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

Supplementary Materials

On the distinct binding modes of expansin and carbohydrate-binding module proteins on crystalline and nanofibrous cellulose: implications for cellulose degradation by designer cellulosomes

Adam Orłowski[†], Lior Artzi', Pierre-Andre Cazade[†], Melissabye Gunnoo†, Edward A. Bayer', Damien Thompson[*†]

[†] *Department of Physics, Bernal Institute, University of Limerick, V94 T9PX, Ireland*

'Department of Biomolecular Sciences, The Weizmann Institute of Science, Rehovot, Israel

[*]Correspondence requested to be sent to: Damien.Thompson@ul.ie

LEGENDS TO SUPPLEMENTAL FIGURES

Figure S0. Synergistic effect of *Ccl*EXL1 and cellulases on filter paper hydrolysis. *C. clariflavum* cellulosomal enzymes GH48 and GH9 were added to filter paper strips that were pretreated either with the expansin-like protein *Ccl*EXL1 or CBM3a of scaffoldin CipA from *C. thermocellum*. A significant synergistic effect was obtained by treatment of the filter paper with *Ccl*EXL1 prior to enzymatic hydrolysis, while treatment with CBM3a did not make a statistically significant contribution to enzymatic hydrolysis. Three asterisks (***) represent statistical significance (p-value <0.001) of two-tailed t-test, p-value=$8.03 \cdot 10^{-5}$.
Figure S1. Sequence alignment of BsEXLX1 (pdb structure 3d30) used in the simulation with the CclEXL1 used in the experiments. A. Homology model structure built by aligning the CclEXL1 sequence with structure 3d30 with the colors corresponding to the identity in sequence (blue – the same amino acid, red – different amino acid). B. The two structures aligned and coloured in gray and green.
Figure S2. Changes of CBM secondary structure in time in simulation CBM-c1.
Figure S3. Changes of CBM secondary structure in time in simulation CBM-c2.
Figure S4. Changes of CBM secondary structure in time in simulation CBM-c3.
Figure S5. Changes of CBM secondary structure in time in simulation CBM-c4.
Figure S6. Changes of CBM secondary structure in time in simulation CBM-c5.
Figure S7. Changes of CBM secondary structure in time in simulation CBM-aI1.
Figure S8. Changes of CBM secondary structure in time in simulation CBM-aI2.
Figure S9. Changes of CBM secondary structure in time in simulation CBM-aII1.
Figure S10. Changes of CBM secondary structure in time in simulation CBM-aII2.
Figure S11. Changes of expansin secondary structure in time in simulation Exp-c1.
Figure S12. Changes of expansin secondary structure in time in simulation Exp-aI1.
Figure S13. Changes of expansin secondary structure in time in simulation Exp-aI2.
Figure S14. Changes of expansin secondary structure in time in simulation Exp-aI3.
Figure S15. Changes of expansin secondary structure in time in simulation Exp-aI4.
Figure S16. Changes of expansin secondary structure in time in simulation Exp-aII1.
Figure S17. Changes of expansin secondary structure in time in simulation Exp-aII2.
Figure S18. Root mean square deviation (RMSD) of CBM C-alpha atoms in simulations: CBM-c1 (A), CBM-c2 (B), CBM-c3 (C), CBM-c4 (D), CBM-c5 (E), CBM-aI1 (F), CBM-aI2 (G), CBM-aII1 (H), CBM-aII2 (I).
Figure S19. Root mean square deviation (RMSD) of expansin C-alpha atoms in simulations: Exp-c1 (A), Exp-aI1 (B), Exp-aI2 (C), Exp-aI3 (D), Exp-aI4 (E), Exp-aII1 (F), Exp-aII2 (G).
Figure S20. Root mean square fluctuation (RMSF) of C-alpha atoms of CBM in simulations: CBM-c1 (A), CBM-c2 (B), CBM-c3 (C), CBM-c4 (D), CBM-c5 (E), CBM-aI1 (F), CBM-aI2 (G), CBM-aII1 (H), CBM-aII2 (I).
Figure S21. Root mean square fluctuation (RMSF) of expansin C-alpha atoms in simulations: Exp-c1 (A), Exp-aI1 (B), Exp-aI2 (C), Exp-aI3 (D), Exp-aI4 (E), Exp-aII1 (F), Exp-aII2 (G).
Figure S22. Initial steps of CBM-cellulose binding from simulation CBM-c2 at: 397 ns (A), 398 ns (B), 399 ns (C), 400 ns (D), 406 ns (E), 410 ns (F), 420 ns (G), 425 ns (H). Cellulose is depicted as an orange surface, protein is presented in "New Cartoon" representation and colored accordingly to the secondary structure, water and ions are not shown for clarity.
Figure S23. Initial steps of CBM-cellulose binding from simulation CBM-c5 at: 222 ns (A), 226 ns (B), 245 ns (C), 250 ns (D). Cellulose is depicted as an orange surface, protein is

presented in “New Cartoon” representation and colored accordingly to the secondary structure, water and ions are not shown for clarity.

Figure S24. Initial steps of CBM-cellulose binding from simulation CBM-c2 with the depicted residues that make first contact with cellulose: SER29 and ASN64 in simulation CBM-c2 (A), ASN64 in simulation CBM-c5 (B), TYR67 in simulation CBM-c5 (C). Cellulose is depicted as an orange surface, protein is presented in “New Cartoon” representation and colored accordingly to the secondary structure, water and ions are not shown for clarity.

Figure S25. Potential of mean force (PMF) for CBM-cellulose (black line), and expansin-cellulose (red li Potential of mean force (PMF) for CBM-cellulose (black line), and expansin-cellulose (red line) binding. The center of the mass distance (COM) corresponds to the protein cellulose distance where 0 nm mean starting (protein bound) distance).

LEGENDS TO SUPPLEMENTAL TABLES

Table S1. Number of molecules and atoms in the simulated systems.
Table S2. CBM-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation CBM-c1.
Table S3. CBM-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation CBM-c2.
Table S4. CBM-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation CBM-c3.
Table S5. CBM-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation CBM-c4.
Table S6. CBM-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation CBM-c5.
Table S7. CBM-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation CBM-aI1.
Table S8. CBM-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation CBM-aI2.
Table S9. CBM-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation CBM-aII1.
Table S10. CBM-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation CBM-aII2.
Table S11. Expansin-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation Exp-c1.
Table S12. Expansin-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation Exp-aI1.
Table S13. Expansin-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation Exp-aI2.
Table S14. Expansin-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation Exp-aI3.
Table S15. Expansin-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation Exp-aI4.
Table S16. Expansin-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation Exp-aII1.
Table S17. Expansin-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation Exp-aII2.
Table S18. Percentage of frames in the trajectory where any atoms of the CBM residues are within 0.35 nm radius from any of cellulose atoms (any occurrence in a frame for a given residue results in a positive result) in the simulations with CBM.
Table S19. Percentage of frames in the trajectory where any atoms of the expansin residues are within 0.35 nm radius from any of cellulose atoms (any occurrence in a frame for a given residue results in a positive result) in the simulations with expansin.

LEGENDS TO SUPPLEMENTAL MOVIES

Movie S1. Trajectory from the simulation CBM-c1. Cellulose is depicted as an orange surface, protein is presented in "New Cartoon" representation and colored accordingly to the secondary structure, water and ions are not shown for clarity.
Movie S2. Trajectory from the simulation CBM-c2. Cellulose is depicted as an orange surface, protein is presented in "New Cartoon" representation and colored accordingly to the secondary structure, water and ions are not shown for clarity.
Movie S3. Trajectory from the simulation CBM-c3. Cellulose is depicted as an orange surface, protein is presented in "New Cartoon" representation and colored accordingly to the secondary structure, water and ions are not shown for clarity.
Movie S4. Trajectory from the simulation CBM-c4. Cellulose is depicted as an orange surface, protein is presented in "New Cartoon" representation and colored accordingly to the secondary structure, water and ions are not shown for clarity.
Movie S5. Trajectory from the simulation CBM-c5. Cellulose is depicted as an orange surface, protein is presented in "New Cartoon" representation and colored accordingly to the secondary structure, water and ions are not shown for clarity.
Movie S6. Trajectory from the simulation CBM-aI1. Cellulose is depicted as an orange surface, protein is presented in "New Cartoon" representation and colored accordingly to the secondary structure, water and ions are not shown for clarity.
Movie S7. Trajectory from the simulation CBM-aI2. Cellulose is depicted as an orange surface, protein is presented in "New Cartoon" representation and colored accordingly to the secondary structure, water and ions are not shown for clarity.
Movie S8. Trajectory from the simulation CBM-aII1. Cellulose is depicted as an orange surface, protein is presented in "New Cartoon" representation and colored accordingly to the secondary structure, water and ions are not shown for clarity.
Movie S9. Trajectory from the simulation CBM-aII2. Cellulose is depicted as an orange surface, protein is presented in "New Cartoon" representation and colored accordingly to the secondary structure, water and ions are not shown for clarity.
Movie S10. Trajectory from the simulation Exp-c1. Cellulose is depicted as an orange surface, protein is presented in "New Cartoon" representation and colored accordingly to the secondary structure, water and ions are not shown for clarity.
Movie S11. Trajectory from the simulation Exp-aI1. Cellulose is depicted as an orange surface, protein is presented in "New Cartoon" representation and colored accordingly to the secondary structure, water and ions are not shown for clarity.
Movie S12. Trajectory from the simulation Exp-aI2. Cellulose is depicted as an orange surface, protein is presented in "New Cartoon" representation and colored accordingly to the secondary structure, water and ions are not shown for clarity.
Movie S13. Trajectory from the simulation Exp-aI3. Cellulose is depicted as an orange surface, protein is presented in "New Cartoon" representation and colored accordingly to the secondary structure, water and ions are not shown for clarity.
Movie S14. Trajectory from the simulation Exp-aI4. Cellulose is depicted as an orange surface, protein is presented in "New Cartoon" representation and colored accordingly to the secondary structure, water and ions are not shown for clarity.
Movie S15. Trajectory from the simulation Exp-aII1. Cellulose is depicted as an orange surface, protein is presented in "New Cartoon" representation and colored accordingly to the secondary structure, water and ions are not shown for clarity.

Movie S16. Trajectory from the simulation Exp-aII2. Cellulose is depicted as an orange surface, protein is presented in "New Cartoon" representation and colored accordingly to the secondary structure, water and ions are not shown for clarity.
Movie S17. Trajectory from the simulation Exp-c1 – with the residues: GLU75 and ASP82 depicted as blue and red VdW spheres accordingly. Cellulose is depicted as an orange surface, protein is presented in "New Cartoon" representation and colored in green, water and ions are not shown for clarity.

Experimental protocols

(a) Cloning

The CBM3a gene was amplified from the genomic DNA of *C. thermocellum* ATCC 27405 (forward primer: 5'-TTATAACCATGGCAAATACACCGGTATCAGGC−3', reverse primer: 5'-TATAATCTCGAGTTATACTACACTGCCACCGGG-3') by polymerase chain reaction (PCR), using Phusion High-Fidelity DNA polymerase (New England Biolabs Inc, Ipswich, MA, USA). The gene was cloned into a pET28a plasmid using the FastDigest restriction enzymes NcoI and XhoI (Thermo Scientific, Fermentas UAB,Vilnius, Lithuania). The plasmid was purified by QIAprep Spin Miniprep Kit (QIAGEN GmbH, D-40724 Hilden, Germany). The genes of the two enzymes GH48 (Clocl_4007) and GH9 (Clocl_2225) and the *Ccl*EXL1 gene were amplified from the genomic DNA of *C. clariflavum* DSM 19732 (Leibniz Institute DSMZ- German Collection of Microorganisms and Cell Cultures), and the genes were cloned into pET28a plasmid as described previously (Artzi et al. 2016).

(b) Protein expression and purification

The pET28a plasmids containing the four genes were transformed into *E. coli* BL21 (DE3) cells. The transformed cells were grown in LB (Luria-Bertani broth) medium, containing 50 µg/mL Kanamycin (Sigma-Aldrich, Rehovot, Israel) and 2 mM $CaCl_2$. The cells were grown for 2.5 h at 37°C to A_{600} 0.8-1, and 0.2 mM of isopropyl-1-thio-β-D-galactoside (IPTG) was added to induce protein expression, and the cells were then transferred for over-night growth at 16°C for 16 h. The cells were harvested by centrifugation (4192 g, 15 min) and sonicated. The sonicate was centrifuged at 26915 g for 30 min, and the supernatant fluids containing the CBM3a protein were added to 2 g of macroporous bead cellulose preswollen gel (IONTO-SORB, Usti nad Labem, Czech Republic). The binding was performed for 1 h at 4°C. The suspension was uploaded onto a gravity column, and the beads were washed with (i) 100 mL of Tris buffer saline (TBS; 137 mM NaCl, 2.7 mM KCl, 25 mM Tris-HCl, pH 7.4) supplemented with 1 M NaCl, and (ii) 100 mL of TBS. The protein was eluted from the beads using 30 mL of 1% trimethylamine (TEA), and the elution fractions were brought to neutral pH by adding 1 M 2-(N-morpholino)ethanesulfonic acid (MES) buffer, pH 5.5. The purity of the protein was estimated by SDS-PAGE, and the protein was dialyzed in TBS containing 5 mM $CaCl_2$. The proteins GH48, GH9 and *Ccl*EXL1 were cloned and expressed with a His-tag, and their purification was performed in batch by Ni-NTA beads purification as described previously (Vazana et al. 2010). All proteins were stored at -20°C with glycerol in 50% vol/vol ratio.

(c) Filter paper hydrolysis assay

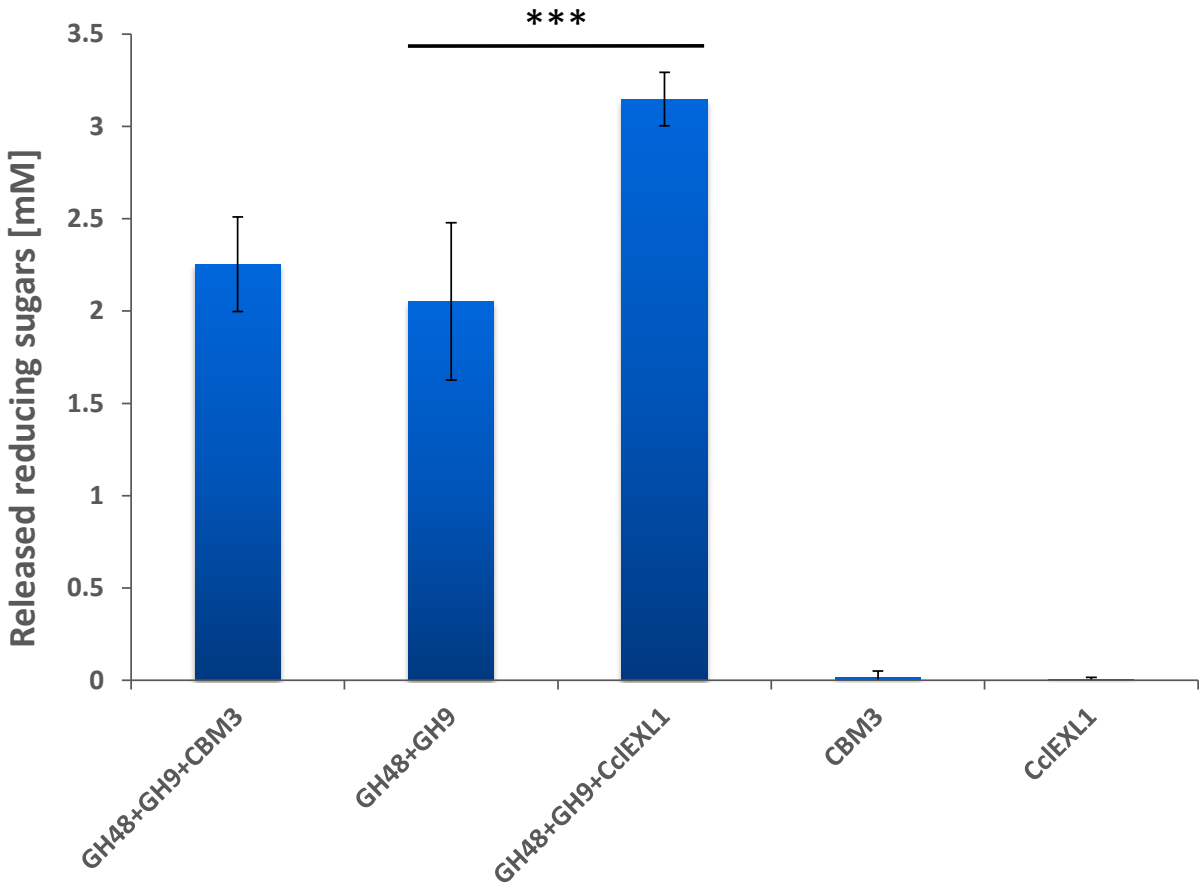
Hand-cut strips (1 cm^2) of filter paper no. 3 (Whatmann, Little Chalfont, Buckinghamshire, UK) were preincubated for 1 h at 55°C with 50 mM acetate buffer, pH 5.5, containing 0.6 mg/mL *Ccl*EXL1 or 0.3 mg/mL CBM3a (i.e., the same molar ratio; MW of *Ccl*EXL1 is 35.12 kDa, MW of CBM3a is 18.37 kDa), or acetate buffer only as a control. After incubation in the

respective solution, the filter paper strips were transferred into tubes containing 50 mM acetate buffer, pH 5.5, supplemented with 12 mM $CaCl_2$ and 2 mM EDTA, containing either 0.5 µM of each of the enzymes GH48 and GH9 or without enzymes. The final volume of each reaction was 200 µL. Tubes were incubated for 24 h, 55°C, rotating (300 rpm). The analysis of the enzymatic activity was performed by using the DNS method (Miller 1959): 150 µL of dinitrosalicylic acid (DNS) were supplemented with 100 µL of reaction volume from each tube, and the samples were boiled for 10 min. The tubes were cooled on ice, and the resultant absorbance was measured at 540 nm to determine the amount of released reducing sugars. The experiment was repeated three times in triplicate.

