Binding pockets in proteins induced by mechanical stress

Matteo Tiberti[1†], Bob-Dan Lechner[1§], Arianna Fornili[1,2]

[1]School of Biological and Chemical Sciences, Queen Mary University of London, London, E1 4NS, United Kingdom

[2]The Thomas Young Centre for Theory and Simulation of Materials, London, United Kingdom

Present addresses:

† The Danish Cancer Society Research Center, DK-2100 Copenhagen, Denmark

§ School of Physics, University of Exeter, Exeter, EX4 4QL, United Kingdom

Supplementary Methods

General MD simulation parameters

All molecular dynamics (MD) simulations were performed using GROMACS 4.6.7[1]. The equations of motion were integrated using the leap-frog method. Unless otherwise stated, a 2-fs time step was used. Periodic boundary conditions were applied. The protein was described using the AMBER99SB*-ILDN force field[2] and the TIP3P[3] model was used for water. The charge of the ionisable residues was set to that of their standard protonation state at pH 7. Counterions were added to neutralise the system and reach an ionic strength of 50 mM. The electrostatic interactions were calculated with the Particle Mesh Ewald method[4] using a 0.9-nm cutoff for the direct space sums, a 0.12-nm Fourier transform grid spacing and a 4-order interpolation polynomial for the reciprocal space sums. A 0.9-nm cutoff was used for van der Waals interactions and long-range corrections to the dispersion energy were included. Unless otherwise stated, all the covalent bonds in the protein were constrained using the LINCS[5] algorithm, while SETTLE[6] was used for water molecules. The Berendsen algorithm was used for temperature and pressure regulation, with coupling constants of 0.2 ps and 1.0 ps, respectively. Coordinates in the production runs were saved every 1 ps unless otherwise stated.

Initial system setup

The X-ray structure of the C1 domain of human cardiac MyBP-C (PDB ID: 3CX2, UniProt ID: Q14896) was used to generate the initial system. A missing residue

(S182) was modelled using ModLoop[7]. The protein was solvated with a truncated octahedral box of water molecules and counterions with a minimal distance of 1 nm between the protein and the box walls, resulting in a system with 10,565 water molecules and a total number of 33,397 atoms.

The Berendsen algorithm[8] was used for both temperature and pressure regulation with coupling constants of 0.2 and 1 ps, respectively. Positional restraints were initially set on heavy atoms with a constant of 2000 kJ/mol/nm^2. Each system was first energy-minimised through 3 stages with 2000 (positional restraints on heavy atoms) + 5000 steps of steepest descent, followed by 2000 steps of conjugate gradient. In the subsequent equilibration, heavy-atom positional restraints were gradually decreased from 2000 to 0 kJ/mol/nm^2 in 1.5 ns, while the temperature was increased from 200 to 300 K at constant volume. The system was then allowed to move freely and was subjected to 1-ns equilibration in NVT conditions at T = 300 K. This was followed by a 4-ns equilibration in NPT conditions with T = 300 K and p = 1 bar. A production NPT simulation was then run for 500 ns, saving the coordinates every 100 ps. A cluster analysis was performed on the production trajectory with the 'gromos'[9] method implemented in GROMACS, with a cutoff of 0.2 nm. The representative structure from the most populated cluster was then used as starting structure for the following simulations.

Constant force simulations (CFX)

The starting structure was initially placed at the centre of a rectangular box, with the vector connecting the N- and C-terminal C$^\alpha$ atoms aligned along the *x* axis and a minimum distance of 1.0 nm between the protein and the walls of the box. The box was then extended by 1.2 nm along the *x* axis to allow for possible structural modifications due to the force. The protein was solvated with TIP3P water molecules, resulting in a total number of ~22,000 atoms (Table S1).

An equilibration protocol similar to the one described above (*Initial system setup*) was adopted, except that only one minimisation stage (steepest descent with positional restraints) was performed and shorter NPT equilibration steps (2 ns) were used. A constant force was subsequently applied for 100 ns using the pull code implemented in GROMACS[10]. The position of the N-terminal C$^\alpha$ atom (D151) was restrained using a harmonic potential with a force constant of 2500 kJ/mol/nm^2. The

external force was applied using a linear potential along the *x* component of the distance vector between the C-terminal (E258) and N-terminal (D151) C^{α} atoms. To avoid using LINCS in combination with the pull code[10], the constraints on the protein covalent bonds were removed and the time-step was correspondingly reduced to 0.5 fs. Force magnitudes ranged from 187.5 to 450 pN (Table S1). Higher forces were also tested (≥ 600 pN) but were not further considered since they induced the unfolding of the domain.

