Deciphering the “Fuzzy” Interaction of FG Nucleoporins and Transport Factors Using Small-Angle Neutron Scattering

Highlights
- Characterization of an FG Nup interaction by SANS with contrast matching
- Interaction of Kap95 and NTF2 increases FG Nup size and imparts local rigidity
- Ensemble analysis reveals conformational properties adopted by the interacting FG Nup
- Ensembles reveal a potentially functionally relevant loop confirmation

Authors
Samuel Sparks, Deniz B. Temel, Michael P. Rout, David Cowburn

Correspondence
cowburn@cowburnlab.org

In Brief
Sparks et al. use small-angle neutron scattering experiments to describe the fuzzy interactions between disordered FG Nups and transport factors (TFs) from the nuclear pore complex, thus elucidating the subtle conformational changes that TFs induce within FG Nup chains. The study helps to define “fuzziness” at intermediate length scales.
Deciphering the “Fuzzy” Interaction of FG Nucleoporins and Transport Factors Using Small-Angle Neutron Scattering

Samuel Sparks,1 Deniz B. Temel,1,3 Michael P. Rout,2 and David Cowburn1,4,*
1Departments of Biochemistry and of Physiology & Biophysics, Albert Einstein College of Medicine, Bronx, NY, USA
2Laboratory of Cellular and Structural Biology, Rockefeller University, New York, NY, USA
3Present address: Amgen Inc., Thousand Oaks, CA, USA
4Lead Contact
*Correspondence: cowburn@cowburnlab.org
https://doi.org/10.1016/j.str.2018.01.010

SUMMARY

The largely intrinsically disordered phenylalanine-glycine-rich nucleoporins (FG Nups) underline a selectivity mechanism that enables the rapid translocation of transport factors (TFs) through the nuclear pore complexes (NPCs). Conflicting models of NPC transport have assumed that FG Nups undergo different conformational transitions upon interacting with TFs. To selectively characterize conformational changes in FG Nups induced by TFs we performed small-angle neutron scattering (SANS) with contrast matching. Conformational-ensembles derived from SANS data indicated an increase in the overall size of FG Nups is associated with TF interaction. Moreover, the organization of the FG motif in the interacting state is consistent with prior experimental analyses defining that FG motifs undergo conformational restriction upon interacting with TFs. These results provide structural insights into a highly dynamic interaction and illustrate how functional disorder imparts rapid and selective FG Nup-TF interactions.

INTRODUCTION

The selective permeability barrier of the nuclear pore complex (NPC) relies on a group of Phe-Gly-rich nucleoporins (FG Nups) that contain large intrinsically disordered domains to generate an entropic barrier to nonspecific diffusion. Small molecules can freely diffuse through the NPC while larger macromolecules are impeded in a size-dependent manner (Knockenhauer and Schwartz, 2016; Timney et al., 2016). Larger macromolecules such as ribosomal subunits bypass the selectivity barrier by interaction with transport factors (TFs), which can permeate the disordered FG meshwork by virtue of making specific contacts with FG Nups. Exactly how FG Nup/TF interactions lead to both rapid and selective transport is unresolved and several conflicting NPC transport models have been proposed (Grunwald et al., 2011; Schmidt and Gorlich, 2016). Despite significant effort (Beck and Hurt, 2017), there is no clear consensus of the underlying conformational preferences of FG Nups, the strengths of the FG Nup/TF interactions, and any associated conformational changes that FG Nup undergo upon interactions with TFs. A molecular description, albeit reductionist, of these fundamental properties is required to comprehend the selectivity and rapidity of facilitated translocation of TFs through the NPC.

Recent studies utilizing nuclear magnetic resonance (NMR) and molecular dynamics (MD) simulations have revealed essential features that promote selective diffusion of TFs. Isolated FG Nups are highly dynamic, random coil polymers that remain disordered while engaged to TFs (Hough et al., 2015; Milles et al., 2015; Raveh et al., 2016). FG Nups interact with TFs using predominantly their FG motifs, and minimally their intervening spacer residues (Hough et al., 2015; Milles et al., 2015). By virtue of multiple TF-interaction sites, FG Nups can make multivalent contacts with TFs, although whether or not high-avidity interactions are critical to the NPC barrier function is still under debate (Hayama et al., 2017; Lim et al., 2015; Schmidt and Gorlich, 2016). FG Nups are members of the unique class of intrinsically disordered proteins (IDPs) that form fuzzy interactions (Sharma et al., 2015; Wu and Fuxreiter, 2016). The interactions formed by FG Nups and TFs are consistent with the random complex classification (Sharma et al., 2015); i.e., an ensemble of rapidly interconverting conformers, with multiple ligand sites (the FG motifs) dynamically contacting multiple TF-interaction sites without forming a stable secondary structure.

Prior NMR studies (Hough et al., 2015; Milles et al., 2015) have provided basic physicochemical properties of FG Nups, and their interaction with TFs at high resolution but in combination with small-angle scattering can provide a comprehensive description of the dynamic ensembles at different spatial and temporal scales (Hennig and Sattler, 2014). Indeed, this combination is particularly useful for characterizing IDPs (Receveur-Brechot and Durand, 2012). While sensitive to local dynamics and structural perturbations, NMR does not characterize global structural properties of IDPs. For example, the use of NMR data alone in ensemble modeling was not sufficient to differentiate between collapsed and extended conformers from disordered ensembles (Brookes and Head-Gordon, 2016). Therefore, it is unclear from our previous measurements (Hough et al., 2015) how TFs alter large-scale FG Nup dynamic structures. Previous reports, using different methods, offer conflicting observations of FG Nups undergoing (1) a reversible collapse upon interaction...
RESULTS AND DISCUSSION

