Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Table S1. List of the simulated systems with the lengths (in ns) of simulations. To reproduce physiological salt conditions, the simulations were carried out with salt concentrations of 137 mM NaCl and 10 mM $CaCl_2$ with counter ions introduced to neutralize the total charge of each simulation cell.

System name	Description	Lengths (ns)
ct_I_300	*Clostridium thermocellum* binding mode I at 300 K Cohesin PDB: 1anu resolution: 2.15 Å Dockerin PDB: 2mte resolution: NA (NMR)	500
ct_I_325, ct_I_350, ct_I_400	*Clostridium thermocellum* binding mode I at 325 K, 350 K and 400 K	500
ct_II_300, ct_II_325, ct_II_350, ct_II_400	*Clostridium thermocellum* binding mode II at 300 K, 325 K, 350 K and 400 K	500
cc_I_300	*Clostridium cellulolyticum* binding mode I at 300K (Residues 16 and 17 were mutated to Ala and Leu respectively) PDB: 2vn6 resolution: 1.9 Å	500
cc_I_325, cc_I_350, cc_I_400	*Clostridium cellulolyticum* binding mode I at 325K, 350K and 400K	500
cc_II_300, cc_II_325, cc_II_350, cc_II_400	*Clostridium cellulolyticum* binding mode II at 300 K, 325 K, 350 K and 400 K	500
ac_I_300	*Acetivibrio cellulolyticus* binding mode I simulated at 300 K (Residues 15 and 16 were mutated to Ile and Asn respectively) PDB: 4uyp resolution: 2.8 Å	500
ac_I_325, ac_I_350, ac_I_400	*Acetivibrio cellulolyticus* binding mode I simulated at 325 K, 350 K and 400 K	500
ac_II_300, ac_II_325, ac_II_350, ac_II_400	*Acetivibrio cellulolyticus* binding mode II simulated at 300 K, 325 K, 350 K and 400 K	500

Figure S1. Secondary structure changes throughout simulations of wildtype **ct** in binding mode I: A, B, C, D – cohesin; E, F, G, H – dockerin at 300 K, 325 K, 350 K and 400 K, respectively.

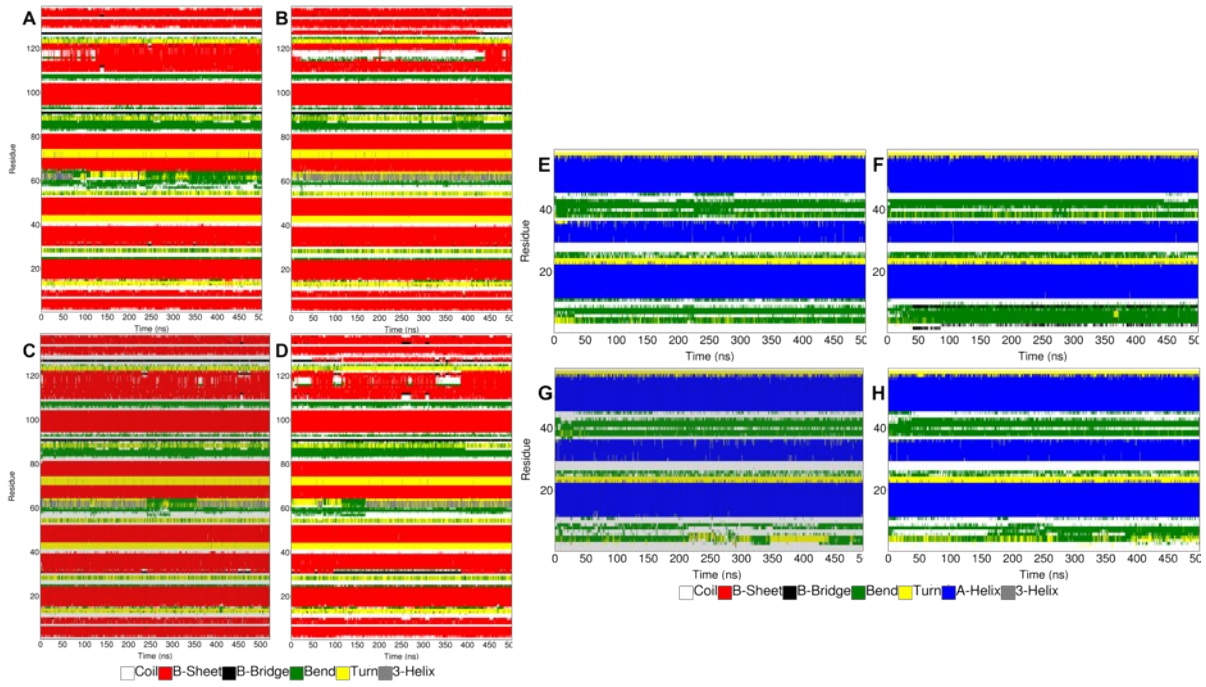


Figure S2. Secondary structure changes throughout simulations of wildtype **ct** in binding mode II: A, B, C, D – cohesin; E, F, G, H – dockerin at 300 K, 325 K, 35 0K and 400 K, respectively.

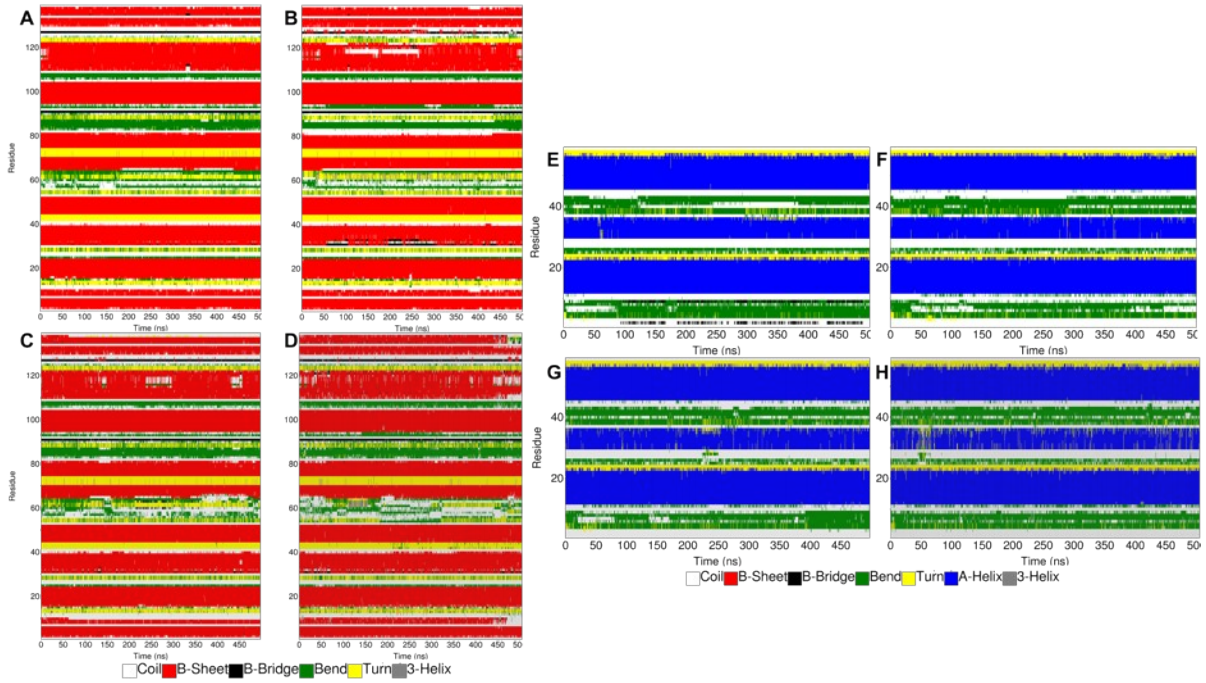


Figure S3. Secondary structure changes throughout simulations of wildtype **ac** in binding mode I: A, B, C, D – cohesin; E, F, G, H – dockerin at 300 K, 325 K, 350 K and 400 K, respectively.

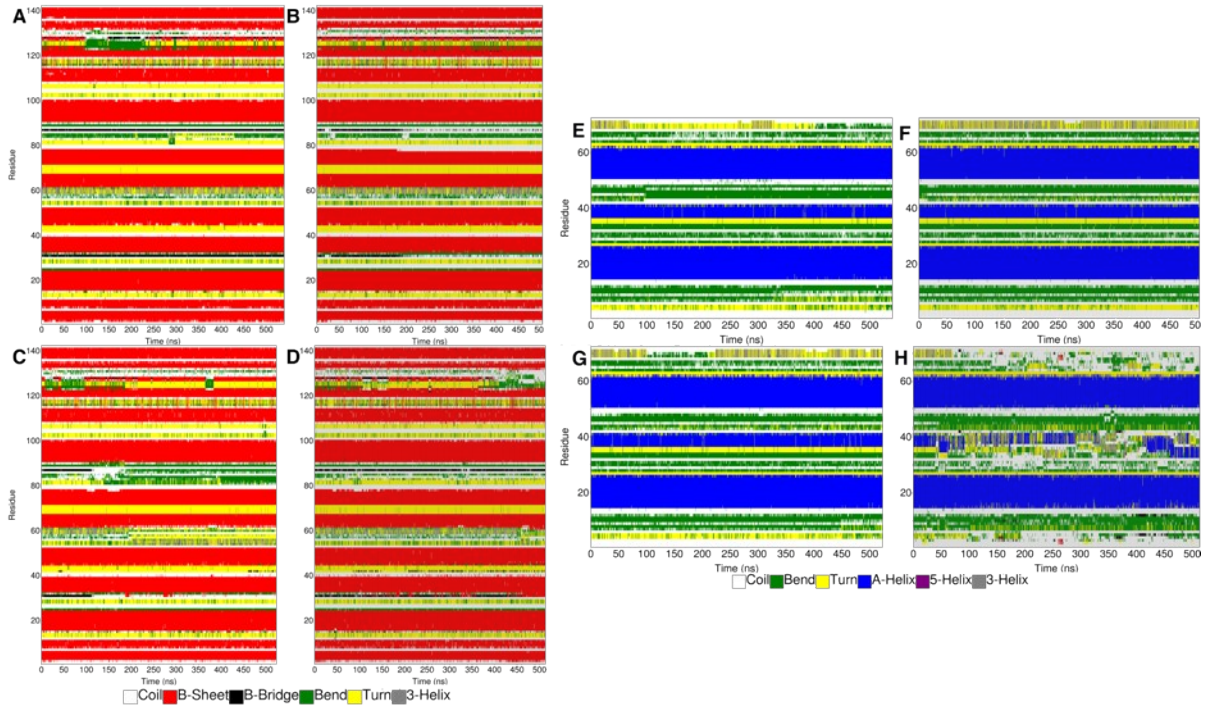


Figure S4. Secondary structure changes throughout simulations of wildtype **ac** in binding mode II: A, B, C, D – cohesin; E, F, G, H – dockerin at 300 K, 325 K, 350 K and 400 K, respectively.

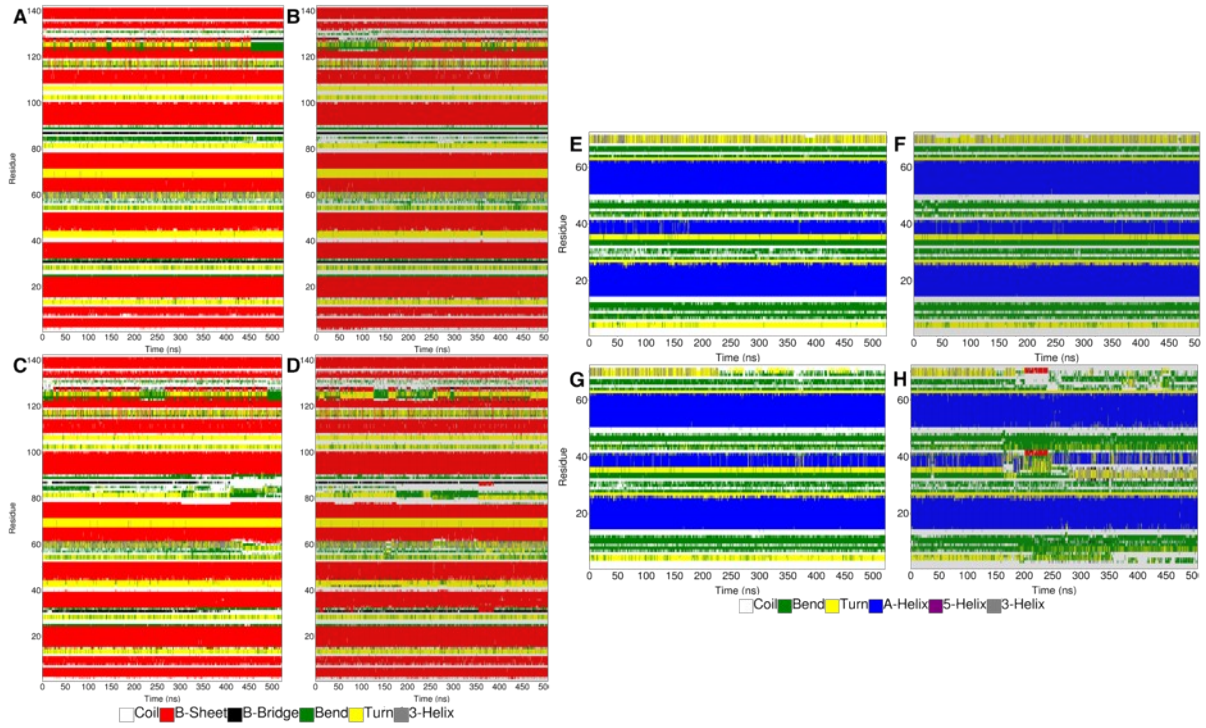


Figure S5. Secondary structure changes throughout simulations of wildtype **cc** in binding mode I: A, B, C, D – cohesin; E, F, G, H – dockerin at 300 K, 325 K, 350 K and 400 K, respectively.

