Structural Dynamics ARTICLE scitation.org/journal/sdy

Solution characterization of the dynamic conjugative entry exclusion protein TraG

Cite as: Struct. Dyn. **9**, 064702 (2022); doi: 10.1063/4.0000171
Submitted: 27 September 2022 · Accepted: 7 December 2022 ·
Published Online: 27 December 2022

View Online Export Citation CrossMark

Nicholas Bragagnolo and Gerald F. Audette[a]

AFFILIATIONS

Centre for Research on Biomolecular Interactions, Department of Chemistry, York University, Toronto, Ontario M3J 1P3, Canada

[a]Author to whom correspondence should be addressed: audette@yorku.ca

ABSTRACT

The R100 plasmid and the secretion system it encodes are representative of F-like conjugative type IV secretion systems for the transmission of mobile DNA elements in gram-negative bacteria, serving as a major contributor to the spread of antibiotic resistance in bacterial pathogens. The TraG protein of F-like systems consists of a membrane-bound N-terminal domain and a periplasmic C-terminal domain, denoted TraG*. TraG* is essential in preventing redundant DNA transfer through a process termed entry exclusion. In the donor cell, it interacts with TraN to facilitate mating pair stabilization; however, if a mating pore forms between bacteria with identical plasmids, TraG* interacts with its cognate TraS in the inner membrane of the recipient bacterium to prevent redundant donor–donor conjugation. Structural studies of TraG* from the R100 plasmid have revealed the presence of a dynamic region between the N- and C-terminal domains of TraG. Thermofluor, circular dichroism, collision-induced unfolding–mass spectrometry, and size exclusion chromatography linked to multiangle light scattering and small angle x-ray scattering experiments indicated an N-terminal truncation mutant displayed higher stability and less disordered content relative to full-length TraG*. The 45 N-terminal residues of TraG* are hypothesized to serve as part of a flexible linker between the two independently functioning domains.

© 2022 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/). https://doi.org/10.1063/4.0000171

I. INTRODUCTION

Horizontal gene transfer (HGT) methods, such as transformation of extracellular DNA, transduction of viral particles, and conjugation, are utilized by bacteria to acquire and integrate novel genes into their host chromosome through homologous recombination; if these genes provide the bacteria a selective advantage, HGT often increases the rate of dissemination of these virulence genes.[1–4] In this way, HGT enhances the rate of bacterial evolution by increasing the plasticity of the bacterial genome.[1,5–7] Extrachromosomal DNA, such as plasmids, can have broad species-host ranges and do not require gene transposition events for proper gene function, thus amplifying the propagation of virulence genes in bacterial communities.[8–10] Conjugation is enabled by a type IV secretion system (T4SS) transcribed and translated from the transfer (*tra*) or virulence (*vir*) gene region of a conjugative plasmid.[10,11] As genes enhancing bacterial virulence, including antibiotic resistance genes, can integrate into conjugative plasmids, many of which have broad-host ranges, conjugation is considered to be the most prominent contributor to the spread of virulence genes in bacterial pathogens.[12–14]

In conjugation, plasmids or chromosomally integrated conjugative elements (ICEs) are replicated in a donor cell and are transferred into a recipient cell in a donor-controlled fashion using a T4SS.[3–6,15] T4SSs are the most ubiquitous secretion system in prokaryotes; natural selection has favored conjugation as donor-mediated HGT appears to be a necessity for survival based on the requirement for rapid evolution in the competitive environments that unicellular organisms find themselves in.[6,7,11,16] The role of the T4SS is (a) to process and secrete self-polymerizing pilin monomers to form a pilus that will extend to contact a neighboring recipient cell and (b) depolymerize at the base to retract the pilus to bring the cells in proximity, allowing for cell membrane fusion and (c) DNA transfer (Fig. 1). The T4SS is a large and dynamic multi-protein complex; in gram negative bacteria, it spans both inner (IM) and outer membranes (OMs).[16–19] The F T4SS is the representative member of a subset of gram negative bacterial T4SS known as the thick flexible T4SS complexes, and the F plasmid which encodes it is the best characterized conjugative plasmid in gram negative bacteria.[6,17,20] The F plasmid was the first conjugative plasmid discovered, as characterized by Lederburg and Tatum and termed the "Fertility" or "F" factor.[21] F and F-like plasmids are unique in their host range as they are most commonly found in gram negative, facultative anaerobic bacterial genera such as *Klebsiella*, *Salmonella*, and

© Author(s) 2022

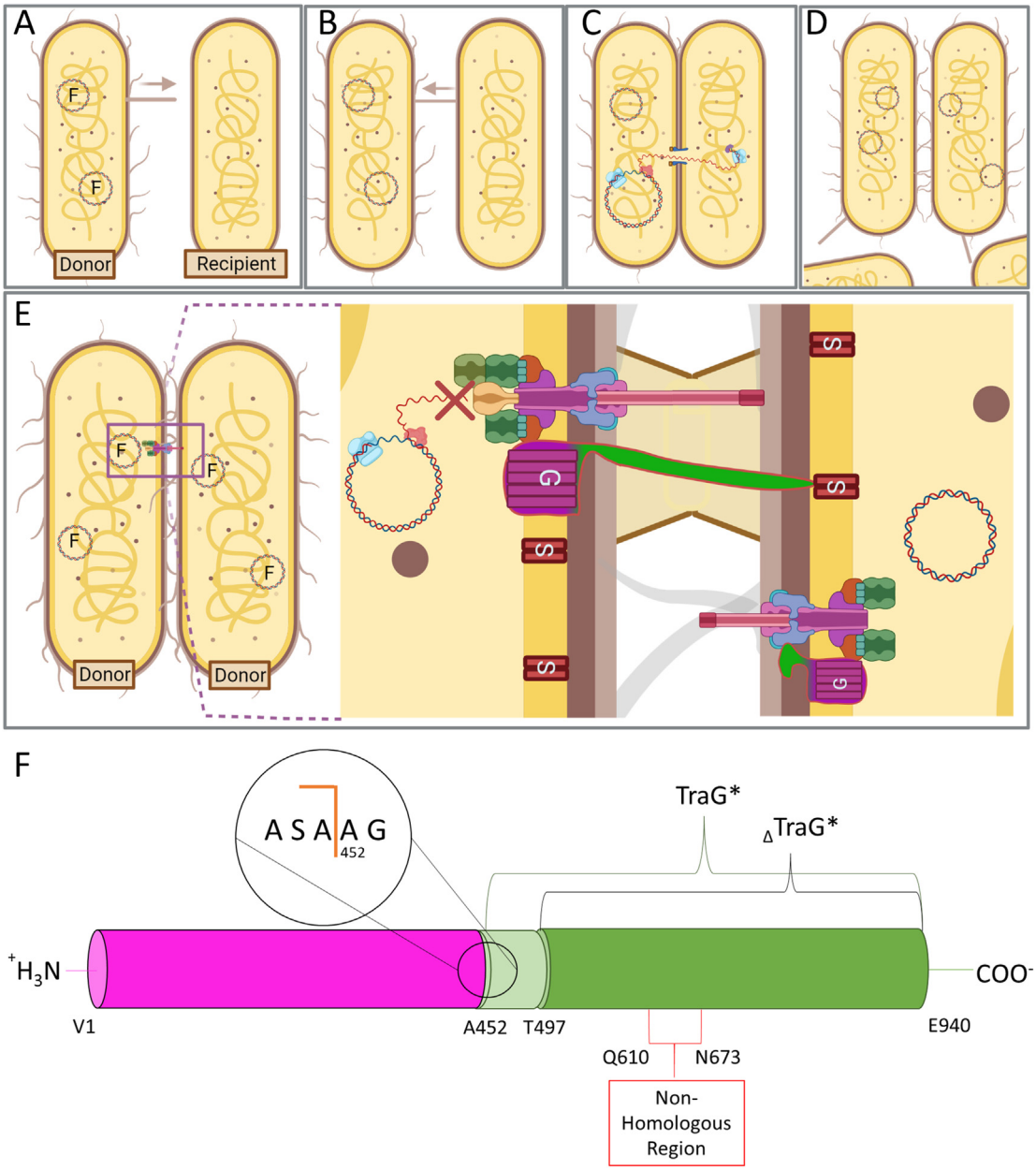

FIG. 1. A simplified depiction of gram-negative bacterial conjugation and entry exclusion by TraG in F-like T4SSs. (a) Donor bacteria possessing F-like plasmids coding for conjugative T4SS are piliated and will repeatedly extend and retract conjugative pili to contact a neighboring bacterium.[101] (b) Pilus retraction occurs when a neighboring recipient bacterium's cell wall has been contacted by the pilus tip of the donor cell, allowing for adhesion.[102] (c) Donor and recipient bacteria are brought into close proximity, and after surface and entry exclusion steps, a mating junction forms resulting in the fusion of lipid bilayers. The synchronous transfer and replication of a single strand of nicked plasmid DNA from the donor to recipient is then mediated by the relaxasome (pink) and transferosome (purple).[6] Rolling circle replication by DNA replication machinery (light blue) in the recipient allows for synthesis of the complementary strand of the plasmid DNA. (d) The conjugative transfer event ends after the single-stranded DNA is transferred to the recipient cell and the mating junction has closed. The recipient cell, following synthesis of the complementary strand of the transferred plasmid DNA, is now a donor cell for the conjugative plasmid. (e) In a colony of bacteria carrying plasmids with identical transfer (*tra*) regions, conjugation between donor cells is prevented through entry exclusion (Eex), proposedly in the event of faulty surface exclusion.[32,100] The formation of a complete mating pore and the transfer of redundant plasmid DNA is prevented through interaction of TraG in the donor cell with a cognate TraS embedded in the recipient's IM. [Figures (a)–(e) made with Biorender]. (f) Generalized domain organization of TraG. The N-terminal transmembrane region of the protein responsible for aiding in pilus assembly is colored in magenta and the C-terminal periplasmic region termed TraG* is colored in green. These regions are separated by a signal I peptidase cleavage site at residue A451.[32] The 45-residue truncation mutant construct is labeled as $_{\Delta}$TraG* and contains the region responsible for interaction with TraS. The non-homologous region, which differs between $TraG_F$ and $TraG_{R100}$, is responsible for Eex. The region shaded in light green from A452 to A496 represents a flexible linker region between the two domains.

© Author(s) 2022

Escherichia.[22] These species can survive harsh conditions and are common human pathogens. The R100 plasmid is from the F-like plasmid family based on genetic phylogeny to relaxase and coupling protein genes, and differs from F at the level of relaxosome formation, regulation, pilus serology, pilus-specific phage sensitivity, surface and entry exclusion between the two transfer systems, and in replicon type.[23–25] The R100 plasmid is one of the classic F-like plasmids; it was determined to be widespread among gram negative bacterial species due to its high propensity to harbor antibiotic resistance genes.[22,26] These systems are important to study as increasing numbers of multi-drug resistant bacterial pathogens have been observed globally as attributed to poor nosocomial hygiene, excessive use of antibiotics in animal agriculture, over-prescription of broad spectrum antibiotics, and furthered by the COVID-19 pandemic due to de-prioritization of bacterial infection prevention and control, over-sanitization with biocides, shortage of personal protective equipment, and crowding of infected hospital patients.[27–31]

In the F-like T4SS superstructure, two IM-bound proteins TraG and TraS are responsible for an event known as entry exclusion (Eex), which functions to prevent redundant donor–donor DNA transfer [Fig. 1(e)].[32] TraG in the donor cell is thought to scan the IM of the recipient cell and interact with the cognate TraS to prevent conjugation. If the TraG–TraS interaction does not occur then TraG interacts with TraN in the OM of the donor cell, resulting in mating pair stabilization (Mps) and advancing the conjugation process.[17,33] Interestingly, Eex systems of the closely related plasmids F and R100 retain plasmid specificity; TraG of the T4SS transfer apparatus present in donor cells will interact with the cognate TraS in the recipient cell to prevent redundant conjugation, but will not interact with TraS from a different plasmid and therefore allow conjugative DNA transfer.[32] Additionally, TraG will not interact with TraS in the same IM, and it functions in a trans fashion only.

