Interdomain linkers tailor the stability of immunoglobulin repeats in polyproteins

Tanuja Joshi, Surbhi Garg, Alejandro Estaña, Juan Cortés, Pau Bernadó, Sayan Das, Anjana R Kammath, Amin Sagar, Sabyasachi Rakshit

HAL Id: hal-03161535

https://laas.hal.science/hal-03161535

Submitted on 7 Mar 2021

Interdomain linkers tailor the stability of Immunoglobulin repeats in polyproteins

Tanuja Joshi[1], Surbhi Garg[1], Alejandro Estaña[3,4], Juan Cortés[4], Pau Bernadó[3], Sayan Das[1], Anjana R. Kammath[1], Amin Sagar[1,3]*, Sabyasachi Rakshit[1,2]*

Author Affiliation

[1]Department of Chemical Sciences, Indian Institute of Science Education and Research Mohali, Punjab, India

[2]Centre for Protein Science Design and Engineering, Indian Institute of Science Education and Research Mohali, Punjab, India

[3]Centre de BiochimieStructurale. INSERM, CNRS, Université de Montpellier, Montpellier, France

[4]LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France

Abstract

Linkers in polyproteins are considered as mere spacers between two adjacent domains. However, a series of studies using single-molecule force spectroscopy have recently reported distinct thermodynamic stability of I27 in polyproteins with varying linkers and indicated the vital role of linkers in domain stability. A flexible glycine rich linker (-$(GGG)_n$, $n \geq 3$) featured unfolding at lower forces than regularly used arg-ser (RS) based linker. Interdomain interactions among I27 domains in Gly-rich linkers were suggested to lead to reduced domain stability. However, the negative impact of inter domain interactions on domain stability is thermodynamically counter-intuitive and demanded thorough investigations. Here, using an array of ensemble equilibrium experiments and in-silico measurements with I27 singlet and doublets with two aforementioned linkers, we delineate that the inter-domain interactions in

fact raise the stability of the polyprotein with RS linker. More surprisingly, a highly flexible Gly-rich linker has no interference on the stability of polyprotein. Overall, we conclude that flexible linkers are preferred in a polyprotein for maintaining domain's independence.

Key word list:

Inter domain linkers (IDLs), Polyproteins, Domain stability, Protein thermodynamics, Structural propensity, Single molecule force spectroscopy(SMFS).

Introduction

Domains as independent evolutionary protein units, have diversely combined in tandem through interdomain linkers (IDLs) and evolved into a repertoire of polyprotein or multidomain proteins (MDP)[1],[2]. Polyproteins constitute more than two-thirds of proteins and control the majority of the function in both eukaryotes and prokaryotes[3],[4],[5]. The *in vitro* synthesis of polyproteins via conjugation often introduces unwanted and uncontrolled diversity in the chemical compositions[6], lengths[7], and secondary structures[8] of IDLs in polyprotein. The diversity in IDL was overlooked by considering their role as mere spacers that demarcate neighboring domains spatially. Recently, a series of single-molecule force spectroscopy (SMFS) studies reported an anomaly in the thermodynamic stability of I27 domains in polyproteins of varying IDLs and thus, additional roles of IDLs in polyproteins were realized[9],[10],[11]. Lower stability of I27 domain was reported in polyprotein with relatively longer, flexible, and Gly-rich IDLs in comparison to short -Arg-Ser-linker (RS linker)[12]. It was reported that the Gly-rich flexible linkers facilitate inter domain interactions, and as a result reduce the domain stability. However, domain-domain interactions reducing domain-stability is contradictory to thermodynamics of natural evolution. These experimental outcomes thus raised doubts on the true estimation of the stability of the individual domains in

polyproteins, the role of the IDLs beyond spacers, and the molecular mechanisms in context of IDLs that alter the stabilities of constituent domains in polyproteins. Here using doublets of I27 domain with RS and poly-Gly linkers, and single domain of I27, we first estimate the effect of linkers on domain stability and subsequently, decipher the molecular mechanism that regulates the domain stability.