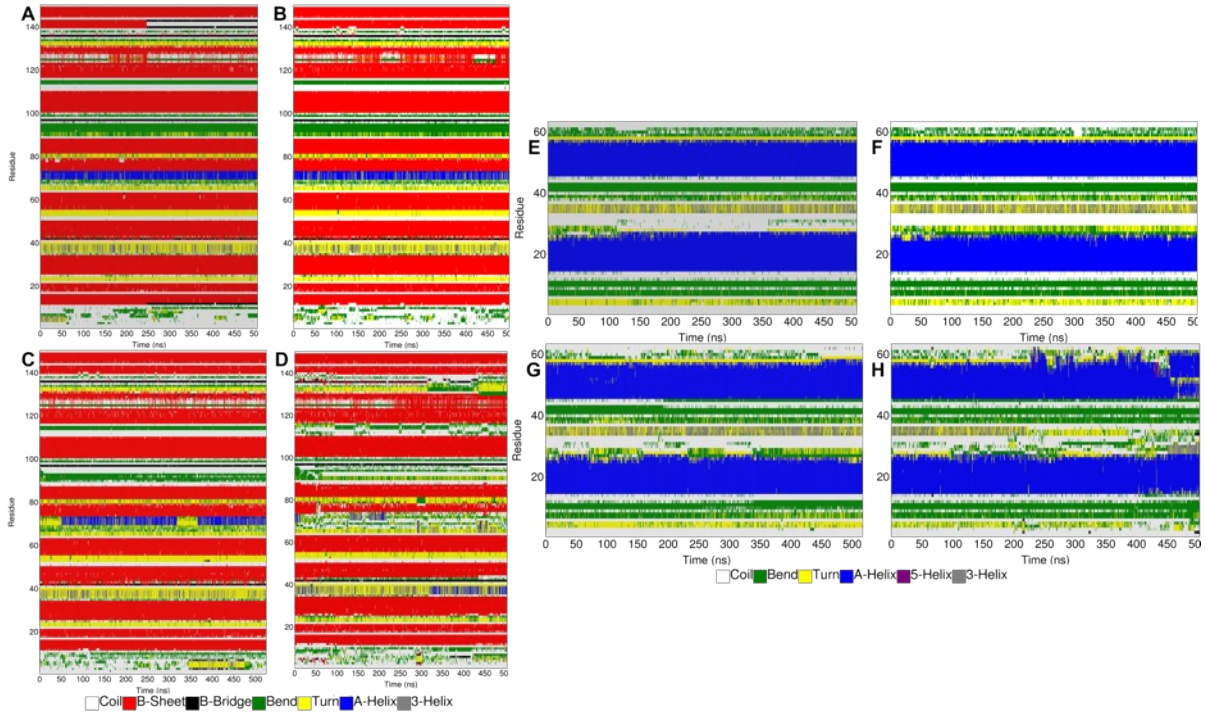


Figure S6. Secondary structure changes throughout simulations of wildtype **cc** in binding mode II: A, B, C, D – cohesin; E, F, G, H – dockerin at 300 K, 325 K, 350 K and 400 K, respectively.

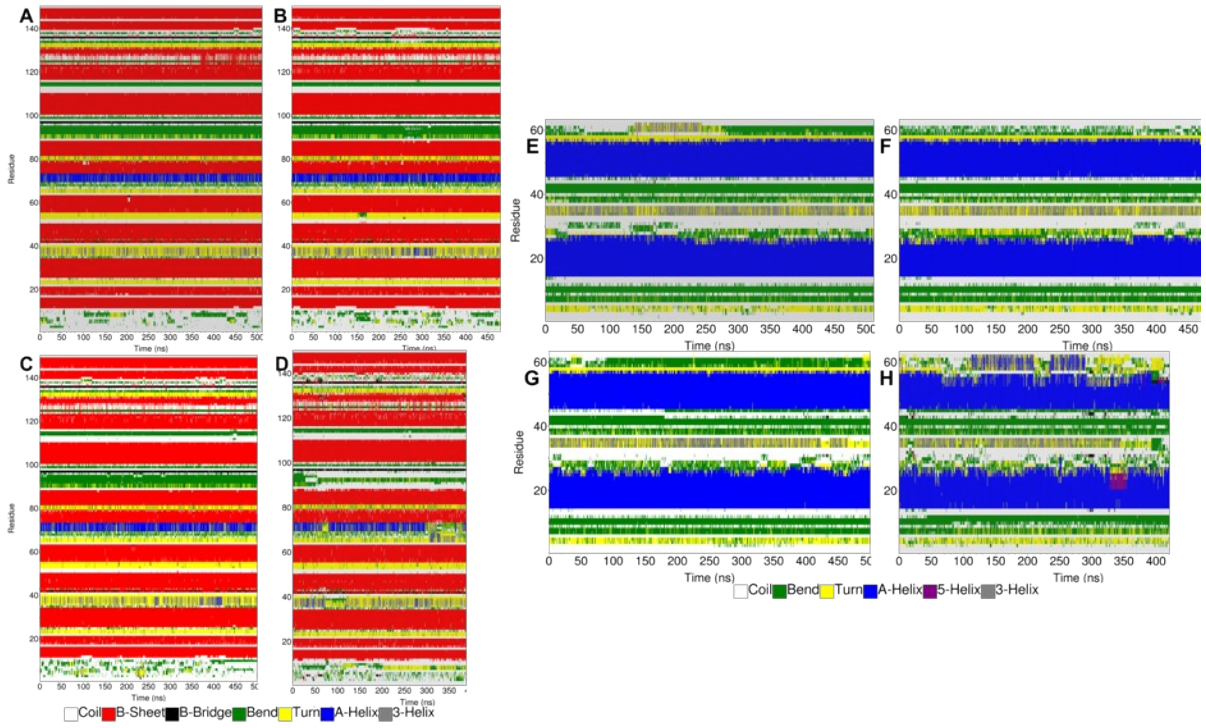


Figure S7. Root mean square deviation of protein backbone in simulations of wildtype **ct**: A, B – binding mode I (cohesin and dockerin respectively); C, D – binding mode II (cohesin and dockerin, respectively).

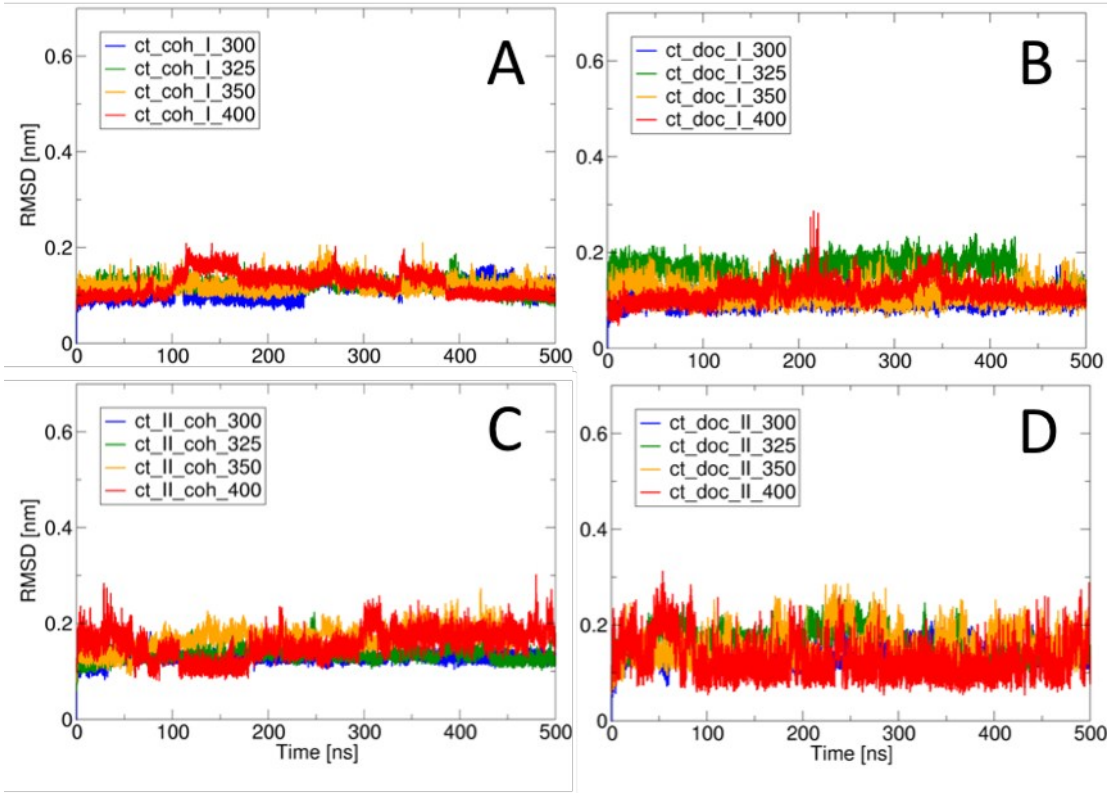


Figure S8. Root mean square deviation of protein backbone in simulations of wildtype **ac**: A, B – binding mode I (cohesin and dockerin respectively); C, D – binding mode II (cohesin and dockerin, respectively).

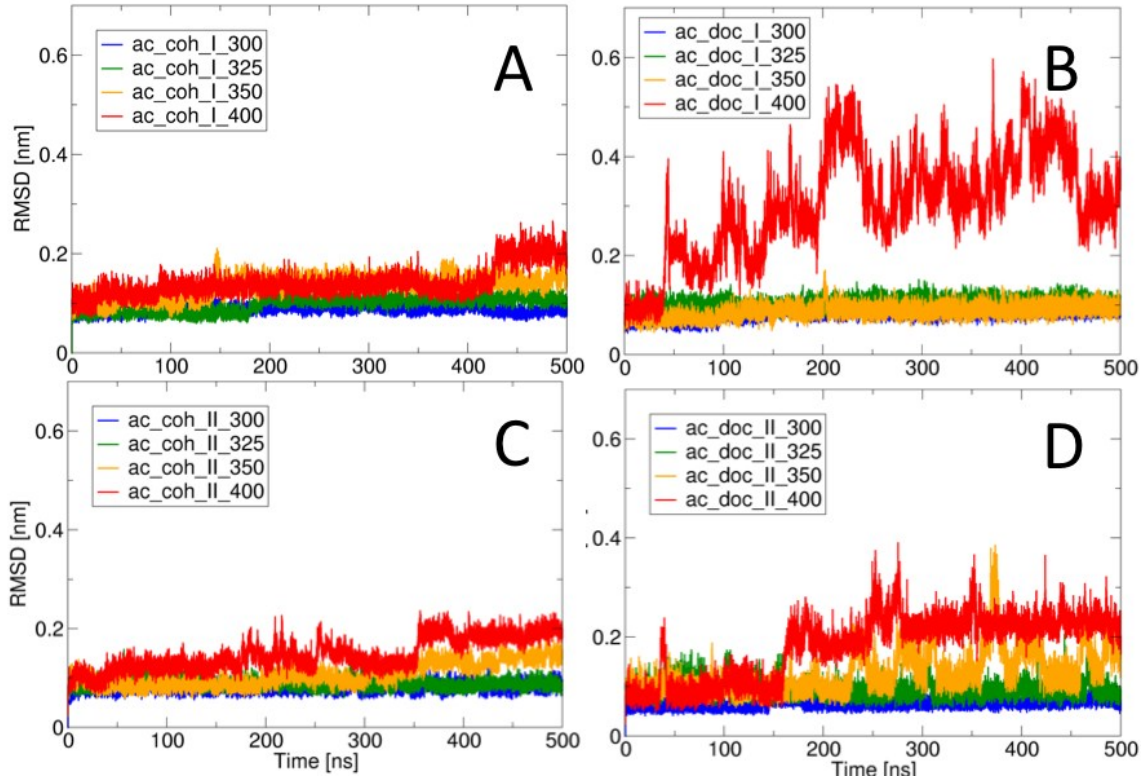


Figure S9. Root mean square deviation of protein backbone in simulations of wildtype **cc**: A, B – binding mode I (cohesin and dockerin, respectively); C, D – binding mode II (cohesin and dockerin, respectively).