TraG is a ∼103 kDa protein with an N-terminal polytopic IM-bound domain and a periplasmic C-terminal domain from amino acid residues A452-Q940 referred to as TraG* [Fig. 1(f)].[32,34] TraG is multifunctional as the N-terminal domain is responsible for pilus assembly, and the periplasmic TraG* is required for Mps and Eex; the protein must be intact for proper function in both pilus assembly and Eex/Mps.[32,35,36] TraG from the F ($TraG_F$) and R100 ($TraG_{R100}$) plasmids have an overall sequence identity of 93%; however, the region in TraG* from residues 610 to 673 shows only 17%. This region of TraG was shown to be responsible for plasmid specificity in entry exclusion, indicating the role of the region for interacting with their cognate TraS.[32]

A high-resolution structure of TraG* has the potential to elucidate the mechanism by which TraG interacts with the cognate TraS of a recipient cell to affect Eex. Disrupting the Eex process to promote lethal zygosis through ceaseless conjugation, or to activate TraG and the Eex process to prohibit conjugation, could allow for the development of a novel class of antibiotics.[37,38] In this study, we report on the structural dynamicity of TraG* in solution. The molecular weight of TraG* (∼50 kDa) complicates nuclear magnetic resonance studies, and the structural dynamicity of the protein complicates crystallization. Protein modeling and analysis using the phase separating protein predictor (PSP)[39] software identified a highly flexible N-terminal domain that is hypothesized to serve as a linker region between the two functional TraG domains, which is supported with predicted models from RoseTTAFold and AlphaFold.[39–42] An N-terminal truncation of TraG* from the R100 plasmid ($TraG^*_{R100}$) was designed based on these data, henceforth referred to as $_{\Delta}TraG^*_{R100}$. The thermal, chemical, and conformational stability of this mutant was confirmed to be improved relative to $TraG^*_{R100}$ using differential scanning fluorimetry (also referred to as Thermofluor assays), circular dichroism (CD), and collision induced unfolding–mass spectrometry (CIU–MS). Size exclusion chromatography linked to multiangle light scattering and small angle x-ray scattering (SEC–MALS–SAXS) studies shows the improved solubility of $_{\Delta}TraG^*_{R100}$ based on a lowered propensity for aggregation resulting in improved data and a more reliable bead model reconstruction.

II. MATERIALS AND METHODS

Details into the expression and purification of TraG* proteins in this study can be found in Sec. S1 of the supplementary material.[105]

A. Differential scanning fluorimetry/Thermofluor assay

To quantify the beneficial properties of buffers and salts in aiding the stability of proteins, and to compare the thermal stability of $TraG^*_{R100}$ and $_{\Delta}TraG^*_{R100}$, Thermofluor assays were performed. Thermofluor involves the use of a fluorescent hydrophobic probe to determine optimal solvent-protein interactions using the melting temperature (T_m) of the protein as an indicator of stability.[43–46] SYPRO orange dye was used as the probe, and it was prepared through serial dilution from a 5000× commercial stock (ThermoFisher) to a final concentration of 10× in the buffers of interest, with 1 mg/ml of $TraG^*_{R100}$ (purified with 5% glycerol) in a total of 50 μl. Fluorescence output was measured on a RotorGeneQ thermocycler (Qiagen) with output set to 470 nm and detected at 610 nm, with a gain of 7; samples were heated from 25 to 99 °C at a ramp rate of 1 °C/min. The choice of buffers used for this assay was inspired by Seabrook and Newman.[46] A total of 14 buffers were tested at 50 mM concentration with either 50 mM NaCl or 200 mM NaCl, and each condition was performed in triplicate, with a lysozyme positive control in 1 × PBS (pH 7.4), dye-only negative control and a protein-only negative control in every experiment. This was repeated for both proteins to gather a total of six data sets for each buffer condition. The data were analyzed using DSFworld[47] which allowed for crude normalization, averaging of the curves to produce more accurate T_m values and aided in creating graphs of the output thermal melting curves.

B. Circular dichroism

The molar ellipticity (mdeg) of $TraG^*_{R100}$ mutants was measured by CD spectroscopy at a protein concentration of 2 μM over wavelengths 190–260 nm. To normalize the CD spectra of the proteins, the spectrum of the solvent [10 mM 2-(N-morpholino)ethanesulfonic acid (MES) pH 6.5, 10% glycerol] was subtracted from the experimental spectra. All spectra were obtained on a J-815 CD spectrometer (Jasco) through a continuous scan performed at 100 μm/min with molar ellipticity measurements every 0.1 nm, with an accumulation factor of 8. Urea denaturation experiments were performed to provide an indication of protein chemical stability when comparing $TraG^*_{R100}$ with and without the putative intrinsically disordered region (IDR), as denaturing studies are common for determining the relative stability of protein mutants.[48–50] $TraG^*_{R100}$ variants were added to solutions with

© Author(s) 2022

different urea concentrations such that the final concentration of protein was 2 μM. Samples were incubated at 25 °C for 1 h prior to CD measurements. Urea concentrations of 0.5, 1, 2, 3, 4, and 5 M to determine the approximate range for protein denaturation. Spectra measurements were performed with the same scanning protocol as stated above, and all results were normalized to their respective solvent conditions. Deconvolution of all CD data was performed using BeStSel,[51] an algorithm for protein fold recognition and secondary structural determination using input CD spectra, unique in its ability to distinguish parallel from antiparallel β-sheets and useful in converting raw data measurements of molar ellipticity to $\Delta\varepsilon$ based on concentration, molecular weight of the protein, and path length (0.1 cm).

C. Collision induced unfolding mass spectrometry

TraG*$_{R100}$ samples were prepared for MS analysis using a 5 kDa molecular weight cutoff (MWCO) dialysis cassette to buffer exchange the protein at 4 °C into 100 mM ammonium acetate (MS grade), pH 6.6. Native MS analysis was performed on a Synapt G2S (Waters) with 5 μM TraG*$_{R100}$ or $_{\Delta}$TraG*$_{R100}$ flowing at a rate of 5 μl/min. In the collision induced unfolding (CIU)–MS experiments, data were collected from 5 V trap collision energy (CE) to 150 V in 5 V increments, with an acquired m/z range of 2000–5000 with no manual trapping. Capillary voltage was set to 3.0 kV, the sampling cone was 150.0, with source and desolvation temperatures at 120 and 250 °C, respectively, cone gas flow 71.0 l/h, nanoflow gas pressure at 2 bar, desolvation gas flow at 600 l/h, the transfer CE was 10.0 V, the trap gas flow was 4.0 ml/min, and the ion mobility spectrometry (IMS) wave delay was 1 ms with a wave height start of 10 V and a wave end height of 40 V.

D. SEC–MALS–SAXS

TraG*$_{R100}$ and $_{\Delta}$TraG*$_{R100}$ were purified as described (supplementary materials Sec. S1[105]); however, a desalting step was not performed. The protein was dialyzed using a 25 kDa MWCO dialysis membrane into 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 100 mM NaCl, 5% glycerol, (and 0.05% NP40 for TraG*$_{R100}$) pH 7.0, and further dialyzed into 30 kDa MWCO concentrator using a 10× dilution of a buffer stock [200 mM HEPES, 1 M NaCl, 50% glycerol (0.5% NP40 for HisTraG*$_{R100}$), pH 7.0, 0.2 μm membrane filtered]. Three hundred milliliters of this buffer stock (200 ml for the NP40 buffer) was chilled at −20 °C and sent to the BioCAT facility at the Advanced Photon Source along with 1000 μl of 5 mg/ml HisTraG*$_{R100}$, and 1000 μl of 5 mg/ml His$_{\Delta}$TraG*$_{R100}$, as quantified using the Edelhoch method. SAXS diffraction data were collected using the parameters seen in Table I. Samples were injected onto a Superdex 200 10/300 Increase SEC column (Cytiva) using an Agilent Infinity II HPLC, then sent through a Wyatt DAWN Heleos II MALS instrument and a Wyatt Optilab T-rEX direct refractive index (dRI) detector, and finally into the SAXS flow cell in the path of the synchrotron beamline. Data were analyzed using the RAW and ATSAS software packages.[52,53] Bead models were generated from the output files of 20 cycles using the integrated DAMMIF tool with DAMAVER averaging and refinement, and clustering using DAMCLUST. SAXS data for TraG* from the R100 plasmid was submitted to SASBDB (https://www.sasbdb.org) under accession codes SASDQG6 (ΔTraG*$_{R100}$) and SASDQH6 (TraG*$_{R100}$).[104]

TABLE I. SEC–MALS–SAXS data collection parameters for HisTraG*$_{R100}$, and His$_{\Delta}$TraG*$_{R100}$.

Data collection parameters		
Instrument	BioCAT (Sector 18, APS)	
Detector	Eiger2 XE 9M	
Wavelength (Å)	1.033	
q-measurement range (1/Å)	0.0028–0.42	
Exposure time (s)	0.5	
Size exclusion column	Superdex 200 10/300 increase	
Flow rate (ml/min)	0.6	
Temperature (°C)	20	
Protein	HisTraG*$_{R100}$	His$_{\Delta}$TraG*$_{R100}$
Concentration (mg/ml)	5.0	4.8
Loaded volume (μl)	300	300
Buffer	20 mM HEPES, 100 mM NaCl, 5% glycerol, 0.05% NP40, pH 7.0	20 mM HEPES, 100 mM NaCl, 5% glycerol, pH 7.0
Structural parameters		
$I(0)$ (cm^{-1}) [from P(r)]	$0.79 \pm 9.46 \times 10^{-3}$	$0.02 \pm 3.9 \times 10^{-5}$
R_g (Å) [from P(r)]	91.23 ± 3.4	44.14 ± 0.23
$I(0)$ (cm^{-1}) (from Guinier)	$0.71 \pm 4.13 \times 10^{-3}$	$0.02 \pm 3.11 \times 10^{-5}$
R_g (Å) (from Guinier)	57.68 ± 0.73	41.27 ± 0.81
D_{max} (Å)	450	175

III. RESULTS

A. Structural prediction software demonstrated large loop regions in TraG models

The phase separation predictor (PSP) software for predicting intrinsically disordered regions (IDRs), which relies upon the propensity of a primary protein sequence for long range pi–pi contacts,[39] provided insight in determining which regions of TraG, the conjugative entry exclusion protein found in F-like T4SS, should be targeted for mutation or deletion to improve protein stability. The region from residues 447–498 is predicted to have phase separating qualities in both TraG from the F plasmid (TraG$_F$) and TraG from the R100 plasmid (TraG$_{R100}$), with 27 of these residues surpassing the PSP algorithm's score threshold for the designation of residues as intrinsically disordered (4.0),[39] thus providing TraG$_F$ and TraG$_{R100}$ with overall scores of 4.06 and 3.74, respectively [Figs. 2(a) and 2(b)]. The lower score of TraG$_{R100}$ deemed it more favorable to undertake further experiments.

The predicted model of TraG$_F$ in the AlphaFold database (code: AF-B1VCA9-F1) indicated high–medium confidence in the prediction of the N-terminal membrane bound domain with medium-low confidence throughout the C-terminal periplasmic domain [Fig. 2(f)].[40,41] Notably, the hypothesized flexible linker region from A452 to A496 and the non-homologous region from R610 to D673 were predicted to be loop regions in the RoseTTAFold [Fig. 2(d)] and AlphaFold [Fig. 2(f)] models. However, a portion of the putative linker region was modeled as a pseudo β-sheet in RoseTTAFold and a β-sheet in