The control simulations (CF0) were performed using the same protocol except for the fact that the external force and the positional restraint on the N-terminal C^{α} atom were not applied. Moreover, a truncated octahedral box was used with a minimum distance of 1.2 nm between the protein and the walls of the box to allow for rotational motions of the protein.

Two replicas were run for each force magnitude (Table S1) to check for reproducibility of the results. Replicas were started from different sets of initial velocities.

Constant force simulations with steering of CV_{pF} *(SCV-CFX)*

SCV-CF*X* simulations were performed by combining the application of an external force with the steering of the collective variable CV_{pF} to induce the opening of the pF pocket. The CV_{pF} variable was defined as the distance from the centre of mass of the F157 side-chain to the bottom of the pocket (centre of mass of the C239, E240, V241, D248, C249 and S250 C^{α} atoms).

All SCV-CF*X* simulations were performed with the PLUMED 2.2.5 plugin[11] coupled with GROMACS 4.6.7. The protein structure generated in the initial setup stage was solvated with an octahedral box with a minimum distance of 1.6 nm to the box walls in order to allow for both protein rotation and pocket opening (Table S1). The same minimisation and equilibration protocol as for the CF simulations was followed. An additional 1-ns NPT step was performed and ten snapshots were extracted at regular intervals of 100 ps. These were used as starting structures for the subsequent simulations, where a harmonic steering potential was applied on CV_{pF}. The reference value in the potential was linearly increased from the starting value of CV_{pF} to a target value of 1.0 nm in 45 ns, using a force constant of 5000 kJ/mol/nm^2. The restraint was then kept fixed at 1.0 nm for 5 ns.

Two sets of 10 replicas were performed, one in the absence (SCV-CF0) and the other in the presence (SCV-CF187.5) of an external force, for a total of 1 μs of simulation time. The external force was applied as described above (*Constant force simulations*). The work performed during the CV_{pF} steering (from the first to the last frame) in each SCV-CFX simulation was calculated using the PLUMED plugin[11].

Pocket detection

The opening of binding pockets during the simulations was detected using the MDpocket[12] software (stand-alone version). This method is based on the repeated application of the fpocket[13] program on snapshots extracted from a MD trajectory. In fpocket, pockets are detected as sets of α-spheres, i.e. spheres in contact with four atoms and with no atom inside. In the MDpocket pipeline, a 0.1-nm spaced grid is first defined on the initial frame of the simulation. For each frame, the α-spheres are identified, filtered by size and clustered by fpocket. Each α-sphere is then assigned to the grid point closest to its centre. For each grid point, a frequency value is calculated as the fraction of frames in which at least one α-sphere is assigned to that grid point.

MDpocket frequency maps were calculated for each 100-ns CFX trajectory, using the default fpocket parameters for α-sphere detection (parameter set 1)[12]. The analysis was performed on trajectories containing protein structures saved every 10 ps. A common frame (initial CF0 structure) was added as the first frame of each trajectory in order to have the same grid definition for all the MDpocket analyses. All frames in each trajectory were aligned to the first one by best fit superimposition. Grid points having frequency ≥ 0.2 for each of the main four pockets (p1-3 and pF) were extracted manually from the grid. For each pocket, a final set of representative grid points was generated by selecting all the grid nodes that were part of the pocket in at least one simulation. These pocket definitions were used to calculate the pocket descriptors (volume and solvent accessible surface) with MDpocket.

Detection of ligand-binding hotspots

Ligand-binding hotspots were detected running the FTMap web server[14] on selected protein structures. FTMap uses small-molecule probes to identify the regions of a protein surface that are more likely to bind drug-like ligands. In particular, regions that can bind different probes are considered as binding hotspots, with binding

affinity proportional to the number of low-energy binding poses of the probes (probe clusters). The 16 small molecules used by FTMap as probes include the most common chemical groups found in drug-like molecules and range in size from ethane to benzaldehyde (Table S4).