We initially characterized the solution state of free FSFG-K. SANS profiles from $[^{2}H]$-FSFG-K, at 42% D$_{2}$O, and natural abundance FSFG-K, at 92% D$_{2}$O, display featureless scattering curves with a power-law decay at $q > 0.05 \text{Å}^{-1}$, typically observed for disordered proteins (Figure 1B) (Cordeiro et al., 2017). SANS measurements of a dilution series of FSFG-K, the corresponding linear Guinier plots, and molecular mass estimates derived from the forward scattering, $I_0$, indicate a lack of significant inter-particle interference or aggregation (Figures S1A and S1B; Table 1) (Trewhella et al., 2017). The radius of gyration, $R_g$, was determined by fitting the scattering data to Debye’s law (Debye, 1947), which is valid over a larger $q$ range relative to standard Guinier analysis (Receveur-Brechot and Durand, 2012) (see STAR Methods for a discussion on extracting $R_g$ values from IDPs by small-angle scattering). The Debye analysis yielded $R_g$ values of 36.2 ± 0.3 Å and 36.3 ± 1.0 Å (average value from the dilution series) for free $[^{2}H]$-FSFG-K (in 42% D$_{2}$O) and natural abundance FSFG-K (in 92% D$_{2}$O), respectively. These values are in good agreement with values obtained previously for FSFG-K from our MD simulation using the Tip4P-D water ($R_g$ 31.9 Å) (Raveh et al., 2016). In addition, a dimensionless Kratky plot (Durand et al., 2010) displayed an initial increase, followed by a plateau at higher $q$ range, characteristic of an IDP (Figure 1B, inset). FSFG-K is then a fully disordered, random coil polymer, in full agreement with our previous NMR chemical shift analysis (Hough et al., 2015) and MD simulations (Raveh et al., 2016).

Contrast matched and inverse contrast matched experiments (Sugiyama et al., 2014) were performed (1) at 42% D$_{2}$O, to match natural abundance TFs (NTF2 and Kap95) observing $[^{2}H]$-FSFG-K, and (2) at 92% D$_{2}$O, to match partially deuterated Kap95 observing natural abundance FSFG-K. In both experiments, excellent suppression of scattering from the TFs was observed (Figures 2A and S2A). Upon addition of TF, the scattering profiles of $[^{2}H]$-FSFG-K displayed reduced intensity at $q = 0.06 \text{Å}^{-1}$ and a shoulder at $q = 0.12 \text{Å}^{-1}$ (Figure 2B). The observed shoulder was less pronounced at lower TF concentration indicating these features are dependent on a high population of FSFG-K in the interacting state (Figures S2B–S2D). Due to poor yield of $[^{2}H]$-Kap95 in minimal D$_{2}$O medium, inverse contrast matching experiments were performed at a low concentration of partially $[^{2}H]$-Kap95. The apparent shoulder was, thus, not as pronounced, but there was some reduced intensity at intermediate scattering angles (Figures S2E and S2F). Nevertheless, the presence of these features indicates there is a shift in the ensemble-averaged conformations of FSFG-K in the interaction state.

The overall size of FSFG-K increased upon addition of TFs. Debye analysis showed a TF concentration-dependent increase in the apparent $R_g$ upon addition of either NTF2 or Kap95, increasing ~15%-30% from 36.2 ± 0.3 Å to a maximum of...
Table 1. Comparison of the $R_g$ and $I_0$ Values Computed from Guinier, Debye, and the Radial Distribution Function, P(r), Analyses

<table>
<thead>
<tr>
<th>Sample</th>
<th>Guinier Analysis (Forster et al., 2010)</th>
<th>Debye Analysis (Debye, 1947)</th>
<th>P(r) (Svergun, 1992)</th>
<th>Expected $I_0$ (cm$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R_g$ (Å) ($q_{max}$$/R_g$)</td>
<td>$I_0$ (cm$^{-1}$)</td>
<td>$R_g$ (Å)</td>
<td>$I_0$ (cm$^{-1}$)</td>
</tr>
<tr>
<td>[2H]-FSFG-K [0.6 mM]</td>
<td>32.0 ± 1.4 (1.33)</td>
<td>0.110 ± 0.004</td>
<td>36.2 ± 0.3</td>
<td>0.115 ± 0.001</td>
</tr>
<tr>
<td>(8.5 mg/mL)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[2H]-FSFG-K [0.6 mM]</td>
<td>–</td>
<td>–</td>
<td>39.6 ± 0.8</td>
<td>0.087 ± 0.002</td>
</tr>
<tr>
<td>Kap95 [0.25 mM]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[2H]-FSFG-K [0.6 mM]</td>
<td>–</td>
<td>–</td>
<td>41.8 ± 1.2</td>
<td>0.097 ± 0.003</td>
</tr>
<tr>
<td>Kap95 [0.5 mM]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[2H]-FSFG-K [0.6 mM]</td>
<td>–</td>
<td>–</td>
<td>41.2 ± 0.3</td>
<td>0.125 ± 0.001</td>
</tr>
<tr>
<td>NTF2 [0.3 mM]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[2H]-FSFG-K [0.6 mM]</td>
<td>–</td>
<td>–</td>
<td>45.6 ± 0.5</td>
<td>0.143 ± 0.002</td>
</tr>
<tr>
<td>NTF2 [0.6 mM]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[2H]-FSFG-K [0.6 mM]</td>
<td>–</td>
<td>–</td>
<td>47.9 ± 0.6</td>
<td>0.143 ± 0.002</td>
</tr>
<tr>
<td>NTF2 [1.2 mM]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSFG-K [375 μM] +</td>
<td>–</td>
<td>–</td>
<td>53.2 ± 1.0</td>
<td>0.042 ± 0.001</td>
</tr>
<tr>
<td>[2H]-Kap95 [75 μM]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSFG-K (20.2 mg/mL)</td>
<td>31.5 ± 0.8 (1.04)</td>
<td>0.1097 ± 0.002</td>
<td>34.8 ± 0.2</td>
<td>0.117 ± 0.001</td>
</tr>
<tr>
<td>FSFG-K (10.6 mg/mL)</td>
<td>32.6 ± 1.4 (1.04)</td>
<td>0.069 ± 0.001</td>
<td>36.6 ± 0.3</td>
<td>0.071 ± 0.001</td>
</tr>
<tr>
<td>FSFG-K (5.3 mg/mL)</td>
<td>34.3 ± 2.9 (0.98)</td>
<td>0.038 ± 0.001</td>
<td>36.2 ± 0.4</td>
<td>0.038 ± 0.001</td>
</tr>
<tr>
<td>FSFG-K (2.7 mg/mL)</td>
<td>33.6 ± 2.9 (1.15)</td>
<td>0.019 ± 0.001</td>
<td>37.5 ± 1.1</td>
<td>0.019 ± 0.001</td>
</tr>
</tbody>
</table>