© Author(s) 2022

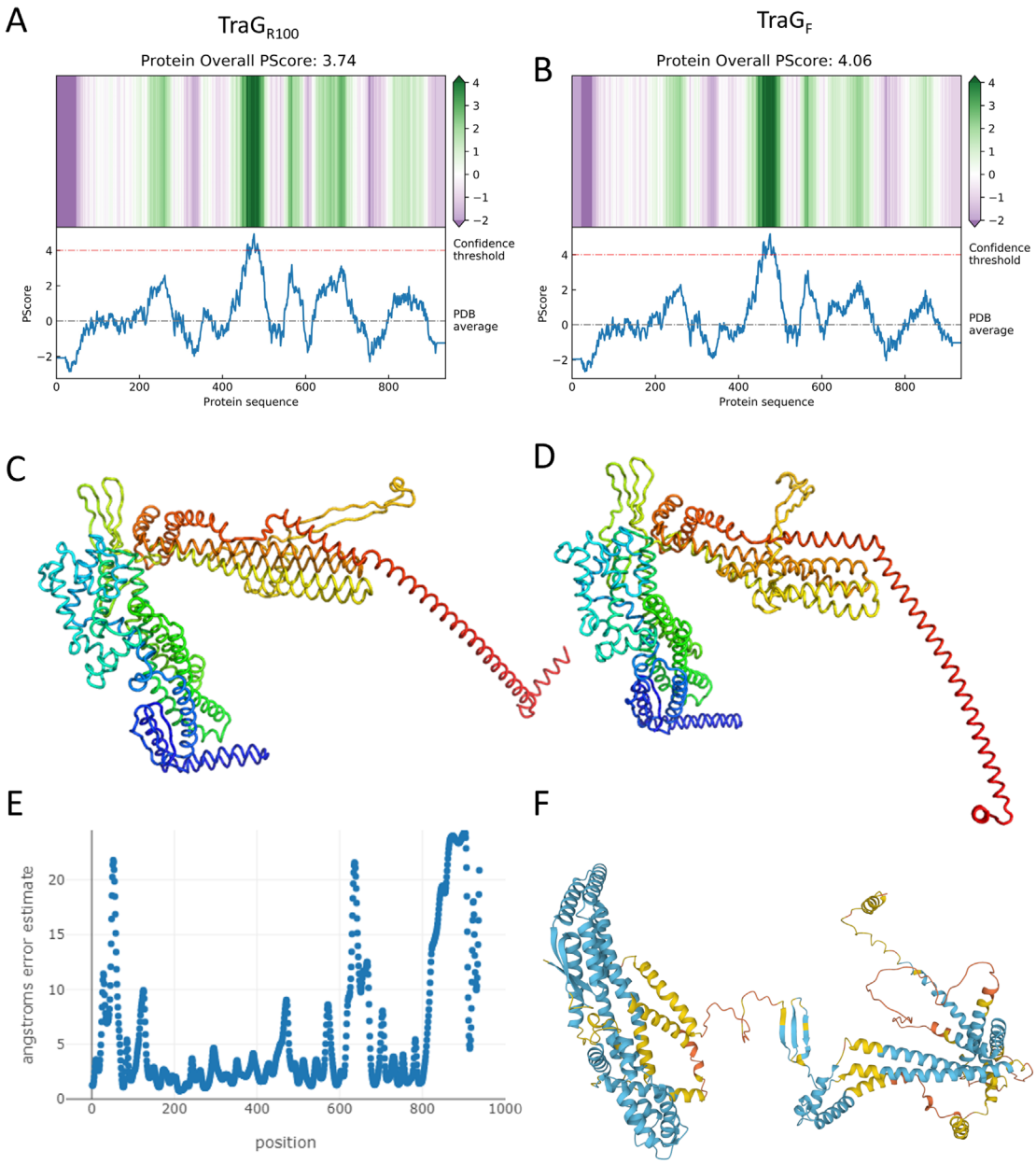

FIG. 2. Predicted disorder content and structural models of TraG constructs from F and R100 plasmids using the PSP, RoseTTAfold and AlphaFold.[39–42] The upper images show the resultant 2D heat maps and graphs produced by entering the (a) TraG$_{R100}$ and (b) TraG$_F$ sequence into the PSP.[39] The only region predicted to be a phase separating IDR is the putative linker from residues 447 to 498, as it is the only region that surpasses the designated confidence threshold. (c) TraG$_{R100}$ as predicted by RoseTTAfold, color mapped by chainbow from N- to C- terminus, with a confidence value of 0.55. (d) TraG$_F$ as predicted by RoseTTAfold, with a confidence value of 0.52; the corresponding chart displaying error estimate per modeled residue is shown in (e). (f) Full-length TraG$_F$ as predicted by AlphaFold, color mapped based on per residue modeling confidence, where blue is 90%–70%, yellow is 70%–50%, and orange is <50%. No residue in the model is mapped with high confidence (>90%). RoseTTAfold models were displayed using Pymol.[103]

AlphaFold. In all models, the final 100 residues of the protein are poorly predicted; the plot displaying the per-residue error estimate in Fig. 2(e) illustrates these observations. Similar predictions were made in the TraG$_{R100}$ RoseTTAFold models; however, the confidence value of the model was slightly higher at 0.55 vs TraG$_F$ at 0.52. The modeling software provided some distinctions in the TraG models, in that the C-terminus is largely unstructured in the AlphaFold TraG$_F$ model, while using RoseTTAFold it is predicted to be a large helix modeled in different orientations depending on *in silico* N-terminal deletions and sequence changes (supplementary material Fig. S1).

© Author(s) 2022

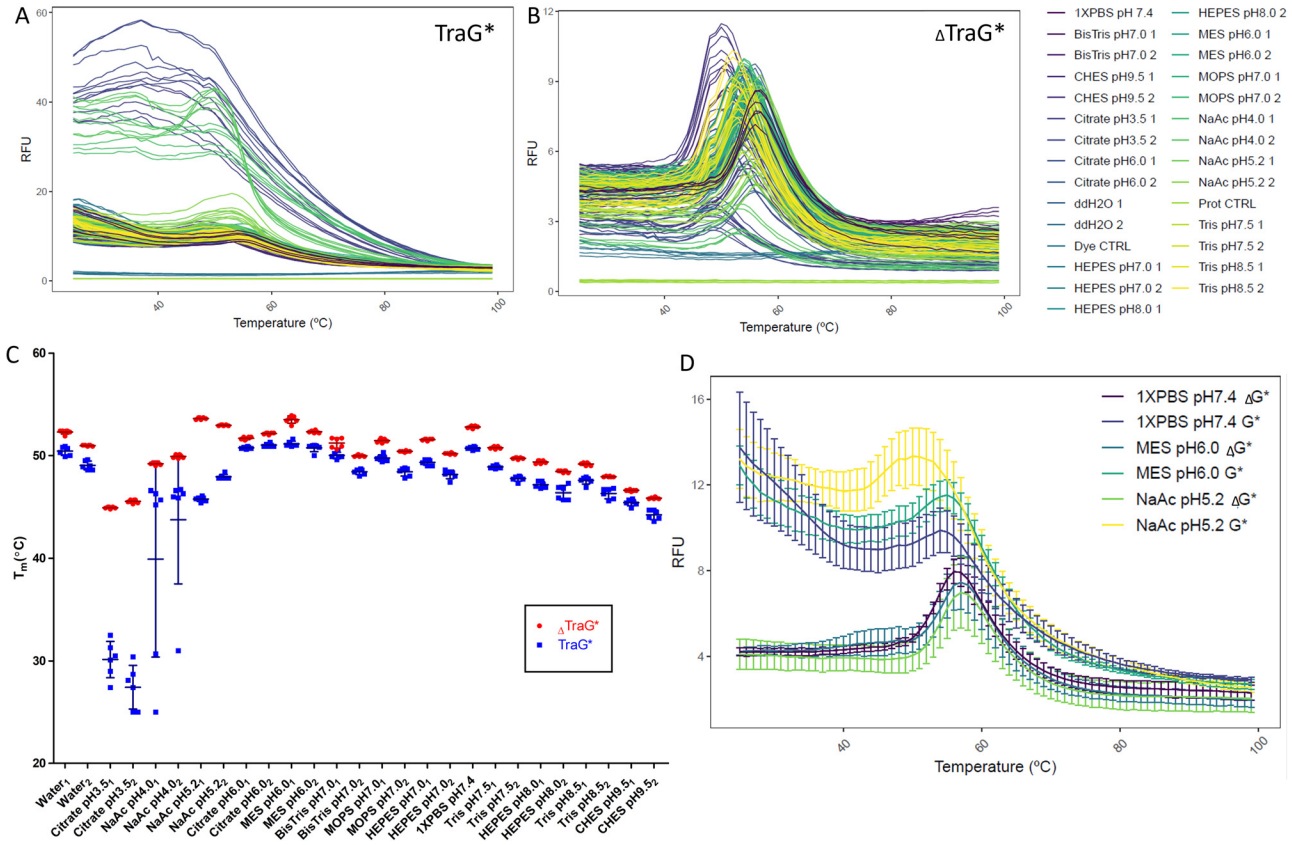

FIG. 3. Melting curve plots displaying the temperature dependent unfolding of (a) HisTraG*$_{R100}$ and (b) $_{\Delta}$TraG*$_{R100}$ in various buffer conditions [seen in the legend right of (b)] as detected using SYPRO orange. Each buffer was at a concentration of 50 mM [except Nanopure water (ddH_2O) and PBS], where one represents the addition of 50 mM NaCl in the buffer and two includes 200 mM NaCl. Six (6) replicates of each experiment are shown. (c) Mean melting temperatures with ±SD of the TraG* variants in all tested buffer conditions. (d) Average melting curve plots of the top three stabilizing buffer conditions for both $_{\Delta}$TraG*$_{R100}$ and TraG*$_{R100}$, as determined by the T_m of these average curves.

B. Thermofluor indicates enhanced thermal stability of $_{\Delta}$TraG*$_{R100}$

Thermofluor experiments were performed to compare the thermal stability of the TraG* variants. These assays involved the use of a hydrophobic fluorescent probe that is quenched upon addition to an aqueous sample of protein in a buffer.[54] Heat is applied to the samples in a gradient fashion; as the protein begins to unfold it exposes hydrophobic residues that the probe binds to and fluoresces. The fluorescence signal is measured through spectrophotometry, as the heat continues to increase the protein begins to aggregate and hydrophobic residues become excluded from solvent, resulting in a loss of probe binding and a decrease in fluorescence. The produced profile is a melting curve that can allow for characterization of thermal stability through comparison of the melting temperature (T_m), which is the midpoint of the unfolding transition, as determined through the highest value on the second derivative of the function. The assay can be altered to optimize buffer conditions to aid in protein stability, characterize the effect of mutations on the thermal stability of proteins, predict the crystallization propensity of proteins, and determine quantitative binding affinities between a protein and a ligand.[54–56]

Thermofluor experiments demonstrated that $_{\Delta}$TraG*$_{R100}$ is more stable than TraG*$_{R100}$ based on lower average T_m values in 25 of the 27 tested buffer conditions (Fig. 3). In the remaining two conditions the error bars overlap, indicating there is no difference in their T_m when in the buffer. The difference in T_m between TraG*$_{R100}$ constructs is minimal in many cases; however, the shapes of TraG*$_{R100}$ melting curves are poorer than $_{\Delta}$TraG*$_{R100}$, as the high fluorescence in the lower temperatures (25–45 °C) and the shallow peak indicates there are some TraG*$_{R100}$ protein species present which have hydrophobic regions exposed immediately upon SYPRO addition [Fig. 3(a)].[43] The variability in the TraG*$_{R100}$ melting curves is higher than the truncated protein as demonstrated by the larger error, which may further indicate the instability of the protein as different batches produce protein samples with different thermal stability, while $_{\Delta}$TraG*$_{R100}$ remains consistent between batches [Fig. 3(c)].[47] Sodium acetate pH 5.2 and MES pH 6.0 were observed to be optimal buffers for TraG*$_{R100}$ variants based on this assay [Fig. 3(d)]. As well, buffers with 50 mM

© Author(s) 2022

NaCl appeared to have a more stabilizing effect than buffers with 200 mM NaCl.

C. Improved chemical stability of $_{\Delta}$TraG*$_{R100}$ as observed by circular dichroism

Circular dichroism (CD) is a widely used and well-developed technique for the determination of protein secondary structure.[51,57] The technique relies upon the differential absorption of left- and right-hand circular polarized light. In this case, the electric field of a photon has a circularly rotational direction relative to the direction of propagation, where the photon vector remains constant in magnitude.[58,59] When these photons pass through an asymmetric (chiral) molecule, the speed, absorbance and the wavelengths differ depending on the direction of their polarization. As the sum of the vectors for the right- and left- handed polarized light forms an ellipse, the change in ellipticity (measured as $\Delta\varepsilon$) of a substance as a function of the wavelength of incident light is reported in performing CD experiments, although the change in absorbance of the differentially polarized light is what is measured in the CD spectrometer. As proteins are chiral molecules and secondary structures result in specific absorbance patterns at certain emission wavelengths, CD is widely used to determine the secondary structure content of proteins and can be used to determine the structural stability of proteins in solution.[58,60,61]

The CD spectra of both TraG*$_{R100}$ variants appear to be mainly α-helical based on the characteristic negative $\Delta\varepsilon$ peaks at 222 and 208 nm (Fig. 4).[59] Deconvolution in BeStSel (Table II) indicates differences in the extent to which the proteins fold into canonical α helices.[51] The model providing the best fit for TraG*$_{R100}$ indicated that 72.3% of the protein was predicted to be α-helical, albeit 20.9% of helices are of a bent or imperfect topology (supplementary material Fig. S2). Turns were of low abundance in the overall structure of the protein at 7.4%, and the remaining 20.3% of the protein was deemed as "other," which include 3_{10} and π helices, but also loops that may not change the direction of polarized UV light and are therefore undetectable by CD.[51,62,63] The spectrum of $_{\Delta}$TraG*$_{R100}$ appears to have better defined peaks at 222 and 208 nm, and the protein is predicted to have 85.1% α-helices, with 67% of optimal topology contrasting with the 51.4% seen in TraG*$_{R100}$. $_{\Delta}$TraG*$_{R100}$ has only 14.8% other character and no strongly structured turns. This strongly indicates that the 45 residues removed from the protein are disordered; the flexible N-terminal region in TraG*$_{R100}$ could be causing distortions in the remaining protein fold based on the predicted high relative change in the percentage of assigned secondary structure.