The protein structures to be used for the FTMap analysis were extracted from the CFX simulations where the pF opening was observed (CF375 and CF450). The protein structures where the entrance to the pF pocket was maximally open were selected. The F157-R160 and M159-D248 distance values were first calculated as the minimum distance over any pair of residue atoms. The pocket rim opening was then calculated as the average between these distance values. The frame with the maximum opening was identified for each simulation and used as input for FTMap. FTMap calculations were performed using default parameters.

Hydrogen bonds analysis

The hydrogen bonds analysis was performed with the GROMACS tool g_hbond using default parameters for hydrogen bonds detection (donor-acceptor distance ≤ 0.35 nm and donor-H-acceptor angle ≥ 150^{o}). Secondary structures were determined with DSSP[15].

Sequence and structure analysis of the I-set family of Ig domains

The sequence analysis was performed on a multiple sequence alignment of I-set domains from Pfam 31.0[16] (Pfam ID: PF07679). The Reference Proteomes sequence alignment with 75% co-membership threshold (RP75) was selected (66,871 sequences). The frequency of amino acids occurring at the alignment position corresponding to F157 in the MyBP-C C1 domain (X157) was then calculated, both including and excluding gaps from the total (Table S5).

A subset composed by I-set sequences with known 3D structures (Pfam3D) was generated using the PDB/Pfam cross-mapping available from the Pfam website. Domains with structure coverage lower than 60%, a gap in the alignment position X157, or sequence identity between the Uniprot and PDB sequence lower than 95% were not considered, resulting in a total number of 89 unique sequences. These were ranked according to their sequence identity to the MyBP-C C1 domain, calculated using the Pfam RP75 alignment. Representative structures were then selected to explore the whole range of sequence identity values (Table S6). The

solvent accessible surface (SAS) of the X157 residue (F or I) was calculated using DSSP[15] and normalised by the corresponding maximum SAS value[17]. A sequence logo of the Pfam3D sequences (RP75 alignment) was generated with the Seq2Logo server[18] selecting the Kullback-Leibler logo type and using default parameters. Co-localisation of Cysteine residues in the A- and G-strands around the pocket was determined by identifying the sequences with at least one Cysteine residue in both the 157-160 and the 246-250 residue intervals (24 out of 66,871 sequences).

References

1. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19-25.
2. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010,** *78*, 1950-1958.
3. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79*, 926-935.
4. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *J. Chem. Phys.* **1998,** *103*, 8577-8593.
5. Hess, B.; Bekker, H.; Berendsen, H. J. C., LINCS: a linear constraint solver for molecular simulations. *J. Comp. Chem.* **1997,** *18*, 1463–1472.
6. Miyamoto, S.; A., K. P., Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comp. Chem.* **1992,** *13*, 952--962.
7. Fiser, A.; Sali, A., ModLoop: automated modeling of loops in protein structures. *Bioinformatics* **2003,** *19*, 2500-2501.
8. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1998,** *81*, 3684-3690.
9. Daura, X.; van Gunsteren, W. F.; Mark, A. E., Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed.* **1999,** *38*, 236-240.
10. van der Spoel, D.; Lindahl, E.; Hess, B.; the GROMACS development team *Gromacs User Manual version 4.6.7*, 2014.
11. Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M., PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009,** *180*, 1961-1972.
12. Schmidtke, P.; Bidon-Chanal, A.; Luque, F. J.; Barril, X., MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics* **2011,** *27*, 3276-3285.
13. (a) Le Guilloux, V.; Schmidtke, P.; Tufféry, P., Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics* **2009,** *10*, 168-11; (b) Schmidtke, P.; Le Guilloux, V.; Maupetit, J.; Tufféry, P., fpocket: online tools for

protein ensemble pocket detection and tracking. *Nucleic Acids Res.* **2010,** *38*, W582-9.
14. Kozakov, D.; Grove, L. E.; Hall, D. R.; Bohnuud, T.; Mottarella, S. E.; Luo, L.; Xia, B.; Beglov, D.; Vajda, S., The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat. Protoc.* **2015**.
15. Kabsch, W.; Sander, C., Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983,** *22*, 2577-2637.
16. Finn, R. D.; Coggill, P.; Eberhardt, R. Y.; Eddy, S. R.; Mistry, J.; Mitchell, A. L.; Potter, S. C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; Salazar, G. A.; Tate, J.; Bateman, A., The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **2016,** *44*, D279-D285.
17. Tien, M. Z.; Meyer, A. G.; Sydykova, D. K.; Spielman, S. J.; Wilke, C. O., Maximum Allowed Solvent Accessibilites of Residues in Proteins. *PLoS ONE* **2013,** *8*, e80635-8.
18. Thomsen, M. C. F.; Nielsen, M., Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. *Nucleic Acids Res.* **2012,** *40*, W281-W287.

