Electronic Supplementary Material (ESI) for Soft Matter.

Supplementary Information: Molecular Dynamics-Based Strength Estimates of Beta Solenoid Proteins

Amanda S. Parker, Krishnakumar M. Ravikumar, Daniel L. Cox

Department of Physics, University of California, Davis, California

July 15, 2017

1 Validation of the beam model

It is not obvious that simple beam theory can accurately describe the behavior of our beta solenoid proteins. As a check, we examined the relationship between spring constant and protein length. Based on the model, $L = Ck^{-1/3}$, where C is a constant ($C = (3(EI))^{1/3}$) and L is the length of the protein from the fixed end to the end at which the force is being applied. We tested this relationship by varying the pulling location along the length of the protein and determining the spring constant to be the slope of the resulting force vs. displacement data. In effect, we roughly estimated the spring constants for different lengths of the protein. We then plotted length L versus $k^{-1/3}$. An example is shown in Fig. 1. We saw generally linear behavior for all the proteins and directions tested, even at small lengths.

2 Packing fraction calculation

To calculate the packing fraction for each protein, we first determined whether each amino acid residue was inward-facing or outward-facing in its turn. We did this by first defining each turn as a path connecting the c-alpha atoms of the amino acid residues of that turn. For each residue in that turn, we calculated the average location of all its atoms. If that location lay inside the path defined by the backbone c-alphas, the residue was counted as being inward-facing. If instead it lay outside, it was counted as outward-facing. Once these were tabulated, the total volume of the inward-facing residues were calculated and divided by the total volume of all the residues in the protein. This ratio was taken as the packing fraction. The packing fraction value for each protein is listed in Table 1.

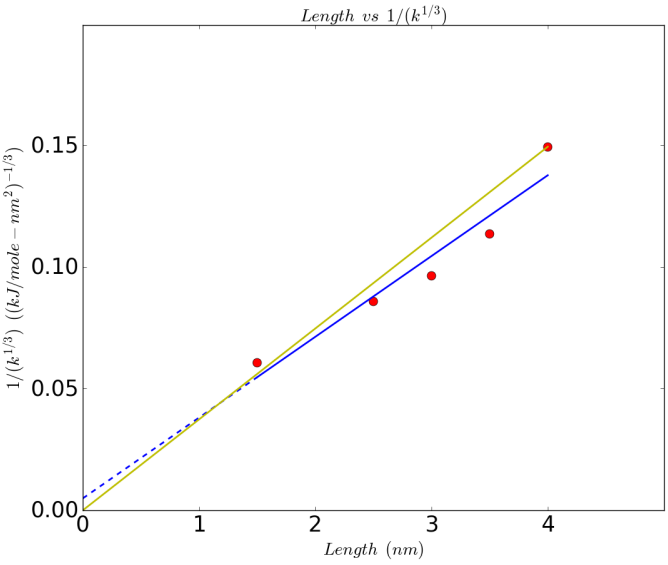

Figure 1: The relationship between one over the cube root of the spring constant and length for pulling of RiAFP in one of its bending directions is shown. The red dots are from simulation data. The yellow line is a fit between the original length of the protein and the origin. The blue line is a fit to the data points. The dashed blue line is the extrapolation of this fit to show the small, but non-zero, y intercept.

Protein	Packing fraction
SQBSP-m1	0.46
TRBSP-m1	0.46
SBAFP-m1	0.43
RiAFP	0.43
SBAFP	0.38
RGAFP-m1	0.35
TMAFP-m1	0.20

Table 1: The packing fraction (fraction of inward-facing amino acid reside volume to total amino acid residue volume) for each protein is shown in descending order.

Protein	Hydrogen bonds per turn
SBAFP-m1	6.2
SQBSP-m1	5.9
SBAFP	5.5
RiAFP	5.3
TRBSP-m1	4.9
RGAFP-m1	4.4
TMAFP-m1	2.8

Table 2: The average number of hydrogen bonds per turn in each protein are shown, in descending order. Each value was computed by taking the total hydrogen bonds in the protein (computed as the average over a 50 ps simulation) and dividing it by the number of total turns.