Results and Discussion

Here, we generated two constructs of I27 doublet connected with RS and poly-Gly (GGGSGGGG) linkers (Figure 1A & 1B) and compared the thermodynamic stabilities of the doublet variants withsingletI27. IDLs do not alter the intrinsic secondary structures of the I27 domains as noticed from the nearly overlapping far-UV CD spectra of doublets with peak-minima at 215 nm, revealing a β dominant system for all proteins (Figure 1C). However, we noticed a unique positive peak at 230 nm for I27 doublet with RS linker. A positive peak at 230 nm usually occurs either from the tertiary interactions of aromatic residues or from the disulfide bonds[13]. To quantify the sulfhydryl groups in the proteins, we performed Ellman's test and detected no disulfide bond in either of the doublets. We collected CD spectra in the presence and absence of disulfide reducing agent. We found no difference, indicating that an aromatic interaction may be responsible for the 230 nm peak in I27 (RS)-doublet. Identical CD spectra for doublets with the I27-singlet in the near-UV range (250 – 320 nm) further indicated that the fusions of two I27 domains had not disturbed the tertiary arrangement of the aromatic residues (SI Figure 1). Overall, our CD data reassures the maintenance of native I27 fold for both the doublets; however, it indicated aromatic stacking or cation-π interactions[14] in I27- RS-I27.We performed analytical size exclusion chromatography where I27 singlet eluted, at last followed by I27-RS-I27 and I27-(GGG)-I27, respectively. I27-RS-I27 eluted at

a higher elution volume compared to I27-(GGG)-I27 (Figure 1D). We expect RS may have a compact structure (SI Figure 2 and SI Table 1).

To evaluate the effect of IDLs (RS, GGGSGGGG) on the equilibrium properties of I27 doublets, we performed chemical and thermal denaturation experiments. Since the critical transition points (d_{50} or T_m) are the most reliable and model-free estimates of thermodynamic stability for polyproteins with intrinsically complex distributions of folded multimers and unfolded monomers in equilibrium[15], we restricted our analysis to the critical transition points and tallied the differences. All protein constructs featured apparent two-state unfolding with Guanidinium chloride (GdmCl) (Figure 2A), indicating a cooperative nature in domain unfolding. The f_D as a denaturant concentration function featured identical behavior with the I27 singlet and the doublet GGGSGGGG (Table 1). However, I27 (RS)-doublet measured a significantly higher d_{50}. The estimation of the thermodynamic m-value that measures the solvent accessible surface area[16] (SASA) also featured a similar trend as d_{50} (Table 1). A higher m-value indicates less SASA of the protein than the singlet-I27 in line with the CD and SEC estimations of a compact structure for I27-RS-I27.

The thermal melts monitored from the changes in mean residue ellipticity (MRE) at 229 nm and 205 nm using CD (Figure 2B) and tryptophan fluorescence (SI Figure 3) also featured similar trends in the f_D with cooperative two-state unfolding. We observed isodichroic and isoemissive points for all transitions reinstating the cooperative two-state thermal unfolding phenomena for proteins (SI Figure 4). I27-RS-I27 withstood a higher transition temperature than I27doublet with GGGSGGGG and I27 singlet (Figure 2B), (Table1). Overall, we noticed higher resistance towards chemical and thermal perturbations for doublet with RS as an IDL. RS linker imparted a stabilizing effect, whereas GGGSGGGG maintained the domain independence of I27 and unfolded as the single-domain. This is an interesting observation and defies the existing reports of lower stability with poly-Gly linkers. Therefore, we rectify the

existing knowledge and report that RS-linker induces higher stability to domains whereas poly-Gly linker maintains domain individuality. However, the molecular origins for stronger thermodynamic stability of I27-(RS) doublet is not yet clear.

To decipher the intramolecular and inter-domain interactions responsible for distinct equilibrium properties of I27-doublet with the different linkers, we performed all-atom Gaussian Accelerated Molecular Dynamics (GAMD) simulations. Four independent GAMD simulations were performed for all the doublets (2*50 ns, 1*100 ns, 1*200ns) leading to 400ns of simulation time for each doublet. While GGGSGGGG maintains the domain boundaries as spacer throughout the simulations, RS forms extensive intermittent contacts with the domains and gets buried between the two domains, essentially ceasing to act like a linker. We calculated the reweighted two-dimensional potential of mean force (PMF) profiles based on angle and distance between the domains (Figure 2C). Expectedly, I27-(GGG)-I27possesses a wide and shallow PMF, indicating that the two domains sample a wide variety of the orientations with small energy barriers, leading to the formation of transiently stable domain-domain interactions multiple times. No domain-linker interactions are identified for I27-(GGG)-I27 in the simulations.