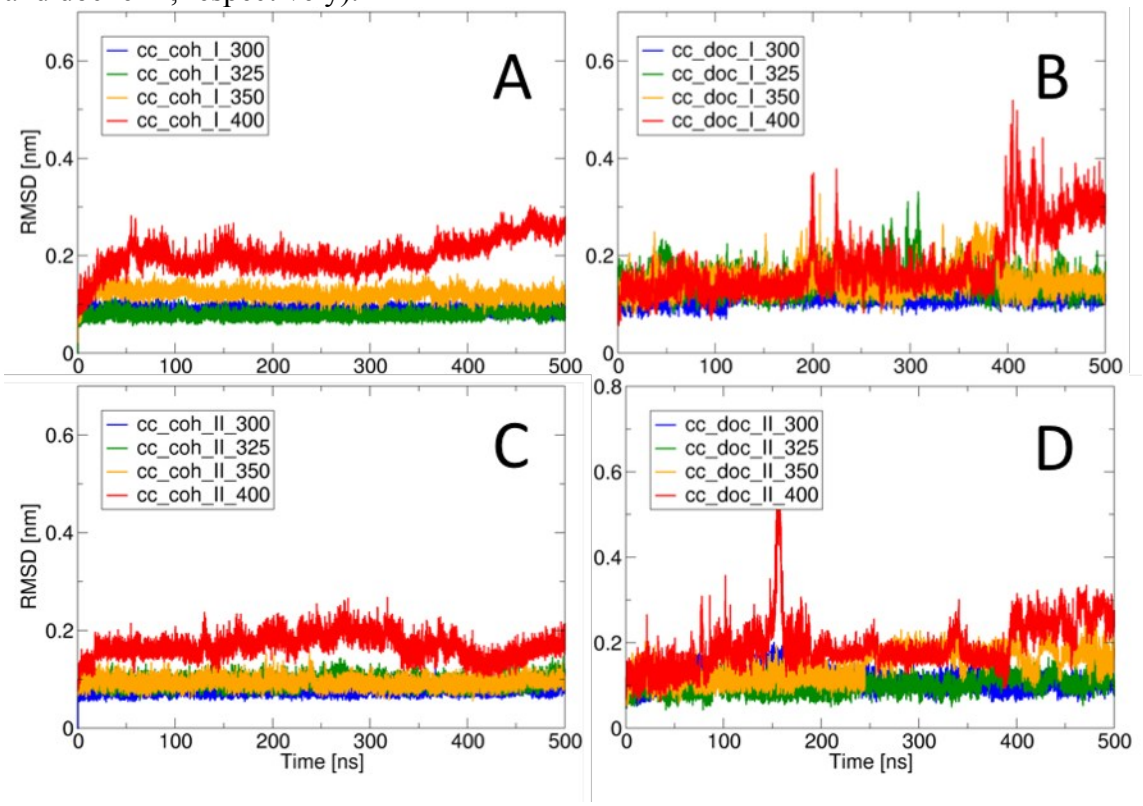


Figure S10. The average solvent accessible surface area (SASA) of each protein was computed and presented in the graphs below: A, B – ct cohesin and ct dockerin, respectively.

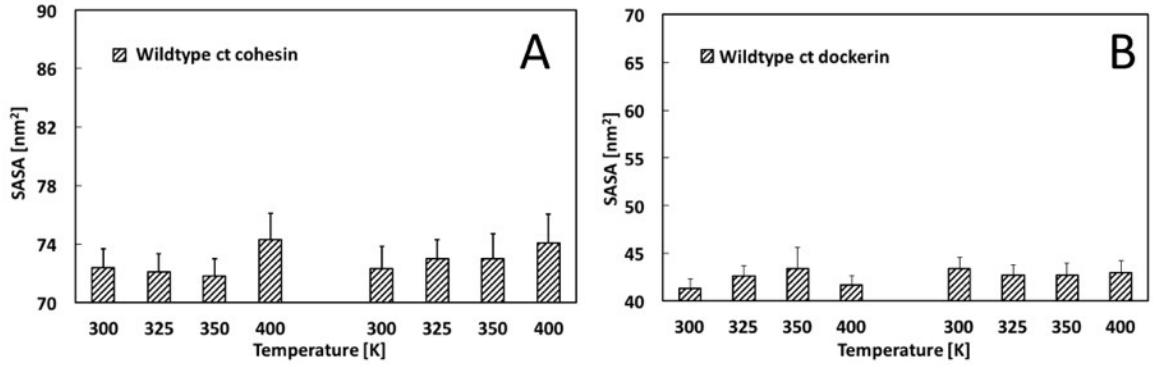


Figure S11. Comparing the average solvent accessible surface area (SASA) of each protein in wildtype and mutated **ac**: A, B – cohesin and dockerin, respectively.

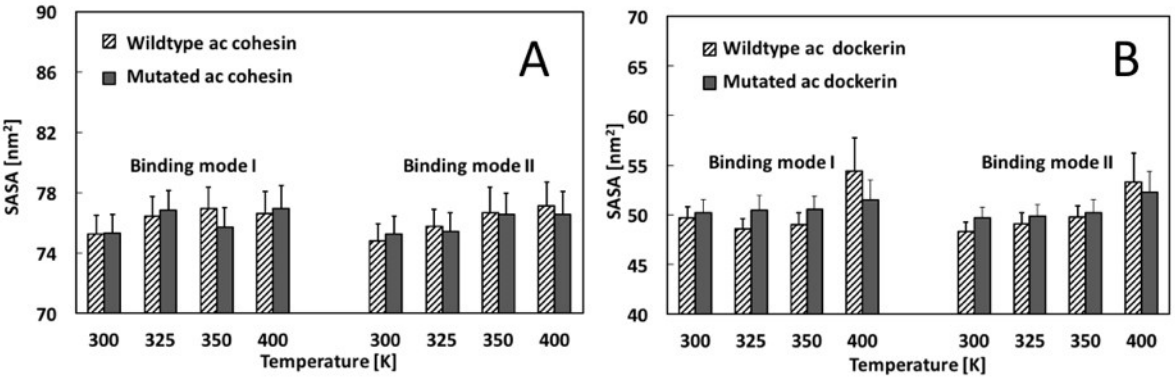


Figure S12. Comparing the average solvent accessible surface area (SASA) of each protein in wildtype and mutated **cc**: A, B – cohesin and dockerin, respectively.

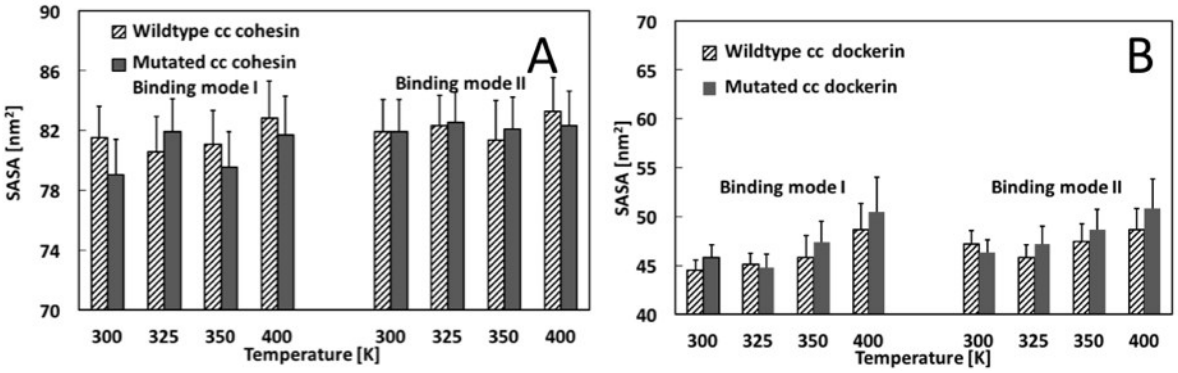


Figure S13. Root mean square fluctuations of C-alpha atoms of individual protein in the simulations at 300 K, 325 K, 350 K and 400 K: A, B represent **ct** in binding mode I (cohesin and dockerin, respectively) while C, D show RMSF in **ct** binding mode II (cohesin and dockerin, respectively).

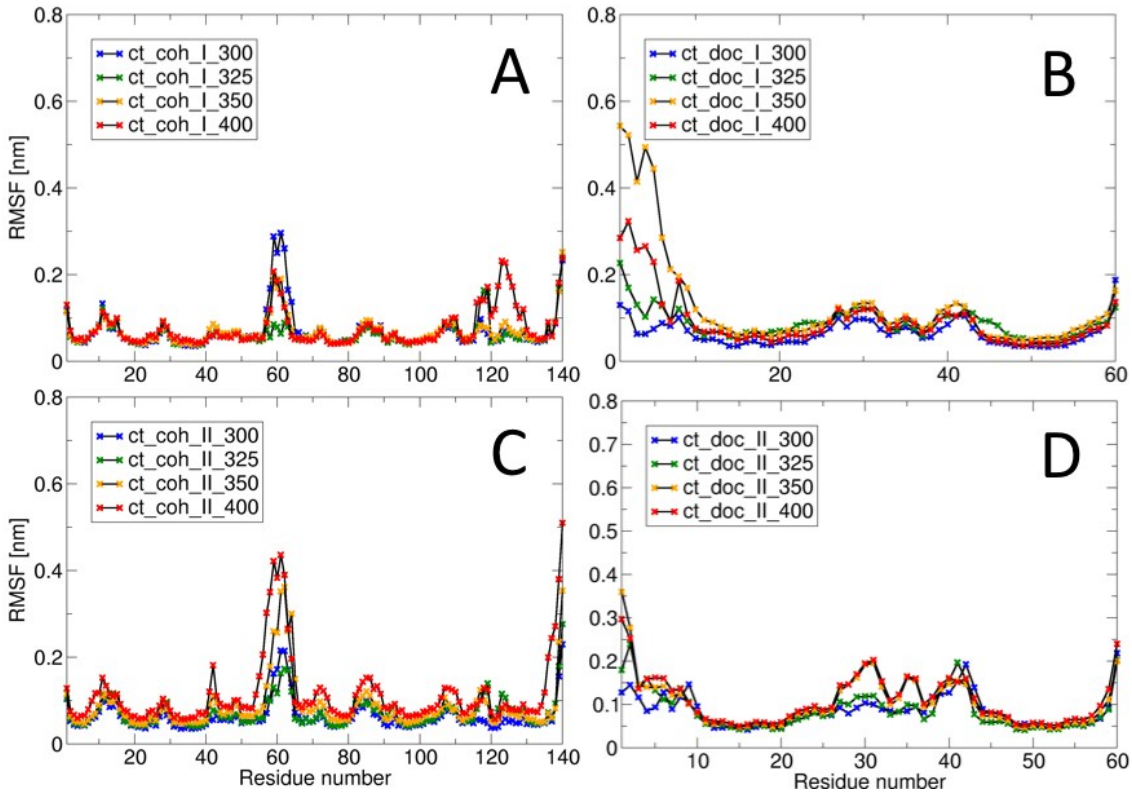


Figure S14. Root mean square fluctuations of C-alpha atoms of individual protein in the simulations at 300 K, 325 K, 350 K and 400 K: A, B represent wildtype **ac** in binding mode I (cohesin and dockerin, respectively) while C, D show RMSF in wildtype **ac** binding mode II (cohesin and dockerin, respectively).

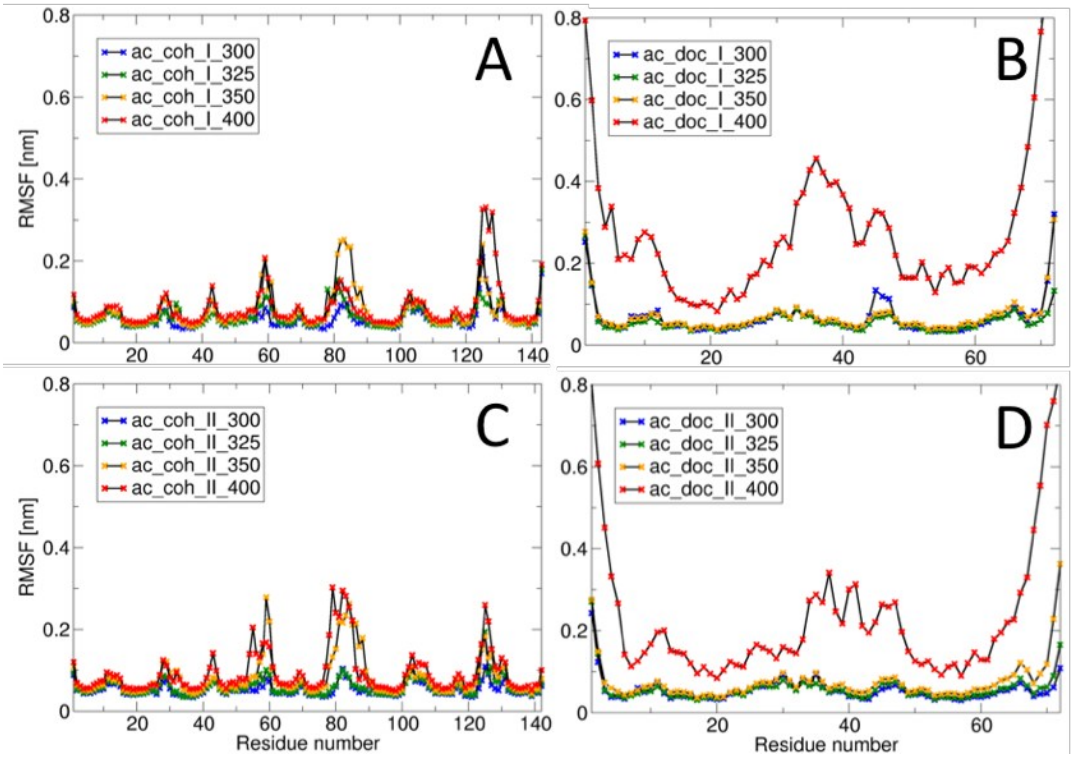


Figure S15. Root mean square fluctuations of C-alpha atoms of individual protein in the simulations at 300 K, 325 K, 350 K and 400 K: A, B represent wildtype **cc** in binding mode I (cohesin and dockerin, respectively) while C, D show RMSF in wildtype **cc** binding mode II (cohesin and dockerin, respectively).