3 Values of hydrogen bonds per turn

In addition to the packing fraction, we investigated the relationship between numbers of hydrogen bonds per turn and persistence length. We first used the Hydrogen Bonds extension in VMD [1] to determine the total number of hydrogen bonds in the protein at each frame of a 50 ps simulation trajectory. The donor-acceptor distance used was 3.0 Angstroms and the angle cutoff was 20 degrees. Then we took the mean number of total hydrogen bonds over the trajectory and divided it by the number of turns in the protein.

The values for average numbers of hydrogen bonds per turn for each protein are listed in Table 2.

4 Beta content

We also checked whether the beta content of each protein (beta sheets and beta bridges) correlated with persistence length. To do this, we used DSSP [2] to calculate the amount of beta content in each protein. By this we mean the fraction of residues involved in beta strand or beta bridge secondary structures to the total number of residues. These values can be seen in Table 3 and the relationship between beta content and persistence length can be seen in Fig. 2 and Fig. 3.

From the plots, we can see that higher persistence length does not positively or negatively correlate with higher beta content.

5 Cross-sectional area and density calculations

The cross-sectional areas and densities of the proteins were calculated using a grid-sampling approach. Each protein was aligned so that its helical axis was along the x-axis. Each atom of each turn was represented as a sphere defined by its van der Waals radius. Each turn was then projected onto the y-z plane, so that it

Protein	Beta content
RiAFP	0.79
SBAFP-m1	0.58
SBAFP	0.54
RGAFP-m1	0.43
TRBSP-m1	0.41
TMAFP-m1	0.25
SQBSP-m1	0.19

Table 3:

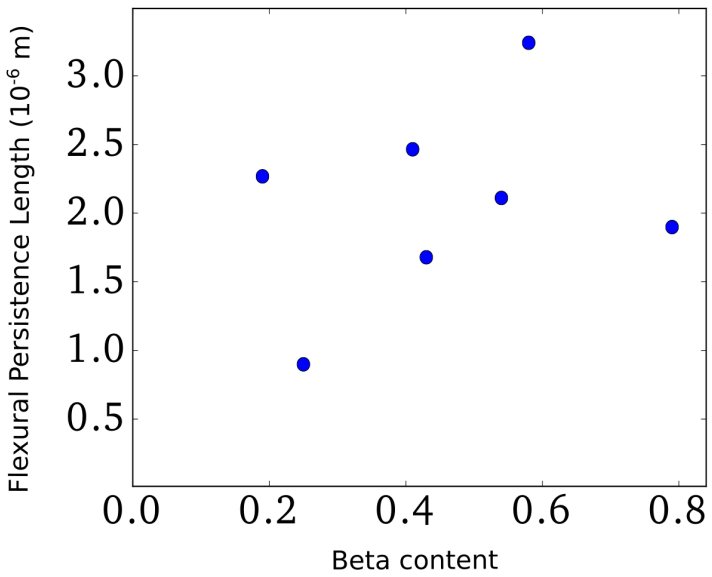

Figure 2: Flexural persistence length versus percent beta secondary structure.

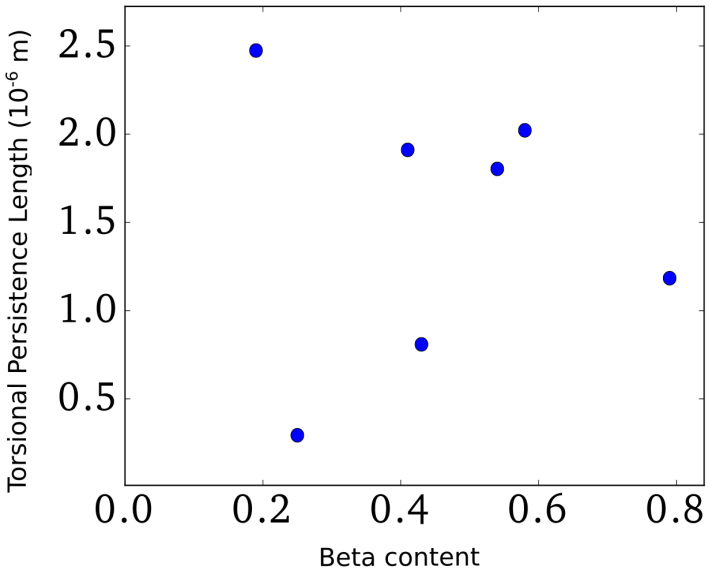

Figure 3: Torsional persistence length versus percent beta secondary structure.