The expected $I_0$ values were calculated assuming the molecular weight derived from the amino acid sequence (14,135.2 Da), a partial specific volume of 0.724 cm$^3$ g$^{-1}$, and a scattering contrast of either 3.357$^{10}{/C}0$ for [2H]-FSFG-K at 42% D$_2$O or 2.404$^{10}{/C}0$ at 92% D$_2$O. In addition, the assumption that 100% of exchangeable Hs are available to solvent was made. Contrast calculations were made using the Web service MULCh (Whitten et al., 2008). See STAR Methods section for a discussion on the calculation of $R_g$ from scattering methods. We note that the $R_g$ derived from the SANS data of FSFG-K [375 μM] with [2H]-Kap95 [75 μM] likely disagrees with the values obtained from [2H]-FSFG-K in 42% D$_2$O buffer due to a slight mismatch in the inverse contrast match at low q (Figure S2A).

47.9 ± 0.6 Å and 41.8 ± 12 Å at the highest concentrations of NTF2 (1.2 mM) and Kap95 (0.5 mM), respectively (Table 1). The increasing $R_g$ of FSFG-K with increasing TF concentration indicates that the observed changes are dependent on the degree of bound FSFG-K. Computing the radial distribution function, $P(r)$, further confirmed changes in the $R_g$ upon addition of TF and showed good agreement with the values obtained by Debye analysis (Figures 2C and Table 1). $P(r)$ curves additionally showed an increase in the maximum dimensions, $D_{max}$, from 127.0 Å in the free state to 168.0 Å and 143.5 Å in the presence of the highest concentrations of NTF2 and Kap95, respectively (Figures 2C and Table 1). An increase in the forward scattering, $I_0$, with increasing NTF2 concentration was also observed (Figures 2B and Table 1). While the increase in $I_0$ could be interpreted as subtle aggregation, dynamic light scattering (DLS) measurements of similarly prepared samples, both of the free components and as well as their complexes, are devoid of aggregate species (note that FSFG-K is unlabeled in the DLS experiments (Figures S2G–S2J). Furthermore, the radius of hydration, $R_h$, determined for FSFG-K in the presence of either TF indicated that the molecular stoichiometry of the interaction is 1:1 (in terms of the number of molecules participating in the interaction). This implies that the increase in $R_h$ and $I_0$ at high TF concentrations is not due to FSFG-K interacting with multiple TFs or aggregation. We therefore can alternatively interpret the increase in $I_0$ as a change in the partial specific volume, $v_s$, of FSFG-K in the interaction state. $I_0$ is related to ($\Delta v_s$)$^2$, where $\Delta v_s$ is the scattering contrast, thus a ~12% increase in the volume of FSFG-K could account for the ~25% change in $I_0$ (i.e., the observed increase from 0.115 cm$^{-1}$ to 0.142 cm$^{-1}$ for [2H]-FSFG-K and [2H]-FSFG-K + NTF2 (1.2 mM), respectively). Taken together, SANS data indicate that FSFG-K adopts larger, more extended conformations in the interaction state and does not undergo collapse upon binding as observed for other IDPs interactions (Green et al., 2016) as well as the reversible collapse observed for FG Nups using other approaches (Cardarelli et al., 2012; Lim et al., 2007).

Interestingly, in presence of either TF, the $P(r)$ curves displayed two peaks at internuclear distances of ~15 Å and ~50 Å (Figure 2C). These real-space interatomic distances reflect the reciprocal space features present in the scattering profile (Figure 2B). The peak at ~50 Å would be the expected distance for the $i$, $i+2$ FSFG motif (separation of 34 amino acids, see STAR Methods) and the peak at ~15 Å may reflect an increased persistence length, $L_p$, of the polymer chain due to the local restriction of the “strongly interacting” FSFG motif (Raveh et al., 2016). An increased persistence length is supported by modeling using the worm-like chain (WLC) model (Sharp and Bloomfield, 1968) (see STAR Methods). The expected range for persistence length of a typical IDP is 9–11 Å, which corresponds with 2–3 amino acid residues that are locally rigid (Perez et al., 2001). For [2H]-FSFG-K, the derived values for $L_p$ increased upon addition of TFs from 8.9 ± 2.3 Å to 16.7 ± 1.2 Å (corresponding with 4–5 amino acids that are locally rigid) for the sample with 1.2 mM NTF2 (Table S1). Additionally for many samples some fitting parameters required constraint (see STAR Methods), good fits to...
expansive data were obtained and the resulted $R_g$ calculated from the model were in excellent agreement with Debye and $D(\rho)$ analysis (Table 1 and Table S1).

Ensemble modeling was performed to quantitatively assess changes in the ensemble-averaged conformation of FSFG-K upon interaction with TFs. The ensemble optimization method (EOM), which uses a genetic algorithm, was used for the selection of an ensemble of conformers whose weighted average scattering curve best reproduces the experimental data (Bernardo et al., 2007; Tria et al., 2015). The structures within the selected ensemble are interpreted as containing the most prominent “features” within the actual sample (Cordeiro et al., 2017). EOM selects these representative structures from a large initial pool of possible conformers. Two starting pools were used for selection to experimental SANS data: (1) a “random pool” of 100,000 all-atom, random coil models produced by the program TraDES (Feldman and Hogue, 2000), and (2) an “MD pool” of ~19,000 conformers obtained directly from the coordinates of FSFG-K in the presence of NTF2 (NTF2 coordinates removed) from our previous MD simulation on Anton (Shaw et al., 2009) with the TIP4P-D water model (Raveh et al., 2016). Using this MD pool enables direct comparison of our MD simulations with the TIP4P-D water model (Raveh et al., 2016). Using this MD pool display a shift in the $R_g$ distributions toward larger values, as expected, as FSFG-K forms more extended conformations in the interaction state, in agreement with both $D(\rho)$ and Debye analysis (Figures 3B and Table S2). We next computed the weighted average Cα-Cα distance maps from the conformers within the selected ensembles (Figures 3C and S3). The maximum dimensions of FSFG-K observed from the averaged distance maps increased from 87.4 Å in the free state to 137.4 Å and 141.0 Å for [H]-FSFG-K in the presence of Kap95 (0.5 mM) and NTF2 (1.2 mM), respectively, comparable with the $D_{\text{max}}$ values computed from $D(\rho)$ (Figures 3C and Table S2).