Urea denaturation studies were performed to support the conclusion surmised from the Thermofluor data that $_{\Delta}$TraG*$_{R100}$ has higher stability than TraG*$_{R100}$. In comparing spectra and tables displaying BeStSel[51] predicted secondary structure topology of TraG*$_{R100}$ to $_{\Delta}$TraG*$_{R100}$, it is evident that the presence of residues A452-A496 destabilizes TraG*$_{R100}$ (Fig. 4, Table II). TraG*$_{R100}$ displayed higher susceptibility to unfolding based on the direct decrement of $\Delta\varepsilon$ at 222 and 208 nm as the urea concentration was increased. The secondary structure of $_{\Delta}$TraG*$_{R100}$ remained relatively intact until incubation with 3.0M urea, while the secondary structure of TraG*$_{R100}$ was partially unfolded at 1.0M urea based on the topology analysis from 260 to 200 nm by BeStSel (Table II).[51]

D. CIU–MS demonstrates differences in conformational stability of TraG*$_{R100}$ and $_{\Delta}$TraG*$_{R100}$

Collision-induced unfolding (CIU) mass spectrometry is a reliable technique for studying and comparing a protein's conformational stability.[64–67] CIU–MS employs an ion mobility spectrometry (IMS) cell within the mass spectrometer that is filled with an inert gas and provides a weak electric field gradient that filters ionized species separated in the previous trap cell. Drift time in the IMS cell, which is defined by the time an ion takes to travel the distance from the drift tube to the detector, is dependent on mass and size as well as the conformational shape of the fragment. In CIU–MS, the trap collision energy (CE) is increased in a stepwise fashion, allowing for breakages of non-covalent bonds that can be visualized by IMS as a shift to a slower moving, unfolded species. Therefore, CIU provides an indication of the conformational stability of proteins; a protein that has a lower number of drift time distributions (representing different

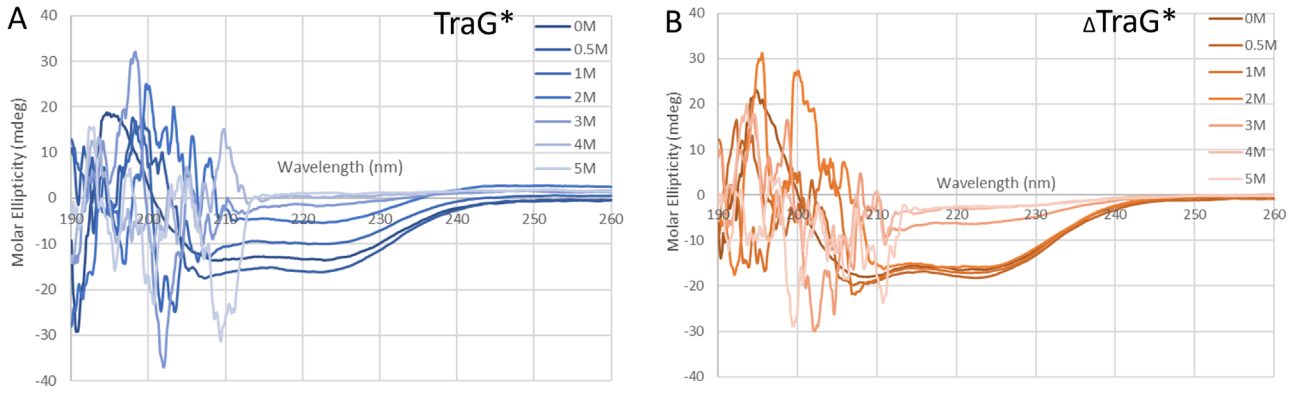

FIG. 4. Urea denaturation of (a) TraG*$_{R100}$ and (b) $_{\Delta}$TraG*$_{R100}$ as measured by CD spectroscopy. All experiments were performed with a final protein concentration of 2 μM and were incubated in the respective concentrations of urea (shown in the figure legends) for 1 h at 25 °C prior to UV measurement. The resultant data were normalized by subtracting using the blank spectra, wherein the buffer conditions in the same concentration of urea were replicated and CD data were collected. Molar ellipticity measurements were obtained every 0.1 nm from 260 to 190 nm. Data from 195–260 nm was entered into BeStSel for deconvolution,[51] and the resultant analyses from triplicate runs is shown in Table II; spectra shown are from single representative experiments. Baseline spectra for both proteins are shown in supplementary material Fig. S2.

© Author(s) 2022

TABLE II. Secondary structure of $TraG^*_{R100}$ and $\Delta TraG^*_{R100}$ under urea denaturation. Data presented from triplicate measurements and deconvoluted with BeStSel. Representative CD spectra shown in Fig. 4.

$TraG^*_{R100}$								
	α-Helix (%)		Antiparallel β-sheet (%)					
(Urea) (M)	Regular	Distorted	Left-twisted	Relaxed	Right-twisted	Parallel β-sheet (%)	Turn (%)	Loop (%)
0.0	40.70	17.10	0.00	2.97	0.00	0.00	8.47	30.77
0.5	18.37	16.80	0.00	0.00	14.73	1.03	9.10	40.03
1.0	19.13	19.07	6.97	3.60	12.33	0.00	12.37	26.57
2.0	8.50	9.00	5.07	2.60	9.90	6.70	11.27	47.00
3.0	5.07	11.17	2.33	0.00	12.43	0.00	15.63	53.37
4.0	0.00	0.00	3.50	21.13	22.80	0.00	19.90	32.70
5.0	8.10	10.80	3.40	5.93	8.80	0.00	15.20	47.77
$\Delta TraG^*_{R100}$								
	α-Helix (%)		Antiparallel β-sheet (%)					
(Urea) (M)	Regular	Distorted	Left-twisted	Relaxed	Right-twisted	Parallel β-sheet (%)	Turn (%)	Loop (%)
0.0	64.20	17.13	0.00	0.03	0.03	1.47	6.77	10.30
0.5	34.63	16.57	0.00	0.00	26.83	0.87	7.07	14.10
1.0	28.10	19.40	9.03	0.00	7.23	0.77	8.77	26.70
2.0	24.43	10.80	10.20	0.00	8.57	10.03	5.87	30.17
3.0	7.50	8.27	0.00	6.77	17.77	0.00	18.40	41.33
4.0	9.37	0.00	0.00	15.50	12.27	3.57	20.67	38.63
5.0	1.37	9.00	0.90	3.30	5.90	0.00	18.83	58.63

conformations) and/or remains in a lower drift time distribution at a higher CE has higher conformational stability. The native mass spectra of $TraG^*_{R100}$ and $_{\Delta}TraG^*_{R100}$ were highly reproducible in biological replicates, where $TraG^*_{R100}$ had highly abundant charge states of 18+ (3144 m/z), 17+ (3329 m/z), and 16+ (3537 m/z) while $_{\Delta}TraG^*_{R100}$ had more abundant 17+ (3054 m/z), 16+ (3245 m/z), and 15+ (3462 m/z) charge states (supplementary material Fig. 3). The propensity for $TraG^*_{R100}$ to occupy higher energy charge states more frequently indicates that the presence of the putative linker region results in lower stability.[68] The mass of each protein as determined using ESIprot was 56 588 ± 8.71 Da for $TraG^*_{R100}$ and 51 906 ± 7.54 Da for $_{\Delta}TraG^*_{R100}$.[69]

CIU data were analyzed using charge states 17+ and 16+ for both proteins, as the 15+ and 18+ charge states were of insufficient abundance in some CEs to be compared. To visualize changes in drift time of these species as a function of CE, CIUSuite2[70] was used to generate CIU heat maps (Fig. 5). The replicates of each plot have some differences in the intensity of the transitions; however, there are common shifts in drift time for the same species (supplementary material Fig. S5). The 16+ charge state of $TraG^*_{R100}$ displays two shifts, one at ∼35 V CE and one at ∼70 V CE. The 16+ charge state of $_{\Delta}TraG^*_{R100}$ has a single shift at ∼25 V CE; however, there appears to be a low abundance of an unfolded state from ∼60 V CE onward. Although $_{\Delta}TraG^*_{R100}$ may transition to a conformational state with slightly less energy, the lower abundance of the species with the longest drift time seen in $TraG^*_{R100}$ indicates a significantly unfolded conformation is achieved more commonly when the putative linker region is present. A similar trend is seen in the 17+ charge state as well. This can also be visualized by observing the trends in the IMS spectra, which are more reproducible and comparable between replicates (supplementary material Fig. 4).

E. SEC–MALS–SAXS aids in defining $_{\Delta}TraG^*_{R100}$ monomer shape

Size exclusion chromatography linked to multi-angled light scattering and small angle x-ray scattering (SEC–MALS–SAXS) is an equilibrium based SAXS method, which links multiple biophysical techniques to accurately inform on particle shape and size.[71] SAXS involves the illumination of an aqueous sample by a collimated monochromatic x-ray beam, where the particles in solution will scatter the x rays.[72] The intensity of this scattering is detected; then, the scattering of the solvent without the particles of interest is determined and subtracted to provide the scattering pattern of the particles of interest. This is simplified through the inclusion of SEC as particles will be separated by size, and highly accurate solvent frames can be collected prior to the void volume and after the total column volume.[71] Coupling SEC –SAXS with MALS, which is the use of monochromatic incident light to cause Rayleigh scattering that is detected upon 17 angles, and a direct refractive index (dRI) detector, which measures refractive indices of a solution in a flow cell and the solvent in a reference cell to determine the RI produced from the particles of interest, has become established as a robust method for acquiring highly

© Author(s) 2022

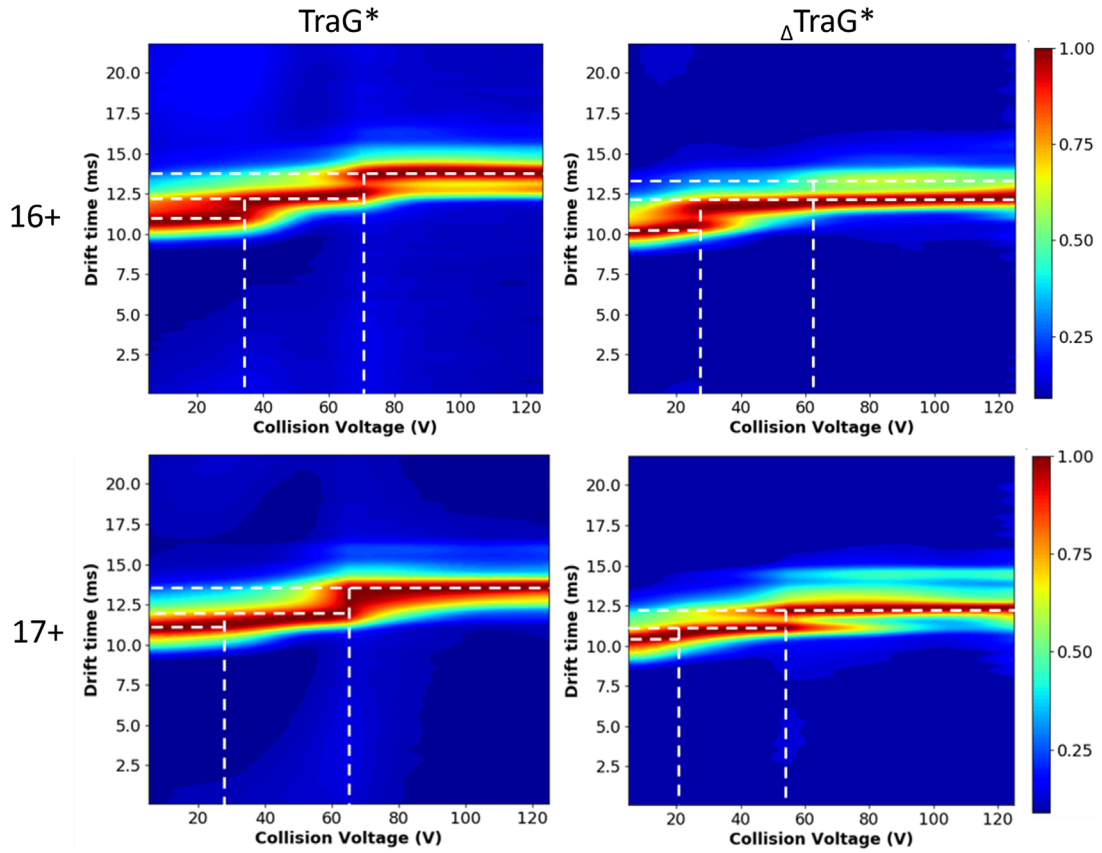

FIG. 5. CIU–MS heat maps displaying the drift time of the 16+ and 17+ charge states of TraG*$_{R100}$ and $_{\Delta}$TraG*$_{R100}$ as trap collision energy is increased from 5 to 125 V. These maps were plotted with CIUSuite2[70] and smoothed with default settings from the software. The m/z range set for plotting the IMS of the 17+ charge states were 3320–3355 for TraG* and 3040–3090 for $_{\Delta}$TraG*, and the m/z range used for the 16+ charge states were 3525–3575 for TraG* and 3230–3280 for $_{\Delta}$TraG*. A biological replicate for each experiment is shown in supplementary material Fig. S5.

accurate biophysical data.[71] MALS and dRI detectors provide a more accurate molecular weight determination than SAXS and provide radius of hydration R_h values (the radius of the hypothetical sphere of aqueous solvent that diffuses at the same rate as the macromolecule) to couple with radius of gyration R_g (the radius obtained from the rotationally averaged volume of the macromolecule) values obtained from SAXS.[71–73]

SEC–MALS–SAXS data of 5 mg/ml TraG*$_{R100}$ indicate that it had aggregated prior to experimentation [Fig. 6(a)]. A highly intense peak at frames 700–1100 is confirmed to result from high molecular weight aggregates of the protein as it was eluted in the SEC column's void volume. The low intensity of the monomeric protein peak at frames 1400–1480 results from loss of monomeric TraG* $_{R100}$ due to aggregate formation. This is also seen in the MALS data by the large light scattering peak at 10.5–12.5 ml and the lower light scattering from the $UV_{280\,nm}$ and the dRI peak seen in 20.5–21.5 min of elution (supplementary material Fig. S6). The resultant SAXS data were buffer subtracted using frames 315–400 and the sample region was chosen as frames 928–939. Evolving Factor Analysis (EFA) was performed on the region 1300–1520 to separate the peak associated with the aggregate from the monomeric protein peak. Two (2) significant singular values were used, with component ranges from frames 1300–1481 and 1340–1520, with the latter range plotted as the scattering profile for further analysis [Fig. 6(b)]. $_{\Delta}$TraG*$_{R100}$ at 5 mg/ml shows the presence of monomeric protein at frames 1409–1460, with no apparent aggregation seen [Fig. 6(a)]. The buffer region used for subtraction was 980–1157, and a linear baseline correction was performed using frames 26–70 and 2550–2594.