Experimental findings – CBM vs. expansin promotion of cellulose hydrolysis

CBM3a is known to bind to cellulose chains very efficiently (Morag et al. 1995). We were interested to explore whether simple binding of this CBM causes disruption in the cellulose structure that results in a synergistic effect on cellulose hydrolysis when working together with cellulases, as reported before for several expansins and expansin-like proteins (Artzi et al. 2016; Chen et al. 2016). To test this, we preincubated filter paper strips with CBM3a or *Ccl*EXL1, and transferred the preincubated strips into tubes containing two recombinant cellulases originated from the *C. clariflavum* genome, exoglucanase GH48 and processive endoglucanase GH9. These cellulases were shown to work in synergistic manner in the past, and moreover, to benefit from a synergistic effect from preincubation of cellulosic substrates with *Cc*lEXL1 (Artzi et al. 2016). The enzymatic assay (Fig. S0) demonstrates that preincubation of filter paper with *Ccl*EXL1 indeed results in a synergistic effect on the cellulose hydrolysis, whereas preincubation with CBM3a did not contribute to the hydrolysis.

Figure S0. Synergistic effect of *Ccl*EXL1 and cellulases on filter paper hydrolysis. *C. clariflavum* cellulosomal enzymes GH48 and GH9 were added to filter paper strips that were pretreated either with the expansin-like protein *Ccl*EXL1 or CBM3a of scaffoldin CipA from *C. thermocellum*. A significant synergistic effect was obtained by treatment of the filter paper with *Ccl*EXL1 prior to enzymatic hydrolysis, while treatment with CBM3a did not make a statistically significant contribution to enzymatic hydrolysis. Three asterisks (***) represent statistical significance (*p*-value <0.001) of two-tailed t-test, *p*-value=$8.03 \cdot 10^{-5}$.



The findings suggest that the binding of CBM3a does not disrupt the substrate to the extent of exposing more cellulose chains for the enzymes to hydrolyze more efficiently. However, *Ccl*EXL1, which is composed of the domains D1 (considered as the "active" part of the protein) and D2 (the CBM63 responsible for the binding of the expansin to the substrate), exhibits a significant effect on the action of the cellulases, probably owing to domain D1 (see also MD simulation data, below). This brings us to the conclusion that the cellulose-binding activity of CBMs *per se* is not sufficiently disruptive for enhancing cellulose degradation by enzymes, and the action of the additional D1 “active” domain is required in order to provide significant disruption in the cellulose crystalline structure.

Driven by the experimental results we performed the atomistic classical molecular dynamics to deduce the molecular mechanisms of binding of the proteins to cellulose and propose molecular mechanisms responsible for the improvement of cellulose degradation by catalytic enzymes caused by addition of expansin.

Figure S1. Sequence alignment of BsEXLX1 (pdb structure 3d30) used in the simulation with the CclEXL1 used in the experiments. A. Homology model structure built by aligning the CclEXL1 sequence with structure 3d30 with the colors corresponding to the identity in sequence (blue – the same amino acid, red – different amino acid). B. The two structures aligned and coloured in gray and green.

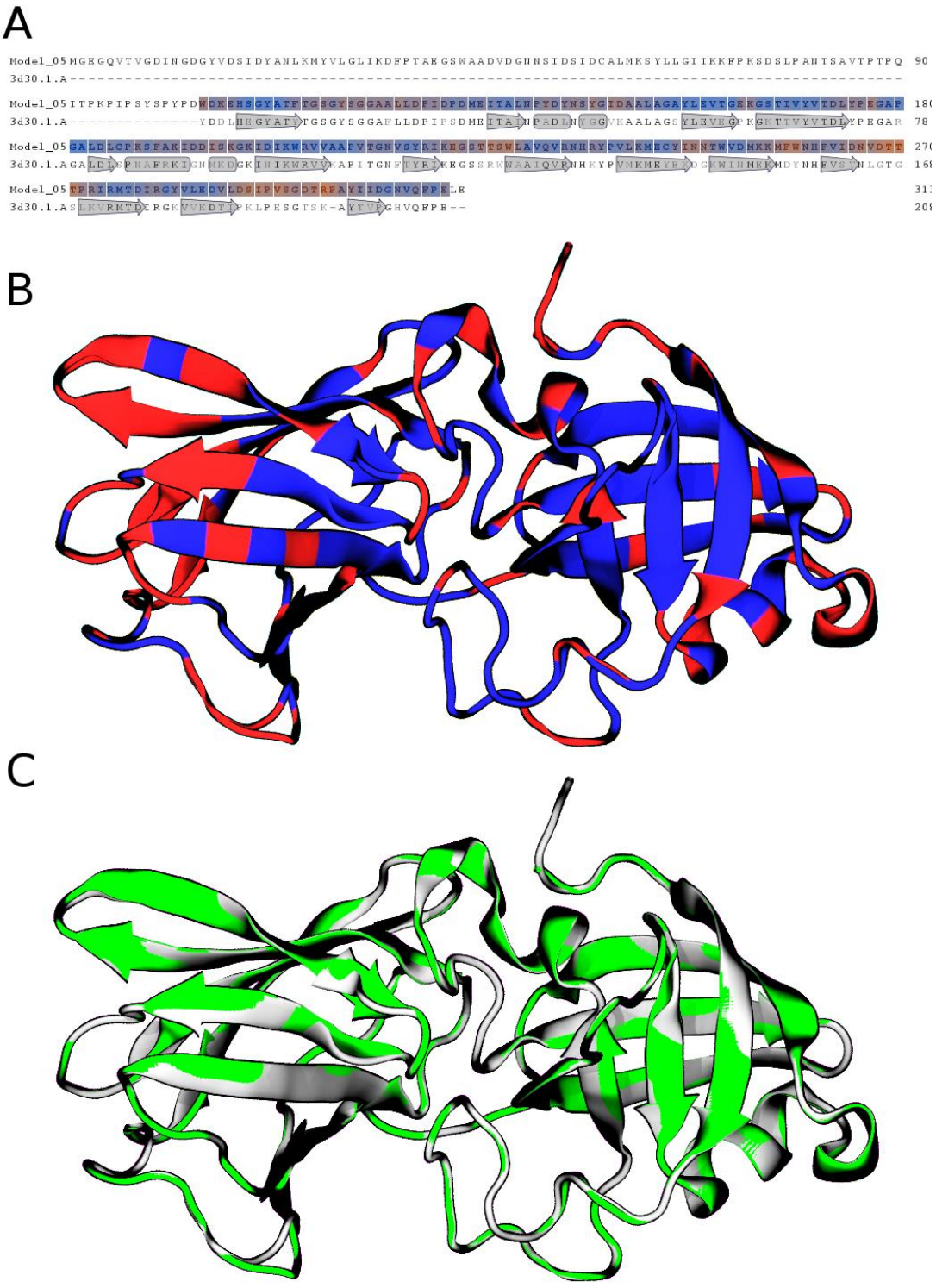


Figure S2. Changes of CBM secondary structure in time in simulation CBM-c1.

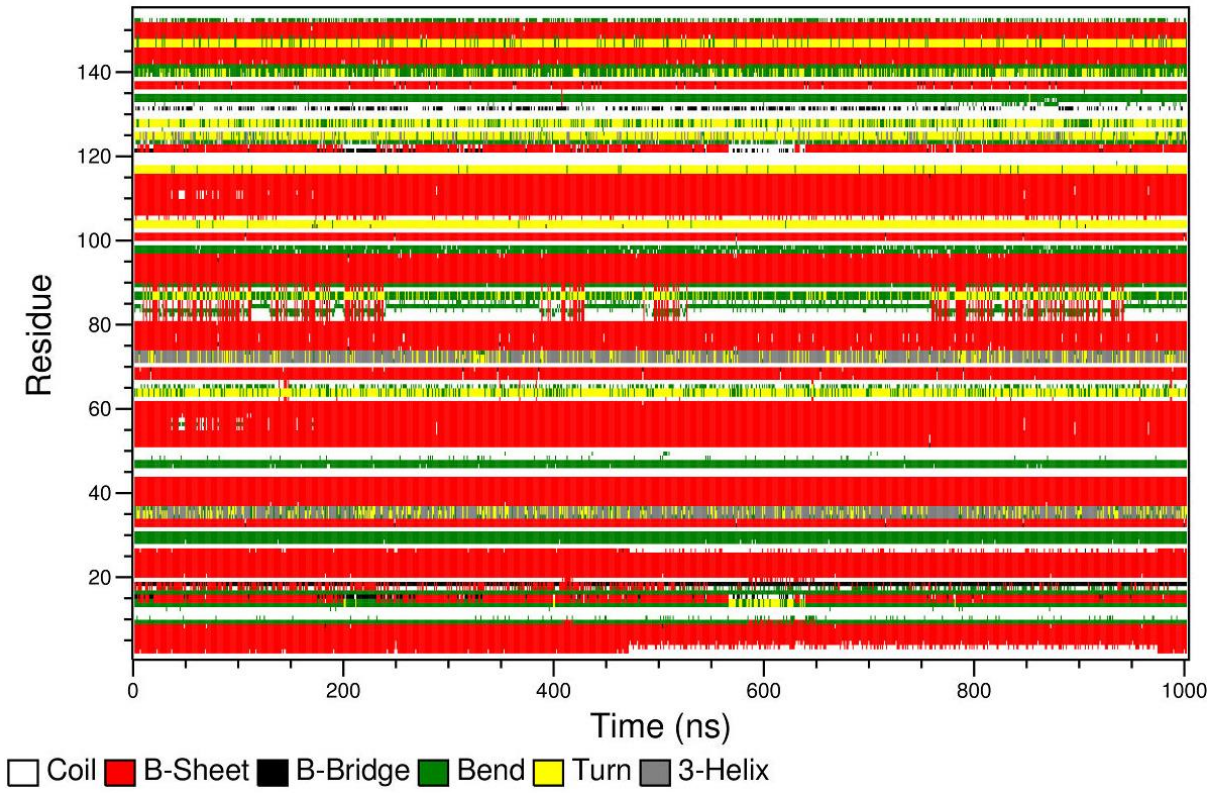


Figure S3. Changes of CBM secondary structure in time in simulation CBM-c2.

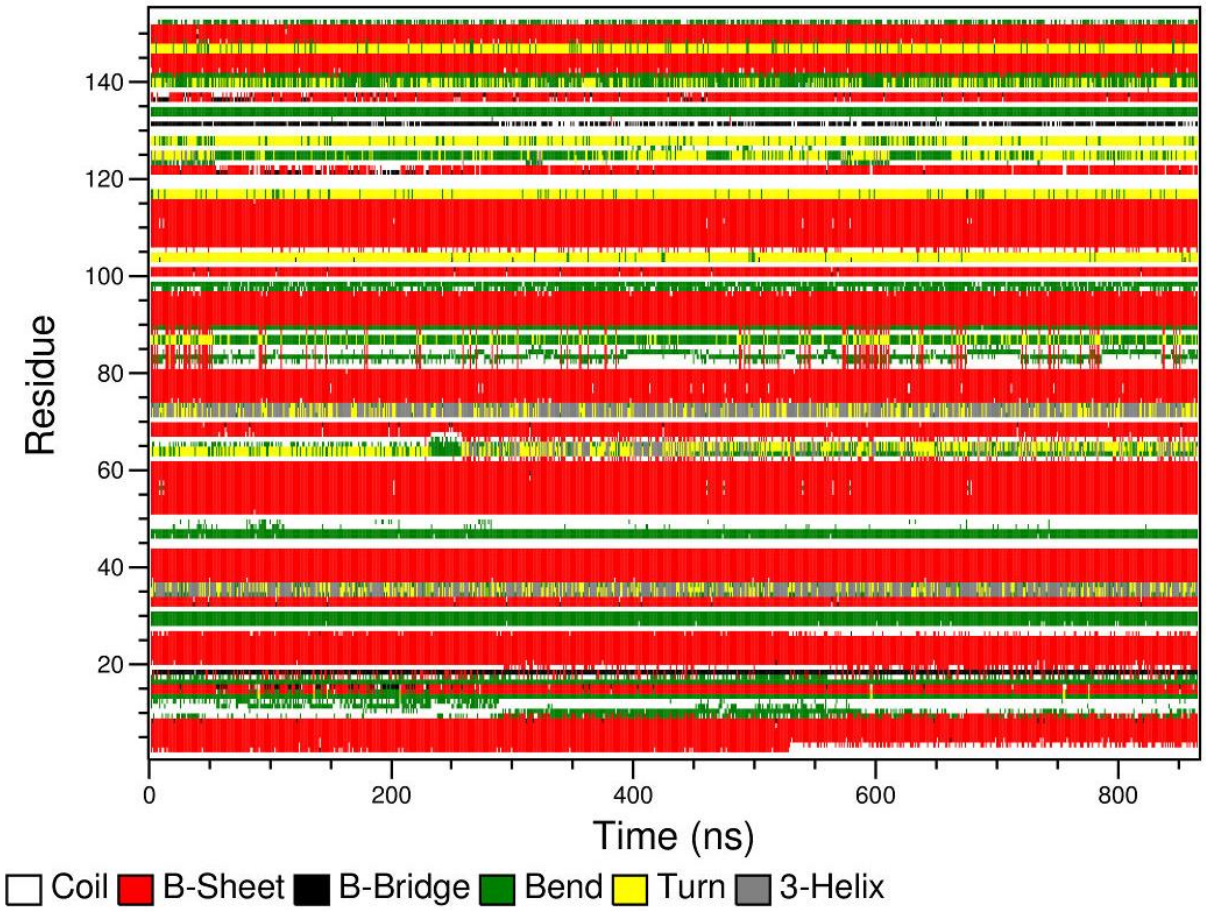


Figure S4. Changes of CBM secondary structure in time in simulation CBM-c3.

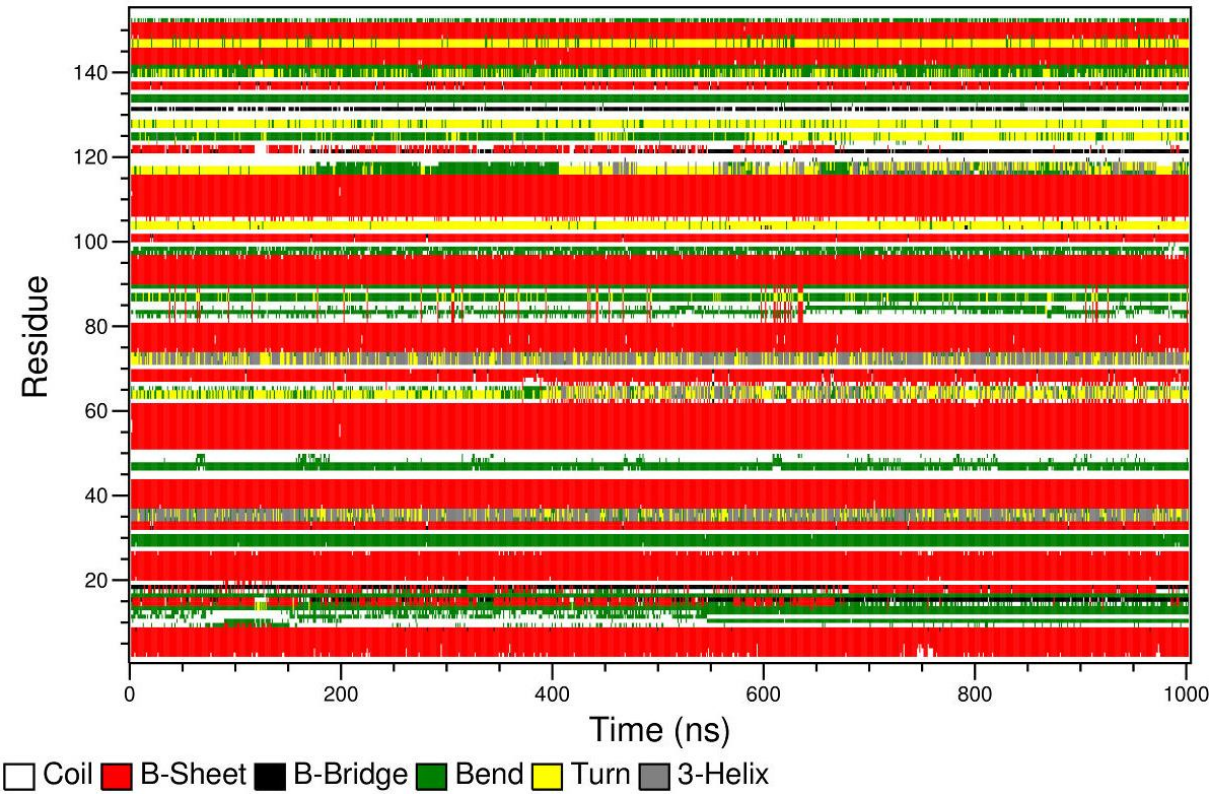


Figure S5. Changes of CBM secondary structure in time in simulation CBM-c4.