If the conformers within the selected ensembles accurately reflect the experimental data, the organization of these structures should result in interatomic distances that mimic the real-space distance computed from $D(\rho)$ in Figure 2C. Indeed, comparing the Cα-Cα distances between every Cα atom within a given conformer, plotted as a histogram over the selected ensemble, with that of the $D(\rho)$, produces comparable profiles (Figure S4A). Therefore, EOM appears to have successfully generated ensembles whose conformations reproduce the real-space distances in the actual ensemble.

In the presence of high concentration of either TF, the conformers of FSFG-K selected from the random pool appear to adopt a highly extended, brush-like morphology (Figures 3D and S3). However, despite the extended conformations, the conformers selected from fits to the scattering data of [H]-FSFG-K with high concentration TFs appear to contain “loops” or “kink” features followed by an extended segment in the chain. The Cα-Cα distance map indicates these features as a clustering of approximately 15 Å at the position of the loop conformation. These interatomic distances were also present at high frequency in the $D(\rho)$ (Figure 2C). For example, when VOM fitting was performed on [H]-FSFG-K in the presence of NTF2 (0.6 mM), kinks in the structure (indicated by arrows in Figure 3D, middle) generally occurred in the middle of the extended conformation, forming a loop of ~10–20 residues with the apex of the loop
composed of 3–6 residues. However, as evident by the distances maps of the individual conformers from the sub-ensembles (Figure S3), these conformers have no selective pressure to occur at the same residue position during fitting, so these kink features can appear blurred on an average distance map.

The individual distances maps and the structures of the selected models indicate that in the presence of either TF, sharp turns are consistently present in the selected ensembles when fit to the data using higher TF concentration (Figure S3). Although the selected ensembles derived from the free [H]-FSFG-K show some similar kink features, the most prominent conformer within the ensemble (comprising 60% of the ensemble) appears as a typical crumpled coil with a low relative $R_g$ (30.9 Å). However, we caution that the specific conformers that comprise the selected ensembles are likely degenerate, and the theoretical scattering from different combinations of conformers could reproduce the experimental data. While it is challenging to place significance on the specific conformers, we attempted to validate selected ensembles by demonstrating that the SANS data of [H]-FSFG-K in the presence of the lowest concentrations of either TF can be adequately fit by a combination of previously optimized models selected from the free FSFG-K and FSFG-K in the presence of the highest concentration of TF.

The underlying assumption is that features within the scattering profile reflect conformations of the bound FSFG-K that increase with a greater proportion of the interacting state occurring at greater TF concentration. Indeed, one model from the free [H]-FSFG-K selected ensemble (comprising 35.2% of the ensemble) and three selected models from the [H]-FSFG-K with NTF2 (0.6 mM) could accurately reproduce the scattering profile of [H]-FSFG-K with NTF2 (0.3 mM) with a $R^2 = 0.33$ (Figure S4B and S4C). Similarly, for Kap95, one of the selected models from the free [H]-FSFG-K selected ensemble (comprising 18.3%) and two of the three models from the [H]-FSFG-K with Kap95 (0.5 mM) selected ensemble reproduce the data of [H]-FSFG-K with Kap95 (0.25 mM) with a $R^2 = 0.26$ (Figures S4B and S4C). Thus, while the specific models may be degenerate, the conformational features reproduce the experimental data.
appear conserved in the selected ensembles from the data of 
$[^2H]$-FSFG-K in the presence of high concentration of TF and
likely reflect functionally relevant conformations.

Strikingly, a similar feature was also observed on the Cx-Cx
distance map calculated from the structures selected from
the MD pool, which reproduced SANS data of $[^2H]$-FSFG-K in
the presence of high concentration of NTF2 (Figures 3C and
S4D). The performance of the EOM fit to the SANS data of
$[^2H]$-FSFG-K and NTF2 (0.6 mM) as well as the $R_g$ distribution
of the optimized ensemble was nearly identical with respect to
the two starting pools (random pool; $\chi^2 = 0.15$; ensemble
average, $R_g = 40.4 \AA$; MD pool, $\chi^2 = 0.19$; ensemble average
$R_g = 39.4 \AA$ (Figure 3 and Table S2). Interestingly, three of
the four conformers that comprise the sub-ensemble that best-fit
the SANS data of $[^2H]$-FSFG-K + NTF2 (0.6 mM) with a selection
from the MD pool are each in a similar pose and were derived from
frames at a similar point in the MD trajectory (Figure 3D).
The other conformer, which accounts for 30% of the ensemble
(Figure 3D, model colored tan), also contains a loop feature
(Figure 3D). This conformer was derived early in the simulation (after
57 ns) where the third FSFG motif of FSFG-K is bound to NTF2 in
a conformation nearly identical to the conformation adopted in
the crystal structure of NTF2 N77Y and a small FSFG peptide
(PDB: 1GYB) (Bayliss et al., 2002). This is illustrated in Figure 3C,
where the gray arrow indicates the position of the bound FSFG
motif. The MD simulations began with the coordinates of the
third FSFG motif and the residues that interact with NTF2 con-
strained to their crystallographic positions and were released
at the start of the simulation. The bound FSFG in the crystal
structures with NTF2 N77Y and a small FSFG peptide
(PDB: 1GYB) (Bayliss et al., 2002) and Kap95 (Bayliss et al., 2000)
form similar conformations with the latter structure, showing six residues (KPAFSF) forming an extended
conformation, which was denoted as “β-strap-like”. Furthermore,
in both crystal structures, the Gly formed a tight turn
configuration, directing the rest of the FG Nup chain away from
contacting other regions of the TFs (Bayliss et al., 2002). Similar
loop features were also observed for each conformer in the
selected ensemble for $[^2H]$-FSFG-K with 1.2 mM NTF2 (Fig-
ure S4D). Therefore, based on ensemble modeling, we propose
that the features in the SANS profiles for FSFG-K interacting with
TFs reflect the bound FSFG motif forming extended conforma-
tions, contributing to the increased persistence length observed,
and that the Gly of FG facilitates the formation of the loop fea-
tures observed in the selected ensembles. The loop confor-
mation may help to avoid steric clashes between residues down-
stream of the FSFG motif and TFs surface enabling the Phe
motifs to probe for interaction sites efficiently. Previous chemical
shift analysis of residues from FSFG-K closely matched their
random coil values, indicating no propensity for this confor-
mation (Hough et al., 2015). We therefore rule out a conformational
selection mechanism and suggest that the bound state confor-
mation is associated with a significant entropy loss, providing
a basis for the weak per-FG motif affinity observed from studies
of similar FG Nup constructs (Milles et al., 2015).