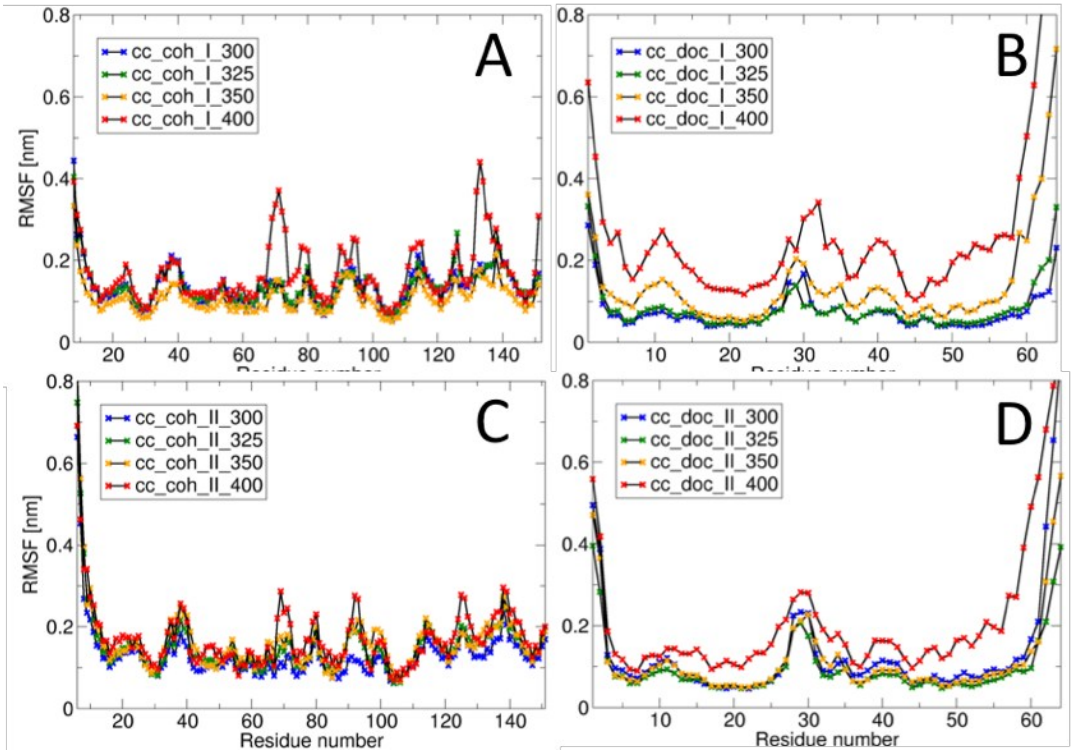


Figure S16. Protein sequences

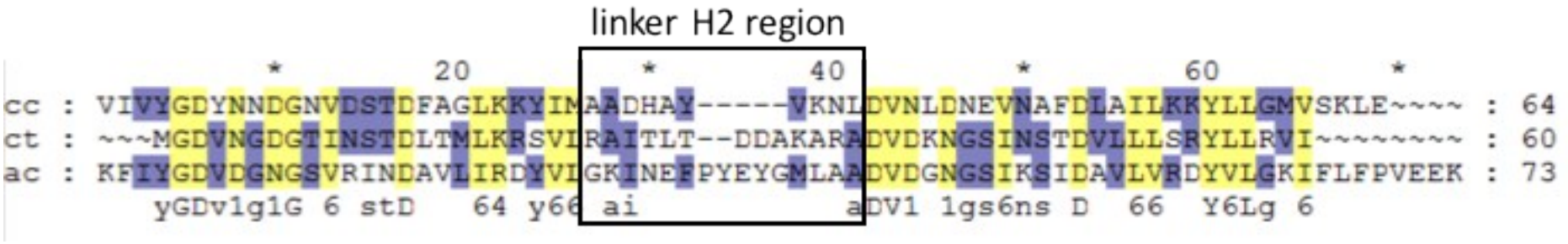


Figure S17. Number of hydrogen bonds with time throughout the simulations of wildtype **ct**: A, B, C, D – binding mode I; E, F, G, H – binding mode II at 300 K, 325 K, 350 K and 400 K, respectively. The hydrogen bond length and angle cutoffs used in this analysis, were 3.5 Å and 30°, respectively.

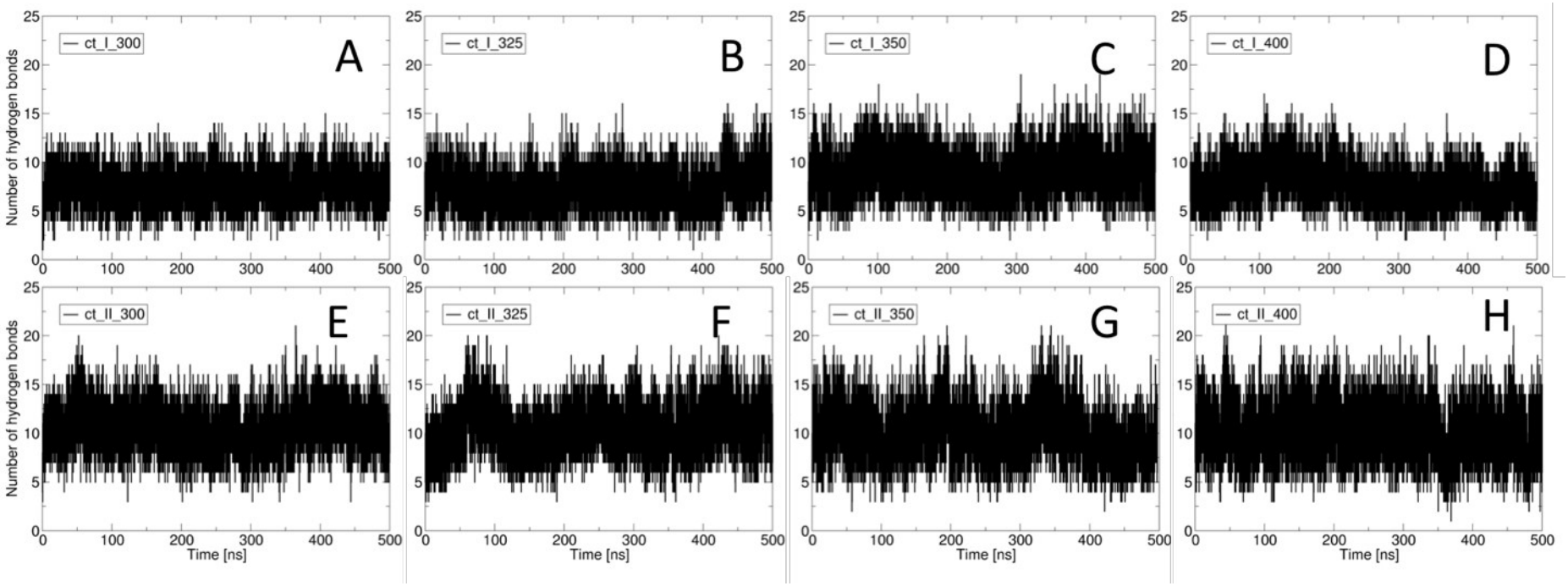


Figure S18. Number of hydrogen bonds with time throughout the simulations of wildtype **ac** A, B, C, D – binding mode I; E, F, G, H – binding mode II at 300 K, 325 K, 350 K and 400 K respectively. The hydrogen bond length and angle cutoffs used in this analysis, were 3.5 Å and 30°, respectively.

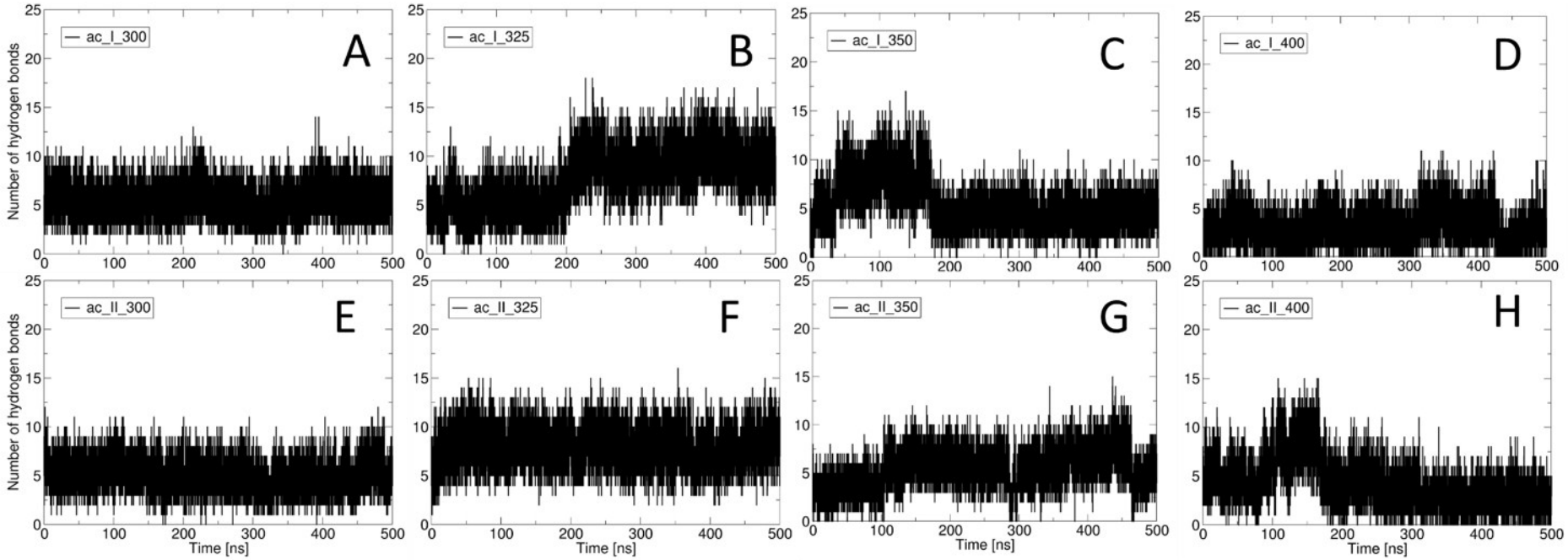


Figure S19. Number of hydrogen bonds with time throughout the simulations of wildtype **cc**: A, B, C, D – binding mode I; E, F, G, H – binding mode II at 300 K, 325 K, 350 K and 400 K respectively. The hydrogen bond length and angle cutoffs used in this analysis, were 3.5 Å and 30°, respectively.