Guinier analysis (a function which allows for observation of scattering at the smallest q values) of the SAXS scattering profiles seen in Fig. 6(b) results in R_g and correlated molecular weight values of 57.68 ± 0.73 Å and 76.6 kDa, 41.27 ± 0.18 Å and 56.4 kDa for TraG*$_{R100}$ and $_{\Delta}$TraG*$_{R100}$, respectively (Guinier analyses seen in supplementary material Fig. S7). The R_g of $_{\Delta}$TraG*$_{R100}$, as calculated from the pair-distance distribution P(r) function from the GNOM program,[52,74] was reasonably close to the R_g from Guinier analysis at 44.14 ± 0.23 Å. However, the R_g of TraG*$_{R100}$ from the P(r) function was 91.23 ± 3.40 Å and the maximum dimension Dmax [the largest paired-distance achieved between points on the molecule as determined by the P(r) function][72] was 450 Å, which is highly overestimated and likely due to the presence of residual aggregated protein [Fig. 6(c)]. In SAXS, the P(r) function is used to describe the distances

© Author(s) 2022

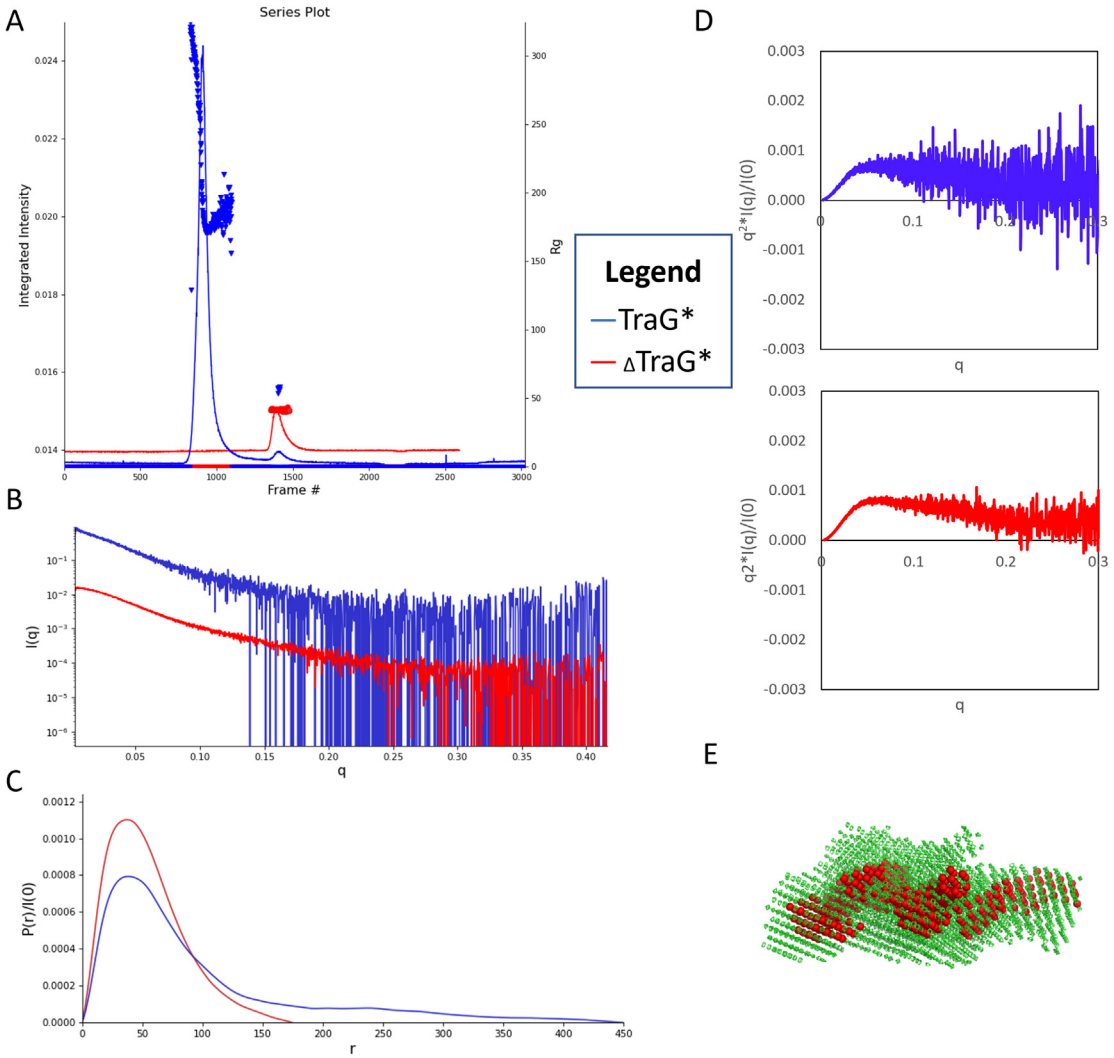

FIG. 6. (a) SEC–MALS–SAXS chromatogram displaying the integrated intensity of diffraction and radius of gyration (R_g) resulting from each frame of data collected for TraG*$_{R100}$ (blue) and $_{\Delta}$TraG*$_{R100}$ (red). (b) Scattering profiles of TraG*$_{R100}$ (blue) and $_{\Delta}$TraG*$_{R100}$ (red) show the abundance of noise in the TraG*$_{R100}$ profile and noise in the high q regions for $_{\Delta}$TraG*$_{R100}$. (c) P(r) functions of TraG*$_{R100}$ (blue) and $_{\Delta}$TraG*$_{R100}$ (red) were produced using GNOM(IFT) and optimized for R_g values comparable to those resulting from Guinier analysis; Dmax of TraG* was truncated at 450 Å while $_{\Delta}$TraG* was cut at 175 Å to allow for some. (d) Normalized Kratky plots resulting from the scattering profile of TraG*$_{R100}$ (blue) and $_{\Delta}$TraG*$_{R100}$ (red). (e) A bead model of $_{\Delta}$TraG*$_{R100}$ which displays the smallest shell (red) with mesh models (green) of the highest shell from DAMCLUST refinement, produced from the P(r) function. This reconstruction had an ambiguity score of 1.940 with 87 ambiguity categories based on AMBIMETER. All SAXS data images were produced using RAW, and bead models were visualized using Pymol.[52,103]

between all scattering points on the macromolecule, grouped into paired-sets, in order to predict the overall shape of the molecule.[75] In comparing the molecular weight estimates from the Guinier analyses to the estimates acquired from MALS data, the estimate for TraG*$_{R100}$ is proximal to the MW from the Guinier analysis at 75 710 ± 53.83 Da; however, the MW estimate for $_{\Delta}$TraG*$_{R100}$ is lower at 50 650 ± 1.06 Da, which is closer to the molecular weight determined from MS (51 906 ± 7.54 Da). The R_h values, as obtained from MALS, were 67.7 ± 0.014 and 47.8 ± 0.007 Å for TraG*$_{R100}$ and $_{\Delta}$TraG*$_{R100}$, respectively.

Kratky plots are used to qualitatively assess the flexibility and/or the degree of unfolding in samples and are mathematically expressed as $q^2I(q)$ vs q.[72,76,77] The shape of a Kratky plot dictates the protein's folding, globular proteins will have a Gaussian peak, unfolded proteins will have a plateau at high q, while prolate or partially unfolded proteins will have a combination of bell-shape and plateau. A normalized Kratky plot normalizes scattering profiles by mass and concentration by plotting $q^2I(q)/I(0)$ vs q. A normalized Kratky analysis was performed on both scattering profiles, and the $_{\Delta}$TraG*$_{R100}$ curve had a slight bell shape returning to the x-axis with higher q values, indicating some disorder is present [Fig. 6(d)]. The normalized Kratky plot for TraG*$_{R100}$ was highly unfolded based on the lack of a Gaussian peak; however, this qualitative analysis is obscured by significant noise at

© Author(s) 2022

high q, likely due to the low concentration of the monomeric protein. The Guinier analysis of $TraG^*_{R100}$ also showed that the monomeric peak has some residual aggregate, as the scattering profile displayed curvature at low q data (supplementary material Fig. S7). The Guinier analysis showed some minor curvature in low q data for $_{\Delta}TraG^*_{R100}$ (supplementary material Fig. S7) and the scattering profile showed some noise in high q data [Fig. 6(b)]. However, this could not be truncated as the shape of the P(r) function was desirable at a Dmax of 175 Å wherein the curve smoothly approaches the x-axis [Fig. 6(c)]. AMBIMETER was used to determine ambiguity of the potential output models; the $_{\Delta}TraG^*_{R100}$ model was reasonable with an ambiguity score of 1.940 and 87 ambiguity categories, while the $TraG^*_{R100}$ model was ambiguous as it had an ambiguity score of 0 with 0 compatible shape categories [Fig. 6(e)].[78] The $TraG^*_{R100}$ model is not shown as it is ambiguous; however, it can be viewed on the SASBDB (https://www.sasbdb.org) under accession code SASDQH6.[104] The produced $_{\Delta}TraG^*_{R100}$ bead model appears to be a coiled rod-like shape (SASDQG6),[104] whereas the $TraG^*_{R100}$ model is a less-defined and further extended structure. The models cannot be accurately compared as the additional lobes seen in the $TraG^*_{R100}$ bead model may result from disruption by the scattering of the residual aggregate protein rather than from the movement of a flexible N-terminal domain.

IV. DISCUSSION

TraG is a multifunctional protein in T4SS encoded on F-like plasmids, including the R100 plasmid. It is composed of a membrane-bound N-terminal domain required for pilus generation and a C-terminal periplasmic domain ($TraG^*$) required for Eex and Mps. $TraG^*$ is logically presumed to have some plasticity in its structure to facilitate these different functions that are expected to require numerous protein–protein interactions (PPIs). Based on the findings presented, it can be stated that the N-terminal region of $TraG^*$ from the R100 plasmid (aa A452-A496 of $TraG_{R100}$) is highly dynamic and its presence destabilizes the protein. The characteristics of higher thermal and chemical stability of the truncation mutant observed from Thermofluor and urea denaturation CD experiments are sufficient in stating that $_{\Delta}TraG^*_{R100}$ is more stable than $TraG^*_{R100}$. In addition, the reduced unfolded population observed in CIU–MS for $_{\Delta}TraG^*_{R100}$ provides further evidence that the removal of the N-terminal region stabilizes the protein. However, it is not sufficient in identifying the deleted region as an intrinsically disordered region (IDR). IDRs shorter than 50 residues are not uncommon, however they typically serve a functional purpose in the native protein.[79] As this region has not been shown conclusively to be intrinsically disordered in full-length $TraG_{R100}$, and its function is not yet confirmed, the region cannot be classified as an IDR. However, the $TraG^*_{R100}$ constructs explored herein provide evidence that the presence of the region from residues A452-A496 results in a protein which has a higher prevalence for aggregation as shown in SEC–MALS–SAXS (Fig. 6). The region likely serves as a flexible linker in the full-length protein; however, when expressed separately from the N-terminal membrane bound portion, this region of $TraG^*_{R100}$ displays higher dynamicity due to its expression as a synthetic construct.