Tables

Table S1. Summary of the simulations.

Name[a]	Applied constant force (pN)	Steering on CV_{pF}	Length (ns)	Number of replicas	Total length (ns)	Number of water molecules	Number of atoms
CF0	0	No	100	2	200	10,063	31,891
CF187.5	187.5	No	100	2	200	6,769	22,003
CF375	375	No	100	2	200	6,769	22,003
CF450	450	No	100	2	200	6,769	22,003
SCV-CF0	0	Yes	50	10	500	13,955	43,573
SCV-CF187.5	187.5	Yes	50	10	500	13,955	43,573

[a]CF=Constant Force simulations, SCV-CF=Constant Force simulations with steering of the collective variable CV_{pF}.

Table S2. List of pF pocket residues

Residue number	Residue type	atoms
155	GLY	CA
156	LEU	N, CA, CB
157	PHE	C, O, N, CA, CB, CG, CE1, CE2, CZ, CD1, CD2
158	VAL	N, CA, CG1
159	MET	C, O
160	ARG	N, CA, CB, CG, CD, NE, CZ, NH1, NH2, C
161	PRO	N, CD, CG, CB
176	ALA	N, CA, CB, C, O
177	ARG	N, CA, C, O
178	VAL	N, CA, CB, CG1, CG2
189	VAL	CG1
191	TRP	CH2, CZ3, CE3
222	PHE	CD2
238	ARG	C
239	CYS	N, CA, CB, SG, C, O
240	GLU	N, CA, CB, C, O
241	VAL	N, CB, CG1
248	ASP	CB, CG, OD1, OD2, C, O
249	CYS	N, CA, C, O
250	SER	N, CA, CB, OG, C, O

Table S3. Standard deviation of work and HB occurrence values for SVC-CF0 and SCV-CF187.5 replicas

Property	SCV-CF0	SCV-CF187.5
work	8.0 kJ/mol	19.5 kJ/mol
HB occurrence (L156-D248)	0.03	0.05
HB occurrence (F157-D248)	0.04	0.05

Table S4. FTMap fragments binding to the pF pocket

Fragment	CF375 (replica 1)[a]	CF375 (replica 2)[a]	CF450 (replica 1)[a]	CF450 (replica 2)[a]
acetamide	-	2	-	1
acetonitrile	-	2	1	1
acetone	-	1	-	1
acetaldehyde	-	2	-	1
methylamine	-	2	1	1
benzaldehyde	-	2	-	1
benzene	-	1	-	1
isobutanol	-	1	-	1
cyclohexane	-	1	-	1
N,N-dimethylformamide	-	1	-	1
dimethyl ether	-	2	-	1
ethanol	-	2	-	1
ethane	-	1	-	1
phenol	-	1	-	1
isopropanol	-	1	-	1
urea	-	2	-	1
Total	**0**	**24**	**2**	**16**

[a]The number of FTMap probe clusters found in the pF pocket hostspot is reported for each probe molecule

Table S5. Frequency of amino acids types at position X157 in the I-set family

Amino acid[a]	Count[b]	Percentage[c]
A	121	0.3 (0.2)
C	4	0.0 (0.0)
F	24500	64.6 (36.6)
I	9624	25.4 (14.4)
K	1	0.0 (0.0)
L	864	2.3 (1.3)
M	283	0.7 (0.4)
T	11	0.0 (0.0)
V	2163	5.7 (3.2)
W	333	0.9 (0.5)
X	2	0.0 (0.0)
Y	11	0.0 (0.0)

[a] Only amino acids with count > 0 are reported. X indicates an unspecified amino acid in the UniProt sequences.
[b] Number of sequences for which a given amino acid is found in the position of the I-set sequence alignment corresponding to F157 in human cardiac MyBP-C C1 (UniProt ID: Q14896).
[c] Percentage calculated as count over the total number of sequences for which X157 is not a gap in the multiple sequence alignment (37,917). The value calculated taking into account all the sequences (66,871) is indicated in parenthesis.