Conversely, I27(RS)-doublet has a much narrower PMF profile with a distinct energy minimum corresponding to compact structures with inter domain angle of ~50º and distance of ~ 30 Å between their geometric centers. The radius of gyration (R_g) as a simulation time function also clearly features the evolution of compact structure for I27-(RS) doublet. We also noticed an increasing number of hydrogen bonds (H-bond) between the linker and the domain, specifically R90 from the linker forms stable H-bonds with E88 and E118 (Figure 3D, E and Supporting Movie 1) and transient H-bonds with L92, E94, and D120. The H-bond partners for R90 are distributed in both the domains, thus stabilizing a compact conformation. We, therefore, report that domain-linker contacts may contribute to higher stability in RS linker.

How does linker structure induce domain-linker contacts? To decipher this, we calculated the most preferred SS of the linkers using a recently developed computational tool called LS2P[17]. The method predicts the secondary structure propensities of a stretch of amino acids by breaking it into a sequence of overlapping tri-peptides. Next, it performs a statistical analysis from a structural database of tripeptides, while accounting the local sequence context. Local secondary structure preference for RS showed a high propensity to a truncated α-helix. A truncated helix is structurally very similar to turn, which transiently induces bent structures, whereas GGGSGGGG preferred random coil structure. A bent linker likely to bring neighboring domains in close proximity and may promote inter-domain and domain-linker contacts.

Next, to estimate the effect of linkers on the conformational variations among doublets theoretically, we generated 10,000 structural models of each I27 doublets using tripeptide-based ensemble generation method[18].This method, which exploits local sequence-dependent structural propensities of the amino acids, has been shown to generate realistic ensemble models of intrinsically disordered proteins/regions. From the generated structures, we mapped the conformational variations across doublets by tracing the relative orientations of the domains from the inter domain angles and distances in two-dimensional contour maps (Figure 3C& D). We noticed the broadest distribution of the inter-domain orientations for I27-(GGG)-I27, as anticipated from the highest flexibility of the linker. For further clarity in the comparisons, we plotted differential contours for doublet with RS linker by subtracting the normalized two-dimensional histogram (angle vs. distance) from I27-(GGG)-I27 (Figure 3D right panel). Interestingly, RS-doublet explores the narrowest distribution with shortest angles and distances. The map for I27-(RS)-I27 indicates minimal conformational entropy in the native form and more compact conformation, in clear agreement with the analytical SEC and GAMD.

To experimentally verify that the ensembles generated from the tripeptide-based search methods are compatible with the conformations of the doublets in solution, we collected Small Angle X-ray Scattering (SAXS) data for doublets and analyzed the data using Ensemble Optimization Method (EOM). The details of SAXS data collection and structural parameters derived from SAXS are provided in SI Table 2. By treating the linkers as random coils and the high-resolution structures of domains as rigid bodies, we generated 10,000 independent structural models for both doublets. All models were filtered using the genetic algorithm implemented in EOM to derive sub-ensembles that collectively describe the corresponding experimental SAXS data. For comparison, we then overlaid the distribution of the inter-domain angles, and distances of the EOM selected conformations on the corresponding conformational map of doublet directly obtained from the ensemble modeling method (Figure 3C-D, Left column). The SAXS profiles and the fits from EOM are presented in Figure 3C-D (Middle column). The Guinier plots are included in the inset to demonstrate the absence of aggregation or inter-particle interaction. It is important to note that many conformations at a given angle between the long axis of the domains (tilt angle) will have indistinguishable SAXS profile irrespective of inter domain rotation around the long axis (azimuthal angle). However, SAXS is sensitive to the overall shape and size of the protein and thus provides a useful way to test if the structures generated by our approach are compatible with the conformations visited in the solution. For both the doublets, there is an excellent agreement between the interdomain angles and distances of sub-ensembles selected by EOM and ensembles generated using the tripeptide approach. We noticed the broadest distribution of the inter-domain orientations for I27-(GGGSGGGG)-I27.