AlphaFold and RoseTTAFold both predict the linker region of TraG to contain a pseudo β-sheet surrounded by loop regions [Fig. 2(f)]. This region may fold cryptically and uncoil like a spring to extend the periplasmic $TraG^*$ to its interacting partner TraS in the IM of the recipient cell. While AlphaFold and RoseTTAFold predictions provide evidence of a flexible N-terminal linker region in $TraG^*$, they also indicate that a C-terminal deletion may be required as well to make the protein more amenable to crystallization (Fig. 2). Based on the models which predict the protein as mainly consisting of extended α-helices, it is visualized why $TraG^*_{R100}$ aggregates excessively rather than crystallizing; non-covalent interactions promoted by crystallization conditions may induce the formation of nonspecific coiled-coils.[80] Additionally, native mass spectra show a large proportion of unstructured species in both $TraG^*_{R100}$ and $_{\Delta}TraG^*_{R100}$ (supplementary material Fig. S3), demonstrating the dynamic nature of the protein regardless of the N-terminal truncation. Another insight from the artificial intelligence (AI)-predicted structures is the region of $TraG^*$ known to interact with TraS is predicted to be an extended loop. This topology could be required for its functionality wherein this region may become structured only when interacting with its binding partner. Determining the crystal structure of $_{\Delta}TraG^*_{R100}$ that includes this interacting motif will shed light on the role of this extended loop in TraG–TraS interactions; these studies are ongoing.)

As the only deletion mutant tested was from the truncation of 45 amino acids, the size of the disordered region in $TraG^*_{R100}$ has not been defined in full; however, the findings of this study have several implications. Many structural predictive software packages such as Phyre2, IUPred2A, ANCHOR2A, and AlphaFold[40,81–83] improperly identified secondary structure properties of $TraG^*$, overestimating the disordered content of the periplasmic protein to be approximately 50%.[40,81–83] CD analysis of $TraG^*_{R100}$ (supplementary material Fig. S2) indicates the protein is not predominantly disordered, rather some regions are predicted to feature distorted helices, at 20.3% of the protein. This indicates that these programs are incorrectly assigning regions of $TraG^*_{R100}$ as disordered, likely due to an inability to identify homologs with known structures, enforcing the postulation regarding the novelty of this protein structure. The identification of a highly dynamic region by the PSP software[39] suggests that other popular algorithms should be modified to include disorder predictions based on long range pi–pi contact frequencies to improve their accuracy.

The increased thermal stability of $_{\Delta}TraG^*_{R100}$ was demonstrated quantitatively via the Thermofluor assays (Fig. 3). The mean T_m of $_{\Delta}TraG^*_{R100}$ in all buffers tested (excluding 50 mM citrate pH 3.5) was 50.6 °C, while the mean T_m of $TraG^*_{R100}$ was 48.0 °C. The difference in T_m is not the only factor that changes when $TraG^*_{R100}$ is truncated, the shape of the melting curves is improved as the relative fluorescence units (RFUs) at initial temperatures is lowered, indicating a reduced quantity of $_{\Delta}TraG^*_{R100}$ molecules are unfolded at the beginning of the assay relative to $TraG^*_{R100}$, and the peak RFU at the T_m is further pronounced in most of the conditions attempted. The large values of the standard error in many of the $TraG^*_{R100}$ experiments are likely due to differences in abundance of aggregation of the protein prior to the start of the experiment; the thermal stability of the truncation mutant is consistent between replicates as the error bars for all buffers tested is smaller, indicating less batch-to-batch variability. Although $_{\Delta}TraG^*_{R100}$ has higher thermal stability relative to $TraG^*_{R100}$, the mean T_m of the lysozyme control is 69.8 °C, indicating that $_{\Delta}TraG^*_{R100}$ is not as stable as a small, well-folded globular protein.

The increased T_m of conditions where buffer pH was approximate to the respective predicted isoelectric point (pI) of the proteins (from ProtParam,[84] pI $TraG^*_{R100}$: 5.83 and pI $_{\Delta}TraG^*_{R100}$: 5.95) is an

© Author(s) 2022

interesting phenomenon. Maintaining protein charge at net neutral is predicted to aid in the formation of intermolecular interactions, which may be the reason for increased thermal stability seen in these conditions. Protein crystallization is often suggested to be started with the pH of the mother liquor equal or proximal to the pI of the protein,[85,86] suggesting optimal buffers of sodium acetate (pH 5.2) and MES (pH 6.0) be explored. Buffers with 50 mM NaCl concentrations appeared to better maintain folding of both proteins as temperature was increased in comparison to buffers with 200 mM NaCl. This could be caused by an increased heat capacity of the solvent when less ionic strength is present, or from higher ionicity affecting intermolecular interactions by masking charges in $TraG^{*}_{R100}$, or that both phenomena contribute to the observed changes. As the thermocycler employed for the Thermofluor assay measures the internal temperature of the instrument rather than the temperature of the buffers themselves, it is possible the buffers with higher salt content absorbed more kinetic energy at the same temperatures and therefore caused the proteins to unfold at slightly lower temperatures. Additionally, increasing salt concentrations may create "salting out" effects; as ionicity is increased salt ions can mask the inter- and intra- molecular ionic interactions that prevent protein unfolding at increased temperatures.[87]

Circular dichroism revealed the changes in secondary structure that occur when the 45 N-terminal residues of $TraG^{*}_{R100}$ are removed. The increase in definition of characteristic absorbance peaks for α-helices at 222 and 208 nm wavelengths is highly indicative of an increase in α-helical character by the truncation of the flexible region. This is quantified by BeStSel,[51] which predicts an increase in the α-helical character and a decrease in loop character in the absence of the N-terminal region (supplementary material Fig. S2). The predicted change in α-helical topology of 15.6% is higher than expected for a loss of only 8.6% of the protein's residues, thus indicating the presence of the flexible region disrupts the topology of the remaining protein. This provides a biophysical explanation for the observed aggregation propensity for $TraG^{*}_{R100}$ relative to $_{\Delta}TraG^{*}_{R100}$ and explains difficulties in attempts to crystallize $TraG^{*}_{R100}$. The high dynamicity of $TraG_{R100}$ residues A452-A496 when not anchored to the C-terminal intermembrane domain results in changes to the overall topology of $TraG^{*}_{R100}$.

Urea denaturation revealed that $_{\Delta}TraG^{*}_{R100}$ has higher chemical stability than $TraG^{*}_{R100}$, as observed by the loss of characteristic α-helical absorbance at lower urea concentrations. For $TraG^{*}_{R100}$ at 1 M urea, the absorbance peaks of 222 and 208 nm were significantly less pronounced, almost decreasing by 50% of their values from the native spectra (Fig. 4). The truncation mutant shows only a minor change in spectra from incubation with 1 M urea in the respective α-helical regions. Quantification via BeStSel predictions indicated a 19.13% α-helical character after incubating $TraG^{*}_{R100}$ in 1 M urea while $_{\Delta}TraG^{*}_{R100}$ had 28.10%, both from an average of three replicates (Table II). Together with the Thermofluor data demonstrating the lowered thermal stability of $TraG^{*}_{R100}$, the urea denaturation and native CD experiments demonstrate an increase in overall stability of the truncation mutant $_{\Delta}TraG^{*}_{R100}$.

Native ESI–MS confirmed the masses of the proteins. The changes relative to expectations based on sequence ($TraG^{*}_{R100}$: 56 699.30 Da, $_{\Delta}TraG^{*}_{R100}$: 52 010.34 Da via ProtParam[84]) might be the result of salt adducts carried over from the purification and potential degradation from the ESI process (supplementary material Fig. S3). The spectra show that both $TraG^{*}_{R100}$ and $_{\Delta}TraG^{*}_{R100}$ samples had unfolded conformations as seen in the 2000–2700 m/z range. However, the peaks seen in the $TraG^{*}_{R100}$ spectrum are more abundant indicating a higher propensity for unfolding in $TraG^{*}_{R100}$. The spectra also demonstrate that $TraG^{*}_{R100}$ has an increased propensity for occupying higher energy charge states of 18+ and 17+ relative to $_{\Delta}TraG^{*}_{R100}$, further indicating the lowered conformational stability of $TraG^{*}_{R100}$.[88]

CIU-MS of the $TraG^{*}_{R100}$ constructs further identifies the lowered conformational stability of $TraG^{*}_{R100}$. The third $TraG^{*}_{R100}$ species with the longest drift time (~14 ms) is observed at nearly 100% abundance for both the 16+ and 17+ charge states above 70 V CE, whereas for $_{\Delta}TraG^{*}_{R100}$ the third species (~13.5 ms) is only partially occupied above 70 V CE and not well defined for the 17+ charge state (Fig. 5). This indicates a third, highly unfolded conformer is more often induced in $TraG^{*}_{R100}$ than $_{\Delta}TraG^{*}_{R100}$, supporting the interpretation that the flexibility of the N-terminal region affects the conformational dynamics of $TraG^{*}_{R100}$. The consistency between biological replicates is evident when viewing the IMS spectra (supplementary material Fig. S4). $_{\Delta}TraG^{*}_{R100}$ and $TraG^{*}_{R100}$ have three conformers each: approximately 10.5, 12.5, 13.5 and 11, 13, 14 ms, respectively, occupied in differing percentages depending on the CE applied. The differences in drift time between conformers of $TraG^{*}_{R100}$ and $_{\Delta}TraG^{*}_{R100}$ are appropriate based on a 45-residue truncation. The shift to the 14 ms conformer is well defined for $TraG^{*}_{R100}$ in the 16+ charge state at 100 V CE; however, the IMS spectrum of the 17+ charge state has a broad peak at 100 V CE that is centered at a smaller drift time (~13.75 ms) when compared to the drift time for the most unfolded conformer of $_{\Delta}TraG^{*}$ (~14.5 ms). Although there is an unfolded conformer for $_{\Delta}TraG^{*}_{R100}$ with a longer drift time (~14.5 ms), it is of low abundance at ~40% signal, whereas the second conformer (~12.5 ms) is at 100% abundance, and the most folded conformer is present at ~30% signal (~10.5 ms). Therefore, the interpretation that $TraG^{*}_{R100}$ can access a more unfolded conformational state at a higher propensity than $_{\Delta}TraG^{*}_{R100}$ is maintained. However, as each charge state was not isolated during data collection prior to the CIU analysis, charge stripping events may have caused some additional noise in the data.[66,67] Therefore, no quantitative claims can be made from this CIU MS data regarding the degree of disorder in each protein construct.

The N-terminal residues of $TraG^{*}_{R100}$ greatly increase the tendency of the protein to aggregate as was demonstrated via SEC–MALS–SAXS (Fig. 6). The high MW peak representing MDa-sized complexes of aggregates is present in $TraG^{*}_{R100}$ despite the addition of the detergent NP40 to aid in maintaining solubility of the protein. This aggregate was not detectable in the sample of $_{\Delta}TraG^{*}_{R100}$ by SAXS or MALS, furthering the interpretation that the protein's stability is greatly increased by the truncation performed in this study. The SEC–MALS–SAXS data from $_{\Delta}TraG^{*}_{R100}$ provided a good MW determination from MALS, at 50 650 Da it was close to the weight provided from MS (51 906 Da) and the estimated weight from ProtParam (52 010.34 Da). SAXS data overestimated the MW as 57.68 kDa and provided a reasonable R_g determination from Guinier analysis (41.27 Å) that was proximal to the R_g from the P(r) function (44.14 Å) by ~3 Å. The R_h determination from MALS provided a value of 47.8 Å, which allows for a shape factor calculation based on the R_g/R_h ratio, resulting in a value of 0.86–0.92 for $_{\Delta}TraG^{*}_{R100}$.[71] This

© Author(s) 2022

represents a more elongated structure as shape factors of ∼0.75 are very compact, globular proteins and become more prolate as the shape factor increases. Additionally, the low peak of the bell-shaped Kratky plot for $_{\Delta}$TraG*$_{R100}$ [Fig. 6(d)] indicates that the protein has some unstructured properties, which may be the reason for some noise at high q in the Guinier analysis (supplementary material Fig. S7) and the ambiguity score of 1.940 for the P(r) function by AMBIMETER.[52,78] Based on the software structural predictions and the known function of TraG*$_{R100}$, it is reasonable to conclude the overall structure of the protein would require flexibility and, therefore, have some intrinsically disordered properties. In comparing the produced bead model of $_{\Delta}$TraG*$_{R100}$ [Fig. 6(e)] to the AlphaFold and RoseTTAFold predicted structures (supplementary material Fig. S1), an elongated structure is confirmed to be present and TraG*$_{R100}$ is highly likely to consist of extended α-helices as predicted. The prolate structure of $_{\Delta}$TraG*$_{R100}$ is also supported by the biochemical data shown in this study and is likely of significance for its function in extending to contact its protein partners. Due to the disruptive aggregation in the TraG*$_{R100}$ samples, MW analyses, R_g analyses and bead model reconstructions resulting from SEC–MALS–SAXS are difficult to interpret. It may be true that the bead model of TraG*$_{R100}$ would have a further extended structure due to the addition of the flexible N-terminus; however, due to the overestimated R_g and D_{max} the reconstruction cannot be compared to $_{\Delta}$TraG*$_{R100}$ to describe how the addition of the N-terminal region affects these values and the low-resolution structural model.