The conformations of FSFG-K within the selected ensembles
also resulted in a high frequency of interatomic distance of
approximately 50 Å from fitting the SANS data with high TF con-
centration (Figures 2C and S4A). The structural origins are most
clearly evident from the models selected based on fitting to
$[^2H]$-FSFG-K and Kap95 (0.5 mM). The kinks in the middle and
end of another linear polymer chain result in elevated dis-
tances of approximately 50 Å which are observed between the
kink regions (Figure S3). There are six equivalent FSFG motifs
present in our construct, evenly spaced by 15 linker residues,
and only the FSFG motifs are involved in the interaction based
on NMR analysis (Hough et al., 2015). The ~50 Å interatomic dis-
tances could reflect two or more FSFG motifs, from the same
FSFG construct, binding to a single TF, although SANS is likely
unable to definitely resolve whether multivalent interactions are
occurring, especially in a highly dynamic system. Such confor-
mations are reminiscent of those adopted by in the crystal struc-
ture of Nup214 bound to CRM1 where multiple FG motifs from
Nup214 are engaging CRM1 (Port et al., 2015). In the
selected ensemble derived from MD simulation with fit to the
NTF2 at high concentration, internuclear distances of ~50 Å
were observed between two adjacent loop features (Figure S4D).

Taking our recent data together, we suggest the interactions
between FG Nups and TFs are a version of a multisite “encounter
complex” (Kozakov et al., 2014; Xu et al., 2008). The initial
encounter occurs where FSFG motifs search for their specific
TFs interaction sites where they then transition to a more strongly
bound interaction mode (Raveh et al., 2016). In the strongly
bound state, the FSFG motifs have reduced dynamic motion,
as shown by the elevated $15N$ relaxation rates ($R_2$) of the FSFG
motifs in the presence of either Kap95 or NTF2 (Hough et al.,
2015; Raveh et al., 2016). Here, our model based on SANS
data extends previous NMR measurement and suggests that
interaction of TFs induces small changes of local rigidity and
effective persistence length of the FSFG motifs (and nearby
residues). The bound motif is proposed here to adopt a confor-
mation similar to that observed in the X-ray crystal structure
(i.e., Gly forms a tight turn distending away from TF). The interac-
tion of FSFG-K and TFs results in an increase in overall size of
FG Nup and higher population of extended conformations
due to (1) the rigid configuration of the bound FSFG motif, which
restricts compaction, similar to how IDPs containing polyproline
helices impart increased persistence length and larger than expected $R_g$ and $D_{max}$ relative to an IDP of the same number of
residues (Boze et al., 2010); (2) the turn feature adopted by the
Gly of the FSFG motif facilities loop formation, helping to po-
sition the rest of the FG Nup chain away from contact with other
regions of the TF, and (3) the influence of steric hindrance due to
the presence of the TF occupying space. In a fuzzy complex, any
local rigidity would be entropically unfavorable and would restrict
the formation of static and high-avidity FG Nup/TF interactions
(Raveh et al., 2016), although it remains to be seen whether other
flavors of FG Nups (Patel et al., 2007; Yamada et al., 2010) have
similar interaction mechanisms.

Characterizing the nature of fuzzy complexes such as inter-
actions involving FG Nups represents a significant challenge
to the current toolbox of structural biology. Detailed understand-
ing requires novel approaches and integration of a range of
experimental and computational methods. SANS with contrast
matching is an underrepresented technique for characterizing
“unstructured” biology and measurements performed here
provide important guides to how IDP complexes can be
described in more detail than “fuzzy”. By combining large-scale
ensemble analysis, and validation by ensemble testing from two
different data sources, we demonstrate that even flexible fuzzy complexes can be approached using SANS with contrast matching.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **CONTACT FOR REAGENT AND RESOURCE SHARING**
- **METHOD DETAILS**
  - Protein Expression and Purification
  - Contrast Matching Experiments
  - Small Angle Neutron Scattering
  - Dynamic Light Scattering
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Calculation of $R_g$
  - Worm-Like Chain Modeling
  - Ensemble Analysis
  - Generation of a Pool of Conformers from ANTON Simulations

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and two tables and can be found with this article online at https://doi.org/10.1016/j.str.2018.01.010.

ACKNOWLEDGMENTS

The authors thank Dr. Paul Dominic Olinares and Dr. Brian Chait for graciously providing mass spectrometry analysis and Drs. Barak Raveh and Ryo Hayama for their helpful discussion. Dr. Shuo Qian provided technical advice at ORNL. This work was supported by NIH GM117212 (D.C.) and GM112108 (M.R.). This research used resources at the High Flux Isotope Reactor, a DOE Office of Science User Facility operated by the Oak Ridge National Laboratory. Simulations used the Anton special-purpose supercomputer provided by the National Resource for Biomedical Supercomputing (NRBSC), the Pittsburgh Supercomputing Center (PSC), and the Biomedical Technology Research Center Resource for Biomedical Supercomputing (NBRSC), the Pittsburgh Supercomputing Center (PSC), and the Biomedical Technology Research Center.