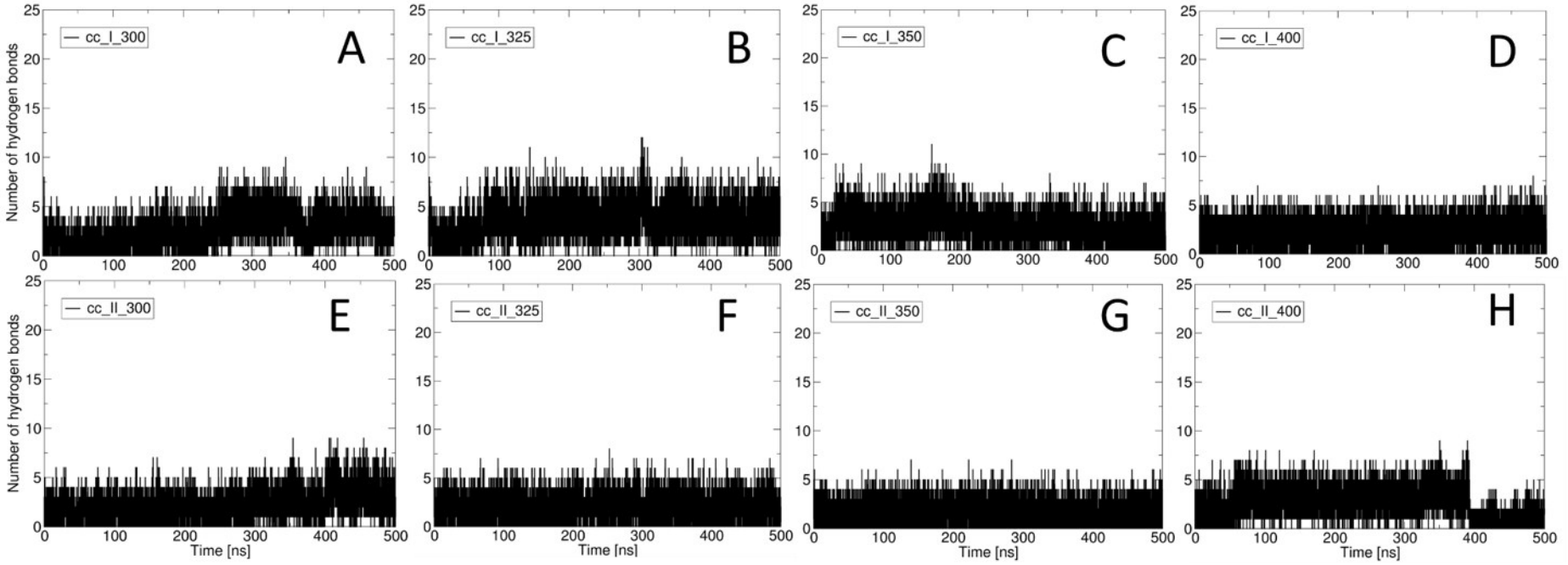


Figure S20. The average number of hydrogen bonds between **ct** cohesin-dockerin pairs was

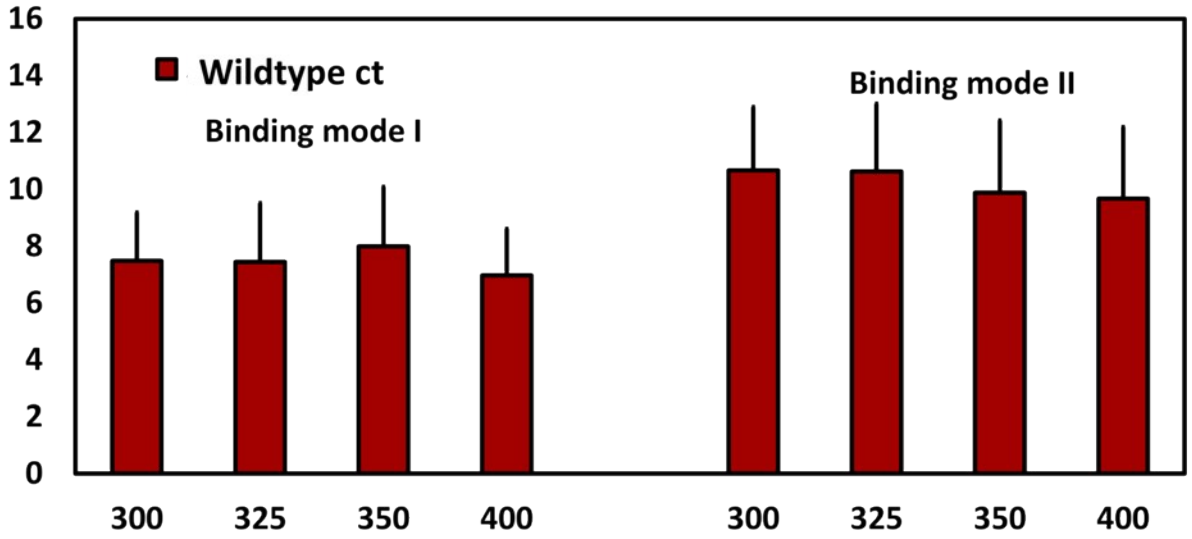


Figure S21. Comparing the average number of hydrogen bonds between coh-doc pair of each wildtype and mutated **ac**, at four temperatures and in binding modes I and II.

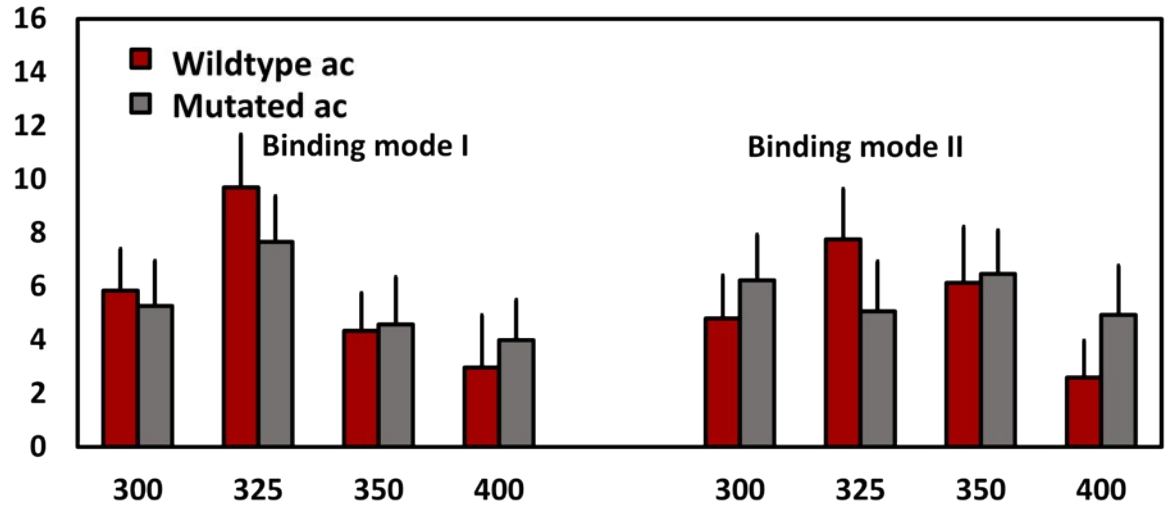


Figure S22. Comparing the average number of hydrogen bonds between coh-doc pair of each wildtype and mutated **cc**, at four temperatures and in binding modes I and II.

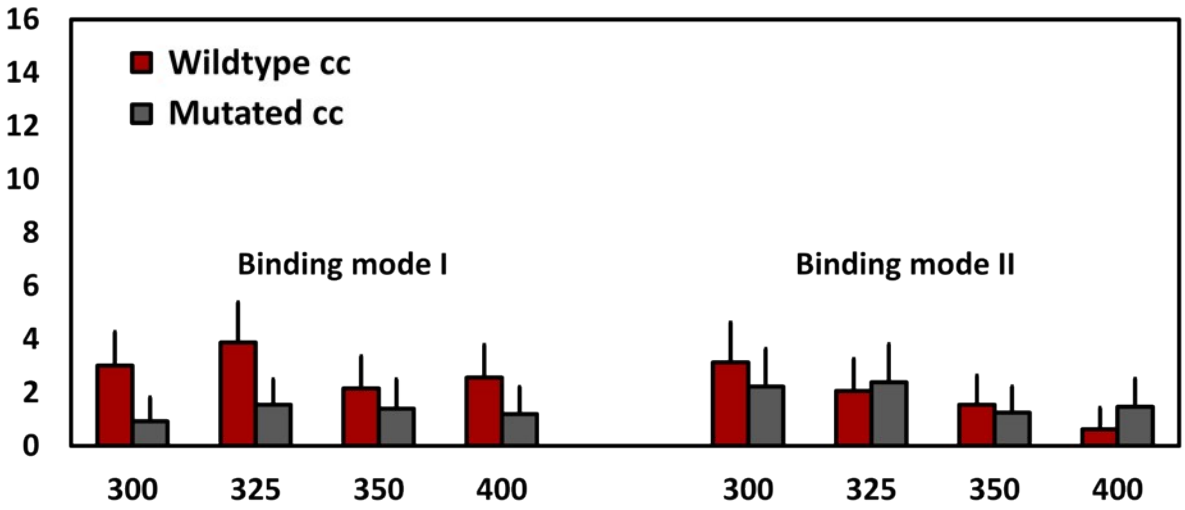


Table S2. List of the main hydrogen bonds involved in cohesin and dockerin binding and their percentage occurrence in the last 200 ns of wildtype **ct** in binding mode I.

Cohesin	Dockerin	Percentage occurrence of H-bonds in the last 200ns of trajectory			
		ct_I_300	***ct_I_325***	***ct_I_350***	***ct_I_400***
ASP35	SER46-Main	97%	73%	96%	94%
TYR70	LEU23	94%	90%	90%	95%
GLU82	ARG54	80%	91%	83%	94%
ASP35	SER46-Side	90%	86%	90%	93%
ASN33	SER46	72%	23%	58%	72%
GLU116	ARG24	71%	80%	74%	31%
ASP35	ARG24	62%	66%	24%	15%
ARG73	ARG24	34%	8%	8%	9%
ASN123	ARG54-Main	0%	79%	35%	32%
ASN123	ARG54-Side	0%	32%	15%	18%
ASP35	ASN45	0%	5%	51%	18%

Table S3. List of the main hydrogen bonds involved in cohesin and dockerin binding and their percentage occurrence in the last 200 ns in wildtype **ct** in binding mode II.

Cohesin	Dockerin	Percentage occurrence of H-bonds in the last 200ns of trajectory			
		ct_II_300	***ct_II_325***	***ct_II_350***	***ct_II_400***
TYR70	LEU57	94%	91%	89%	80%
ASP35	SER12	92%	93%	89%	86%
GLU82	ARG20	91%	89%	75%	78%
ASP35	SER12	90%	92%	88%	87%
ASP83	LYS19	82%	6%	78%	84%
ASN123	ARG20	58%	34%	50%	44%
ASN33	THR13	52%	38%	10%	35%
ASN33	SER12	40%	36%	37%	37%
ALA81	LYS19	39%	0%	26%	50%
ASN123	ARG20	90%	19%	35%	67%
GLU116	ARG58	95%	78%	45%	13%

Table S4. List of the main hydrogen bonds involved in cohesin and dockerin binding and their percentage occurrence in the last 200 ns in wildtype **cc** in binding mode I.

Cohesin	Dockerin	Percentage occurrence of H-bonds in the last 200ns of trajectory			
		cc_I_300	***cc_I_325***	***cc_I_350***	***cc_I_400***
ASN47	ALA47	65%	75%	30%	25%
GLY128	GLU44	60%	0%	0%	0%
SER85	ILE26	90%	34%	24%	27%
TYR49	ILE26	7%	57%	38%	28%
SER85	MET27	0%	29%	25%	25%
LYS125	ASP30	0%	62%	7%	0%

Table S5. List of the main hydrogen bonds involved in cohesin and dockerin binding and their percentage occurrence in the last 200 ns in wildtype **cc** in binding mode II.

Cohesin	Dockerin	Percentage occurrence of H-bonds in the last 200ns of trajectory			
		cc_II_300	***cc_II_325***	***cc_II_350***	***cc_II_400***
SER85	LEU57	75%	49%	32%	0%
LYS137	ASP15	60%	26%	42%	26%
TYR49	LEU57	55%	22%	37%	0%
LYS125	GLU65	50%	0%	52%	10%
ASP15	LYS137	16%	0%	0%	0%

Table S6. List of the main hydrogen bonds involved in cohesin and dockerin binding and their percentage occurrence in the last 200 ns in wildtype **ac** in binding mode I.

Cohesin	Dockerin	Percentage occurrence of H-bonds in the last 200ns of trajectory			
		ac_I_300	***ac_I_325***	***ac_I_350***	***ac_I_400***
GLU78	ARG22	89%	86%	64%	68%
TYR32	ASP59	87%	89%	80%	0%
GLU78	ARG58	74%	78%	89%	67%
ALA79	ARG58	79%	10%	0%	29%
GLU128	LYS50	19%	45%	78%	40%
GLY125	ASN52	68%	0%	0%	0%
ASP82	LYS64	44%	0%	0%	51%
ARG83	ASP59	0%	100%	0%	0%
GLY125	ASN52	0%	83%	0%	0%
ASP82	LYS64	0%	81%	0%	0%
ARG83	ASP59	0%	0%	0%	94%

Table S7. List of the main hydrogen bonds involved in cohesin and dockerin binding and their percentage occurrence in the last 200 ns in wildtype **ac** in binding mode II.

Cohesin	Dockerin	Percentage occurrence of H-bonds in the last 200ns of trajectory			
		ac_II_300	***ac_II_325***	***ac_II_350***	***ac_II_400***
GLU78	ARG58	92%	86%	79%	81%
GLU78	ARG22	78%	54%	44%	5%
GLY125	ASN16	76%	59%	66%	1%
GLU128	ARG14	42%	38%	38%	23%
TYR32	ASP23	38%	0%	53%	0%
ARG83	ASP23	0%	100%	0%	0%
ASP82	LYS28	0%	65%	0%	0%

Figure S23. Interaction energies between the cohesin and dockerin modules in native **ct**: A, B – van der Waal's interactions in binding mode I and II, respectively; C, D – electrostatics interactions in binding mode I and II, respectively.