V. CONCLUSIONS

The findings presented here imply that the N-terminal region of TraG*$_{R100}$ is highly flexible and results in the destabilization of the protein, resulting in lowered thermal and chemical stability, increased conformational dynamics and a higher aggregation propensity, thus making it difficult to characterize structurally. This was determined through comparisons of TraG*$_{R100}$ with an N-terminal truncation mutant $_{\Delta}$TraG*$_{R100}$, which presents a more stable protein. This supports the conclusion that the N-terminal region of TraG*$_{R100}$ acts as a flexible linker that connects the membrane-bound portion of TraG$_{R100}$ and periplasmic TraG*$_{R100}$. This is congruent with known moieties in structural biology as the occurrence of unstructured regions approximately 50 residues in length is common in functional proteins.[89–91] Intrinsically disordered proteins (IDPs) are more common in eukaryotes and viruses due to the requirement of complexity in their morphologies; it is suggested that there exists a link between intrinsic disorder and evolution.[92] There are many examples of prokaryotic IDPs however, such as the proteins that regulate the assembly of large multiprotein complexes such as FlgE in the bacterial flagellum,[93] and Ffh and FtsY of the ribosome.[89,94,95] In both cases there exists an intrinsically disordered region (IDR) that serves as a linker region for two independent functional domains required in achieving a variety of conformations for important protein–protein interactions (PPIs). As well, several bacterial regulatory proteins have conserved short intrinsically disordered linker regions called Q-linkers.[89,96,97] Therefore, it is not uncommon in bacterial proteins for disordered regions to be important flexible linkers of folded domains. It is possible this region is merely a linker and does not contain interacting domains, making it non-essential in the protein's structural solution. However, this is difficult to assess *in vivo* as experiments such as conjugative mating assays to test this theory are not possible as TraG* is nonfunctional when expressed on its own; biological function requires full-length TraG.[32]

TraG and TraS form the Eex system of the F-like T4SS and are heavily relied upon for preventing donor-donor plasmid exchange.[98,99] As excessive conjugation can lead to lethal zygosis, it has been theorized that disruption of Eex systems would be detrimental to bacterial colony survival.[100] Development of novel mechanisms to disrupt TraG–TraS interactions and cause ceaseless conjugation or completely prohibit conjugation is contingent on the structural knowledge of the system's protein subunits. Therefore, it is important to understand the structural characteristics of TraG and TraS, and how they interact within the Eex context. TraG*$_{R100}$ and TraG*$_{F}$ are predicted to be highly similar in structure based on overall sequence similarity; however, the region predicted to interact with TraS exhibits more sequence variation that the rest of the protein, which in turn provides plasmid-specific Eex.[32] This study indicates that TraG*$_{R100}$ is a structurally elongated protein with a highly dynamic N-terminal region that likely plays important functional roles in mediating interactions involved in Eex (with TraS) and Mps (with TraN) within the F-like T4SS conjugative assembly.

ACKNOWLEDGMENTS

The authors wish to thank Christina Rodriguez, Arnold Apostol, and Rana Salem for useful discussions. This research was supported by an operating grant from the Natural Sciences & Engineering Council of Canada (NSERC; Funding No. RGPIN-02618-2018) to G.F.A. N.B. gratefully acknowledges financial support from NSERC and York University. We wish to thank Cristina Lento, Esther Wolf, Vimanda Chow, and Derek Wilson for access to their Waters Synapt G2S instrument for CIU MS and highly engaging discussions. Small angle x-ray scattering data were collected at the BioCAT facility at the Advanced Photon Source (APS). We wish to thank Jesse Hopkins for SAXS data collection. The APS is a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. This project was supported by Grant No. P30 GM138395 from the National Institute of General Medical Sciences of the National Institutes of Health. Use of the Pilatus 3 1M detector was provided by Grant No. 1S10OD018090 from NIGMS. The content is solely the responsibility of the authors and does not necessarily reflect the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Nichoas Bragagnolo: Investigation (equal); Methodology (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Gerald F. Audette:** Conceptualization (equal); Funding acquisition (equal); Methodology (equal); Project administration (equal); Supervision (equal); Validation

© Author(s) 2022

(equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request. SAXS data for TraG* from the R100 plasmid are available in SASBDB (https://www.sasbdb.org) under accession codes SASDQG6 and SASDQH6.[104]

REFERENCES

[1]C. M. Thomas and K. M. Nielsen, "Mechanisms of, and barriers to, horizontal gene transfer between bacteria," Nat. Rev. Microbiol. **3**, 711–721 (2005).
[2]W. Arber, "Horizontal gene transfer among bacteria and its role in biological evolution," Life **4**, 217–224 (2014).
[3]J. P. Gogarten and J. P. Townsend, "Horizontal gene transfer, genome innovation and evolution," Nat. Rev. Microbiol. **3**, 679–687 (2005).
[4]C. Gyles and P. Boerlin, "Horizontally transferred genetic elements and their role in pathogenesis of bacterial disease," Vet. Pathol. **51**, 328–340 (2014).
[5]H. K. Allen *et al.*, "Call of the wild: Antibiotic resistance genes in natural environments," Nat. Rev. Microbiol. **8**, 251–259 (2010).
[6]N. Bragagnolo *et al.*, "Protein dynamics in f-like bacterial conjugation," Biomedicines **8**, 362 (2020).
[7]A. A. Vogan and P. G. Higgs, "The advantages and disadvantages of horizontal gene transfer and the emergence of the first species," Biol. Direct **6**, 1 (2011).
[8]E. Grohmann, G. Muth, and M. Espinosa, "Conjugative plasmid transfer in gram-positive bacteria," Microbiol. Mol. Biol. Rev. **67**, 277–301 (2003).
[9]N. A. Lerminiaux and A. D. S. Cameron, "Horizontal transfer of antibiotic resistance genes in clinical environments," Can. J. Microbiol. **65**, 34–44 (2018).
[10]A. Jain and P. Srivastava, "Broad host range plasmids," FEMS Microbiol. Lett. **348**, 87–96 (2013).
[11]T. R. D. Costa *et al.*, "Secretion systems in gram-negative bacteria: Structural and mechanistic insights," Nat. Rev. Microbiol. **13**, 343–359 (2015).
[12]S. A. Lujan, L. M. Guogas, H. Ragonese, S. W. Matson, and M. R. Redinbo, "Disrupting antibiotic resistance propagation by inhibiting the conjugative DNA relaxase," Proc. Natl. Acad. Sci. **104**, 12282–12287 (2007).
[13]E. Cabezón, F. de la Cruz, and I. Arechaga, "Conjugation inhibitors and their potential use to prevent dissemination of antibiotic resistance genes in bacteria," Front. Microbiol. **8**, 1–7 (2017).
[14]G. Waksman, "From conjugation to T4S systems in gram-negative bacteria: A mechanistic biology perspective," EMBO Rep. **20**, e47012 (2019).
[15]P. M. Bennett, "Plasmid encoded antibiotic resistance: Acquisition and transfer of antibiotic resistance genes in bacteria," Br. J. Pharmacol. **153**, 347–357 (2008).
[16]P. J. Christie, "The mosaic type IV secretion systems," EcoSal Plus **7**, 1–34 (2016).
[17]D. Arutyunov and L. S. Frost, "F conjugation: Back to the beginning," Plasmid **70**, 18–32 (2013).
[18]B. Hu, P. Khara, and P. J. Christie, "Structural bases for F plasmid conjugation and F pilus biogenesis in *Escherichia coli*," Proc. Natl. Acad. Sci. **116**, 14222 (2019).
[19]A. Shala-Lawrence, N. Bragagnolo, R. Nowroozi-Dayeni, S. Kheyson, and G. F. Audette, "The interaction of TraW and TrbC is required to facilitate conjugation in F-like plasmids," Biochem. Biophys. Res. Commun. **503**, 2386–2392 (2018).
[20]D. E. Bradley, "Morphological and serological relationships of conjugative pili," Top. Catal. **4**, 155–169 (1980).
[21]E. L. Tatum and J. Lederberg, "Gene recombination in the bacterium *Escherichia coli*," J. Bacteriol. **53**, 673–684 (1947).
[22]G. Koraimann, "Spread and persistence of virulence and antibiotic resistance genes: A ride on the F plasmid conjugation module," EcoSal Plus **8**, 3 (2018).
[23]K. G. Anthony, W. A. Klimke, J. Manchak, and L. S. Frost, "Comparison of proteins involved in pilus synthesis and mating pair stabilization from the related plasmids F and R100-1: Insights into the mechanism of conjugation," J. Bacteriol. **181**, 5149–5159 (1999).
[24]S. Lang and E. L. Zechner, "General requirements for protein secretion by the F-like conjugation system R1," Plasmid **67**, 128–138 (2012).
[25]M. P. Garcillán-Barcia, M. V. Francia, and F. De La Cruz, "The diversity of conjugative relaxases and its application in plasmid classification," FEMS Microbiol. Rev. **33**, 657–687 (2009).
[26]Y. Hu, R. A. Moran, G. A. Blackwell, A. McNally, and Z. Zong, "Fine-scale reconstruction of the evolution of FII-33 multidrug resistance plasmids enables high-resolution genomic surveillance," mSystems **7**, e00831-21 (2022).
[27]F. Serapide *et al.*, "Is the pendulum of antimicrobial drug resistance swinging back after COVID-19?," Microorganisms **10**, 957 (2022).
[28]M. Seneghini *et al.*, "It is complicated: Potential short- and long-term impact of coronavirus disease 2019 (COVID-19) on antimicrobial resistance—An expert review," Antimicrob. Steward. Healthcare Epidemiol. **2**, 1–8 (2022).
[29]Q. Chang, W. Wang, G. Regev-Yochay, M. Lipsitch, and W. P. Hanage, "Antibiotics in agriculture and the risk to human health: How worried should we be?," Evol. Appl. **8**, 240–247 (2015).
[30]C. Pelat *et al.*, "Hand hygiene, cohorting, or antibiotic restriction to control outbreaks of multidrug-resistant enterobacteriaceae," Infect. Control Hosp. Epidemiol. **37**, 272–280 (2015).
[31]J. M. Thomson and R. A. Bonomo, "The threat of antibiotic resistance in gram-negative pathogenic bacteria: β-lactams in peril!," Curr. Opin. Microbiol. **8**, 518–524 (2005).
[32]G. F. Audette, J. Manchak, P. Beatty, W. A. Klimke, and L. S. Frost, "Entry exclusion in F-like plasmids requires intact TraG in the door that recognizes its cognate TraS in the recipient," Microbiology **153**, 442–451 (2007).
[33]W. A. Klimke and L. S. Frost, "Genetic analysis of the role of the transfer gene, traN, of the F and R100-1 plasmids in mating pair stabilization during conjugation," J. Bacteriol. **180**, 4036–4043 (1998).
[34]L. S. Frost, K. Ippen-Ihler, and R. A. Skurray, "Analysis of the sequence and gene products of the transfer region of the F sex factor," Microbiol. Rev. **58**, 162–210 (1994).
[35]N. Firth and R. Skurray, "Characterization of the F plasmid bifunctional conjugation gene, traG. MGG," Mol. Gen. Genet. **232**, 145–153 (1992).
[36]N. Willetts and M. Achtman, "Genetic analysis of transfer by the *Escherichia coli* sex factor F, using P1 transductional complementation," J. Bacteriol. **110**, 843–851 (1972).
[37]R. A. Skurray and P. Reeves, "F factor mediated immunity to lethal zygosis in *Escherichia coli* K 12," J. Bacteriol. **117**, 100–106 (1974).
[38]J. T. Ou, "Role of surface exclusion genes in lethal zygosis in *Escherichia coli* K12 mating," MGG Mol. Gen. Genet. **178**, 573–581 (1980).
[39]R. M. Vernon *et al.*, "Pi-Pi contacts are an overlooked protein feature relevant to phase separation," Elife **7**, 1–48 (2018).
[40]M. Varadi *et al.*, "AlphaFold protein structure database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models," Nucl. Acids Res. **50**, D439–D444 (2022).
[41]J. Jumper *et al.*, "Highly accurate protein structure prediction with AlphaFold," Nature **596**, 583–589 (2021).
[42]M. Baek *et al.*, "Accurate prediction of protein structures and interactions using a three-track neural network," Science **373**, 871–876 (2021).
[43]U. B. Ericsson, B. M. Hallberg, G. T. DeTitta, N. Dekker, and P. Nordlund, "Thermofluor-based high-throughput stability optimization of proteins for structural studies," Anal. Biochem. **357**, 289–298 (2006).
[44]J. J. Lavinder, S. B. Hari, B. J. Sullivan, and T. J. Magliery, "High-throughput thermal scanning: A general, rapid dye-binding thermal shift screen for protein engineering," J. Am. Chem. Soc. **131**, 3794–3795 (2009).
[45]K. Phillips and A. H. de la Peña, "The combined use of the Thermofluor assay and ThermoQ analytical software for the determination of protein stability and buffer optimization as an aid in protein crystallization," Curr. Protoc. Mol. Biol. **2011**, 1–15.
[46]S. A. Seabrook and J. Newman, "High-throughput thermal scanning for protein stability: Making a good technique more robust," ACS Comb. Sci. **15**, 387–392 (2013).
[47]T. Wu *et al.*, "Three essential resources to improve differential scanning fluorimetry (DSF) experiments," bioRxiv (2020).