Table S6. **Selected I-set domains.**

UniProt ID	UniProt name	Organism	Domain	PDB ID	domain (residue interval)[a]	Seq. Ide. to My-BPC C1 (%)	X157[b]	normalized SAS (%)[c]
Q14896	MyBP-C, cardiac-type	Human	C1	3XC2	156-256	100	F	11.7
O70468	MyBP-C, cardiac-type	Mouse	C1	4EDQ	154-254	91.1	F	8.3
Q00872	MyBP-C, skeletal slow-type	Human	C1	2DAV	59-159	40.6	F	6.3
Q14324	MyBP-C, skeletal fast-type	Human	C1	2EDN	52-153	33.7	F	5.0
Q5VST9	Obscurin	Human	O32	2YZ8	3185-3269	28.2	F	12.1
Q9ET54	Palladin	Mouse	Ig-like 3	2LQR	1026-1117	23.9	F	2.9
Q8WZ42	Titin	Human	I10	1G1C	2078-2168	22.0	I	21.3
P19320	Vascular cell adhesion protein 1	Human	Ig-like 1	1IJ9	25-112	12.5	I	12.2

[a] Pfam domain boundaries (UniProt sequence numbering)
[b] Amino acid at the equivalent sequence position of My-BPC C1 F157
[c] Solvent Accessible Surface (SAS) of X157 (F or I) normalised by the corresponding maximum SAS