AUTHOR CONTRIBUTIONS

S.S. and D.C. contributed the concept and analysis, S.S. and D.B.T. conducted scattering measurement. M.P.R. and D.C. supervised sample preparation. S.S. and D.C. drafted the paper, and all authors reviewed and finalized the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 2, 2017
Revised: October 27, 2017
Accepted: January 12, 2018
Published: February 8, 2018

REFERENCES


STAR METHODS

KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacterial and Virus Strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BL21-Gold(DE3) strain</td>
<td>Agilent Technologies</td>
<td>#230132</td>
</tr>
<tr>
<td><strong>Chemicals, Peptides, and Recombinant Proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deuterium oxide</td>
<td>Cambridge Isotope Lab.</td>
<td>DLM-6-1000</td>
</tr>
<tr>
<td>Glutathione sepharose 4B</td>
<td>GE Health</td>
<td>Cat#17075601</td>
</tr>
<tr>
<td>TALON resin</td>
<td>Clonetech</td>
<td>Cat#635501</td>
</tr>
<tr>
<td>Human alpha thrombin</td>
<td>Enzyme research laboratories</td>
<td>Cat#HT1002a</td>
</tr>
<tr>
<td><strong>Recombinant DNA</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plasmid: pET-FSFG-K-His</td>
<td>(Hough et al., 2015)</td>
<td>N/A</td>
</tr>
<tr>
<td>Plasmid pGEX-GST-Kap95</td>
<td>(Hough et al., 2015)</td>
<td>N/A</td>
</tr>
<tr>
<td>Plasmid: pRSF-NTF2-His</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Software and Algorithms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ensemble from simulation</td>
<td>(Raveh et al., 2016)</td>
<td><a href="https://www.ncbi.nlm.nih.gov/pubmed/27091992">https://www.ncbi.nlm.nih.gov/pubmed/27091992</a></td>
</tr>
<tr>
<td>PDB generation, VMD</td>
<td>U. Illinois</td>
<td><a href="http://www.ks.uiuc.edu/Research/vmd/">http://www.ks.uiuc.edu/Research/vmd/</a></td>
</tr>
<tr>
<td>SCATTER</td>
<td>Biosis</td>
<td><a href="http://www.biosis.net/tutorial/9">http://www.biosis.net/tutorial/9</a></td>
</tr>
<tr>
<td>Contrast match calculation</td>
<td>MULCH</td>
<td><a href="http://smb-research.smb.usyd.edu.au/NCVWeb/">http://smb-research.smb.usyd.edu.au/NCVWeb/</a></td>
</tr>
<tr>
<td>SANS data analysis using the ATSAS software suite</td>
<td>EMBL</td>
<td><a href="https://www.embl-hamburg.de/biosaxs/software.html">https://www.embl-hamburg.de/biosaxs/software.html</a></td>
</tr>
<tr>
<td>Ensemble generation</td>
<td>traDES</td>
<td><a href="http://trades.blueprint.org/">http://trades.blueprint.org/</a></td>
</tr>
<tr>
<td>DYNAMICS 7.1.0.25</td>
<td>Wyatt Technology</td>
<td><a href="https://www.wyatt.com/products/software/dynamics.html">https://www.wyatt.com/products/software/dynamics.html</a></td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SANS measurements</td>
<td>Bio-SANS Oak Ridge</td>
<td><a href="http://neutrons.ornl.gov/biosans">http://neutrons.ornl.gov/biosans</a></td>
</tr>
</tbody>
</table>

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, David Cowburn, cowburn@cowburnlab.org

METHOD DETAILS

Protein Expression and Purification
FSFG-K was expressed and purified as described previously (Hough et al., 2015). Briefly, BL21 (DE3) gold cells (Agilent) containing the expression plasmid (pET24a) were grown to an OD<sub>600</sub> of 0.6-0.8 and induced with 1 mM IPTG for 3 h at 37 °C. To produce [2H]-FSFG-K, samples were grown in M9 minimal media containing 99% D<sub>2</sub>O (Cambridge Isotopes Laboratories, Tewksbury MA) with natural abundance glucose and NH<sub>4</sub>Cl as the sole carbon and nitrogen source, respectively (Shekhtman et al., 2002). Cells were lysed under denaturing conditions (8 M urea) and purified over Talon resin in lysis buffer (20 mM HEPES, 150 mM KCl, 2 mM MgCl<sub>2</sub>, pH 7.4) with a protease inhibitor cocktail tablet, and with additional AEBSF and pepstatin A. The column was successively washed with lysis buffer with 8 M urea and the protein was eluted in elution buffer (20 mM HEPES-KOH, pH 6.8, 150 mM KCl, and 250 mM imidazole) with no urea. The elution was concentrated by centrifugal concentrators with 3 kDa MWCO (EMD Millipore, MA), and gel filtered using a Superdex S-200 column. Yeast NTF2 was expressed using a pRSFduet expression plasmid and purified in an identical manner to FSFG-K except that urea was absent from all the buffers. Kap95 was expressed and purified as previously described (Hough et al., 2015). Cleavage of the GST-tag was performed by incubating Kap95GST with thrombin overnight. The sample was passed through benzamidine sepharose (GE) to remove thrombin, followed by glutathione Sepharose 4B (GE) resin to remove free GST. After thrombin cleavage, Kap95 was further purified by gel filtration using a Superdex S-200 column. Partially deuterated Kap95 was produced by growing Kap95 in M9 minimal media containing 85% D<sub>2</sub>O with natural abundance glucose and NH<sub>4</sub>Cl as the sole carbon and nitrogen source, respectively. MALDI-TOF mass spectrometry was performed to determine the percent deuterium incorporation achieved by comparing the mass of the partially [2H]-Kap95 to a sample prepared under identical
conditions but grown in regular media (H2O). The amount of deuterium incorporation (of non-exchangeable protons) was 62% and this value was used to calculate the scattering length density.

