. Villalta, I.J. Arinze, Multiple nuclear localization signals function in the nuclear import of the transcription factor Nrf2, J. Biol. Chem. 283 (2008) 8984–8994, https://doi.org/10.1074/jbc.M709040200.
[77] A. Strasser, A. Dickmanns, R. Luhrmann, R. Ficner, Structural basis for m3G-cap-mediated nuclear import of spliceosomal UsnRNPs by snurportin1, EMBO J. 24 (2005) 2235–2243, https://doi.org/10.1038/sj.emboj.7600701.
[78] S.N. Yang, A.A. Takeda, M.R. Fontes, J.M. Harris, D.A. Jans, B. Kobe, Probing the specificity of binding to the major nuclear localization sequence-binding site of importin-alpha using oriented peptide library screening, J. Biol. Chem. 285 (2010) 19935–19946, https://doi.org/10.1074/jbc.M109.079574.
[79] A.J. O'Reilly, J.B. Dacks, M.C. Field, Evolution of the karyopherin-beta family of nucleocytoplasmic transport factors; ancient origins and continued specialization, PLoS One 6 (2011) e19308, https://doi.org/10.1371/journal.pone.0019308.