Interestingly, RS-doublet explores a narrower distribution with shorter angles and distances. The map for I27-(RS)-I27 indicates minimal conformational entropy in the native form and more compact conformation, in apparent agreement with the analytical SEC. It is pertinent to

mention here that the ensemble generation method just ensures a lack of steric clashes and does not carry out any energy minimization that may promote interdomain interactions. The technique thus indicates that the compact conformations of I27-(RS)-I27 are driven by the conformational preferences of the IDL. A compact conformation may facilitate aromatic stacking or interaction of aromatic amino acids with charged amino acids across domains or with linkers, as noticed in CD.

It is important to note here that the bent and compact conformation of I27-(RS) doublet is predominantly driven by the native conformation of the RS linker, as estimated from the tripeptide-based ensemble method. The domain-linker H-bond interactions elevate the system's overall enthalpy, featuring increased thermal and chemical stability in equilibrium. Further, F14 of the N-terminal I27 forms frequent contacts with arginine of RS linker, which may be responsible for the peak at 230 nm in CD (SI Figure 5).

The higher stability for RS linker is previously reported for the polyprotein constructs of I27 in single-molecule force spectroscopy (SMFS) studies[12]'[19]. Compared to RS linker, SMFS estimated a low-force unfolding for I27-polyproteins with Glycine rich linkers[12] and factually concluded that Glycine rich linker is destabilizing the domains. In agreement with SMFS, we measured lower resistance to thermal and chemical perturbations for I27-(GGGSGGGG)-I27 than I27-(RS)-I27 in ensemble equilibrium measurements. However, we noticed matching thermodynamic stability of I27-(GGGSGGGG)-I27with I27 singlet.

Therefore, we first rectified the factual error in the literature and restated that the RS-linker in I27-polyprotein further stabilizes the protein cooperatively. In contrast, Glycine rich linker maintains the equilibrium properties of the individual I27 domains in the polyprotein. We next used a series of experimental and computational probes, including analytical SEC, SAXS, conformational ensemble modeling techniques, and GAMD simulation, to decipher the molecular interactions that contribute to the additional stability in I27-(RS)-I27. Our

experimental studies identified that the RS-linker imposes a compact geometry with far less conformational entropy than the GGGSGGGG linker. The conformational maps estimated from the ensemble modeling method that accounts only local structural preferences and steric clash avoidance (disregarding long-range molecular interactions) and GAMD that resolves the structural and dynamic information at the atomic level further ratified the compactness of the I27-(RS)-I27 over I27-(GGGSGGGG)-I27.Multiple GAMD simulations highlighted that the stable inter-domain interactions favor the compact conformation for I27-(RS)-I27. The frequency of the interdomain interactions for I27-(RS)-I27 is somewhat smaller than I27-(GGGSGGGG)-I27, however, significantly long-lived and enthalpically drive the compact conformation as the most populated one within the simulation trajectories while the interactions are transient for GGGSGGGG.

Further, we identified that E88, E118, L92, E94, and D120 residues from the domains and R90 from linker participate in the domain-domain and domain-linker interactions for I27-(RS)-I27. We can observe that the difference in the linker region's secondary structure accounts for the differences in the thermodynamic stability of the connecting domains. It is pertinent to comment here that insufficient sampling in MD can bias the conformational map and may erroneously indicate more compact conformation for I27-(GGGG)-I27 than I27-(RS)-I27, as reported previously[11].

Conclusion

In conclusion, our study outlines the attributes of IDLs with disparate structural features on synthetic polyproteins of Ig-domains. Linkers with a random-coil secondary structure can maintain the thermodynamic property of individual domains in the polyprotein. However, linkers with a high propensity to induce turns can facilitate domain-domain and domain-linker interactions and stabilize the polyprotein by increasing the overall enthalpy of the system.

These outcomes guide us to tune the thermodynamics of polyproteins by systematic designs of IDLs.