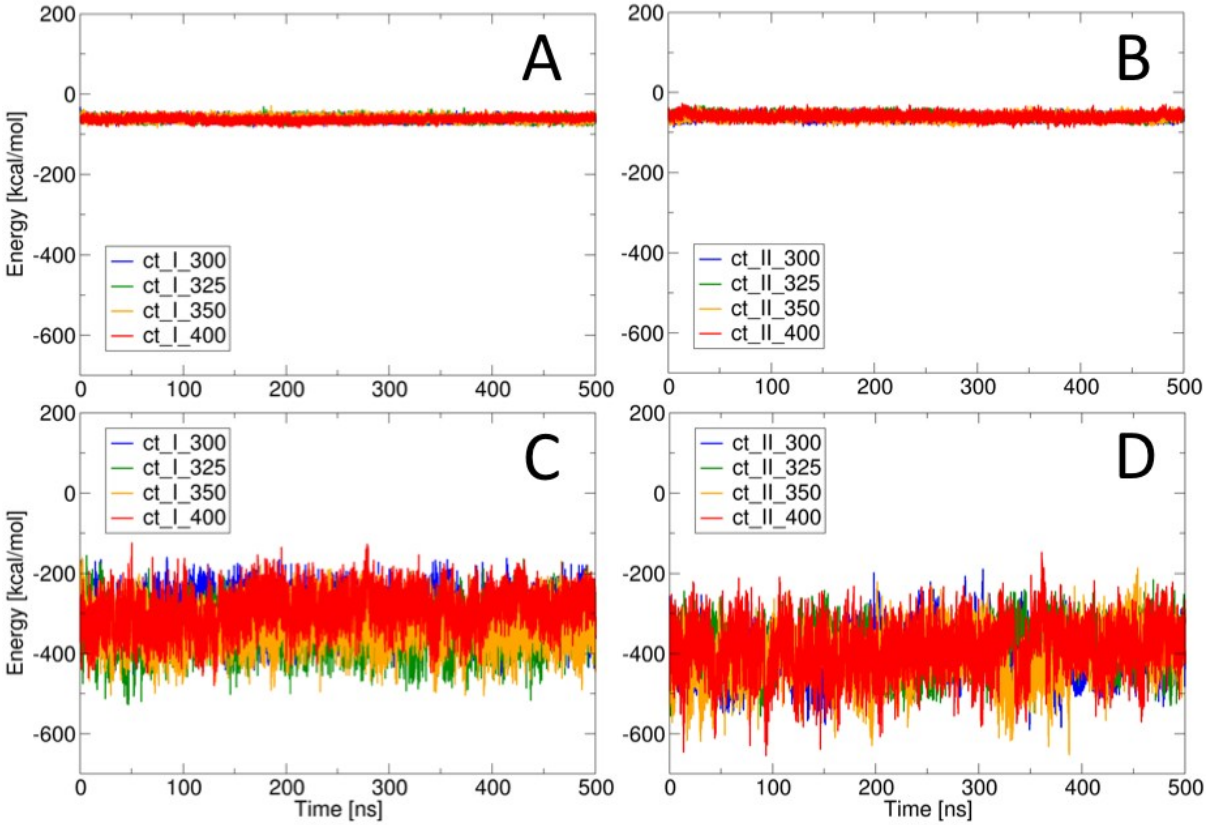


Figure S24. Interaction energies between the cohesin and dockerin modules in **ac**: A, B – van der Waal's interactions in binding mode I and II, respectively; C, D – electrostatics interactions in binding mode I and II, respectively.

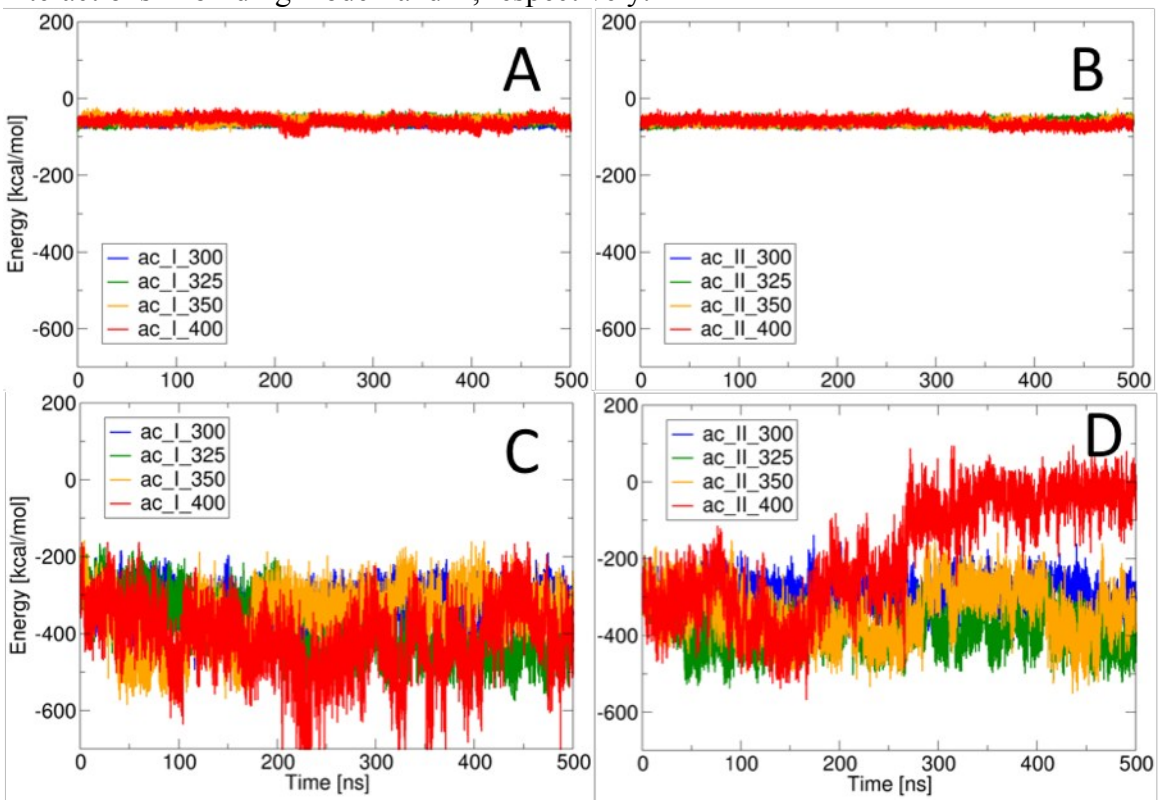
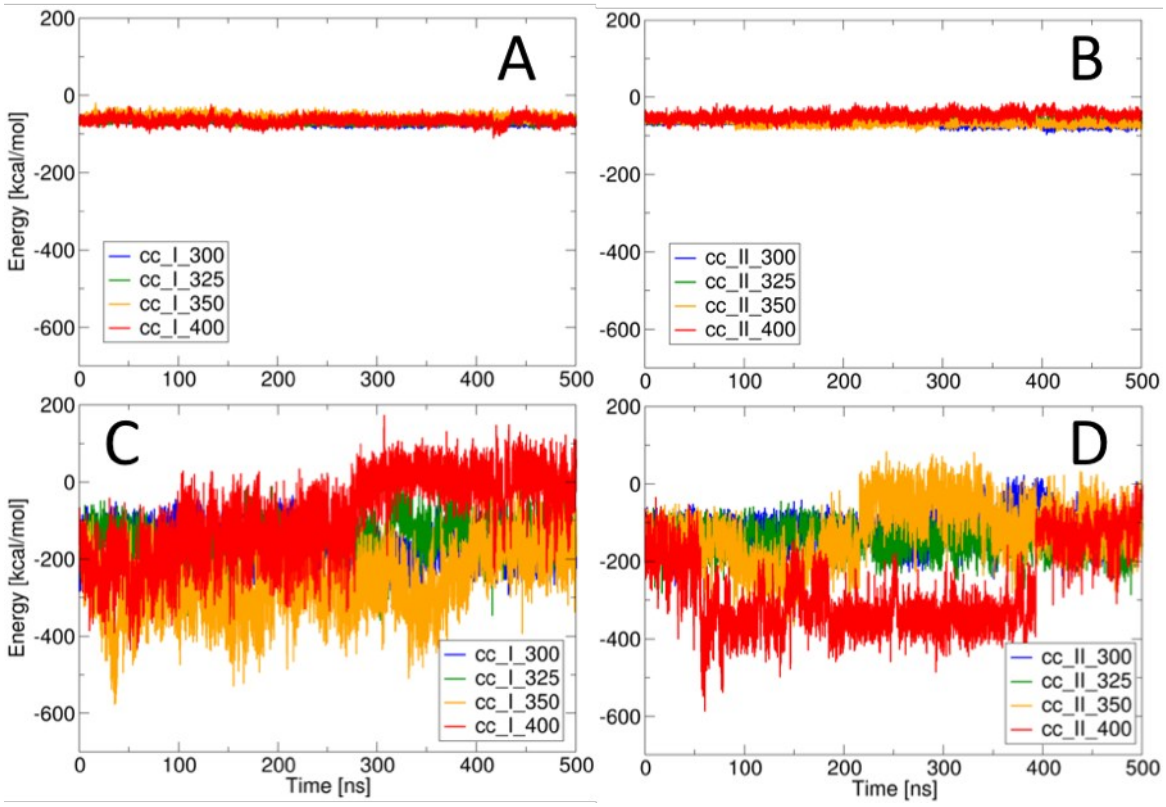


Figure S25. Interaction energies between the cohesin and dockerin modules in **cc**: A, B – van der Waal's interactions in binding mode I and II respectively; C, D – electrostatics interactions in binding mode I and II respectively.



Supplementary note on the effect of water on binding of dockerin to the cohesin domain

Ordered water molecules are known to mediate the hydrogen bond network at protein-protein interfaces. Their small size (and hence mobility) coupled with the H-bond donor hydrogen atoms and H-acceptor oxygen atom, allows water molecules to fill empty spaces between proteins and modulate the binding specificity. In this study, cohesin and dockerin complexes were analyzed for water-mediated hydrogen bonding by computing the number of hydrogen bonds formed with water within 5 Å of the interface. Out of the three species, only **ct** exhibited water-mediated hydrogen bonds, between Coh|Glu82-Arg54|Doc with occurrence of up to 77%. The water-mediated hydrogen bond is present (more than 50% occurrence) in all **ct** simulations except in **ct_II_400**. By contrast, mesophilic complexes do not show any stable water-mediated hydrogen bonds.

Figure S26. Two pathways for the unbinding of dockerin from cohesin: A – **ct_I_300** and **B – ct_I_400**. **ct_I_300** decays comparatively slower than **ct_I_400** due to a slightly different pathway.

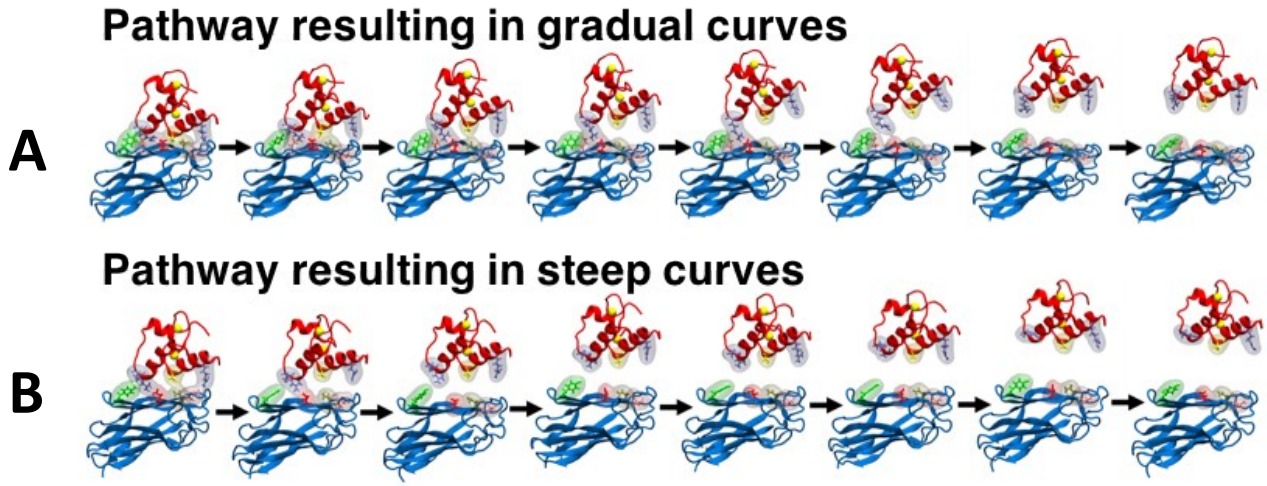


Figure S27. Secondary structure changes throughout simulations of mutated **ac** in binding mode I: A, B, C, D – cohesin; E, F, G, H – dockerin at 300 K, 325 K, 350 K and 400 K, respectively.