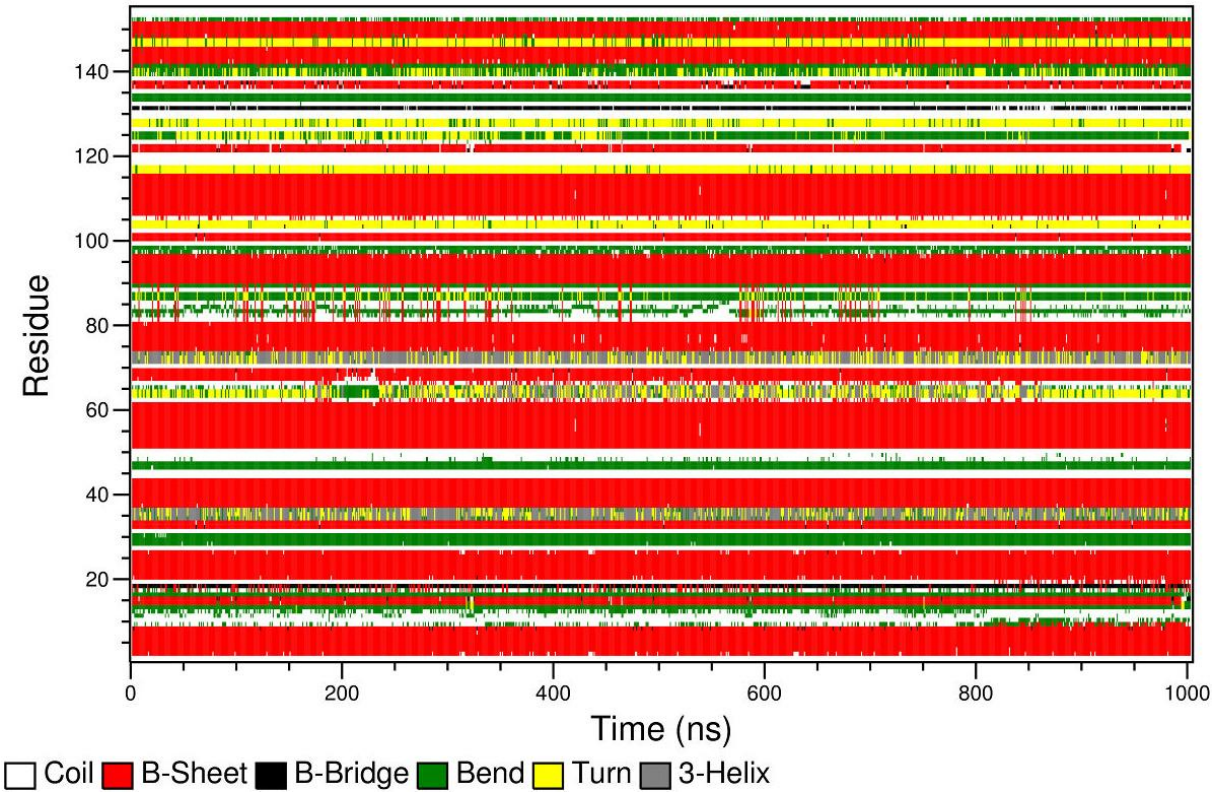


Figure S6. Changes of CBM secondary structure in time in simulation CBM-c5.

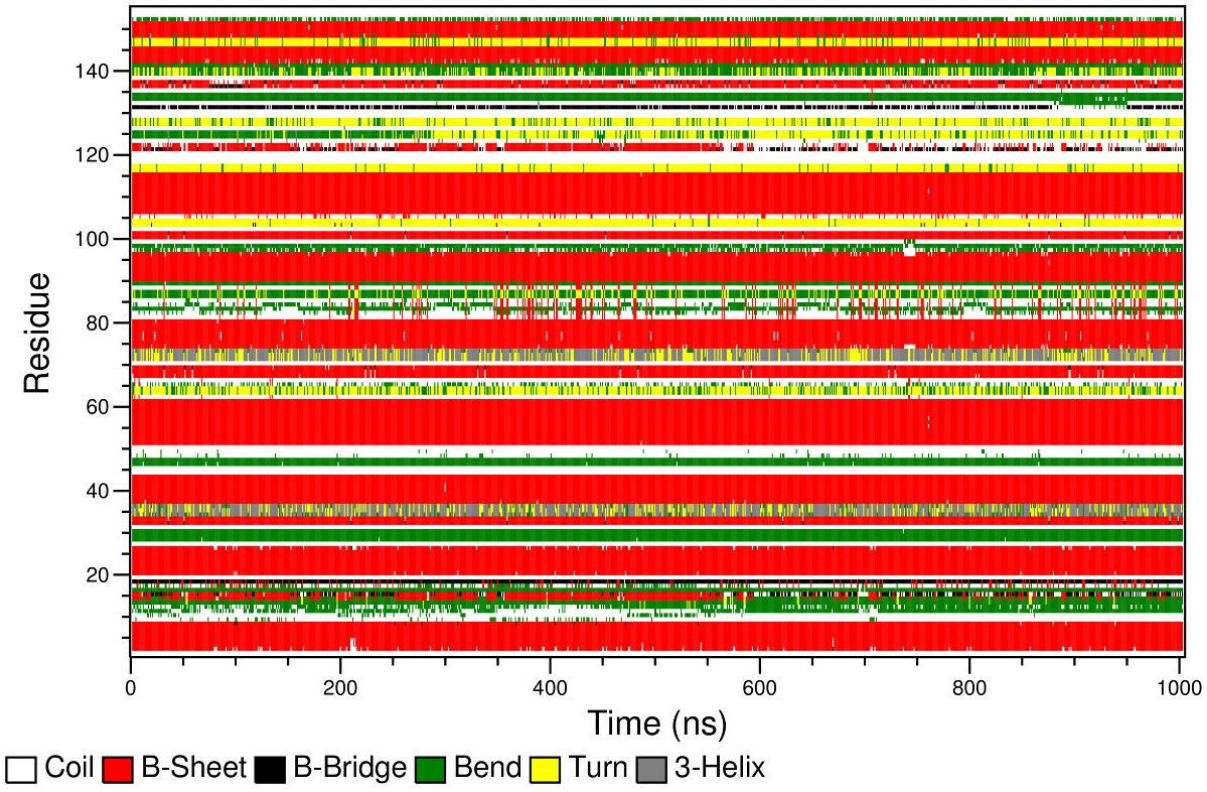


Figure S7. Changes of CBM secondary structure in time in simulation CBM-aI1.

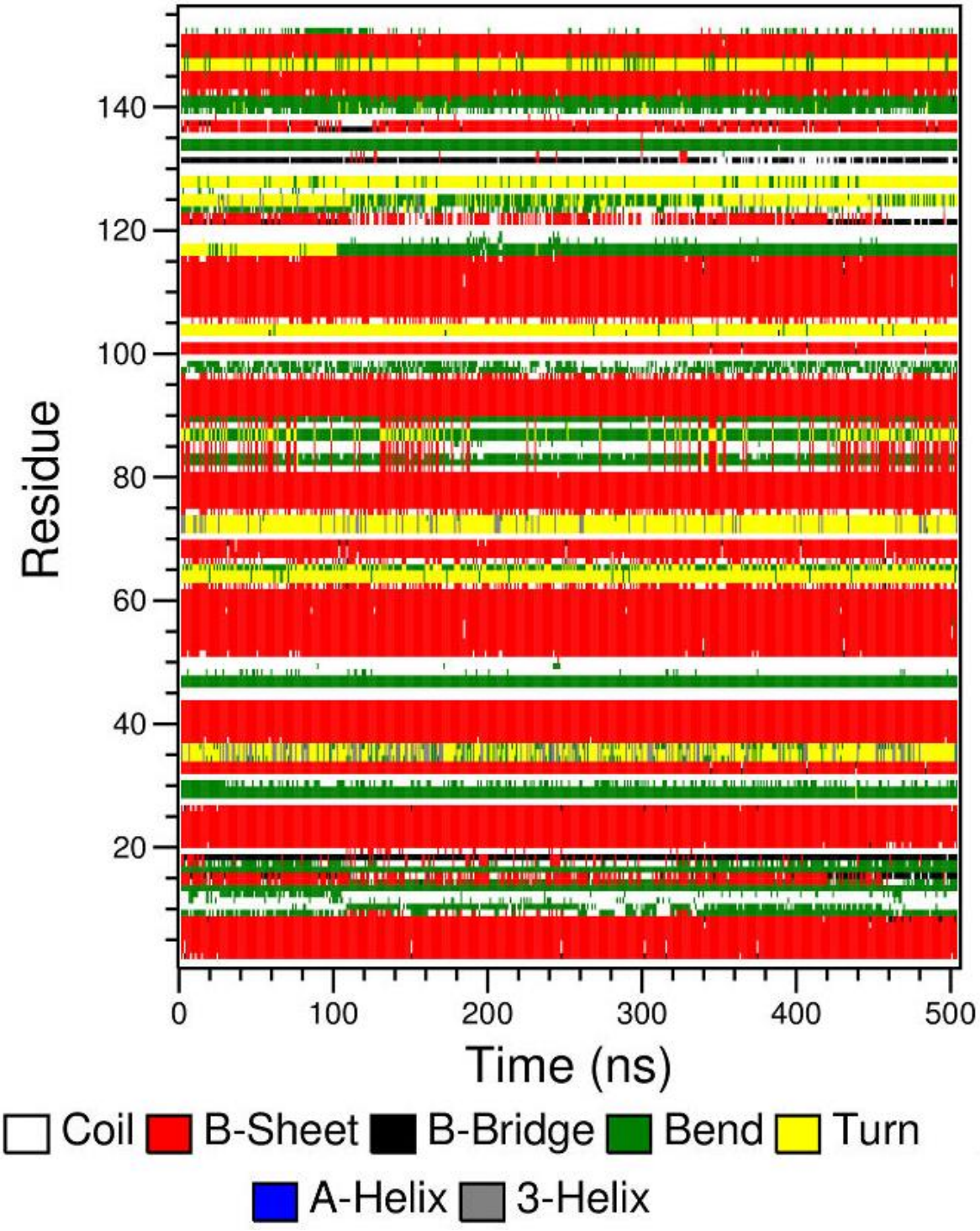


Figure S8. Changes of CBM secondary structure in time in simulation CBM-aI2.

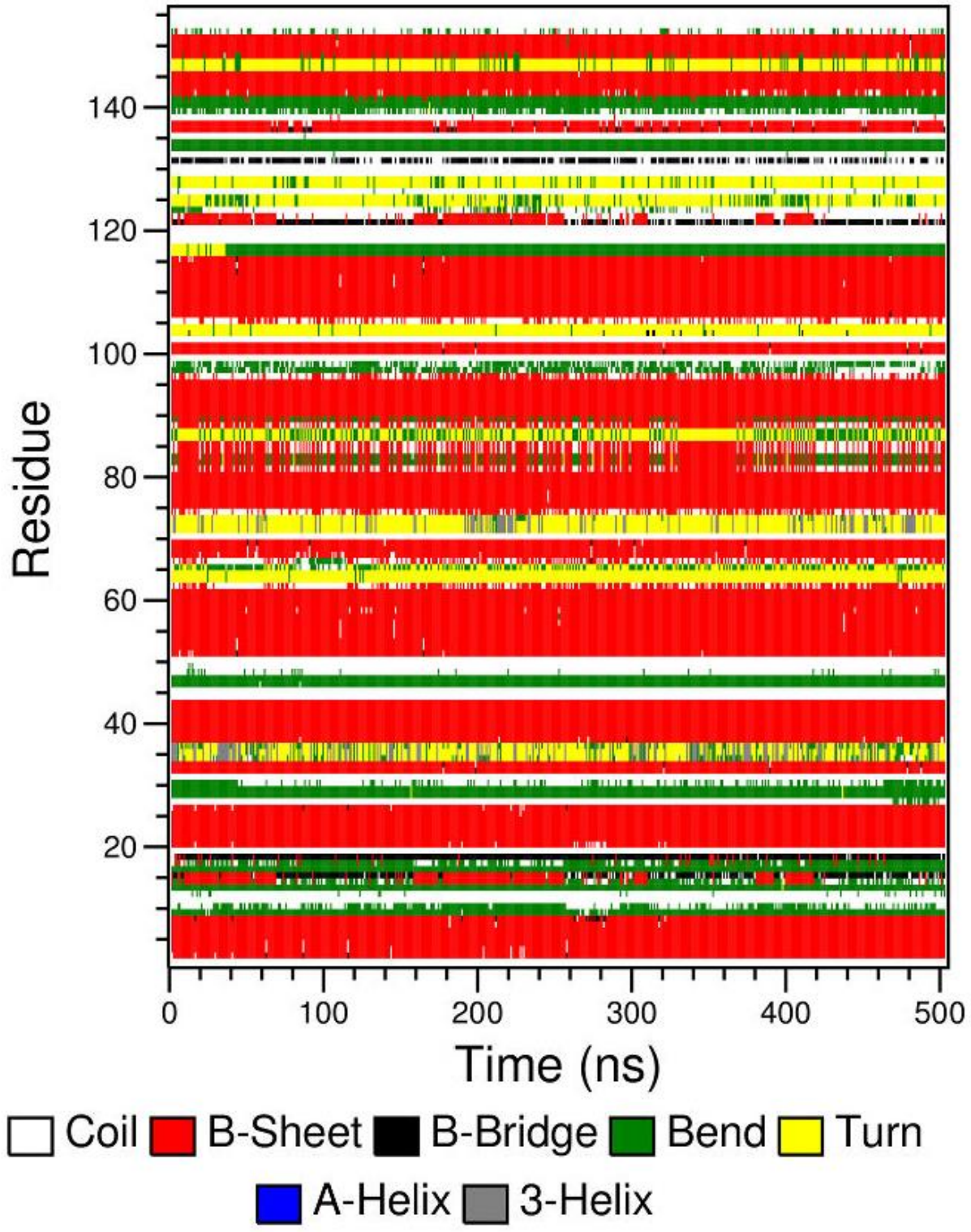


Figure S9. Changes of CBM secondary structure in time in simulation CBM-aII1.