© Author(s) 2022

[48]Y. Griko, N. Sreerama, P. Osumi-Davis, R. W. Woody, and A. Y. Woody, "Thermal and urea-induced unfolding in T7 RNA polymerase: Calorimetry, circular dichroism and fluorescence study," Protein Sci. **10**, 845–853 (2001).
[49]K. Glover, Y. Mei, and S. C. Sinha, "Identifying intrinsically disordered protein regions likely to undergo binding-induced helical transitions," Biochim. Biophys. Acta-Proteins Proteomics **1864**, 1455–1463 (2016).
[50]K. Matsuo, Y. Sakurada, R. Yonehara, M. Kataoka, and K. Gekko, "Secondary-structure analysis of denatured proteins by vacuum-ultraviolet circular dichroism spectroscopy," Biophys. J. **92**, 4088–4096 (2007).
[51]A. Micsonai *et al.*, "Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy," Proc. Natl. Acad. Sci. **112**, E3095–E3103 (2015).
[52]J. B. Hopkins, R. E. Gillilan, and S. Skou, "BioXTAS RAW: Improvements to a free open-source program for small-angle x-ray scattering data reduction and analysis," J. Appl. Crystallogr. **50**, 1545–1553 (2017).
[53]K. Manalastas-Cantos *et al.*, "ATSAS 3.0: Expanded functionality and new tools for small-angle scattering data analysis," J. Appl. Crystallogr. **54**, 343–355 (2021).
[54]N. Bai, H. Roder, A. Dickson, and J. Karanicolas, "Isothermal analysis of ThermoFluor data can readily provide quantitative binding affinities," Sci. Rep. **9**, 2650 (2019).
[55]S. Boivin, S. Kozak, and R. Meijers, "Optimization of protein purification and characterization using Thermofluor screens," Protein Expression Purif. **91**, 192–206 (2013).
[56]F. Dupeux, M. Röwer, G. Seroul, D. Blot, and J. A. Márquez, "A thermal stability assay can help to estimate the crystallization likelihood of biological samples," Acta Crystallogr., Sect. D **67**, 915–919 (2011).
[57]S. W. Provencher and J. Glöckner, "Estimation of globular protein secondary structure from circular dichroism," Biochemistry **20**, 33–37 (1981).
[58]S. M. Kelly, T. J. Jess, and N. C. Price, "How to study proteins by circular dichroism," Biochim. Biophys. Acta-Proteins Proteomics **1751**, 119–139 (2005).
[59]N. J. Greenfield, "Using circular dichroism spectra to estimate protein secondary structure," Nat. Protoc. **1**, 2876–2890 (2007).
[60]S. Benjwal, S. Verma, K.-H. Röhm, and O. Gursky, "Monitoring protein aggregation during thermal unfolding in circular dichroism experiments," Protein Sci. **15**, 635–639 (2006).
[61]R. Usha and T. Ramasami, "The effects of urea and n-propanol on collagen denaturation: Using DSC, circular dicroism and viscosity," Thermochim. Acta **409**, 201–206 (2004).
[62]V. Geetha, "Distortions in protein helices," Int. J. Biol. Macromol. **19**, 81–89 (1996).
[63]H. R. Wilman, J. Shi, and C. M. Deane, "Helix kinks are equally prevalent in soluble and membrane proteins," Proteins **82**, 1960–1970 (2014).
[64]G. Ben-Nissan and M. Sharon, "The application of ion-mobility mass spectrometry for structure/function investigation of protein complexes," Curr. Opin. Chem. Biol. **42**, 25–33 (2018).
[65]M. E. Ridgeway, C. Bleiholder, M. Mann, and M. A. Park, "Trends in trapped ion mobility: Mass spectrometry instrumentation," TrAC Trends Anal. Chem. **116**, 324–331 (2019).
[66]S. M. Dixit, D. A. Polasky, and B. T. Ruotolo, "Collision induced unfolding of isolated proteins in the gas phase: Past, present, and future," Curr. Opin. Chem. Biol. **42**, 93–100 (2018).
[67]D. D. Vallejo *et al.*, "A modified drift tube ion mobility-mass spectrometer for charge-multiplexed collision-induced unfolding," Anal. Chem. **91**, 8137–8146 (2019).
[68]S. Tamara *et al.*, "Symmetry of charge partitioning in collisional and UV photon-induced dissociation of protein assemblies," J. Am. Chem. Soc. **138**, 10860–10868 (2016).
[69]R. Winkler, "ESIprot: A universal tool for charge state determination and molecular weight calculation of proteins from electrospray ionization mass spectrometry data," Rapid Commun. Mass Spectrom. **24**, 285–294 (2010).
[70]D. A. Polasky, S. M. Dixit, S. M. Fantin, and B. T. Ruotolo, "CIUSuite 2: Next-generation software for the analysis of gas-phase protein unfolding data," Anal. Chem. **91**, 3147–3155 (2019).
[71]M. A. Graewert *et al.*, "Adding size exclusion chromatography (SEC) and light scattering (LS) devices to obtain high-quality small angle x-ray scattering (SAXS) data," Crystals **10**, 975 (2020).
[72]A. G. Kikhney and D. I. Svergun, "A practical guide to small angle x-ray scattering (SAXS) of flexible and intrinsically disordered proteins," FEBS Lett. **589**, 2570–2577 (2015).
[73]J. Stetefeld, S. A. McKenna, and T. R. Patel, "Dynamic light scattering: A practical guide and applications in biomedical sciences," Biophys. Rev. **8**, 409–427 (2016).
[74]D. I. Svergun, "Determination of the regularization parameter in indirect-transform methods using perceptual criteria," J. Appl. Crystallogr. **25**, 495–503 (1992).
[75]H. Liu and P. H. Zwart, "Determining pair distance distribution function from SAXS data using parametric functionals," J. Struct. Biol. **180**, 226–234 (2012).
[76]V. M. Burger, D. J. Arenas, and C. M. Stultz, "A structure-free method for quantifying conformational flexibility in proteins," Sci. Rep. **6**, 1–9 (2016).
[77]F. Tondnevis, R. E. Gillilan, L. B. Bloom, and R. McKenna, "Solution study of the *Escherichia coli* DNA polymerase III clamp loader reveals the location of the dynamic $\psi\chi$ heterodimer," Struct. Dyn. **2**, 054701 (2015).
[78]M. V. Petoukhov and D. I. Svergun, "Ambiguity assessment of small-angle scattering curves from monodisperse systems," Acta Crystallogr., Sect. D **71**, 1051–1058 (2015).
[79]R. van der Lee *et al.*, "Classification of intrinsically disordered regions and proteins," Chem. Rev. **114**, 6589–6631 (2014).
[80]F. Fiumara, L. Fioriti, E. R. Kandel, and W. A. Hendrickson, "Essential role of coiled coils for aggregation and activity of Q/N-rich prions and polyQ proteins," Cell **143**, 1121–1135 (2010).
[81]L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. E. Sternberg, "The Phyre2 web portal for protein modeling, prediction and analysis," Nat. Protoc. **10**, 845–858 (2015).
[82]B. Mészáros, G. Erdös, and Z. Dosztányi, "IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding," Nucl. Acids Res. **46**, W329–W337 (2018).
[83]Z. Dosztányi, B. Mészáros, and I. Simon, "ANCHOR: Web server for predicting protein binding regions in disordered proteins," Bioinformatics **25**, 2745–2746 (2009).
[84]V. K. Garg *et al.*, "MFPPI: Multi FASTA ProtParam interface," Bioinformation **12**, 74–77 (2016).
[85]K. A. Kantardjieff and B. Rupp, "Protein isoelectric point as a predictor for increased crystallization screening efficiency," Bioinformatics **20**, 2162–2168 (2004).
[86]J. Kirkwood, D. Hargreaves, S. O'Keefe, and J. Wilson, "Using isoelectric point to determine the pH for initial protein crystallization trials," Bioinformatics **31**, 1444–1451 (2015).
[87]A. C. Dumetz, A. M. Snellinger-O'Brien, E. W. Kaler, and A. M. Lenhoff, "Patterns of protein-protein interactions in salt solutions and implications for protein crystallization," Protein Sci. **16**, 1867–1877 (2007).
[88]S. J. Watt *et al.*, "Effect of protein stabilization on charge state distribution in positive- and negative-ion electrospray ionization mass spectra," J. Am. Soc. Mass Spectrom. **18**, 1605–1611 (2007).
[89]H. J. Dyson and P. E. Wright, "Intrinsically unstructured proteins and their functions," Nat. Rev. Mol. Cell Biol. **6**, 197–208 (2005).
[90]V. N. Uversky, "Natively unfolded proteins: A point where biology waits for physics," Protein Sci. **11**, 739–756 (2002).
[91]V. N. Uversky, "A decade and a half of protein intrinsic disorder: Biology still waits for physics," Protein Sci. **22**, 693–724 (2013).
[92]B. Xue, A. K. Dunker, and V. N. Uversky, "Orderly order in protein intrinsic disorder distribution: Disorder in 3500 proteomes from viruses and the three domains of life," J. Biomol. Struct. Dyn. **30**, 137–149 (2012).
[93]C. S. Barker, I. V. Meshcheryakova, A. S. Kostyukova, P. L. Freddolino, and F. A. Samatey, "An intrinsically disordered linker controlling the formation and the stability of the bacterial flagellar hook," BMC Biol. **15**, 1–14 (2017).
[94]L. F. Estrozi, D. Boehringer, S. O. Shan, N. Ban, and C. Schaffitzel, "Cryo-EM structure of the *E. coli* translating ribosome in complex with SRP and its receptor," Nat. Struct. Mol. Biol. **18**, 88–90 (2011).
[95]P. J. Focia, I. V. Shepotinovskaya, J. A. Seidler, and D. M. Freymann, "Heterodimeric GTPase core of the SRP targeting complex," Science **303**, 373–377 (2004).

© Author(s) 2022

[96]J. C. Wootton and M. H. Drummond, "The Q-linker: A class of interdomain sequences found in bacterial multidomain regulatory proteins," Protein Eng. **2**, 535–543 (1989).
[97]V. P. Reddy Chichili, V. Kumar, and J. Sivaraman, "Linkers in the structural biology of protein-protein interactions," Protein Sci. **22**, 153–167 (2013).
[98]M. Achtman, P. A. Manning, B. Kusecek, S. Schwuchow, and W. Neil, "Genetic analysis of F sex factor cistrons needed for surface exclusion in *Escherichia coli*," J. Mol. Biol. **138**, 779–795 (1980).
[99]M. Achtman, N. Kennedy, and R. Skurray, "Cell-cell interactions in conjugating *Escherichia coli*: Role of traT protein in surface exclusion," Proc. Natl. Acad. Sci. **74**, 5104–5108 (1977).
[100]M. P. Garcillán-Barcia and F. de la Cruz, "Why is entry exclusion an essential feature of conjugative plasmids?," Plasmid **60**, 1–18 (2008).
[101]C. Virolle, K. Goldlust, S. Djermoun, S. Bigot, and C. Lesterlin, "Plasmid transfer by conjugation in gram-negative bacteria: From the cellular to the community level," Genes (Basel) **11**, 1–33 (2020).
[102]K. G. Anthony, C. Sherburne, R. Sherburne, and L. S. Frost, "The role of the pilus in recipient cell recognition during bacterial conjugation mediated by F-like plasmids," Mol. Microbiol. **13**, 939–953 (1994).
[103]L. Schrödinger, *The PyMOL Molecular Graphics System* (PyMOL, 2020).
[104]N. Bragagnolo and G. F. Audette, "Solution characterization of the dynamic conjugative entry exclusion protein TraG," SASBDB. Dataset SASDQG6 and SASDQH6.
[105]See supplementary material at https://www.scitation.org/doi/suppl/10.1063/4.0000171 for including details on the expression and the purification of proteins used in this study and other supporting data and figures.

© Author(s) 2022