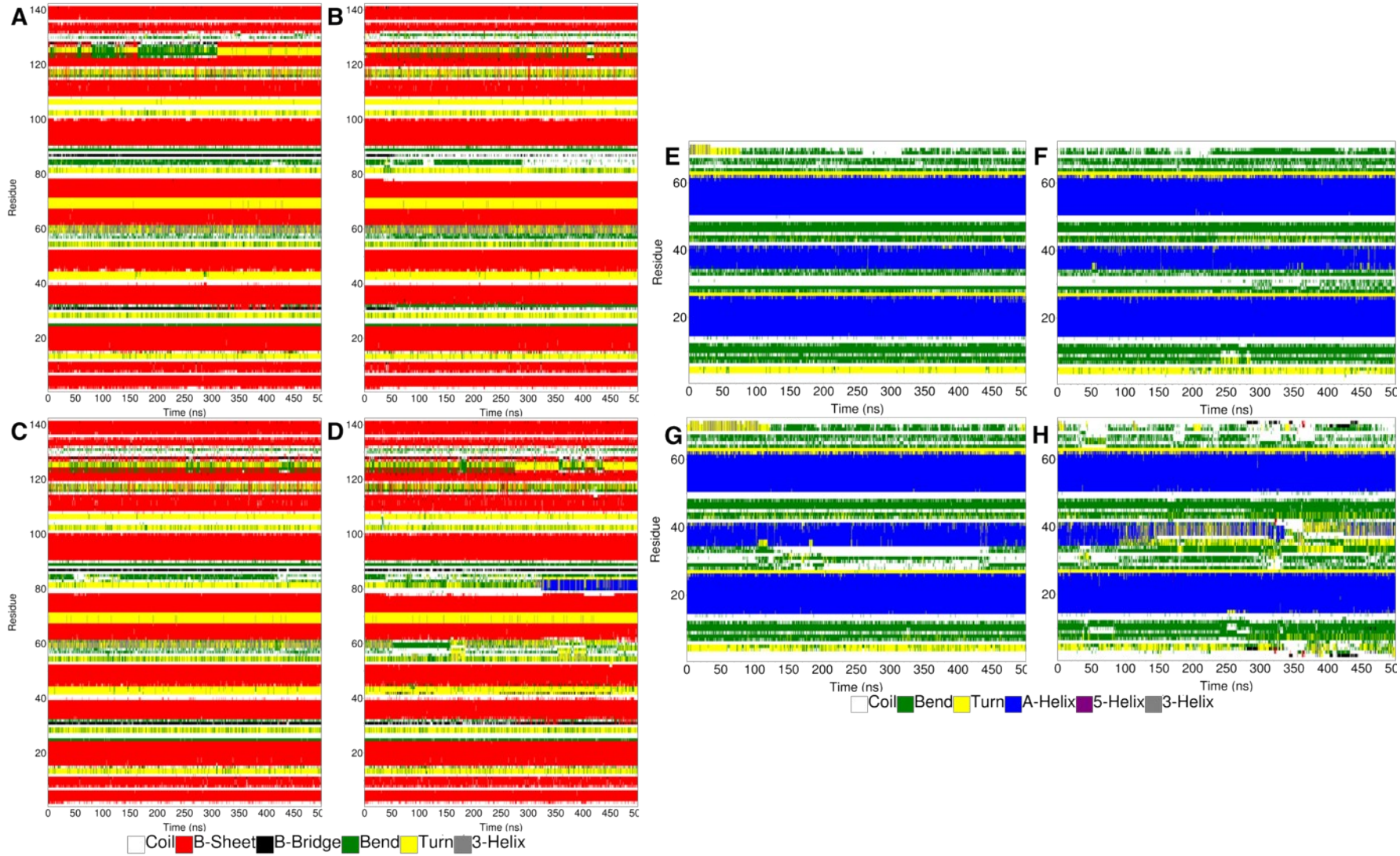


Figure S28. Secondary structure changes throughout simulations of mutated **ac** in binding mode II: A, B, C, D – cohesin; E, F, G, H – dockerin at 300 K, 325 K, 350 K and 400 K, respectively.

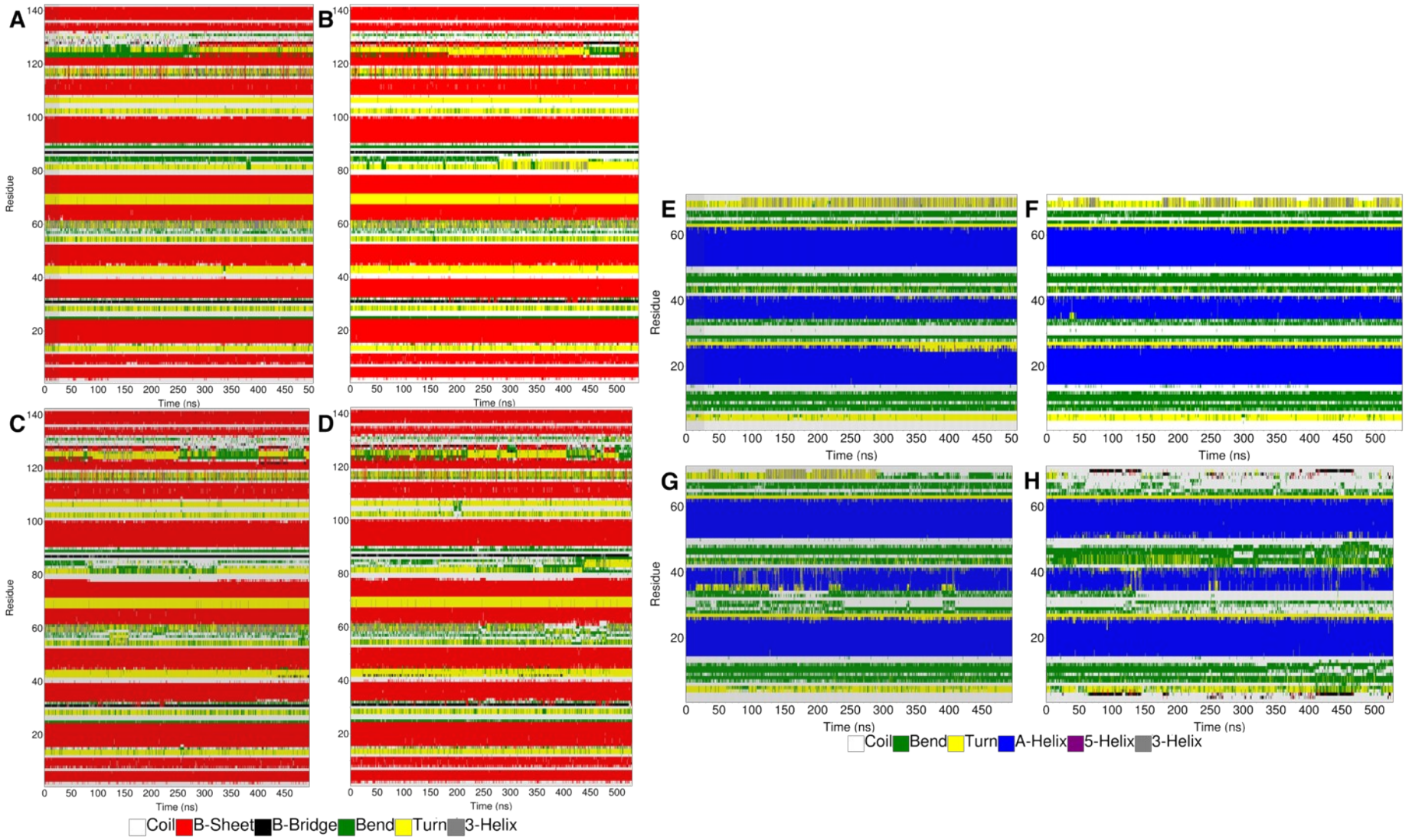


Figure S29. Secondary structure changes throughout simulations of mutated **cc** in binding mode I: A, B, C, D – cohesin; E, F, G, H – dockerin at 300 K, 325 K, 350 K and 400 K, respectively.

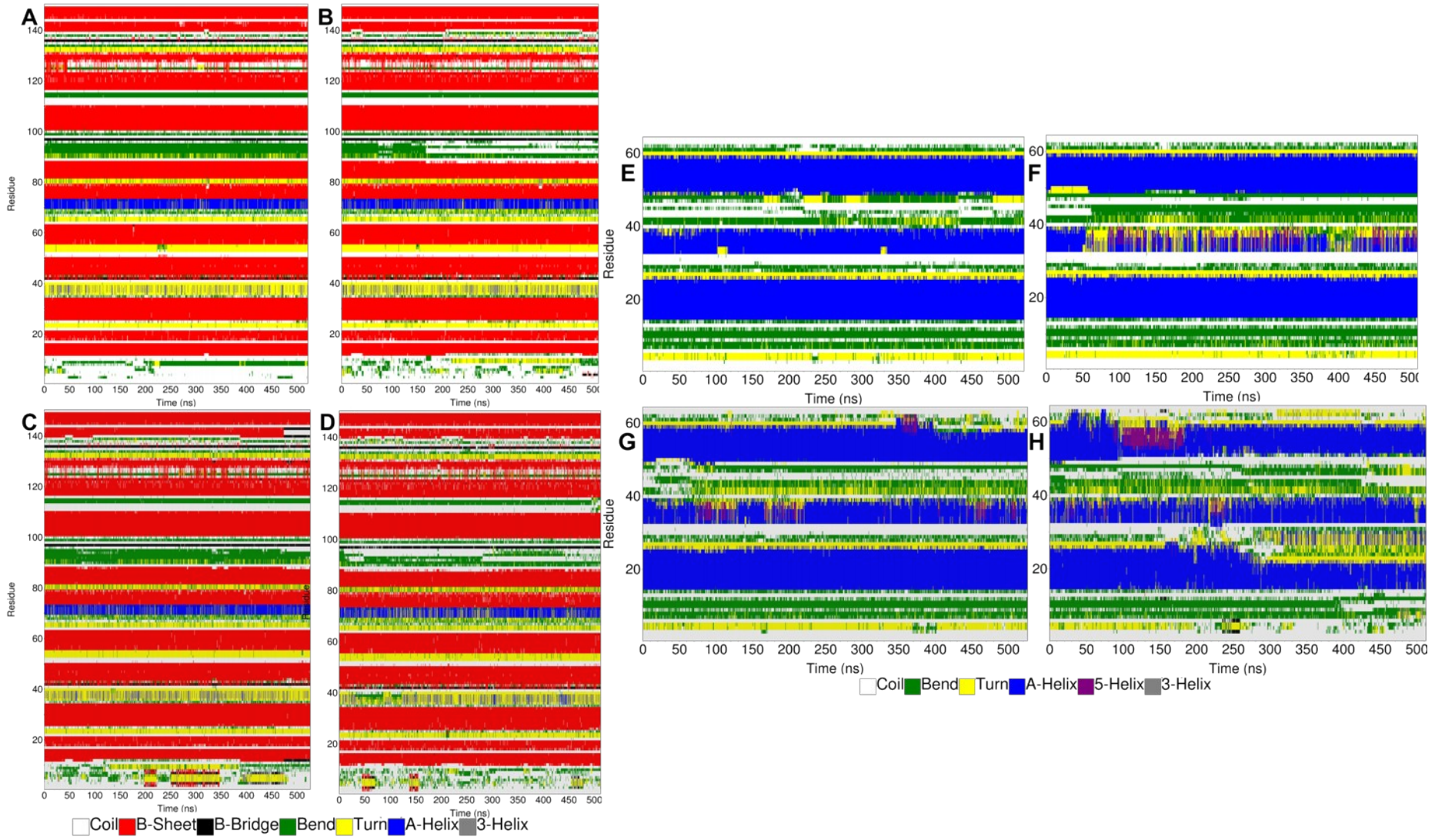


Figure S30. Secondary structure changes throughout simulations of mutated **cc** in binding mode II: A, B, C, D – cohesin; E, F, G, H – dockerin at 300 K, 325 K, 350 K and 400 K, respectively.