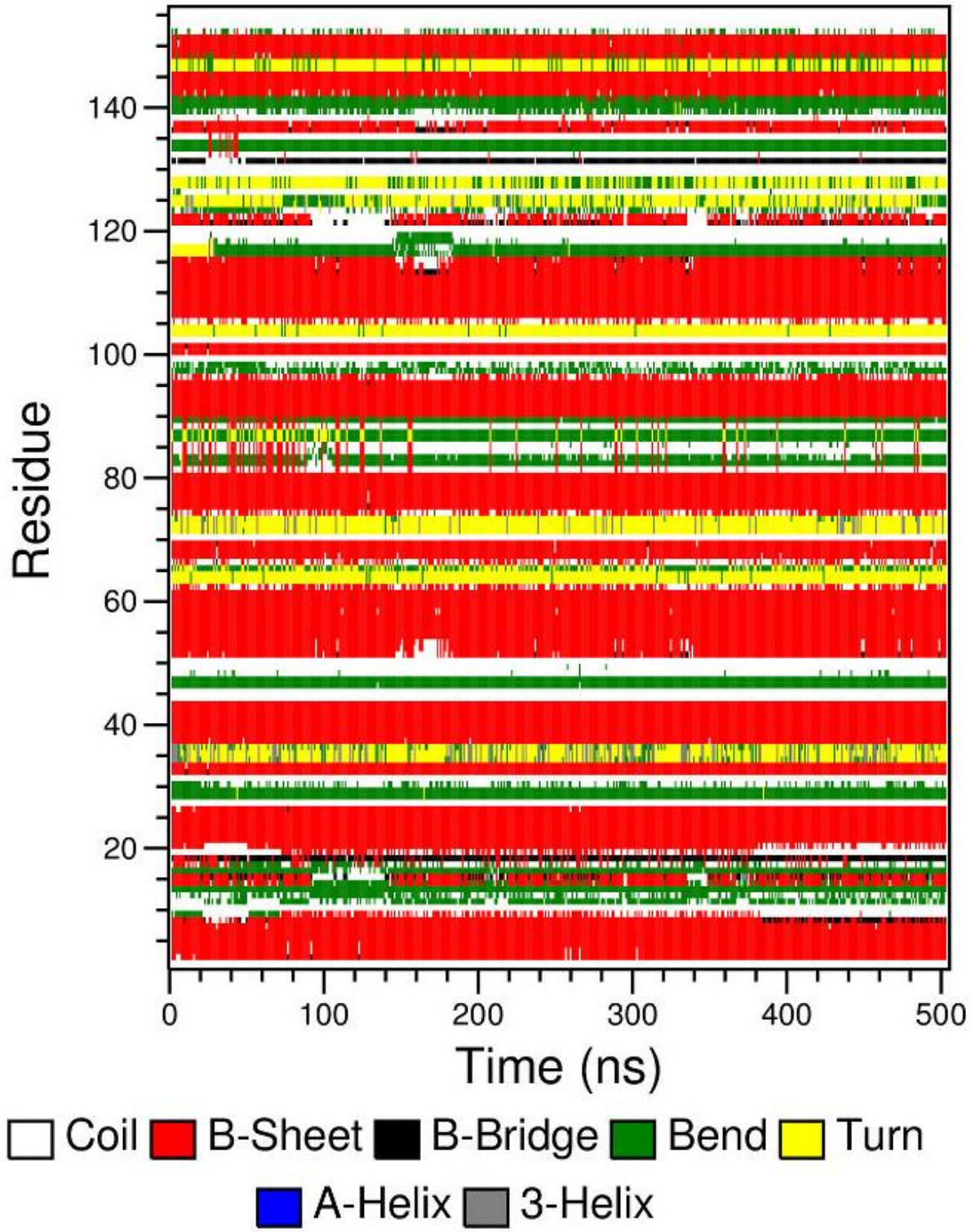


Figure S10. Changes of CBM secondary structure in time in simulation CBM-aII2.

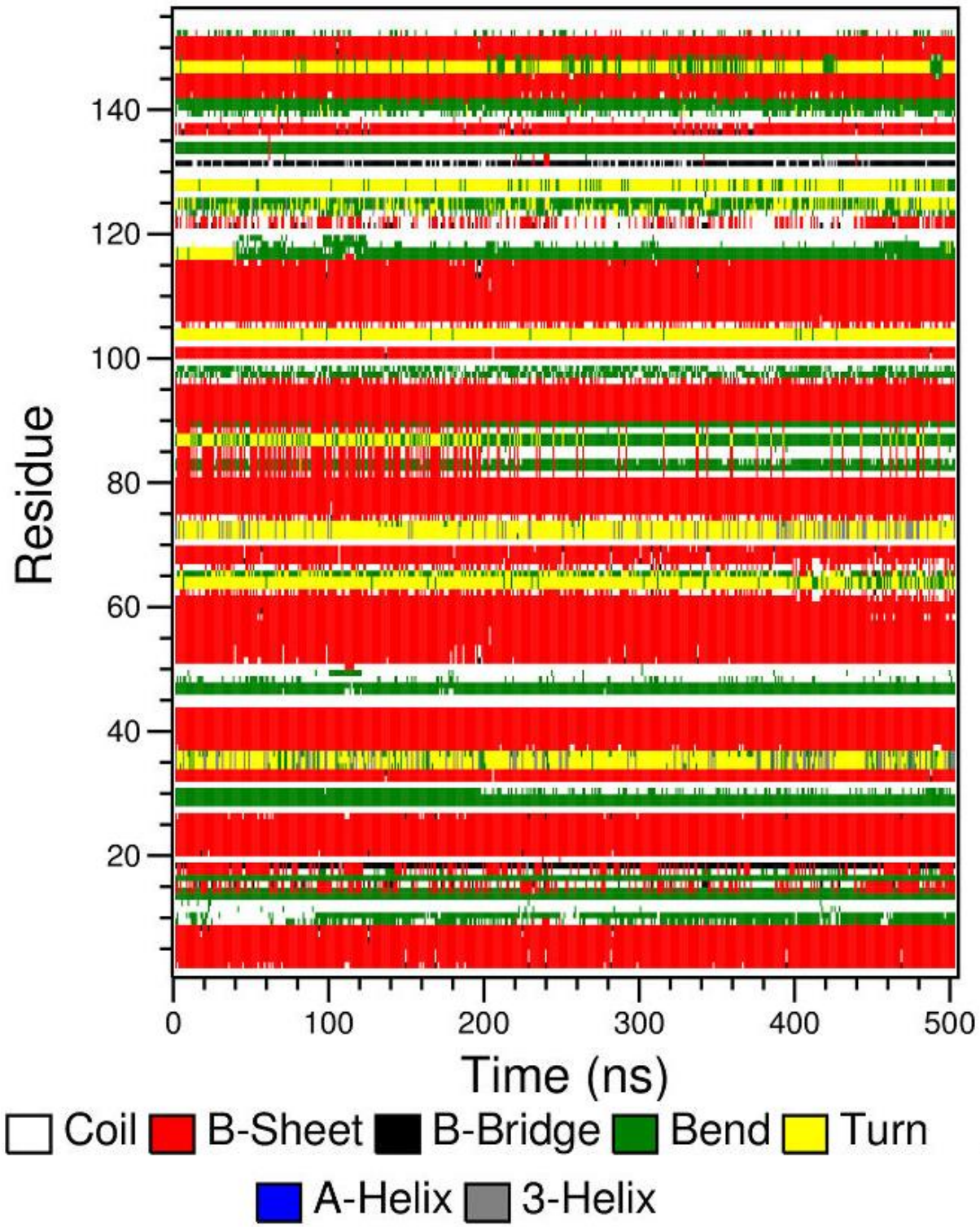


Figure S11. Changes of expansin secondary structure in time in simulation Exp-c1.

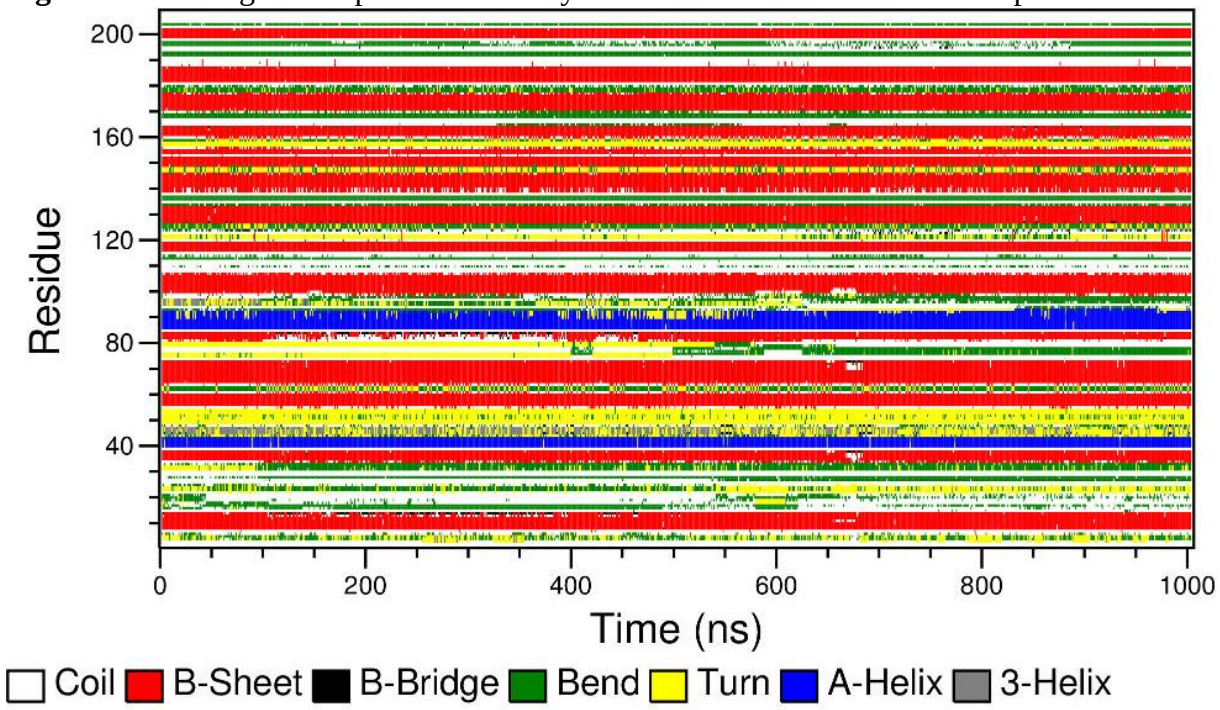


Figure S12. Changes of expansin secondary structure in time in simulation Exp-aI1.

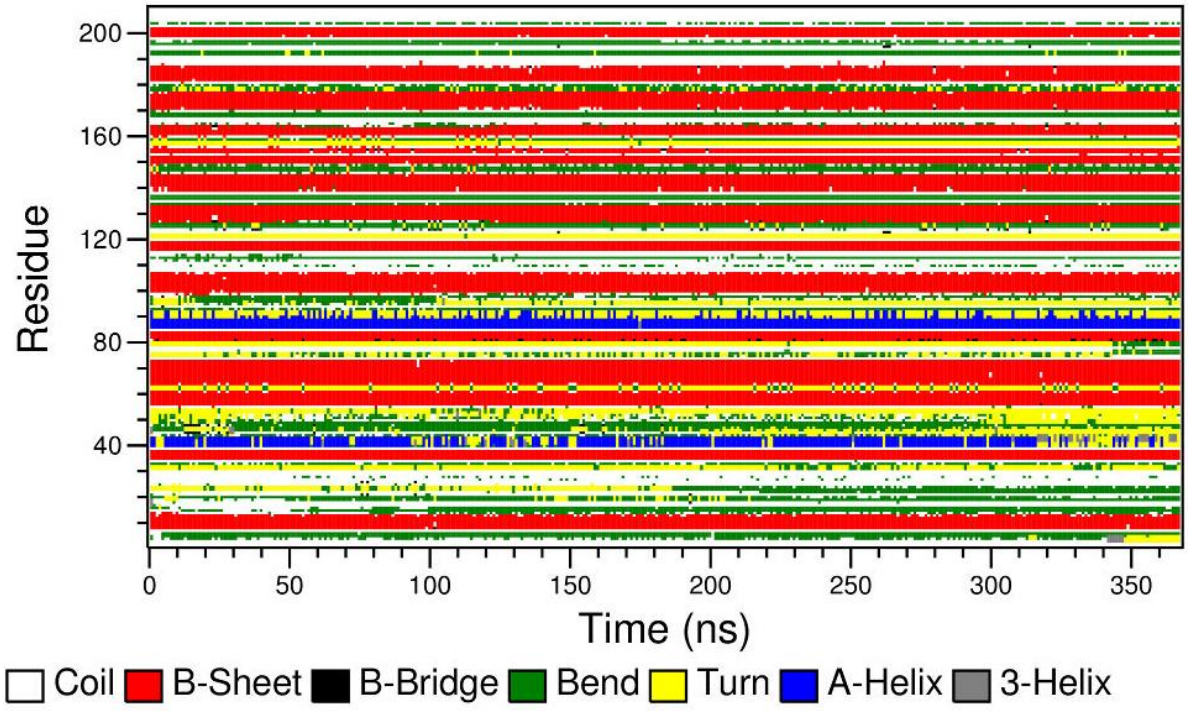


Figure S13. Changes of expansin secondary structure in time in simulation Exp-aI2.

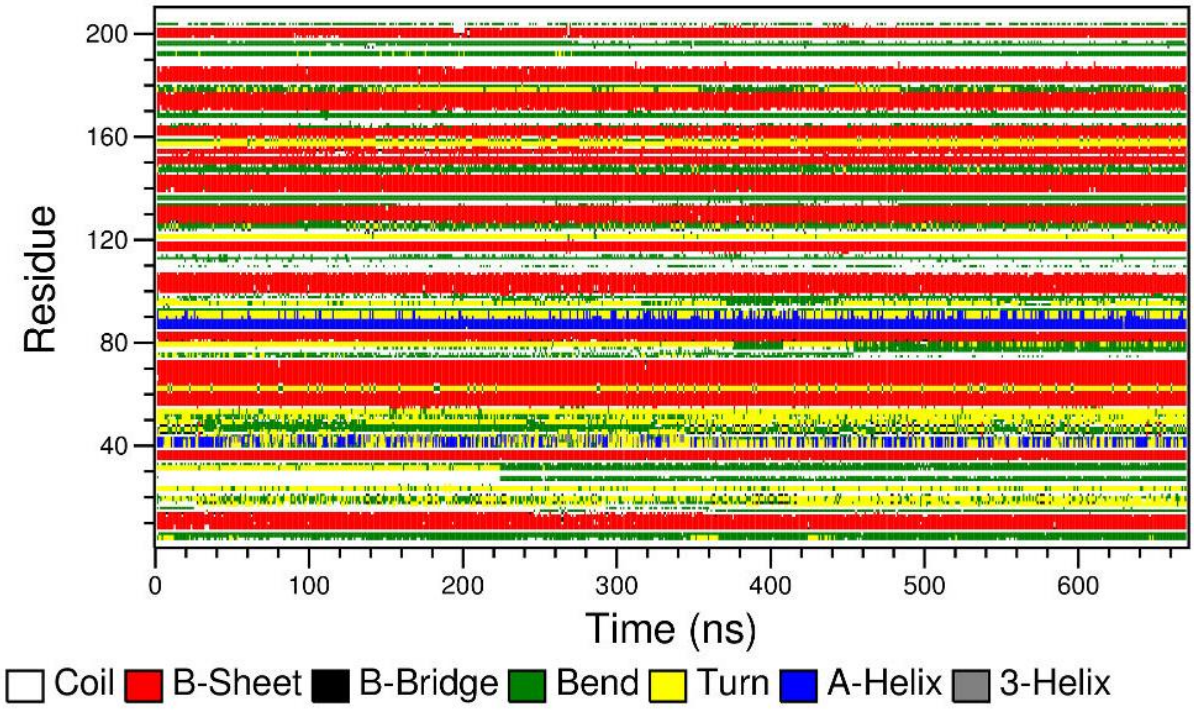


Figure S14. Changes of expansin secondary structure in time in simulation Exp-aI3.

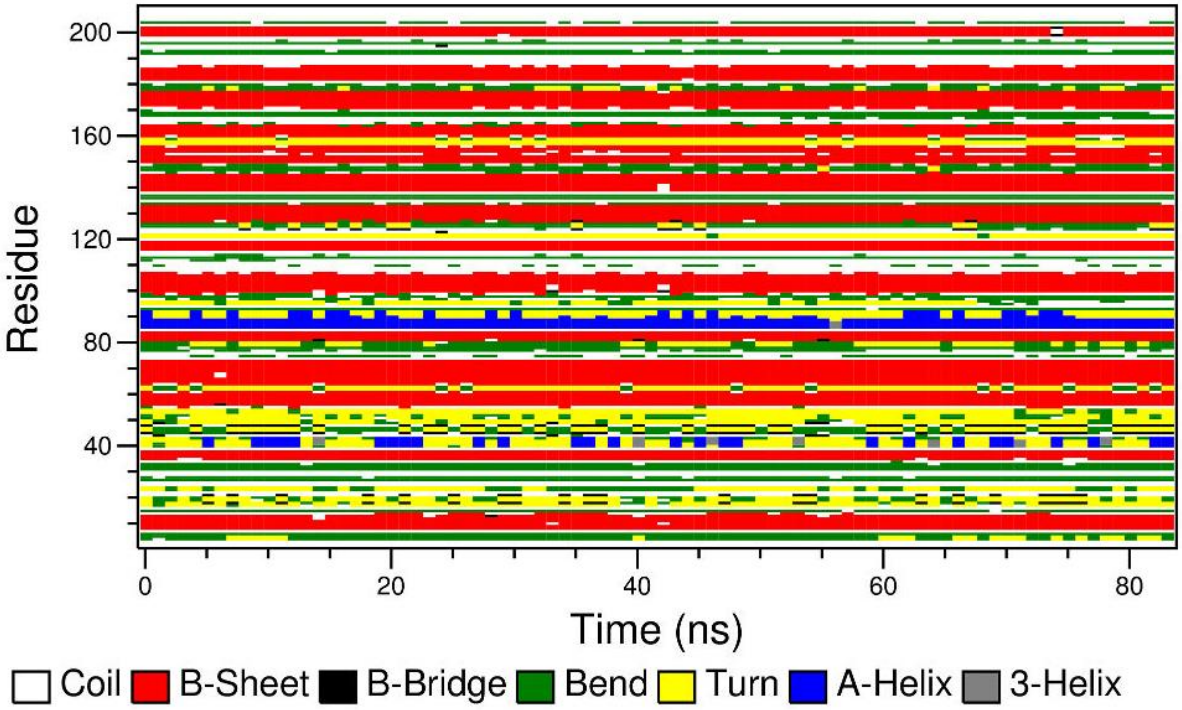


Figure S15. Changes of expansin secondary structure in time in simulation Exp-aI4.