**Contrast Matching Experiments**

Samples were prepared by dialysis of separate stock solutions of [2H]-FSFG-K, NTF2 and Kap95 into 42% D2O buffer (20 mM HEPES, 150 mM KCl, 2 mM MgCl2, pH 6.8). FSFG-K and [2H]-Kap95 was dialyzed, separately, in identical buffer (with additional 0.5 mM TCEP) but with 92% D2O to match [2H]-Kap95. In both cases, the dialysate was used to measure buffer scattering background. The web service MULCH (Whitten et al., 2008) was used to determine the contrast match point by calculating the scattering length density of [62%-2H]-Kap95 and the buffer at different percentages of D2O. The neutron scattering length density computed for 92% D2O buffer was 5.79 x 10^10 cm^-2 and 3.41 x 10^10 cm^-2 for natural abundance FSFG-K. The scattering length density for 42% D2O buffer was 2.36 x 10^10 cm^-2 and 6.15 x 10^10 cm^-2 for [2H]-FSFG-K.

Protein concentrations for FSFG-K and NTF2 were measured by BCA assay kit (ThermoScientific, MA) following the manufacturer’s instructions. The concentration of Kap95 was determined by OD280 based on standard amino acid content (ε280 = 85,260 M^-1 cm^-1).

**Small Angle Neutron Scattering**

SANS measurements were conducted at the Bio-SANS instrument at the High-Flux Isotope Reactor, Oak Ridge National Laboratory (Heller et al., 2014). Data measurements of the contrast match series obtained at 42% D2O were acquired with a sample-to-detector distance of 2.53 m providing a q range of 0.021-0.397 Å^-1. Data measurements made for the inverse contrast matching series at 92% D2O were obtained using a dual-detector setup with sample-to-detector distances of 6 m for the main detector and 1.1 m for the wing detector covering a q range of 0.007 to 0.95 Å^-1. The scattering vector q is defined as \( q = 4\pi\lambda^{-1}\sin(\theta) \), where \( \lambda \) is the neutron wavelength (Å) with a wavelength spread, \( \Delta\lambda \), of 0.15 set by a neutron velocity selector. All samples were measured in a 1 mm path length at 25 °C. Samples containing [2H]-FSFG-K for the contrast match at 42% D2O were run for 3.5 h at 2.53 m. Natural abundance FSFG-K samples in 92% D2O at 20.2 mg/mL and at 10.6 mg/mL were acquired for 1 h, while samples at 5.3 mg/mL and 2.7 mg/mL were acquired for 3 h and 3.5 h, respectively. Inverse contrast match experiments with 5.3 mg/mL FSFG-K and partially-[2H]-Kap95 (1:0.2 molar ratio) were performed for 4.5 h. The scattering intensity profiles were obtained by azimuthally averaged the neutron detector counts, which were normalized to incident beam monitor counts. Detector dark current and pixel sensitivity were used for sensitivity correction and scattering from the quartz cell was subtracted. The software program PRIMUS from the ATSAS suite (Franke et al., 2017) was used to perform data merging as well as perform solvent subtraction. The software program DATGNOM, also from the ATSAS suite, was used to automatically determine the maximum intramolecular distance \( D_{max} \) and compute of the distance distribution function \( P(r) \) using the fitted value of \( R_g \) from Debye analysis as the initial estimate for expected \( R_g \).

**Dynamic Light Scattering**

Dynamic light scattering measurements were made on a Dynapro plate reader (Instruments, Santa Barbara, CA) at 298 K. Samples were centrifuged prior to experiment and were placed in a temperature-regulated cell at a temperature of 25.0 °C. Experiments were run in a 384 well plate format with 10 acquisitions of 5 s were acquired for each sample. The Dynapro software, DYNAMICS version 7.1.0.25, was used to analyze autocorrelation profiles with regularization fitting.

**QUANTIFICATION AND STATISTICAL ANALYSIS**

Calculation of \( R_g \)

Guinier approximation, which is the standard method of determining \( R_g \), extracts the \( R_g \) from scattering data within the small \( q \) regime. In this regime \( I(q) \) relates to \( \exp(-q^2R_g^2/3) \) and thus in the limit \( q \to 0 \), the slope of the scattering data, transformed as \( \log[I(q)] \) versus \( q^2 \), allows for estimation of \( R_g \). For the Guinier approximation an upper limit for the \( q \) range depends on the particle shape and homogeneity. For a sphere of uniform scattering density, Guinier analysis is valid in the limit of \( q_{max}R_g < 1.3 \). However, for highly disordered systems, Guinier analysis is known to be restricted to very narrow \( q \)-range \( (q_{max}R_g < 0.7-1) \) since higher order terms within the Guinier approximation become significant when distances between scattering elements of a polymer chain become large (Borgia et al., 2016). Therefore, determining \( R_g \) by Guinier analysis is an experimental issue where the region of validity is either limited to only a few data points, because of limitation in instrument configuration, or the scattering data is hidden by the beam-stop and is therefore inaccessible. Guinier plots and analysis was performed using the software program SCATTER (Forster et al., 2010). For the Guinier analysis of the contrast matching datasets the instrument configuration (a sample-to-detector distance of 2.53 m) did not allow for sufficient coverage of the low \( q \) Guinier region to allow for analysis within the limit of \( q_{max}R_g < ~1 \). The free [2H]-FSFG-K in 42% D2O, which had a smallest relative \( R_g \) was subsequently the only dataset to be fit albeit with only the first 5 data points. Guinier analysis of this dataset therefore should be considered an inaccurate measurement of the \( R_g \).

The Debye equation (Debye, 1947), which describes the behavior of a Gaussian chain, has been suggested as an alternative as it can be applied to a much larger \( q \) range \( (q_{max}R_g < 3) \) (Perez et al., 2001; Receveur-Brechot and Durand, 2012). Debye analysis was used to determine the \( R_g \) by the equation:

\[
\frac{I(q)}{I(0)} = \frac{2}{x^2} (x - 1) + e^{-x} 
\]

(Equation 1)
where \( q^2 R_g^2 \). The q range used in for fitting was optimized for each sample. However, in some cases it has been noted that the Debye model is known to fit poorly for chains with excluded volume at larger q (Petrescu et al., 1998).