was now represented by a collection of overlapping circles. The turn was bounded by a box and this box was then divided into a grid with spacing of approximately 0.03 nm in each direction. This grid was sampled and points lying inside circles were counted as part of the turn's area. The total area of the turn of the protein was determined as the ratio of points lying inside circles to total grid points, multiplied by the total area of the box. To get a measure of the overall cross-sectional area of a protein, the areas of each turn of the protein were averaged.

The density was determined by giving a weight to each grid point lying inside a circle, based on the atom's weight and radius. This weight was essentially an area density, since it was calculated as the atom's mass divided by the area of the projected van der Waals sphere representing the atom. Points that lay inside multiple circles were weighted as the sum of the weights of each circle.

Examples of both the full cross-sectional area and density map are shown in Fig. 4 for one turn of SBAFP-m1.

6 Area moment of inertia calculations

The area moments of inertia were estimated by approximating the proteins' cross-sections as regular geometries defined roughly by their backbones. The lengths of the sides of the cross-sectional shapes were estimated using bond lengths in VMD [1] and the area moments of inertia were found using these side lengths. The results are listed in table 4. The polar moment of inertia can by found by simply adding the two area moments of inertia for the two bending directions.

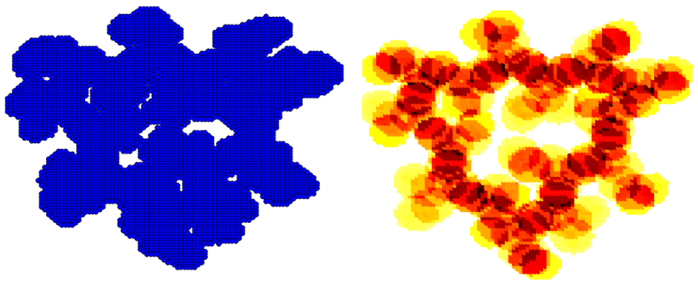

Figure 4: The total area (left) and density map (right) of the eighth turn of SBAFP-m1 is shown. Both were created using Python and Matplotlib by implementing a grid-sampling technique. In the case of the density map, the grid points were weighted by the area densities of the van der Waals circles representing the atoms in the turn.

Protein	Area moment of inertia, I_x	Area moment of inertia, I_y
SBAFP-m1	1.5	1.5
TRBSP-m1	1.8	1.8
RiAFP	0.1	1.6
SQBSP-m1	1.3	1.3
SBAFP	0.9	0.9
RGAFP-m1	0.2	0.7
TMAFP	0.1	0.2

Table 4: Area moments of inertia for each protein, in each bending direction.

7 Protein Sequences

The sequences for the seven proteins studied are listed in tables 5, 6, 7, 8, 9, 10, and 11.

ASP	LEU	SER	ILE	VAL	ASP	LEU	ARG	GLY	ALA
VAL	LEU	GLU	ASN	ILE	ASN	LEU	SER	GLY	ALA
ILE	LEU	HIS	GLY	ALA	MET	LEU	ASP	GLU	ALA
ASN	LEU	GLN	GLN	ALA	ASN	LEU	SER	ARG	ALA
ASP	LEU	SER	GLY	ALA	THR	LEU	ASN	GLY	ALA
ASP	LEU	ARG	GLY	ALA	ASN	LEU	SER	LYS	ALA
ASP	LEU	SER	ASP	ALA	ILE	LEU	ASP	ASN	ALA
ILE	LEU	GLU	GLY	ALA	ILE	LEU	ASP	GLU	ALA
VAL	LEU	ASN	GLN	ALA	ASN	LEU	LYS	ALA	ALA
ASN	LEU	GLU	GLN	ALA	ILE	LEU	SER	HIS	ALA
ASN	LEU	ARG	GLU	ALA	ASP	LEU	SER	GLU	ALA
ASN	LEU	GLU	ALA	ALA	ASP	LEU	SER	GLY	ALA
ASP	LEU	ALA	ILE	ALA	ASP	LEU	HIS	GLN	ALA
ASN	LEU	HIS	GLN	ALA	ALA	LEU	GLU	ARG	ALA

Table 5: Sequence of SQBSP-m1.