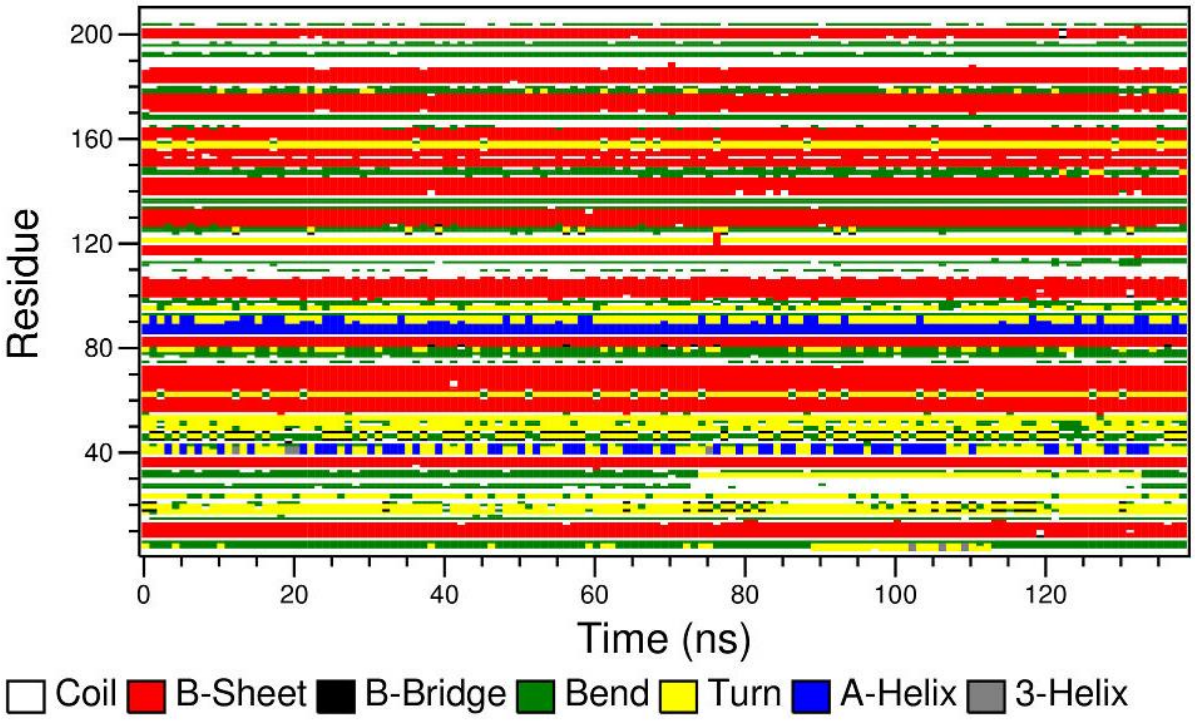


Figure S16. Changes of expansin secondary structure in time in simulation Exp-aII1.

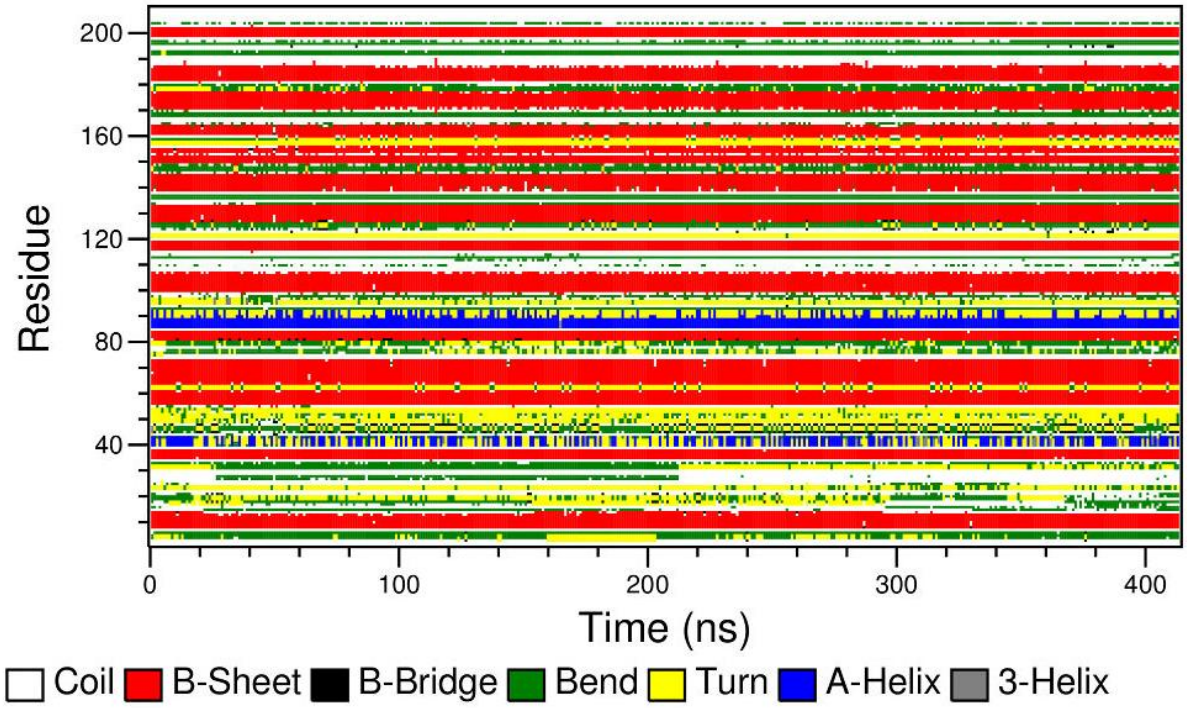


Figure S17. Changes of expansin secondary structure in time in simulation Exp-aII2.

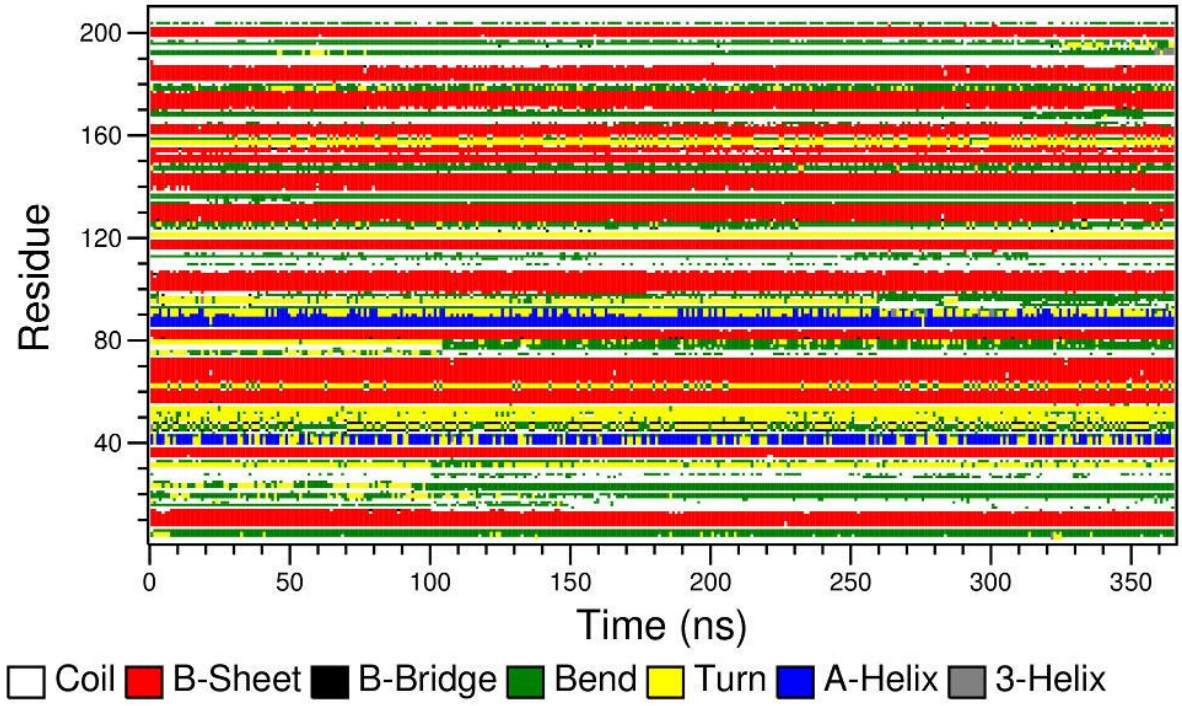


Figure S18. Root mean square deviation (RMSD) of CBM C-alpha atoms in simulations: CBM-c1 (A), CBM-c2 (B), CBM-c3 (C), CBM-c4 (D), CBM-c5 (E), CBM-aI1 (F), CBM-aI2 (G), CBM-aII1 (H), CBM-aII2 (I).

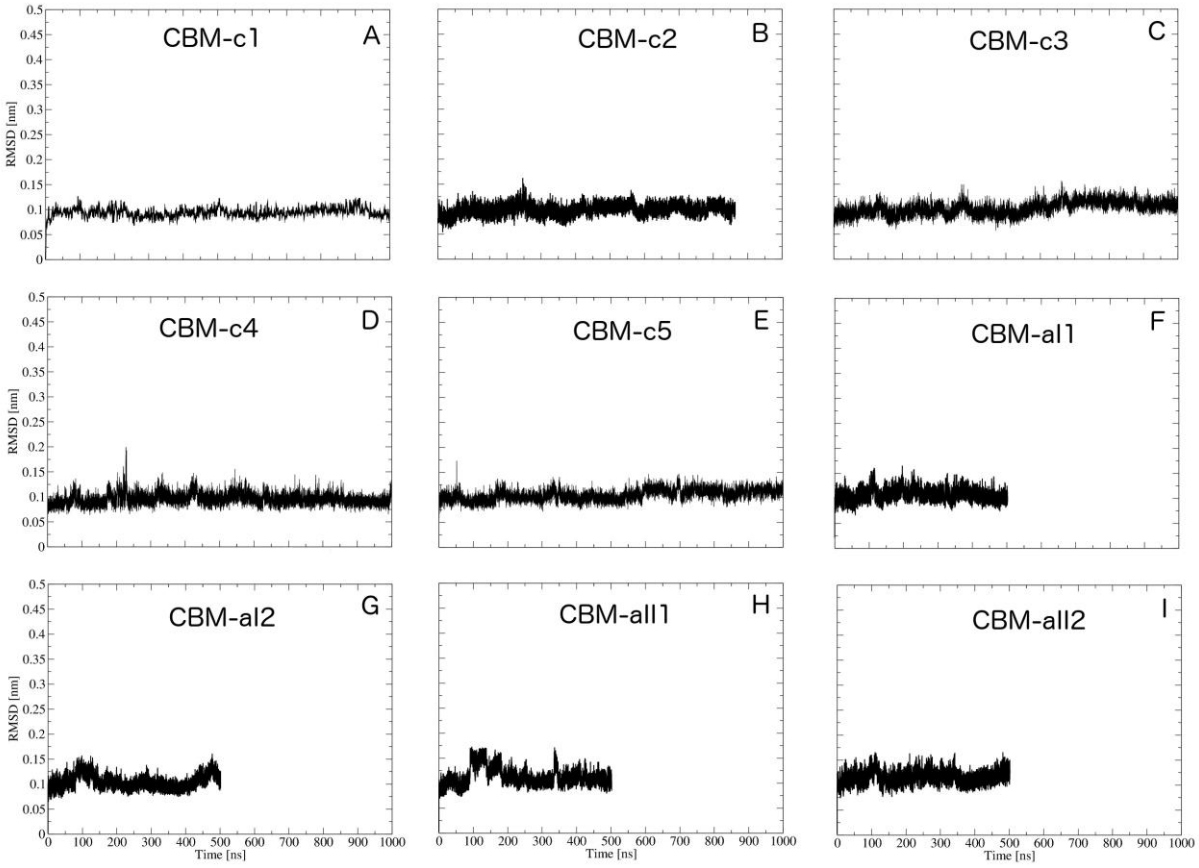


Figure S19. Root mean square deviation (RMSD) of expansin C-alpha atoms in simulations: Exp-c1 (A), Exp-aI1 (B), Exp-aI2 (C), Exp-aI3 (D), Exp-aI4 (E), Exp-aII1 (F), Exp-aII2 (G).

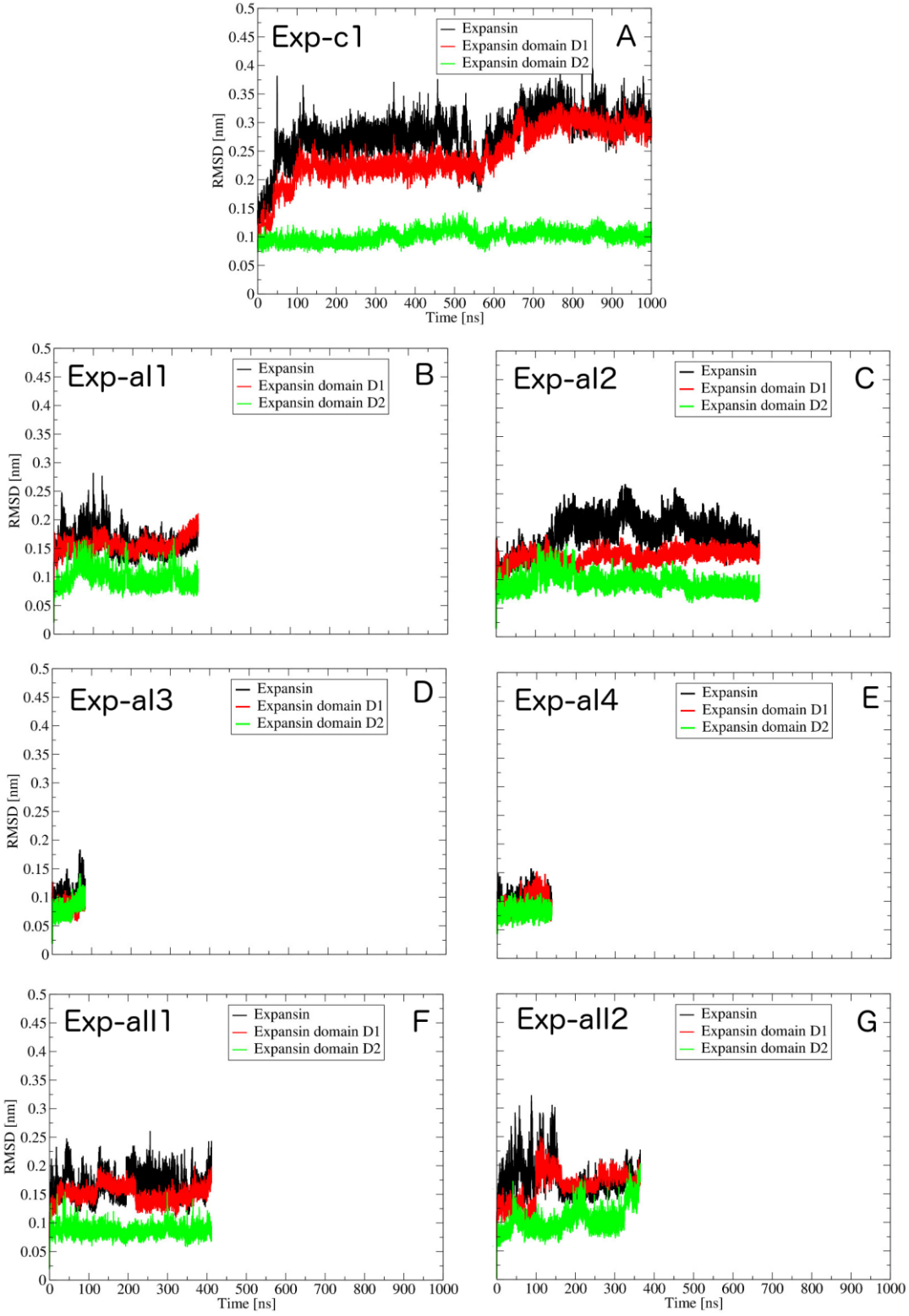


Figure S20. Root mean square fluctuation (RMSF) of C-alpha atoms of CBM in simulations: CBM-c1 (A), CBM-c2 (B), CBM-c3 (C), CBM-c4 (D), CBM-c5 (E), CBM-aI1 (F), CBM-aI2 (G), CBM-aII1 (H), CBM-aII2 (I).