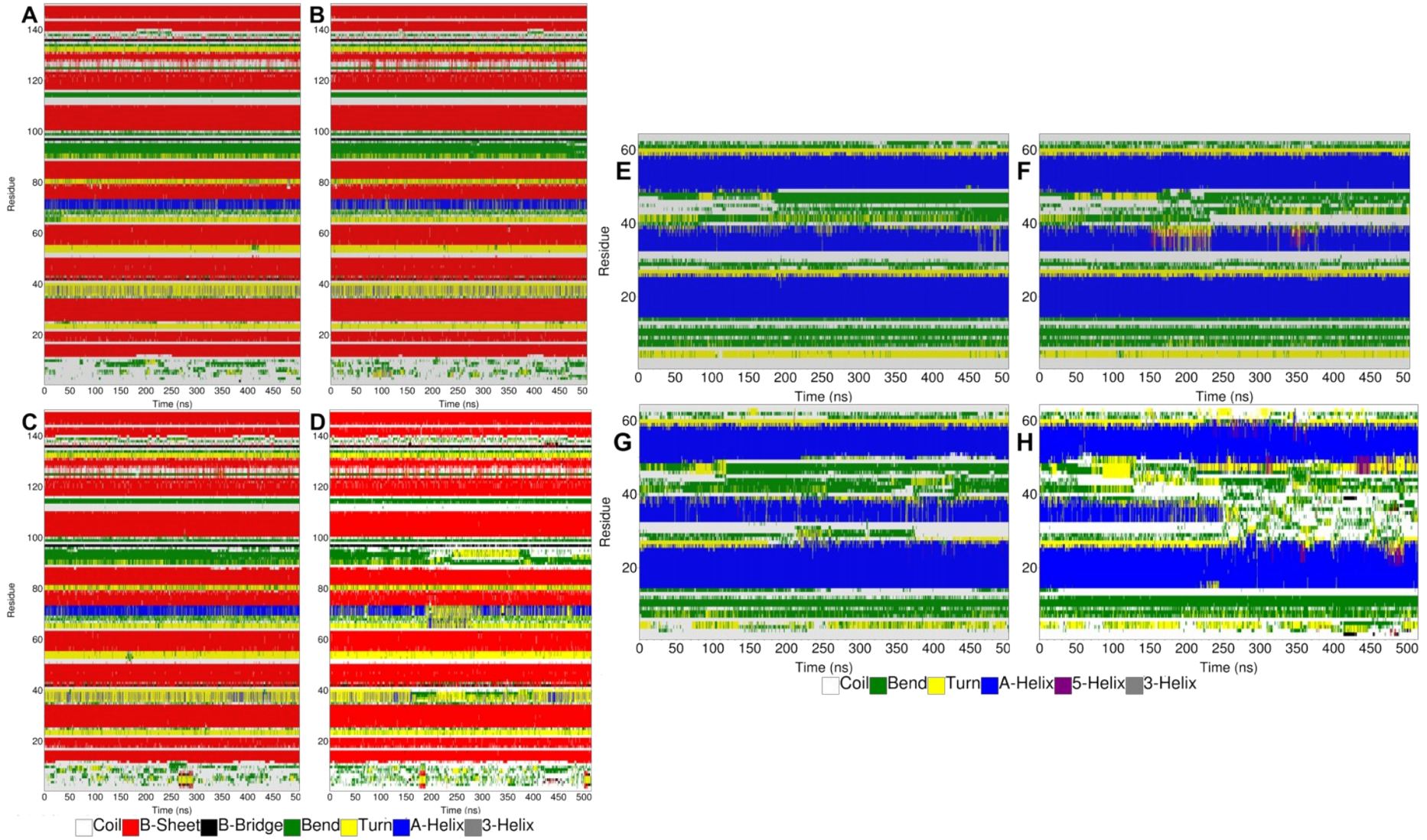


Figure S31. Root mean square fluctuations of C-alpha atoms of individual protein in the simulations at 300 K, 325 K, 350 K and 400 K: A, B represent mutated **ac** in binding mode I (cohesin and dockerin, respectively), while C, D show RMSF in mutated **ac** binding mode II (cohesin and dockerin, respectively).

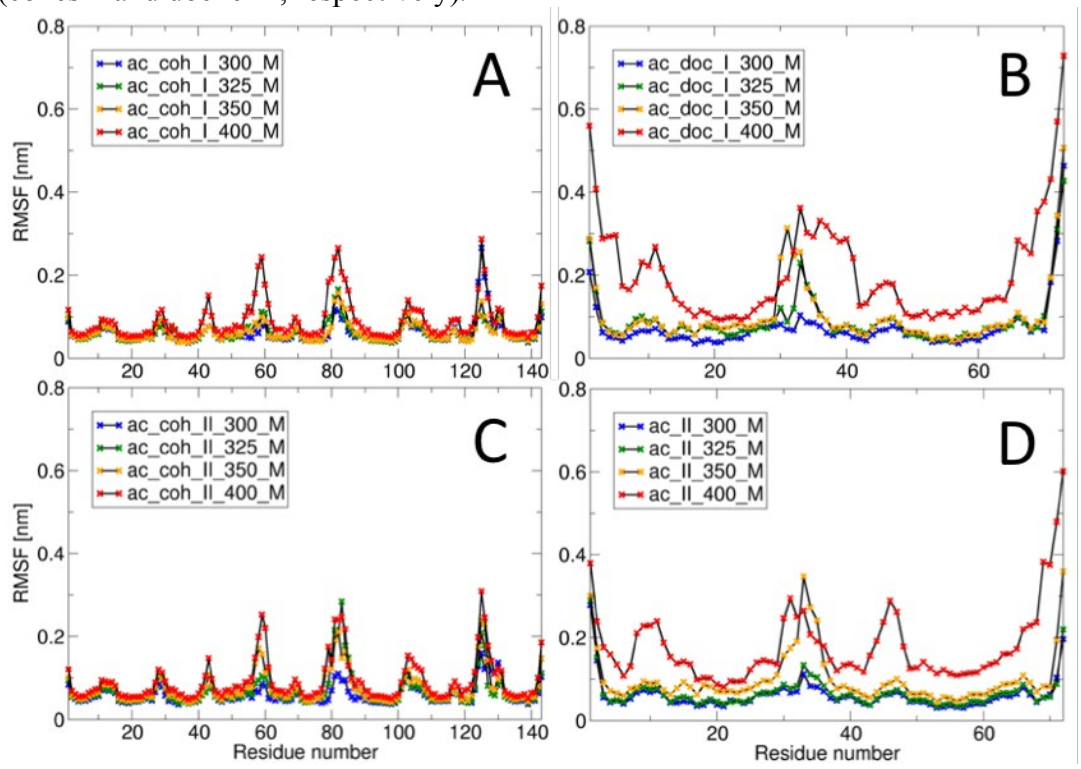


Figure S32. Root mean square fluctuations of C-alpha atoms of individual protein in the simulations at 300 K, 325 K, 350 K and 400 K: A, B represent mutated **cc** in binding mode I (cohesin and dockerin, respectively), while C, D show RMSF in mutated **cc** binding mode II (cohesin and dockerin, respectively).

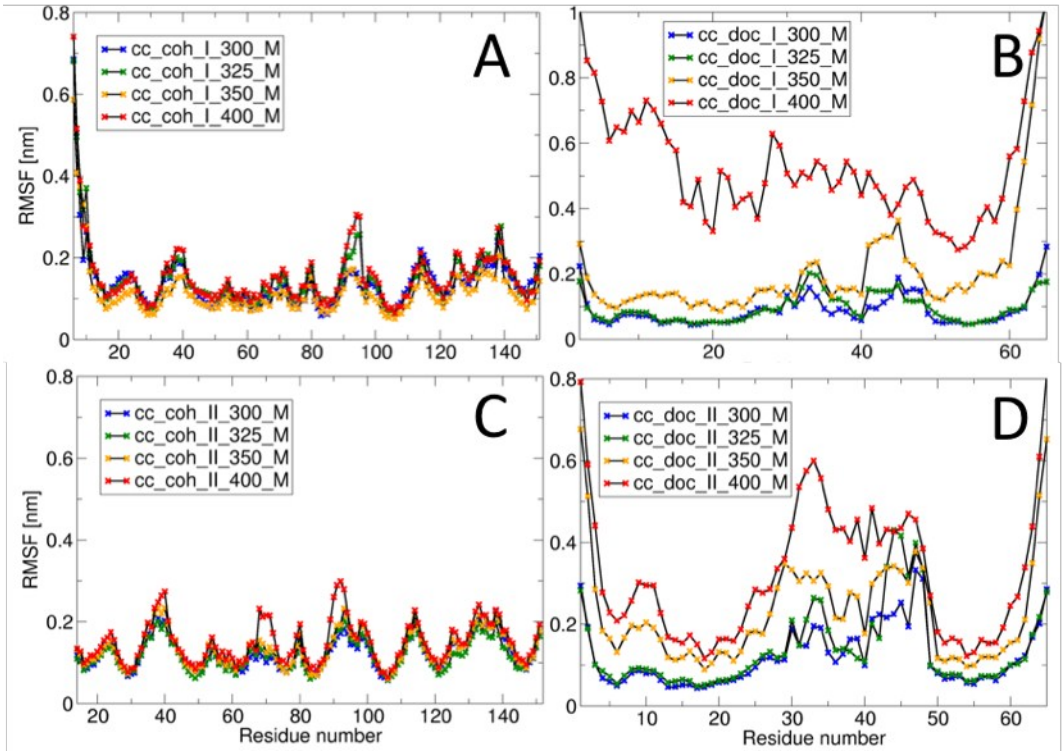


Figure S33. Binding free energy obtained by umbrella sampling for **ct** in binding mode II at 325 K (green) and at 350 K (blue), depicting error bars. Error was estimated by bootstrapping with three different methods: Bayesian histogram mixing, histogram mixing and trajectory mixing. The most favourable, Bayesian histogram mixing (red and orange), and the least favourable, trajectory mixing (green and blue), are shown for both temperatures.

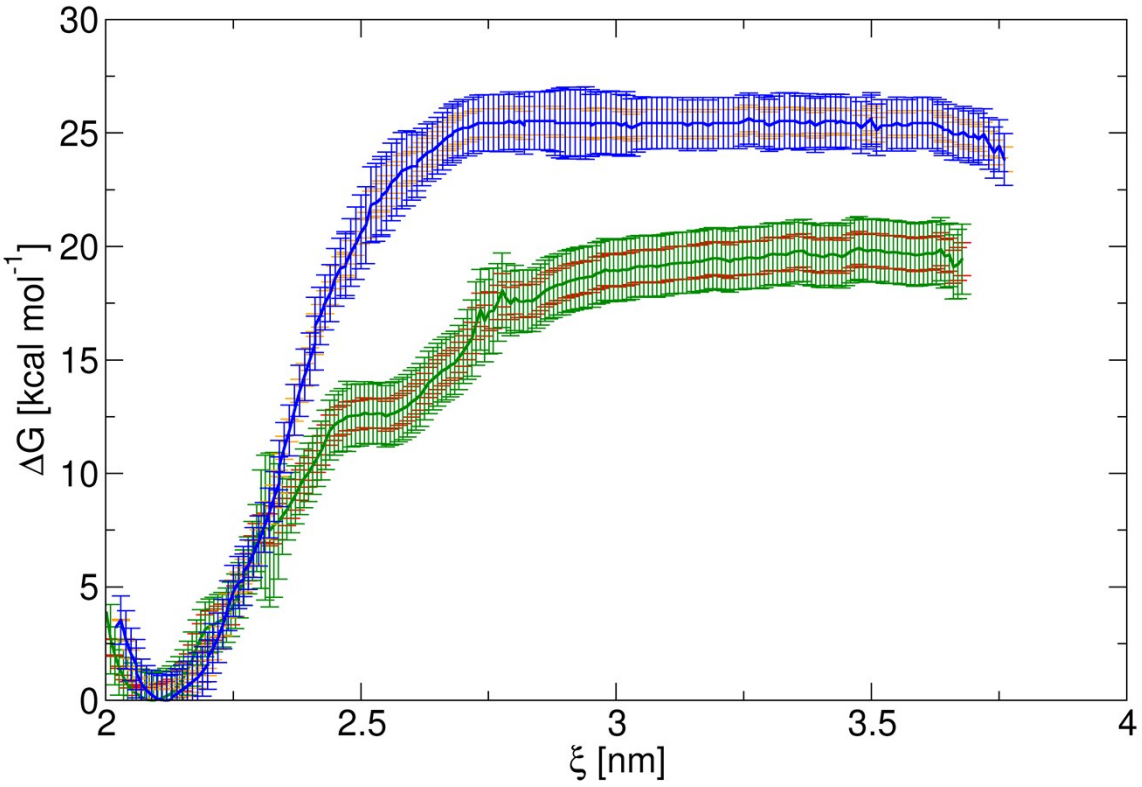


Figure S34. Root Mean Square Deviation of the mutated dockerin (A) of **cc** and of the calcium ions (B) at 300 K (black) and 400 K (red).

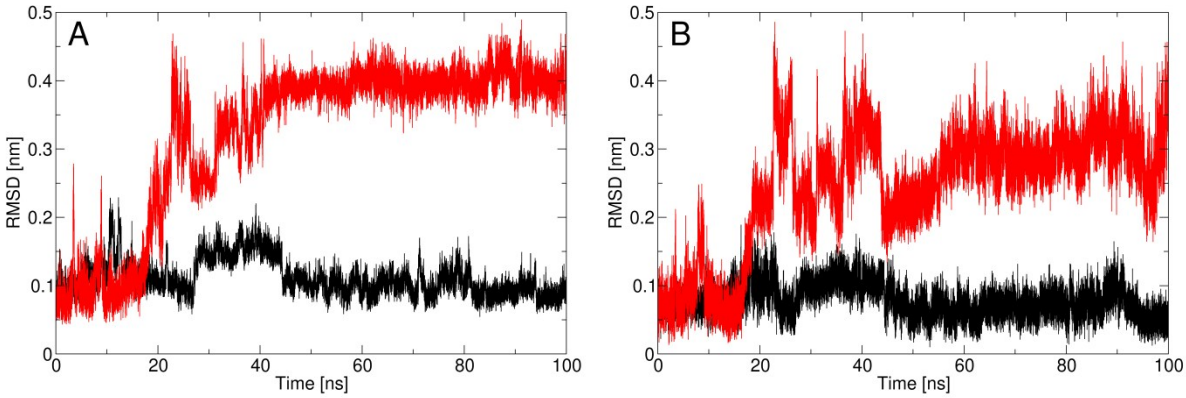


Figure S35. Binding free energy between mutated dockerin and cohesin of the mesophilic **cc** in binding mode I at 300 K (black) and 400 K (red). Error bars are obtained by bootstrapping and trajectory mixing.