ALA	SER	ARG	ILE	THR	ASN	SER	GLN	ILE	VAL
LYS	SER	GLU	ALA	THR	ASN	SER	ASP	ILE	ASN
ASN	SER	GLN	LEU	VAL	ASP	SER	ILE	SER	THR
ARG	SER	GLN	TYR	SER	ASP	ALA	ASN	VAL	LYS
LYS	SER	VAL	THR	THR	ASP	SER	ASN	ILE	ASP
LYS	SER	GLN	VAL	TYR	LEU	THR	THR	SER	THR
GLY	SER	GLN	TYR	ASN	GLY	ILE	TYR	ILE	ARG
SER	SER	ASP	THR	THR	GLY	SER	GLU	ILE	SER
GLY	SER	SER	ILE	SER	THR	SER	ARG	ILE	THR
ASN	SER	ARG	ILE	THR	ASN	SER	GLN	ILE	VAL
LYS	SER	GLU	ALA	THR	ASN	SER	ASP	ILE	ASN
ASN	SER	GLN	LEU	VAL	ASP	SER	ILE	SER	THR
ARG	SER	GLN	TYR	SER	ASP	ALA	ASN	VAL	LYS
LYS	SER	VAL	THR	THR	ASP	SER	ASN	ILE	ASP
LYS	SER	GLN	VAL	TYR	LEU	THR	THR	SER	THR
GLY	SER	GLN	TYR	ASN	GLY	ILE	TYR	ILE	ARG
SER	SER	ASP	THR	THR	GLY	SER	GLU	ILE	SER
GLY	SER	SER	ILE	SER	THR	SER	ARG	ILE	THR

Table 6: Sequence of TRBSP-m1.

ASN	ASP	ILE	ASP	GLY	THR	ASN	ASN	GLU	VAL
ASP	GLY	SER	GLU	ASN	VAL	LEU	ALA	GLY	ASN
ASP	ASN	THR	VAL	SER	GLY	ASP	ASN	ASN	SER
VAL	SER	GLY	SER	ASN	ASN	THR	VAL	SER	GLY
ASN	ASP	ASN	THR	VAL	THR	GLY	SER	ASN	HIS
VAL	VAL	SER	GLY	THR	ASN	HIS	ILE	VAL	THR
ASP	ASN	ASN	ASN	ASN	VAL	SER	GLY	ASN	ASP
ASN	ASN	VAL	SER	GLY	SER	PHE	HIS	THR	VAL
SER	GLY	GLY	HIS	ASN	THR	VAL	SER	GLY	SER
ASN	ASN	THR	VAL	SER	GLY	LYS	ARG	HIS	ARG
VAL	GLN	GLY	THR	ASN	ASN	ARG	VAL	THR	ASP

Table 7: Sequence of RGAFP-m1.

GLY	VAL	GLU	ILE	GLY	GLU	GLY	THR	VAL	LEU
LYS	SER	GLY	VAL	VAL	VAL	ASN	GLY	GLY	THR
LYS	ILE	GLY	ARG	ASP	ASN	GLU	ILE	TYR	GLN
GLY	ALA	SER	ILE	GLY	GLY	GLY	VAL	GLU	ILE
GLY	ASP	ARG	ASN	ARG	ILE	ARG	GLU	SER	VAL
THR	ILE	GLY	GLY	GLY	GLY	VAL	VAL	GLY	SER
ASP	ASN	LEU	LEU	MET	ILE	ASN	ALA	GLY	ILE
ALA	GLY	ASP	CYS	THR	VAL	GLY	ASN	ARG	CYS
ILE	LEU	ALA	ASN	ASN	ALA	THR	LEU	ALA	GLY
GLY	VAL	GLU	ILE	GLY	GLU	GLY	THR	VAL	LEU
LYS	SER	GLY	VAL	VAL	VAL	ASN	GLY	GLY	THR
LYS	ILE	GLY	ARG	ASP	ASN	GLU	ILE	TYR	GLN
GLY	ALA	SER	ILE	GLY	GLY	GLY	VAL	GLU	ILE
GLY	ASP	ARG	ASN	ARG	ILE	ARG	GLU	SER	VAL
THR	ILE	GLY	GLY	GLY	GLY	VAL	VAL	GLY	SER
ASP	ASN	LEU	LEU	MET	ILE	ASN	ALA	GLY	ILE
ALA	GLY	ASP	CYS	THR	VAL	GLY	ASN	ARG	CYS
ILE	LEU	ALA	ASN	ASN	ALA	THR	LEU	ALA	GLY

