Supplementary Information: Stiffness of the C-terminal disordered linker affects the geometry of the active site in endoglucanase Cel8A

Bartosz Różycki and Marek Cieplak
Institute of Physics, Polish Academy of Sciences,
Al. Lotników 32/46, 02-668 Warsaw, Poland

Conformational ensemble of Cel8A-ScafT and Cel8A*-ScafT

Using the Kim-Hummer model [1], as described in the Methods section, we performed REMC simulations of both the wild-type Cel8A-ScafT complex and the quadruple mutant Cel8A*-ScafT mini-cellulosome. In these simulations we assumed that the rigid bodies representing the catalytic domain in Cel8A and Cel8A* were identical, and the four mutations had an influence only on the energy of the inter-domain contacts. In other words, our assumption here is that the structures of the backbone chains of Cel8A and Cel8A* are identical, and the four mutations affect only the chemical composition and orientations of the side chains. Such assumptions are often made in theoretical and computational studies on the effects of mutations on protein stability [2] but do not have to be correct in general.

To characterize the conformational variability of these complexes, we measured their characteristic dimensions as functions of the simulation progression. Specifically, for each of the simulated conformations we computed the radius of gyration

$$R_{\mathrm{g}} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(\vec{r}_i - \vec{r}_{\mathrm{cm}}\right)^2} \tag{1}$$

and the maximum extension

$$D_{\max} = \max_{(i,j)} \left|\vec{r}_i - \vec{r}_j\right| \tag{2}$$

Here, $\vec{r}_i$ and $\vec{r}_j$ are the coordinate vectors of α-C atoms i and j, respectively, M is the total number of amino acid residues, and $\vec{r}_{\mathrm{cm}} = \frac{1}{M}\sum_{j=1}^{M}\vec{r}_j$ is the vector that describes the location of the center of mass.

Panels A and B of Fig. S1 show distributions for $D_{\max}$ and R_{g}, respectively, as obtained in the simulations. The black and red lines correspond to Cel8A-ScafT and Cel8A*-ScafT, respectively. The data shown in Fig. S1A and B show no differences in $D_{\max}$ or R_{g} between the wild-type Cel8A-ScafT complex and the quadruple mutant Cel8A*-ScafT mini-cellulosome.

The Cel8A-ScafT complex contains two disordered linkers. One of the linkers is present in Cel8A and connects the catalytic domain (CD) with the dockerin domain (Doc). Its amino acid sequence is the following SGQPTPPSNPTPSLPPQ. The other linker is localized within ScafT and connects CBM with cohesin (Coh). Its sequence is TPPATTKPPATTKPPATTIPPSDDP. It is interesting to note that these two disordered linkers contain multiple proline and threonine residues.