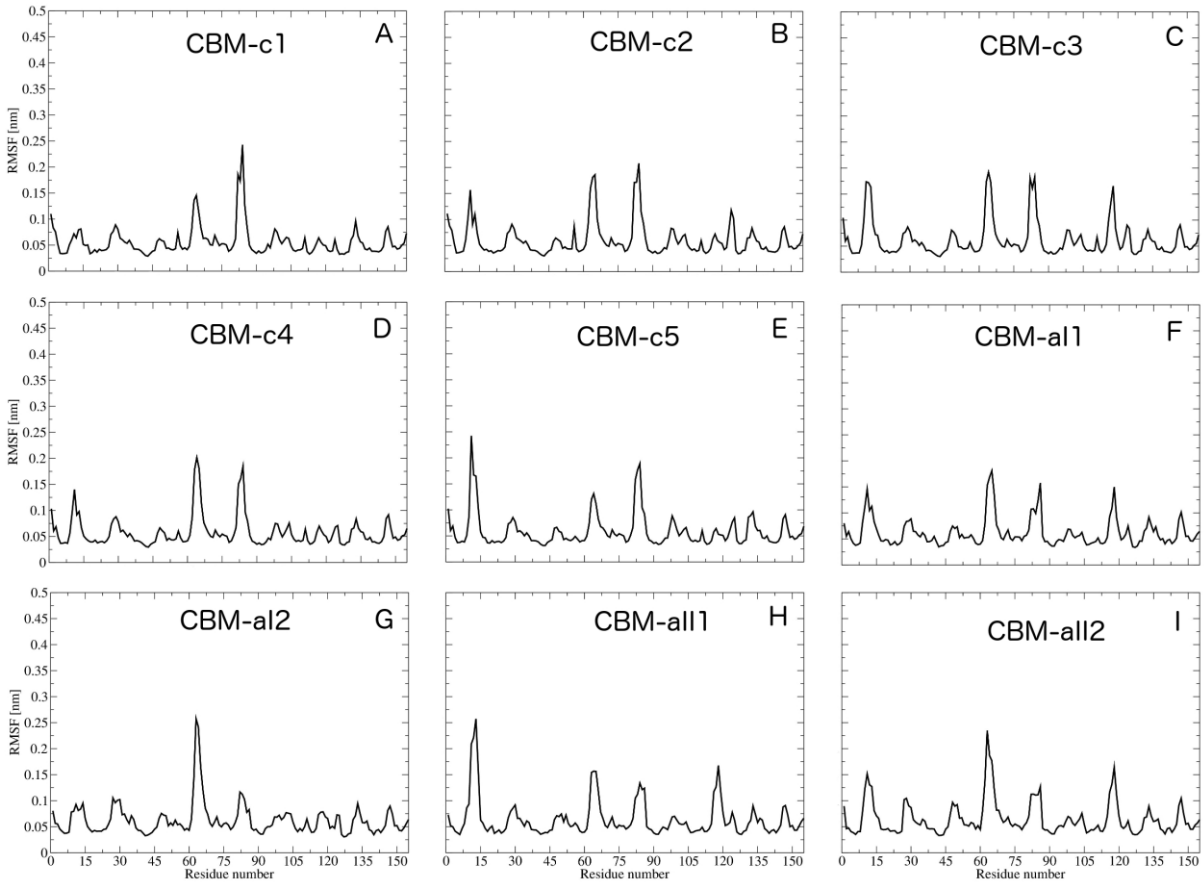


Figure S21. Root mean square fluctuation (RMSF) of expansin C-alpha atoms in simulations: Exp-c1 (A), Exp-aI1 (B), Exp-aI2 (C), Exp-aI3 (D), Exp-aI4 (E), Exp-aII1 (F), Exp-aII2 (G).

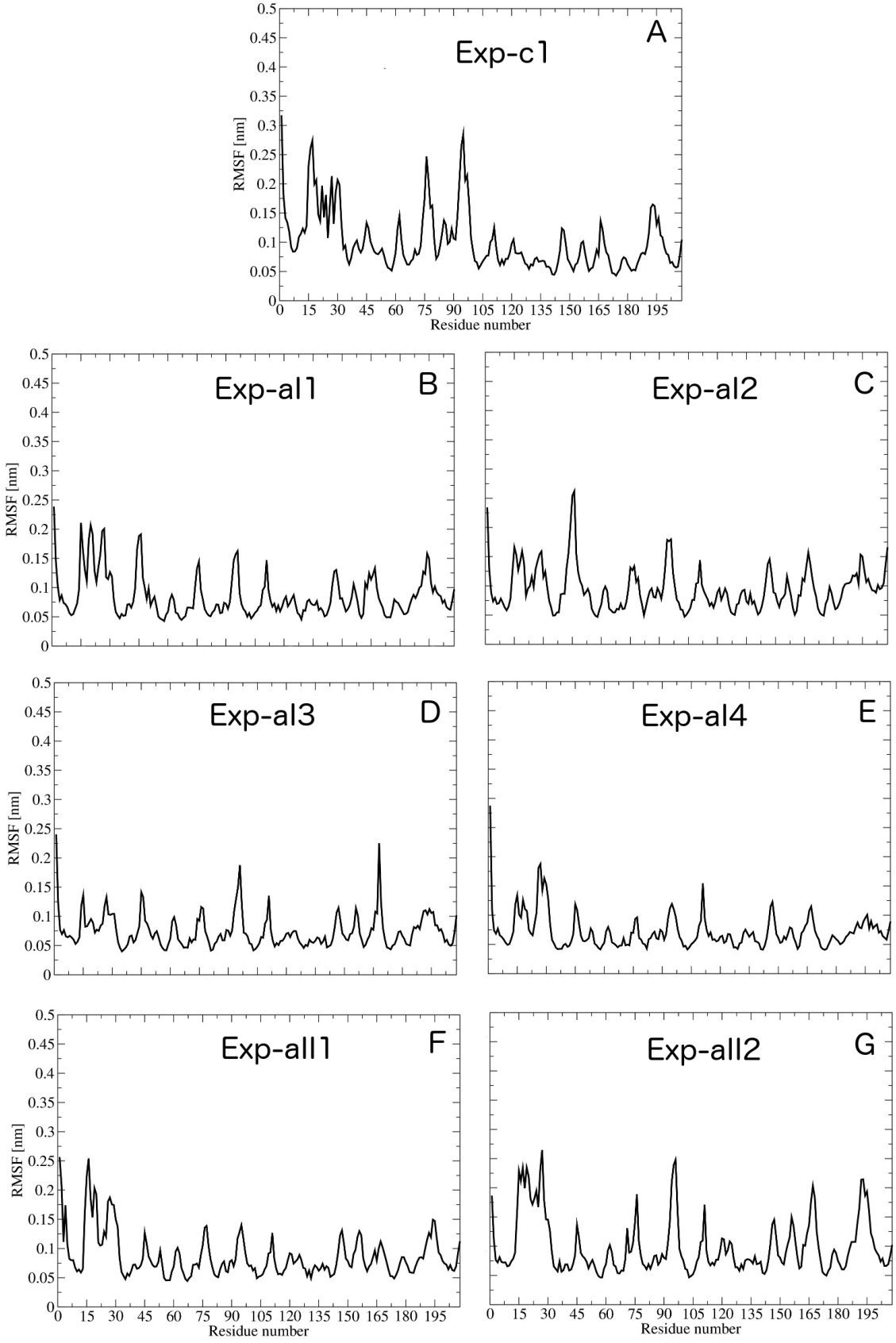


Figure S22. Initial steps of CBM-cellulose binding from simulation CBM-c2 at: 397 ns (A), 398 ns (B), 399 ns (C), 400 ns (D), 406 ns (E), 410 ns (F), 420 ns (G), 425 ns (H). Cellulose is depicted as an orange surface, protein is presented in "New Cartoon" representation and colored accordingly to the secondary structure, water and ions are not shown for clarity.

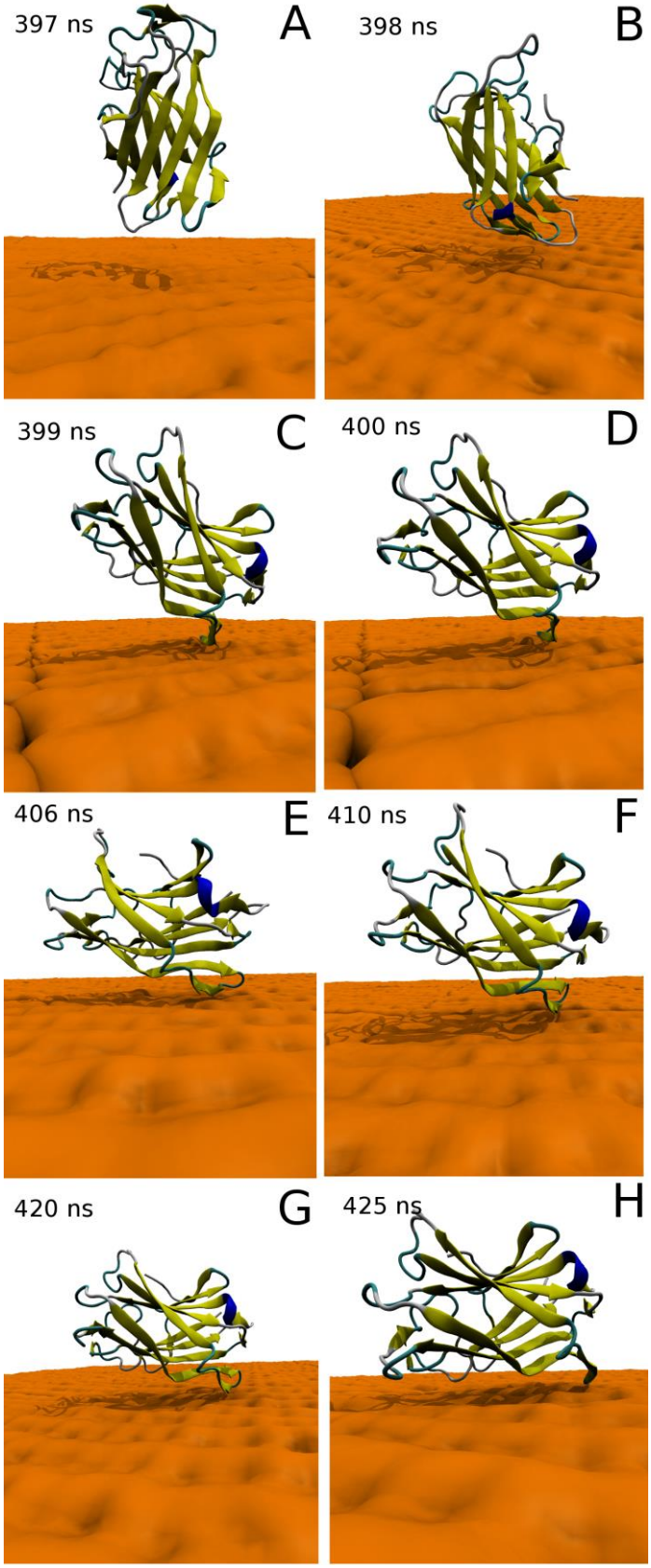


Figure S23. Initial steps of CBM-cellulose binding from simulation CBM-c5 at: 222 ns (A), 226 ns (B), 245 ns (C), 250 ns (D). Cellulose is depicted as an orange surface, protein is presented in "New Cartoon" representation and colored accordingly to the secondary structure, water and ions are not shown for clarity.

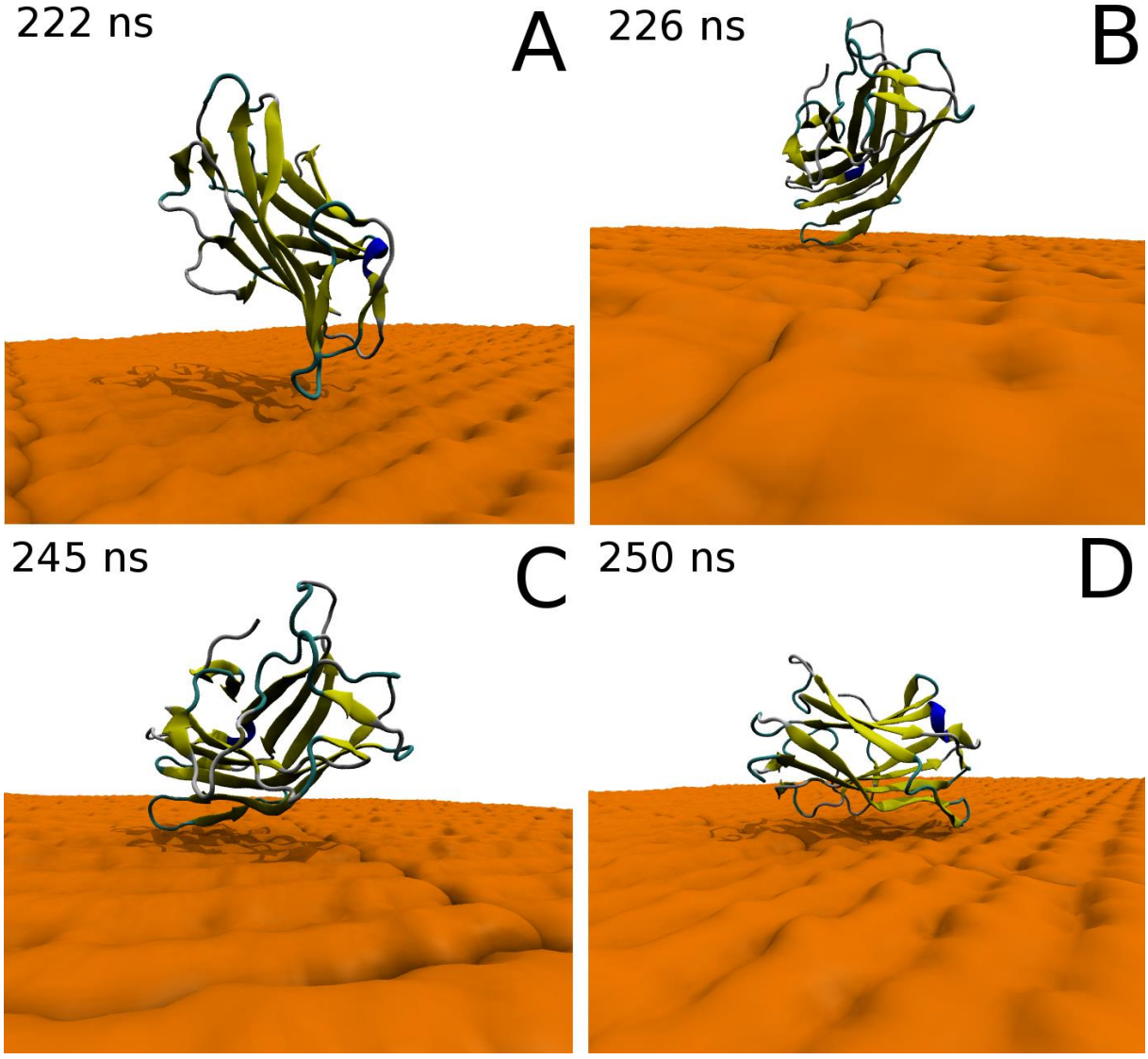


Figure S24. Initial steps of CBM-cellulose binding from simulation CBM-c2 with the depicted residues that make first contact with cellulose: SER29 and ASN64 in simulation CBM-c2 (A), ASN64 in simulation CBM-c5 (B), TYR67 in simulation CBM-c5 (C). Cellulose is depicted as an orange surface, protein is presented in "New Cartoon" representation and colored accordingly to the secondary structure, water and ions are not shown for clarity.

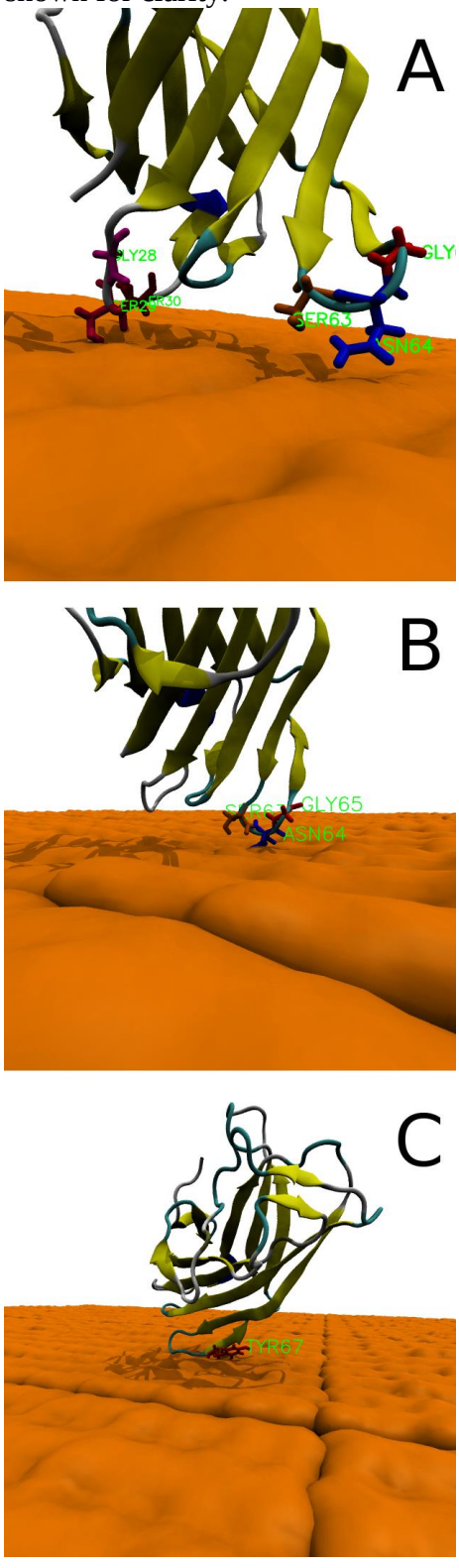


Figure S25. Potential of mean force (PMF) for CBM-cellulose (black line), and expansin-cellulose (red line) binding. The center of the mass distance (COM) corresponds to the protein cellulose distance where 0 nm mean starting (protein bound) distance).