The radial distribution function, \( P(r) \), is obtained by an indirect Fourier transform of the scattering pattern related by the equation (Svergun, 1992):

\[
P(r) = \frac{r^2}{2\pi^2} \int_0^\infty \frac{q^2 I(q) \sin(qr)}{qr} dq
\]

\( P(r) \) is equal to zero at \( r = 0 \) and is expected to decay smoothly to zero at \( r = D_{\text{max}} \). The \( R_g \) can be calculated from the \( P(r) \) by the function:

\[
R_g = \frac{\int_0^{R_g} r^2 P(r) dr}{\int_0^{\infty} r^2 P(r) dr}
\]

Calculation of \( P(r) \) and resulting \( R_g \) was performed using Gnom (Svergun, 1992). For disordered systems, \( R_g \) values obtained from the \( P(r) \) are considered a more reliable estimate relative to the values obtained Guinier approximation (Perez et al., 2001). Durand and co-workers state that the \( R_g \) by Guinier analysis for a completely unfolded protein would yield values that are systematically smaller than the “true” \( R_g \) whereas values calculated from Debye equation and \( P(r) \) were considered more accurate. Our results are in agreement showing smaller \( R_g \) values from fits to the Guinier approximation whereas values from Debye and \( P(r) \) are in better agreement (Table 1). Accurate extraction of \( R_g \) via \( P(r) \) methods requires a minimum q value of \( \sim \pi/D_{\text{max}} \). There is, however, an inherent uncertainty deriving the \( D_{\text{max}} \) and care is usually taken in deriving an optimal value (Trewhella et al., 2017). In some cases, underestimation of the maximum dimensions for an unfolded protein can lead to an underestimation of the \( R_g \) (Borgia et al., 2016). In this study, the \( R_g \) is calculated from the \( P(r) \), the Debye function and the values obtained from ensemble modeling agree reasonably well ensuring confidence in the obtained values.

**Worm-Like Chain Modeling**

The worm-like chain model (Kratky-Porod chain) is a form factor equation used to describe a polymer chain with a contour length, \( L \), and a persistence length, \( L_p \). The expression for the form factor used to describe the Kratky-Porod chain is written as (Sharp and Bloomfield, 1968):

\[
\frac{I(q)}{I(0)} = \frac{2}{\pi^2} \left( x - 1 + e^{-x} \right) + b \left( \frac{4}{15} + \frac{7}{15x} - \left( \frac{11}{15} + \frac{7}{15x} \right) e^{-x} \right)
\]

where \( x = \frac{q^2 L b}{6} \). From the fitted values of \( L \) and \( b \), one can compute the radius of gyration \( R_g \) by:

\[
R_g^2 = b^2 \left[ \left( \frac{y}{6} - \frac{1}{4y} - \frac{1}{8y^2} \right) \left( 1 - e^{-y} \right) \right]
\]

where \( y = \frac{L}{D} \). This function is valid for \( L > 10 \) and \( q < \frac{3}{D} \). Note that for the samples with [\(^{2}H\)]-FSFG + Kap95 at 42% D2O these conditions could not be maintained.

From the calculated \( L_p \), the average number of amino acid residues that are locally rigid, \( n_p \), can be calculated by \( n_p = L_p / l_0 \), where, \( l_0 = 3.78 \) Å, the distance between two adjacent Cα residues. Fitting was performed in the limit of \( q < 0.13 \) Å\(^{-1}\). For many samples, the value of \( L \) needed to be constrained to the theoretical limit. The maximum \( L \), is defined as \( L \equiv n l_{ff} \), where \( n \) is the number of amino acids \( (n = 133) \), \( l_0 = 3.78 \) Å, and \( f \) is a geometric factor \( (f = 0.95) \). FSFG-K, therefore, has a maximum \( L \) of 477.6 Å. As noted previously (Daughdrill et al., 2012; Perez et al., 2001), a mathematical limitation inherent to WLC model, the dependence of the function on the product \( Lb \) can lead to high error in the fitted value \( L \). With the imposed constraint, we were able to observe good fits to experimental data \( (R^2 > 0.99 \text{ for each sample with the exception of the two lowest concentration (5.3 and 2.7 mg/mL) apo FSFG-K samples in 92% D}_2\text{O}) \).

From polymer theory, we estimate the root-mean-square-distance, \( \langle r^2 \rangle^{1/2} \) (in picometers), between ends of an ideal chain undergoing random walk as:

\[
\langle r^2 \rangle^{1/2} = a \sqrt{n}
\]
**Ensemble Analysis**

Ensemble analysis was performed using the EOM 2.0 software from the ATSAS package (Tria et al., 2015). This approach involved generating a pool of random coil protein models. This was performed using trajectory directed ensemble sampling traDES (Feldman and Hogue, 2000, 2002) to produce a pool of 100,000 possible structures. Theoretical scattering of each model was then computed using CRYSON from the ATSAS package (Svergun et al., 1998). This was performed twice, calculating for 42% D$_2$O and 80% deuteration of [2H]-FSFG and calculating for 92% D$_2$O and 0% deuteration. The genetic algorithm method, GAJOE, within EOM 2.0 package was subsequently used to select from the pool of theoretical scattering curves, a subset of structures whose weighted average scattering curve, that best fits the contrast matching data. GAJOE was run with default parameters except for allowing the algorithm to use a minimum of 1 curve per ensemble (default = 5) and not enabling the subtraction of a constant value. GAJOE was repeated 1,000 times and no constant subtraction was used.

**Generation of a Pool of Conformers from ANTON Simulations**

All-atom molecular dynamics simulations of FSFG-K with NTF2 using the TIP4pD water model (Piana et al., 2015) was performed previously using the Anton supercomputer (Raveh et al., 2016). The trajectory consisted of 19,930 frames with a frame rate of 0.06 ns/frame, totaling ~1.2 µs. The trajectory was loaded into VMD (Humphrey et al., 1996) and PDB files of FSFG-K were written for each frame. As above, CRYSON was used to compute the theoretical scattering for each model and GAJOE was used for ensemble selection.