Accession Number for SAXS data

SASDKB2 – Two titin I27 domains linked with RS linker

SASDKC2 – Two titin I27 domains linked with GGGSGGGG linker

Availability of Data and Materials

All data generated or analyzed during this study are included in this article and its supplementary information files. The accession numbers are provided to access the SAXS data.

Corresponding Address

S.R. srakshit@iisermohali.ac.in

A.S. amin.sagar@cbs.cnrs.fr

Competing Interest

The authors declare no competing interest.

References

[1]   J.-H. Han, S. Batey, A.A. Nickson, S.A. Teichmann, J. Clarke, The folding and evolution of multidomain proteins, Nat. Rev. Mol. Cell Biol. 8 (2007) 319–330. https://doi.org/10.1038/nrm2144.

[2]   A. Borgia, K.R. Kemplen, M.B. Borgia, A. Soranno, S. Shammas, B. Wunderlich, D. Nettels, R.B. Best, J. Clarke, B. Schuler, Transient misfolding dominates multidomain protein folding, Nat. Commun. 6 (2015) 8861. https://doi.org/10.1038/ncomms9861.

[3]   M. Levitt, Nature of the protein universe, Proc. Natl. Acad. Sci. 106 (2009) 11079–11084. https://doi.org/10.1073/pnas.0905029106.

[4]   G. Apic, J. Gough, S.A. Teichmann, Domain combinations in archaeal, eubacterial and eukaryotic proteomes., J. Mol. Biol. 310 (2001) 311–25. https://doi.org/10.1006/jmbi.2001.4776.

[5]   S. Batey, A.A. Nickson, J. Clarke, Studying the folding of multidomain proteins, HFSP J. 2 (2008) 365–377. https://doi.org/10.2976/1.2991513.

[6]   R.M. Bhaskara, A.G. de Brevern, N. Srinivasan, Understanding the role of domain–domain linkers in the spatial orientation of domains in multi-domain proteins, J. Biomol. Struct. Dyn. 31 (2013) 1467–1480. https://doi.org/10.1080/07391102.2012.743438.

[7]   V.P. Reddy Chichili, V. Kumar, J. Sivaraman, Linkers in the structural biology of protein-protein interactions., Protein Sci. 22 (2013) 153–67. https://doi.org/10.1002/pro.2206.

[8]   N. Amet, H.-F. Lee, W.-C. Shen, Insertion of the Designed Helical Linker Led to Increased Expression of Tf-Based Fusion Proteins, Pharm. Res. 26 (2009) 523–528. https://doi.org/10.1007/s11095-008-9767-0.

[9]   M. Carrion-Vazquez, A.F. Oberhauser, S.B. Fowler, P.E. Marszalek, S.E. Broedel, J. Clarke, J.M. Fernandez, Mechanical and chemical unfolding of a single protein: A comparison, Proc. Natl. Acad. Sci. 96 (1999) 3694–3699. https://doi.org/10.1073/pnas.96.7.3694.

[10]   B. Yang, Z. Liu, H. Liu, M.A. Nash, Next Generation Methods for Single-Molecule Force Spectroscopy on Polyproteins and Receptor-Ligand Complexes, Front. Mol. Biosci. 7 (2020) 1–19. https://doi.org/10.3389/fmolb.2020.00085.

[11]   F. Tian, G. Li, B. Zheng, Y. Liu, S. Shi, Y. Deng, P. Zheng, Verification of sortase for protein conjugation by single-molecule force spectroscopy and molecular dynamics simulations., Chem. Commun. (Camb). 56 (2020) 3943–3946. https://doi.org/10.1039/d0cc00714e.

[12]   S. Garg, G.S. Singaraju, S. Yengkhom, S. Rakshit, Tailored Polyproteins Using Sequential Staple and Cut, Bioconjug. Chem. 29 (2018) 1714–1719. https://doi.org/10.1021/acs.bioconjchem.8b00163.

[13]   S.M. Kelly, N.C. Price, The use of circular dichroism in the investigation of protein structure and function., Curr. Protein Pept. Sci. 1 (2000) 349–84. https://doi.org/10.2174/1389203003381315.