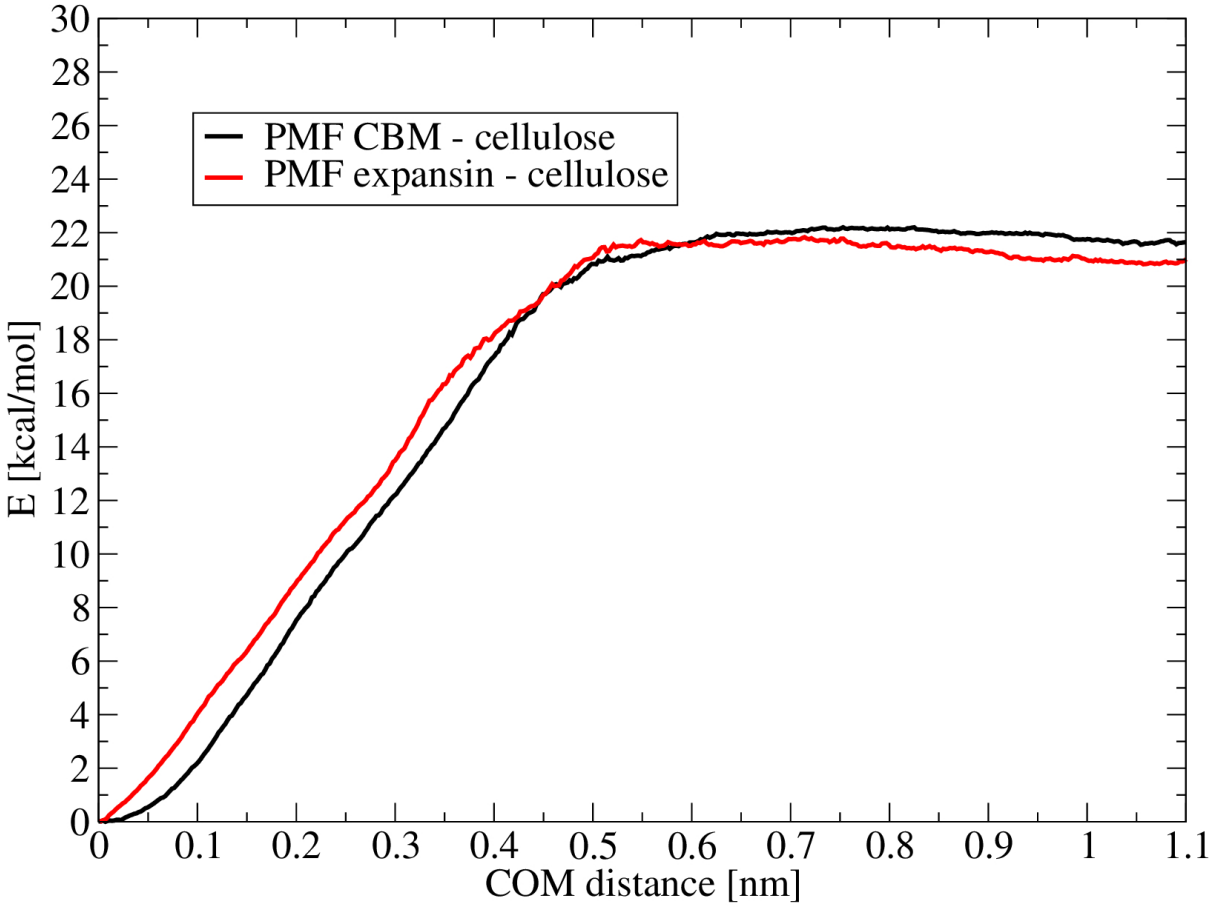


Table S1. Number of molecules and atoms in the simulated systems.

Simulation number - type	Total number of atoms (including water molecules and ions)
1 - PMF-CBM	~ 100 000
2 – PMF-Exp	~ 100 000
3 – CBM-c1	~ 50 000
3 – CBM-c(2-5)	~ 100 000
4 – CBM-aI(1-2)	~ 400 000
5 – CBM-aII(1-2)	~ 400 000
6 – Exp-c1	~ 100 000
7 – Exp-aI(1-4)	~ 400 000
8 – Exp-aII(1-2)	~ 400 000

Table S2. CBM-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation CBM-c1.

Donor residue-atom name	Acceptor residue-atom name	Occupancy
GLN110-NE2	CEL8-O5	34.15%
ASN16-ND2	CEL470-O3	31.59%
GLN110-NE2	CEL8-O3	15.55%
CEL345-O3	ASP56-O	14.39%
ASN16-ND2	CEL445-O2	13.75%
CEL134-C6	SER9-O	9.60%
CEL471-O2	ASP56-OD2	8.14%
HIS57-ND1	CEL345-O2	8.10%
ASN16-ND2	CEL469-O6	7.98%
CEL468-O6	THR14-O	7.48%
CEL446-O6	TRP118-CH2	6.94%
CEL8-C2	GLN110-OE1	6.90%
CEL344-O6	ASP56-O	5.66%
ARG112-NH2	CEL320-O6	5.40%
CEL133-C1	ASN10-OD1	5.36%
CEL133-C5	ASN10-OD1	5.24%

Table S3. CBM-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation CBM-c2.

Donor residue-atom name	Acceptor residue-atom name	Occupancy
CEL229-O3	ASN10-OD1	17.33%
CEL394-O6	ASP56-O	16.81%
TRP118-NE1	CEL374-O6	14.71%
CEL228-O6	ASN10-OD1	9.16%
CEL110-O6	ASN64-O	5.76%

Table S4. CBM-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation CBM-c3.

Donor residue-atom name	Acceptor residue-atom name	Occupancy
CEL302-O6	ASN146-O	30.87%
CEL440-O6	ASN146-O	13.04%
CEL442-O6	LYS80-O	10.48%
CEL303-O3	ASN146-O	7.01%

Table S5. CBM-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation CBM-c4.

Donor residue-atom name	Acceptor residue-atom name	Occupancy
CEL479-O6	ASP46-O	11.14%
GLN48-NE2	CEL352-O2	8.44%
CEL478-C1	THR122-OG1	4.94%

Table S6. CBM-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation CBM-c5.

Donor residue-atom name	Acceptor residue-atom name	Occupancy
ASN16-ND2	CEL159-O2	24.15%
CEL10-O6	ASP56-O	19.75%
TRP118-NE1	CEL135-O2	14.66%
CEL36-O3	HIS57-NE2	5.60%
SER12-OG	CEL80-O3	5.13%
ASN16-ND2	CEL183-O4	4.93%

Table S7. CBM-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation CBM-aI1.

Donor residue-atom name	Acceptor residue-atom name	Occupancy
CEL22-O2	GLU154-O	39.85%
CEL21-O3	PRO155-O2	32.96%
CEL25-O2	ASP139-OD1	27.39%
GLY152-N	CEL25-O3	22.85%
CEL23-O3	ASP139-OD1	20.26%
CEL21-O3	PRO155-C	18.87%
CEL22-O4	GLU154-OE2	16.64%
CEL22-O4	GLU154-CD	16.40%
CEL20-O3	PRO155-O2	15.76%
CEL20-O4	PRO155-O2	15.57%
GLY152-N	CEL26-O6	14.25%
CEL25-O2	GLY152-O	14.17%
SER29-OG	CEL27-O6	13.97%
GLN140-NE2	CEL20-O4	13.65%
SER131-N	CEL20-O3	13.65%
CEL25-O3	ASP139-OD1	12.74%
CEL22-O4	GLU154-OE1	11.94%
ASN86-ND2	CEL18-O3	11.46%
CEL24-O6	ASP139-OD1	11.11%
CEL25-O2	ASP139-CG	9.47%
CEL24-O2	GLU137-OE2	9.43%
CEL24-O6	ASP139-OD2	9.28%
CEL25-O2	ASP139-OD2	9.28%
CEL1-O2	SER29-OG	9.08%
CEL26-O6	GLY152-O	9.04%
CEL21-O3	PRO155-O1	8.92%
CEL24-O3	GLU154-OE2	8.88%
CEL22-O3	PRO155-O2	8.80%
CEL24-O3	GLU154-OE1	8.72%
CEL20-O2	ASN86-OD1	8.28%
CEL26-O6	LEU149-O	7.56%
LYS153-NZ	CEL24-O6	7.36%
CEL24-O3	GLU154-CD	7.32%
CEL25-O3	GLU137-OE1	7.25%
CEL24-O2	GLU137-OE1	7.01%
CEL24-O6	ASP139-CG	6.81%
GLY152-N	CEL25-O1	6.41%
CEL19-O6	ASN86-OD1	6.37%
CEL24-O2	GLU154-OE1	6.25%
GLN140-NE2	CEL20-O3	5.97%
CEL22-O2	PRO155-O2	5.89%
CEL22-O3	PRO155-O1	5.89%
CEL25-O3	GLU137-OE2	5.53%
CEL26-O2	VAL150-O	5.53%
SER29-OG	CEL1-O3	5.45%

Table S8. CBM-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation CBM-aI2.

Donor residue-atom name	Acceptor residue-atom name	Occupancy
ASN16-ND2	CEL1-O1	50.20%
CEL1-O1	ASP117-OD2	46.77%
GLN21-NE2	CEL2-O6	46.53%
CEL1-O1	ASP117-OD1	36.97%
ASN10-ND2	CEL5-O2	33.90%
CEL1-O3	CYS55-O	32.59%
CEL1-O2	ASP117-OD2	32.27%
CEL1-O6	TRP118-O	32.03%
THR15-OG1	CEL5-O2	31.95%
CEL2-O2	ASP56-O	29.48%
ASN120-ND2	CEL1-O1	25.58%
CEL1-O2	ASP117-CG	24.90%
CEL1-O1	ASP117-CG	24.78%
CEL5-O2	THR14-O	22.63%
CEL1-O2	ASP117-OD1	19.80%
TRP118-N	CEL1-O1	19.36%
GLN110-NE2	CEL2-O5	15.74%
GLN110-NE2	CEL2-C1	15.46%
ARG112-NH1	CEL3-O2	15.34%
GLN108-NE2	CEL1-O2	14.06%
CEL22-O3	ASN64-O	13.39%
SER119-OG	CEL1-O1	13.11%
GLN110-NE2	CEL2-O1	12.11%
CEL1-O6	ASN120-ND2	11.87%
GLN108-NE2	CEL1-O1	10.60%
ASN120-ND2	CEL1-C1	10.48%
CEL25-O2	THR97-OG1	10.44%
CEL3-O2	SER9-O	9.52%
CEL24-O3	THR97-OG1	9.32%
GLY99-N	CEL23-O3	8.84%
ASN16-ND2	CEL6-O1	8.76%
CEL2-O6	ASP56-OD2	8.37%
ASN73-ND2	CEL25-O2	7.29%
ASN16-ND2	CEL6-O2	7.09%
TYR67-N	CEL22-O4	7.05%
CEL3-O3	SER9-O	6.81%
CEL22-O3	GLU102-OE2	6.77%
ASN64-ND2	CEL22-O2	6.49%
CEL22-O3	GLU102-OE1	6.49%
ASN64-ND2	CEL22-O3	6.33%
CEL22-O2	GLU102-OE1	6.29%
CEL1-O2	SER9-O	5.86%
TYR7-OH	CEL2-O6	5.86%
HIS57-NE2	CEL2-O6	5.78%

CEL1-O2	GLN21-OE1	5.70%
CEL23-O2	ASP33-OD2	5.62%
CEL22-O2	GLU102-OE2	5.54%
CEL2-O6	TRP54-NE1	5.38%
CEL1-O6	THR71-OG1	5.30%
THR71-OG1	CEL1-O1	5.14%
ASN19-ND2	CEL4-O6	5.10%

Table S9. CBM-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation CBM-aII1.

Donor residue-atom name	Acceptor residue-atom name	Occupancy
CEL19-O2	SER29-OG	22.61%
GLU137-N	CEL21-O6	20.20%
SER29-OG	CEL19-O2	19.98%
LYS153-NZ	CEL23-O3	18.32%
CEL24-O2	GLN140-OE1	15.55%
CEL23-O2	ASP139-OD1	11.49%
SER29-OG	CEL22-O6	11.01%
CEL25-O2	PRO155-O1	10.85%
CEL22-O2	GLU137-O	10.71%
SER29-OG	CEL21-O2	9.50%
CEL23-O3	GLN140-OE1	9.12%
CEL21-O6	GLU137-OE2	8.40%
CEL25-O2	PRO155-O2	6.69%
SER29-OG	CEL20-O6	6.45%
CEL20-O6	SER29-OG	6.17%
ASN1-N	CEL22-O6	5.42%
CEL19-O2	SER29-CB	5.32%
CEL23-O2	ASP139-OD2	5.28%
CEL21-O2	GLY28-O	5.06%

Table S10. CBM-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation CBM-aII2.

Donor residue-atom name	Acceptor residue-atom name	Occupancy
CEL25-O2	GLY152-O	44.02%
ASN1-ND2	CEL23-O5	24.70%
LEU149-N	CEL26-O6	22.37%
GLY152-N	CEL24-O6	22.19%
CEL25-O3	LEU149-O	9.90%
ASN1-ND2	CEL23-C1	9.56%
ASN1-ND2	CEL23-O6	9.26%
ASN1-ND2	CEL23-O2	8.31%
CEL24-O3	ASP139-OD2	8.03%
CEL24-O2	ASP139-OD2	7.97%
ASN1-ND2	CEL23-O3	7.79%
CEL23-O6	ASN1-OD1	7.77%
SER30-OG	CEL22-O6	6.97%
CEL22-O2	SER29-OG	6.95%
GLU154-N	CEL26-O6	6.79%
CEL21-O2	SER29-OG	6.77%
LEU149-N	CEL25-O6	6.21%
GLU154-N	CEL26-O2	6.07%
CEL26-O3	GLU154-OE2	5.04%

Table S11. Expansin-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation Exp-c1.

Donor residue-atom name	Acceptor residue-atom name	Occupancy
GLY193-N	CEL11-O2	38.19%
TYR157-OH	CEL81-O2	23.56%
CEL82-O6	ASP156-OD2	17.71%
CEL82-O6	ASP156-OD1	16.42%
ASN164-ND2	CEL37-O2	15.98%
CEL10-O6	GLY193-O	15.35%
CEL33-O2	GLU120-OE1	9.93%
CEL33-O2	GLU120-OE2	9.29%
SER195-OG	CEL34-O3	7.66%
SER195-OG	CEL9-O2	7.33%
CEL35-C3	GLY121-O	6.84%
CEL82-O6	ASP156-CG	5.45%
CEL106-O6	GLU75-OE2	5.24%
THR163-CG2	CEL61-O4	5.05%
CEL106-O6	GLU75-OE1	4.98%

Table S12. Expansin-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation Exp-aI1.

Donor residue-atom name	Acceptor residue-atom name	Occupancy
CEL18-O6	THR14-O	20.81%
CEL20-O2	TYR13-OH	12.65%
CEL15-O6	GLU75-OE2	7.17%
CEL13-O6	GLU120-OE2	7.01%
CEL13-O6	GLU120-OE1	6.38%
CEL19-O3	ASP96-OD1	5.40%
CEL19-O3	ASP96-OD2	5.37%
CEL17-O3	THR14-O	5.07%
CEL15-O6	GLU75-OE1	4.94%
CEL16-O2	GLY20-O	4.94%

Table S13. Expansin-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation Exp-aI2.

Donor residue-atom name	Acceptor residue-atom name	Occupancy
CEL5-O2	SER16-OG	29.79%
CEL4-O6	THR14-O	25.93%
LYS119-NZ	CEL9-O2	24.92%
THR14-CB	CEL4-O5	11.52%
SER19-OG	CEL6-O2	10.08%
CEL16-O6	SER195-OG	9.48%
GLY193-N	CEL12-O2	8.64%
CEL12-O6	GLU191-O	7.63%
LYS119-NZ	CEL8-O6	6.73%
CEL12-O2	SER192-O	6.34%
CEL9-O6	GLU120-OE2	5.32%
CEL9-O6	GLU120-OE1	5.20%

Table S14. Expansin-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation Exp-aI3.