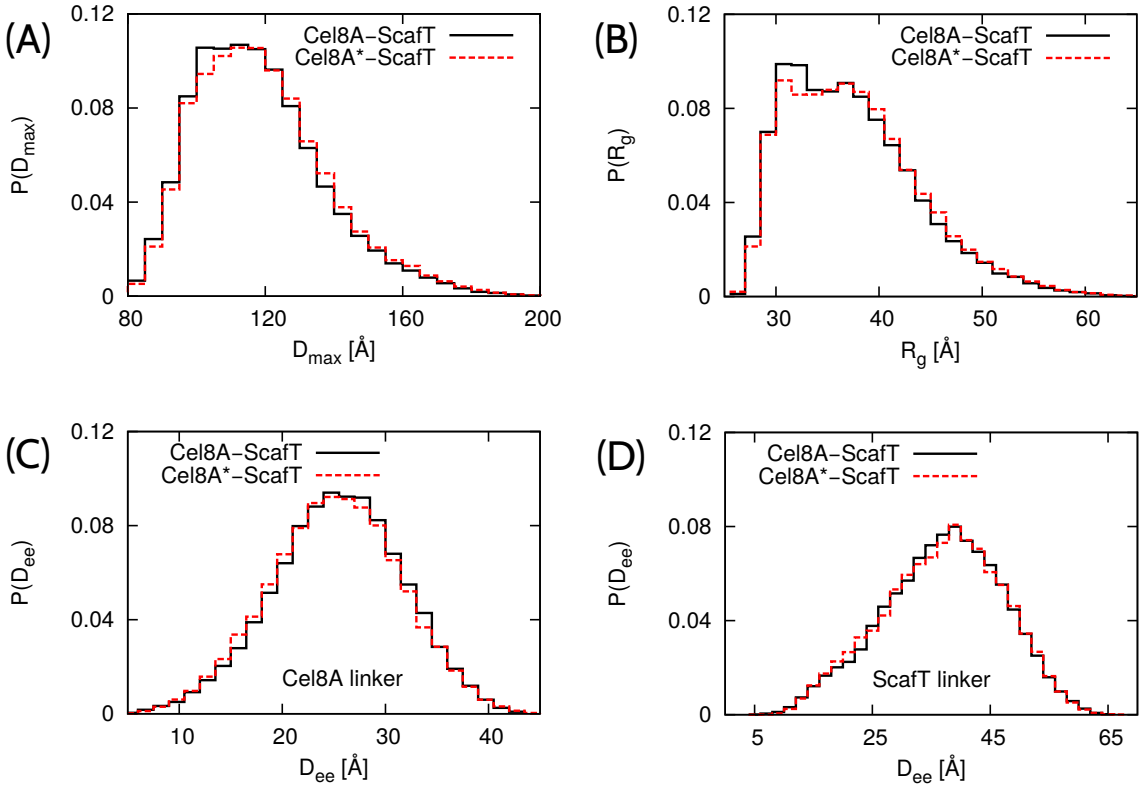

Figure S1: Results of simulations of the Cel8A-ScafT and Cel8A*-ScafT complexes at 300 K. The solid (black) and dashed (red) lines correspond to Cel8A-ScafT and Cel8A*-ScafT, respectively. (A) Distribution for the maximum extension $D_{\max}$. (B) Distribution for the radius of gyration R_g. (C) Distribution for the end-to-end distance D_{ee} of the Cel8A linker joining the catalytic domain with the dockerin domain. (D) Distribution for the end-to-end distance D_{ee} of the ScafT linker joining CBM with cohesin.

domain pair	Cel8A-ScafT	Cel8A*-ScafT
CD-Doc	0.38 ± 0.01	0.39 ± 0.03
CD-CBM	0.14 ± 0.02	0.12 ± 0.02
CD-Coh	0.51 ± 0.03	0.51 ± 0.04
Doc-CBM	0.13 ± 0.02	0.14 ± 0.02
CMB-Coh	0.43 ± 0.03	0.43 ± 0.04

Table SI: Probabilities of having at least one contact between individual domains in Cel8A-ScafT (left column) and Cel8A*-ScafT (right column) at 300 K.

An important quantity that characterizes the extension of the linkers is their end-to-end distance $D_{\rm ee}$. Using the simulated conformations, we calculated the end-to-end distance distributions, $P(D_{\rm ee})$, for the two flexible linkers. The results are shown in panels C and D of Fig. S1.

To characterize the spatial arrangements of the Cel8A-ScafT and Cel8A*-ScafT complexes, we analyzed their inter-domain contacts. We assume that two residues are in contact if their α-C atoms are separated by less than their interaction radius $\tilde{\sigma}_{ij}$. If there is at least one contact between residues in two separate domains, we take the two domains to be in contact. The number of recorded conformations in which two given domains, μ and ν, are found to be in contact, $n_{\mu,\nu}$, divided by the total number of

conformations recorded in the MC simulation trajectory, n_{tot}, yields an estimate for the probability of finding the two domains in contact. Table SI shows the probabilities of having a contact between the different domains. For example, in the wild-type Cel8A-ScafT complex, CD was found to be in contact with Coh and Doc, respectively, in about 51% and 38% of all the conformations generated in the MC simulations. The statistical errors on the contact probabilities were estimated on the basis of five independent simulation runs. Importantly, we could not find any systematic differences neither in the inter-domain contacts as listed in Tab. SI nor in the extensions of the flexible linkers as shown in panels C and D of Fig. S1. The conformational ensembles of Cel8A-ScafT and Cel8A*-ScafT were found to be practically indistinguishable in these simulations.

References

[1] Y.-C. Kim and G. Hummer, *Coarse-grained models for simulations of multiprotein complexes: Application to ubiquitin binding*, J. Mol. Biol., 2008, **375**, 1416-1433.

[2] J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau and L. Serrano, *The FoldX web server: an online force field*. Nucl. Acids Res., 2005, **33**, W382-388.