Table 8: Sequence of SBAFP-m1.

ALA	SER	ARG	ALA	GLU	ALA	ARG	GLY	GLU	ALA
MET	ALA	GLU	GLY	HIS	SER	ARG	GLY	CYS	ALA
THR	SER	HIS	ALA	ASN	ALA	THR	GLY	HIS	ALA
ASP	ALA	ARG	SER	MET	SER	GLU	GLY	ASN	ALA
GLU	ALA	TYR	THR	GLU	ALA	LYS	GLY	THR	ALA
MET	ALA	THR	SER	GLU	ALA	SER	GLY	GLU	ALA
ARG	ALA	GLN	THR	ASN	ALA	ASP	GLY	ARG	ALA
HIS	SER	SER	SER	ARG	THR	HIS	GLY	ARG	ALA
ASP	SER	THR	ALA	SER	ALA	LYS	GLY	GLU	ALA
MET	ALA	GLU	GLY	THR	SER	ASP	GLY	ASP	ALA
LYS	SER	TYR	ALA	SER	ALA	ASP	GLY	ASN	ALA
CYS	ALA	LYS	SER	MET	SER	THR	GLY	HIS	ALA
ASP	ALA	THR	THR	ASN	ALA	HIS	GLY	THR	ALA
MET	ALA	ASP	SER	ASN	ALA	ILE	GLY	GLU	ALA
ARG	ALA	GLU	THR	ARG	ALA	GLU	GLY	ARG	ALA
GLU	SER	SER	SER	ASP	THR	ASP	GLY	CYS	

Table 9: Sequence of RiAFP.

GLY	ALA	CYS	THR	GLY	CYS	GLY	ASN	CYS	PRO	ASN	ALA
VAL	THR	CYS	THR	ASN	SER	GLN	HIS	CYS	VAL	LYS	ALA
ASN	THR	CYS	THR	GLY	SER	THR	ASP	CYS	ASN	THR	ALA
GLN	THR	CYS	THR	ASN	SER	LYS	ASP	CYS	PHE	GLU	ALA
ASN	THR	CYS	THR	ASP	SER	THR	ASN	CYS	TYR	LYS	ALA
THR	ALA	CYS	THR	ASN	SER	SER	GLY	CYS	PRO		

Table 10: Sequence of TMAFP.

GLY	THR	CYS	VAL	ASN	THR	ASN	SER	GLN	ILE	THR	ALA
ASN	SER	GLN	CYS	VAL	LYS	SER	THR	ALA	THR	ASN	CYS
TYR	ILE	ASP	ASN	SER	GLN	LEU	VAL	ASP	THR	SER	ILE
CYS	THR	ARG	SER	GLN	TYR	SER	ASP	ALA	ASN	VAL	LYS
LYS	SER	VAL	THR	THR	ASP	CYS	ASN	ILE	ASP	LYS	SER
GLN	VAL	TYR	LEU	THR	THR	CYS	THR	GLY	SER	GLN	TYR
ASN	GLY	ILE	TYR	ILE	ARG	SER	SER	THR	THR	THR	GLY
THR	SER	ILE	SER	GLY	PRO	GLY	CYS	SER	ILE	SER	THR
CYS	THR	ILE	THR								

Table 11: Sequence of SBAFP.

References

[1] W. Humphrey, A. Dalke, and K. Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[2] W. Kabsch and C. Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, 1983.