Donor residue-atom name	Acceptor residue-atom name	Occupancy
CEL13-O2	SER195-OG	32.41%
SER192-OG	CEL9-O6	15.61%
SER195-OG	CEL7-O6	12.36%
CEL8-O6	GLY121-O	12.36%
CEL6-O2	GLU120-OE1	10.32%
LYS119-NZ	CEL6-O6	9.96%
CEL3-O6	SER19-O	8.64%
CEL9-O6	GLY193-O	8.52%
CEL4-O3	ASP156-OD2	7.92%
CEL4-O6	ALA22-O	7.68%
CEL3-O6	GLY20-O	6.72%
GLY193-N	CEL9-O6	5.76%
SER195-OG	CEL13-O2	5.64%
CEL8-O2	GLY193-O	5.52%

Table S15. Expansin-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation Exp-aI4.

Donor residue-atom name	Acceptor residue-atom name	Occupancy
GLY193-N	CEL18-O2	72.87%
LYS119-NZ	CEL21-O2	52.75%
CEL20-O3	ASP156-OD1	50.43%
CEL20-O3	ASP156-OD2	41.68%
CEL20-O3	ASP156-CG	32.85%
CEL24-O6	GLY20-O	24.10%
CEL26-O2	THR14-O	18.81%
GLY17-N	CEL24-O6	18.67%
CEL18-O2	THR163-OG1	15.70%
CEL17-O6	GLU191-O	12.30%
CEL24-O6	SER19-O	11.07%
ASN164-ND2	CEL17-O6	10.64%
GLY193-N	CEL18-C2	9.41%
CEL17-O6	ASN164-OD1	9.04%
CEL29-O6	MET94-O	8.97%
CEL18-O3	GLY193-O	7.96%
CEL22-C3	ALA22-O	6.87%
CEL16-O6	TRP125-CE2	6.80%
CEL20-O4	ASP156-OD2	5.07%
CEL20-O4	ASP156-OD1	4.99%
CEL24-O6	GLY20-C	4.92%

Table S16. Expansin-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation Exp-aII1.

Donor residue-atom name	Acceptor residue-atom name	Occupancy
TYR157-OH	CEL23-O6	32.95%
CEL19-O2	GLU191-OE2	26.69%
LYS48-NZ	CEL25-O6	21.86%
SER192-OG	CEL20-O2	17.69%
GLY193-N	CEL19-O2	16.67%
CEL19-O3	GLY193-O	15.82%
CEL21-O2	TRP126-NE1	15.70%
CEL19-O2	GLU191-OE1	15.57%
TYR157-OH	CEL22-O6	14.00%
LYS48-NZ	CEL24-O6	10.16%
SER192-OG	CEL19-O6	9.58%
CEL26-O6	GLU75-OE1	7.91%
CEL26-O3	GLU75-O	7.40%
CEL26-O2	GLU75-OE1	6.40%
CEL26-O3	GLU75-OE1	6.33%
THR194-OG1	CEL19-O3	6.19%
CEL26-O6	GLU75-OE2	5.29%
ARG78-NH2	CEL27-O6	5.14%
GLY193-N	CEL20-O2	5.00%
CEL24-O6	ASP156-OD2	5.00%
CEL24-O6	ASP156-OD1	4.97%

Table S17. Expansin-cellulose hydrogen bonds occurrence in time (in percentage of frames where the interaction is present) in simulation Exp-aII2.

Donor residue-atom name	Acceptor residue-atom name	Occupancy
ARG78-NH1	CEL24-O3	42.11%
CEL27-O2	GLU153-OE2	35.01%
CEL25-O6	ALA40-O	29.55%
LYS48-NZ	CEL27-O6	26.40%
CEL26-O2	GLU153-OE1	22.67%
CEL27-O3	GLU153-OE2	20.97%
ASN43-ND2	CEL26-O2	19.49%
CEL27-O2	GLU153-OE1	18.97%
TRP126-NE1	CEL29-O6	16.06%
CEL26-O3	GLU153-OE1	15.21%
CEL27-O3	GLU153-CD	14.99%
CEL27-O3	GLU153-OE1	13.32%
CEL26-O2	GLU153-CD	12.31%
ARG78-NH1	CEL24-O6	12.17%
ARG78-N	CEL25-O6	10.58%
CEL26-O2	GLU153-OE2	10.39%
CEL26-O3	GLU75-OE1	10.03%
ASN151-ND2	CEL27-O2	8.55%
CEL27-O2	GLU153-CD	8.50%
MET1-N	CEL22-O6	8.31%
ASN151-ND2	CEL26-O6	8.25%
MET1-N	CEL23-O6	7.87%
CEL27-O2	GLU75-OE2	7.04%
CEL24-O2	GLY45-O	6.83%
CEL26-O3	GLU75-OE2	6.36%
CEL26-O3	GLU153-CD	6.00%
CEL25-O6	GLY76-O	5.84%
CEL27-O2	GLU75-OE1	5.46%
GLU75-N	CEL27-O6	5.46%
CEL26-O3	GLU153-OE2	5.35%

Table S18. Percentage of frames in the trajectory where any atoms of the CBM residues are within 0.35 nm radius from any of cellulose atoms (any occurrence in a frame for a given residue results in a positive result) in the simulations with CBM.

	CBM-aI1	CBM-aI2	CBM-aII1	CBM-aII2	CBM-c1	CBM-c2	CBM-c3	CBM-c4	CBM-c5
ASN1	75.76	-	53.76	90.73	-	-	11.68	-	4.71
LEU2	5.24	-	-	3.06	-	-	-	-	-
LYS3	3.53	-	8.82	3.40	-	-	-	-	-
TYR7	-	66.84	-	-	22.78	21.84	-	-	1.81
SER9	-	82.55	-	-	99.96	52.29	-	-	3.90
ASN10	-	88.47	-	-	99.99	65.34	-	-	-
PRO11	-	58.44	-	-	99.15	62.42	-	-	70.31
SER12	-	59.95	-	-	99.84	63.01	-	1.36	50.77
ASP13	3.43	-	-	-	1.96	-	-	38.28	-
THR14	-	73.16	-	-	52.72	5.46	-	77.72	2.04
THR15	-	87.15	-	-	58.79	2.36	-	-	2.33
ASN16	-	88.19	-	-	99.60	63.16	-	-	69.23
SER17	-	39.24	-	-	61.51	8.43	-	-	15.94
ASN19	-	82.57	-	-	64.71	45.48	-	-	-
GLN21	-	69.91	-	-	1.60	-	-	-	-
THR27	10.26	3.37	6.13	12.75	-	-	-	-	-
GLY28	63.63	3.47	75.48	43.24	-	-	-	-	-
SER29	68.42	7.42	83.03	75.99	-	-	28.26	-	4.38
SER30	22.41	5.07	53.01	34.50	-	-	28.21	-	1.14
ALA31	2.90	5.78	-	-	-	-	-	-	-
ASP33	-	11.48	-	-	-	-	-	-	-
SER35	-	9.54	-	-	-	-	-	-	-
LYS36	-	-	3.43	-	-	-	47.49	-	-
LEU39	-	5.68	-	-	-	-	-	-	-
ARG40	-	-	-	-	-	-	5.26	-	-
THR44	-	-	-	-	-	-	-	1.79	-
ASP46	-	-	-	-	-	-	-	76.06	-
GLY47	-	-	-	-	-	-	-	77.58	-
GLN48	-	-	-	-	-	-	-	73.22	-
LYS49	-	47.17	-	-	-	-	-	73.72	-
TRP54	-	68.68	-	-	23.51	3.22	-	-	63.37
CYS55	-	45.28	-	-	-	-	-	-	37.35
ASP56	-	78.32	-	-	100.00	64.02	-	-	70.13
HIS57	-	84.41	-	-	99.87	65.13	-	-	72.30
ILE60	-	8.20	-	-	-	-	-	-	-
ILE61	-	11.48	-	-	3.67	13.59	-	-	3.35
GLY62	-	74.44	-	-	-	-	-	-	-
SER63	-	68.69	-	-	-	1.49	-	-	-
ASN64	-	94.40	-	-	16.18	66.97	-	-	12.15
GLY65	-	86.70	-	-	46.77	67.95	-	-	16.99
SER66	-	83.28	-	-	21.30	8.65	-	-	12.88
TYR67	-	34.20	-	-	99.97	67.37	-	-	77.44

ASN68	-	19.82	-	-	-	1.82	-	-	3.74
GLY69	-	12.92	-	-	-	1.20	-	-	45.79
ILE70	-	13.38	-	-	-	-	-	-	-
THR71	-	75.41	-	-	8.37	23.62	-	-	57.92
SER72	-	91.12	-	-	-	1.59	-	-	13.25
ASN73	-	87.72	-	-	-	-	-	-	-
VAL74	-	4.54	-	-	-	-	-	-	-
LYS75	-	34.05	-	-	-	-	-	-	-
THR77	-	-	-	-	-	-	8.13	-	-
VAL79	-	-	-	-	-	-	16.00	-	-
LYS80	-	-	-	-	-	-	15.02	1.78	-
MET81	-	-	-	-	-	-	14.88	-	-
SER82	1.39	-	10.46	2.65	-	-	60.06	1.11	-
SER83	2.44	-	26.20	3.30	-	-	43.70	5.04	-
SER84	1.07	-	7.34	-	-	-	34.67	32.57	-
THR85	39.63	-	64.74	1.33	-	-	-	10.61	-
ASN86	85.91	-	26.79	2.14	-	-	1.95	78.50	-
ASN87	85.52	-	60.17	-	-	-	-	-	-
TYR91	-	-	-	-	-	-	10.44	-	-
THR97	-	93.00	-	-	-	-	-	-	-
GLY98	-	92.15	-	-	-	-	-	-	-
GLY99	-	87.21	-	-	-	-	-	-	-
THR100	-	46.25	-	-	-	-	-	-	-
GLU102	1.82	72.90	-	-	-	-	-	-	-
PRO103	8.51	63.26	-	-	-	-	-	-	-
GLY104	-	82.22	-	-	-	-	-	-	-
ALA105	-	74.82	-	-	-	-	-	-	-
HIS106	-	7.19	-	-	-	-	-	-	-
GLN108	-	58.31	-	-	1.10	1.62	-	-	-
GLN110	-	72.19	-	-	92.28	52.07	-	-	12.02
ARG112	-	72.82	-	-	99.98	63.66	-	-	70.10
LYS115	-	-	-	-	-	-	-	5.40	-
ASN116	-	23.79	-	-	-	-	-	-	23.21
ASP117	-	75.55	-	-	1.94	2.52	-	61.87	43.89
TRP118	-	88.52	-	-	100.00	64.23	-	-	69.87
SER119	-	76.42	-	-	-	-	-	72.64	-
ASN120	-	81.32	-	-	25.69	-	-	79.32	3.45
THR122	-	-	-	-	-	-	-	80.26	-
SER124	51.87	-	-	-	-	-	-	76.99	-
ASN125	18.67	-	-	-	-	-	-	78.34	-
ASP126	6.46	-	-	-	-	-	-	-	-
TYR127	79.40	-	66.87	-	-	-	-	-	-
PHE129	30.13	-	-	-	-	-	-	-	-
LYS130	49.62	-	40.18	-	-	-	-	-	-
SER131	26.74	-	-	-	-	-	-	-	-
ALA132	3.10	4.05	-	-	-	-	-	-	-
SER133	3.08	35.53	-	-	99.58	66.24	-	-	20.46

GLN134	3.04	8.98	14.21	-	73.49	55.07	-	-	6.96
PHE135	-	25.32	-	-	-	-	-	-	-
VAL136	19.14	-	45.26	-	-	-	-	-	-
GLU137	81.55	-	63.25	11.21	-	-	-	-	-
TRP138	58.79	-	72.07	-	-	-	-	-	-
ASP139	89.13	-	31.92	34.04	-	-	-	-	-
GLN140	61.88	-	76.53	-	-	-	-	-	-
TYR144	6.72	-	-	3.52	-	-	-	-	-
LEU145	-	-	3.96	-	-	-	-	-	-
ASN146	-	-	19.50	7.86	-	-	90.04	-	-
GLY147	-	-	-	4.28	-	-	65.62	-	-
VAL148	29.30	-	56.02	86.53	-	-	92.77	-	4.78
LEU149	79.18	-	2.73	97.76	-	-	3.90	-	3.51
VAL150	77.97	-	11.51	59.40	-	-	-	-	-
TRP151	92.18	-	-	87.79	-	-	-	-	-
GLY152	87.04	-	-	66.03	-	-	-	-	-
LYS153	91.67	-	66.11	66.53	-	-	-	-	3.53
GLU154	83.06	-	2.61	53.28	-	-	-	-	-
PRO155	83.91	-	33.25	3.65	-	-	-	-	-

Table S19. Percentage of frames in the trajectory where any atoms of the expansin residues are within 0.35 nm radius from any of cellulose atoms (any occurrence in a frame for a given residue results in a positive result) in the simulations with expansin.

	Exp-aI1	Exp-aI2	Exp-aI3	Exp-aI4	Exp-aII1	Exp-aII2	Exp-c1
MET1	-	-	-	-	-	53.91	-
GLU8	-	-	-	-	-	3.01	-
TYR10	-	-	-	-	-	1.94	-
THR12	-	-	-	2.77	-	-	-
TYR13	77.33	14.64	-	46.86	-	-	-
THR14	78.55	92.80	71.90	93.78	-	-	-
GLY15	74.47	68.70	10.95	95.06	-	-	2.04
SER16	43.35	86.79	20.64	71.98	-	-	3.37
GLY17	72.23	45.01	4.04	31.16	-	-	2.50
TYR18	42.17	32.71	18.41	92.51	-	-	1.08
SER19	69.16	95.63	88.11	76.38	-	-	3.15
GLY20	37.52	44.99	87.46	81.54	-	-	-
GLY21	46.84	78.29	39.36	78.89	-	-	-
ALA22	1.31	81.43	99.98	94.49	-	-	7.96
PHE23	-	-	8.08	-	-	2.23	-
LEU24	30.55	81.51	98.68	92.87	-	-	-
ILE28	-	-	4.11	-	-	-	-
SER30	12.96	4.01	-	1.26	-	-	-
PRO39	-	-	-	7.03	-	89.67	-
ALA40	-	-	-	21.53	1.50	96.25	-
ASP41	-	-	-	-	-	1.80	-
ASN43	-	-	-	4.46	-	85.91	-
GLY45	-	-	-	-	-	65.43	-
GLY46	-	-	-	-	4.81	92.99	-
VAL47	-	-	-	-	-	24.88	-
LYS48	-	-	-	44.47	94.12	87.41	8.86
ASP71	-	28.76	6.93	13.34	-	-	-
LEU72	-	3.06	20.79	3.45	-	29.55	22.67
TYR73	2.17	-	-	-	-	-	-
PRO74	39.47	82.24	94.20	7.77	30.45	12.81	66.66
GLU75	48.70	36.33	-	1.26	96.57	29.56	78.43
GLY76	-	-	-	44.49	89.92	40.30	25.69
ALA77	-	-	-	13.91	4.62	43.43	2.89
ARG78	-	-	-	9.46	44.61	93.34	-
PRO85	73.01	-	-	24.12	-	-	-
ARG89	13.68	5.53	-	-	-	-	-
MET94	72.73	10.31	-	36.96	-	-	-
LYS95	43.18	8.46	-	43.96	-	-	-
ASP96	32.39	3.04	-	35.90	-	-	-
GLY97	20.05	2.98	-	38.53	-	-	-
LYS98	11.29	-	-	-	-	10.37	-
ASN100	-	-	-	-	-	5.39	-
LYS119	21.92	88.89	100.00	92.78	12.81	1.33	2.20

GLU120	29.82	90.30	96.63	98.16	12.46	1.98	75.79
GLY121	1.70	88.64	100.00	96.81	85.66	2.58	79.98
SER123	-	30.86	72.76	63.60	27.82	-	72.93
ARG124	-	1.41	3.25	-	2.02	-	-
TRP125	-	86.08	100.00	99.97	98.16	9.10	80.04
TRP126	-	82.28	100.00	100.00	97.54	58.55	80.04
MET139	-	-	-	-	-	4.53	-
LYS140	-	-	-	-	-	12.08	-
TYR143	-	-	-	-	-	6.09	-
LYS145	-	-	-	-	-	2.11	-
ILE150	-	-	-	-	-	11.54	-
ASN151	-	-	-	-	-	36.98	-
GLU153	-	-	-	-	35.21	76.65	13.60
LYS154	-	-	-	-	-	1.93	-
MET155	-	4.72	22.35	94.60	96.13	16.52	74.88
ASP156	-	6.06	20.52	97.58	59.78	29.87	60.01
TYR157	31.41	82.48	100.00	99.16	87.31	11.13	75.33
VAL161	-	-	-	45.07	35.96	32.32	-
THR163	-	11.31	65.74	98.71	85.61	47.97	79.88
ASN164	-	1.83	7.46	59.50	16.78	-	45.88
GLU191	-	35.63	22.86	-	65.11	-	71.46
SER192	-	80.17	94.39	-	97.69	-	80.03
GLY193	-	84.53	73.12	-	81.60	-	79.91
THR194	-	30.39	28.75	-	52.79	-	67.17
SER195	38.93	72.88	97.88	-	80.52	-	79.99
LYS196	20.08	7.30	-	-	-	-	11.97
GLU208	-	-	-	-	3.45	-	-