[14]   T.W. Craven, M.-K. Cho, N.J. Traaseth, R. Bonneau, K. Kirshenbaum, A Miniature Protein Stabilized by a Cation−π Interaction Network, J. Am. Chem. Soc. 138 (2016) 1543–1550. https://doi.org/10.1021/jacs.5b10285.

[15]   C. Park, S. Marqusee, Analysis of the stability of multimeric proteins by effective Δ G and effective m -values, Protein Sci. 13 (2004) 2553–2558.

https://doi.org/10.1110/ps.04811004.

[16]   J.K. Myers, C. Nick Pace, J. Martin Scholtz, Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding, Protein Sci. 4 (1995) 2138–2148. https://doi.org/10.1002/pro.5560041020.

[17]   A. Estaña, A. Barozet, A. Mouhand, M. Vaisset, C. Zanon, P. Fauret, N. Sibille, P. Bernadó, J. Cortés, Predicting Secondary Structure Propensities in IDPs Using Simple Statistics from Three-Residue Fragments, J. Mol. Biol. 432 (2020) 5447–5459. https://doi.org/10.1016/j.jmb.2020.07.026.

[18]   A. Estaña, N. Sibille, E. Delaforge, M. Vaisset, J. Cortés, P. Bernadó, Realistic Ensemble Models of Intrinsically Disordered Proteins Using a Structure-Encoding Coil Database, Structure. 27 (2019) 381-391.e2. https://doi.org/10.1016/j.str.2018.10.016.

[19]   B. Yang, Z. Liu, H. Liu, M.A. Nash, Next Generation Methods for Single-Molecule Force Spectroscopy on Polyproteins and Receptor-Ligand Complexes, Front. Mol. Biosci. 7 (2020) 1–19. https://doi.org/10.3389/fmolb.2020.00085.

Figure legends

Figure 1. Linkers generate the conformational variations among doublets. A-B) Schematics of I27 doublet variants possessing different inter domain linkers RS (green), and GGGSGGGG (red). **B)** Far UV CD spectra of I27 singlet and I27 doublet variants to monitor the effect of linkers on the secondary structures of polyprotein. **D)**Comparative size-exclusion chromatograms (SEC) of the doublet variants with singlet I27 monitored using absorbance at 280 nm.

Figure 2. **Domain-linker contacts stabilize RS-doublet. A&B)** Chemical and thermal unfolding of I27 doublets and I27 singlet as monitored by intrinsic tryptophan fluorescence and far-UV CD, respectively. The data are normalized to estimate unfolded populations (see Methods). The solid lines are the Boltzmann fits to data. **C)** The 2D (Inter domain angle, distance) potential of mean force (PMF) calculated by reweighting multiple simulations combined for I27 doublet variants. **D)** A plot of the number of H-bonds between the linker and the domains (left X-axis) and the radius of gyration (R_g) (right X-axis) as a function of simulation time. **E)** A representation of the stable conformations visited by RS where the linker forms multiple hydrogen bonds (gray cylinders) with both the I27 domains.

Figure 3. Linkers control the architectures of polyproteins by steering inter-domain orientations. A&B) Different secondary structure propensities of the linkers in the doublets as predicted using LS2P (5 amino acid residues from both ends of the linker are the part of I27 domain). **C&D)** The left panels show the 2D histogram of the inter domain angles and distances estimated from the corresponding ensemble generated structures as visited by I27 doublet variants (panel **C** for GGGSGGGG and **D** for RS). Black circles are the angle-distance distributions of I27 doublets as obtained from the experimental SAXS data after analyzing

using EOM. The SAXS profiles along with the fit by EOM (black line) are shown in the middle panel. The insets show the Guinier plots. The difference in the normalized angle-distance histogram of the I27-(RS)-I27 and I27-(GGGSGGGG)-I27. (Right panel)
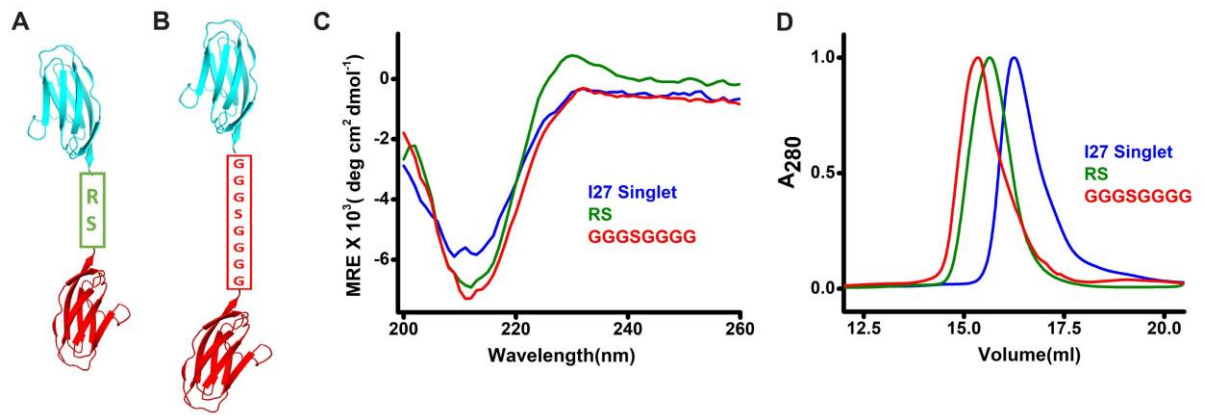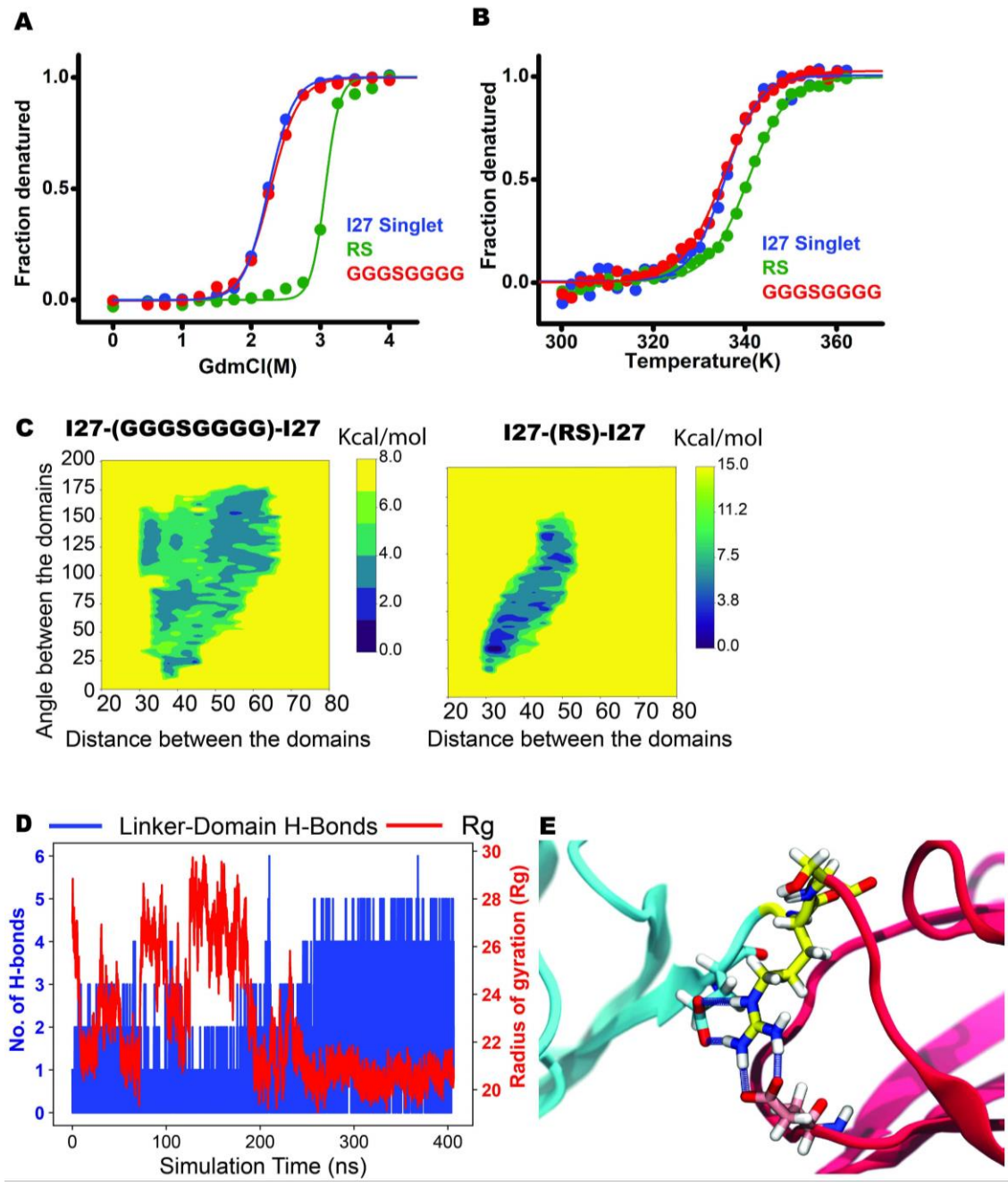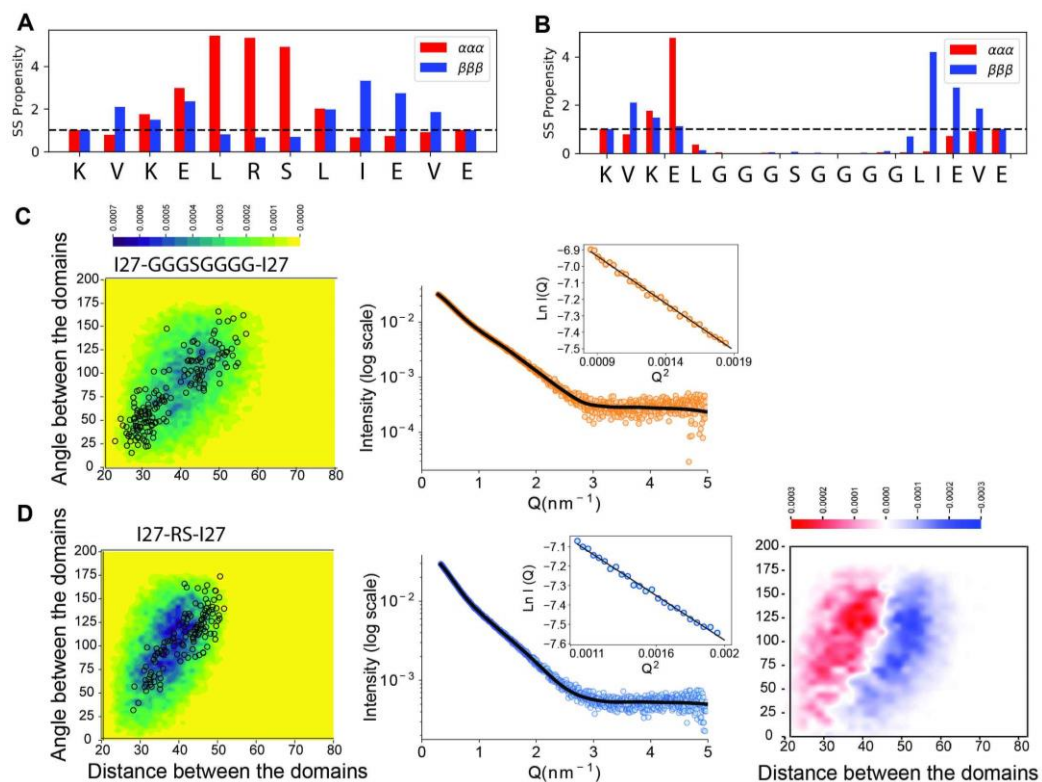
Figure 1.

Figure 2.

Figure 3.

Table 1: Summary of thermodynamic variables obtained from denaturation experiments.

Linker	d_{50} (M)	-m (kcal/mol.M)	T_m(K)
Singlet	2.2±0.02	3.4±1.0	334.3±0.4
RS	3.0±0.01	5.9±1.8	340.2±1.1
GGGSGGGG	2.2±0.18	3.0±0.1	334.5±0.5

The variations in the table are the measure of standard error of mean (SEM) from different measurements.