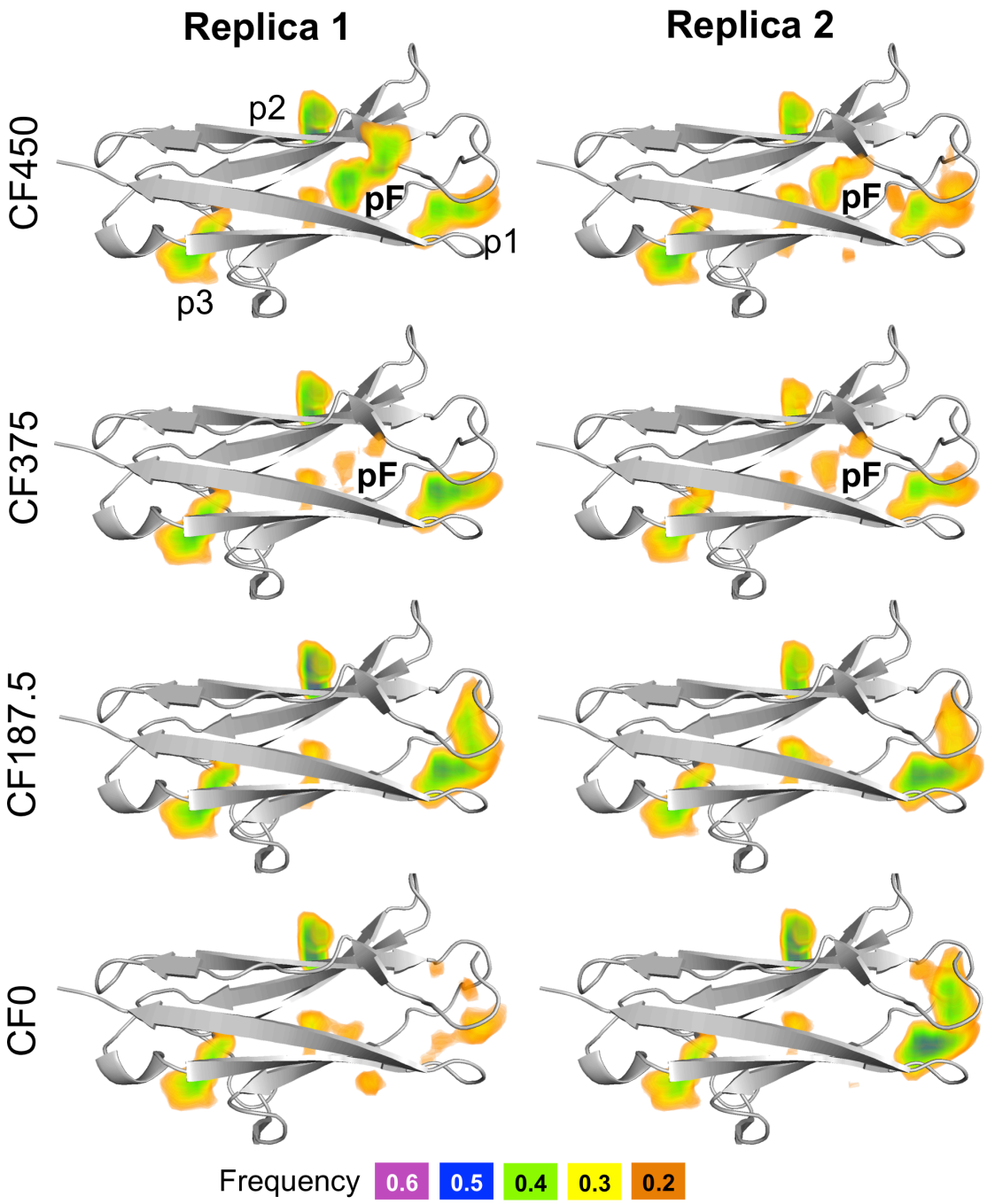


Figure S1. **Pockets identified by MDpocket in the CF simulations.** MDpocket frequency maps (colour surface) are superimposed on the initial structure of the MyBP-C C1 domain (white cartoon). Different isosurfaces are shown, corresponding to different values of pocket frequency. The four main pockets identified in our simulations are labelled in the top left panel. The force-dependent pocket pF was observed only in the CF3750 and CF450 simulations.

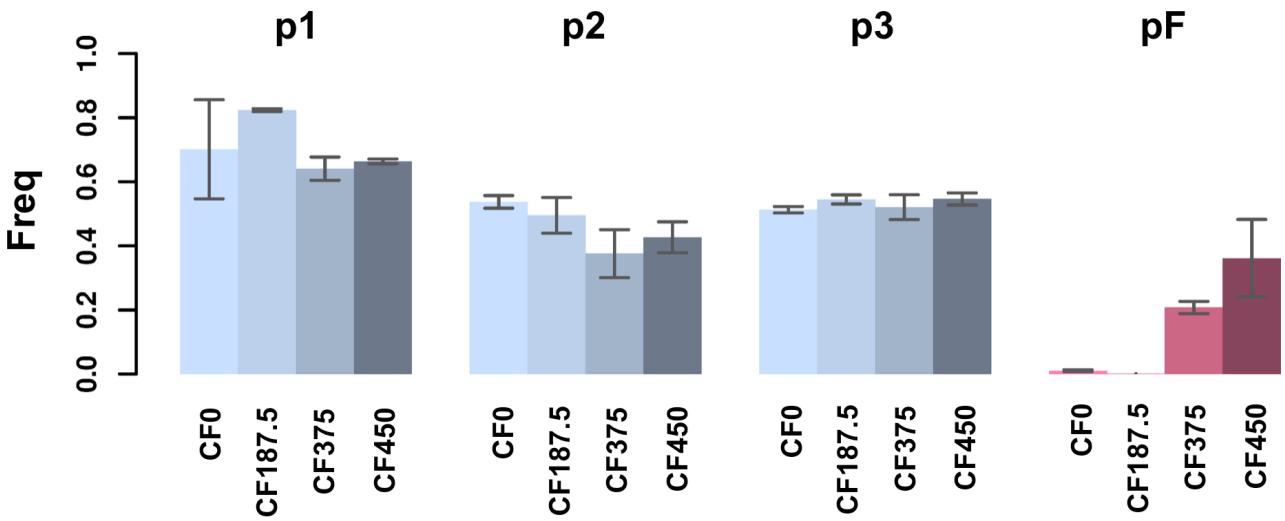


Figure S2. Pocket frequency in the CFX simulations. The fraction of structures with pocket volume > 0.17 nm^3 observed during the CF simulations is reported for pockets p1-3 and pF. The coloured bars show the average fraction calculated over two replicas with the same force magnitude, while the error bars indicate the fraction observed for each replica as minimum and maximum. The threshold value (0.17 nm^3) was determined as giving the best agreement between the fraction values reported here and the frequency values used for the isosurfaces reported in Figure S1, but similar trends were observed for different thresholds.

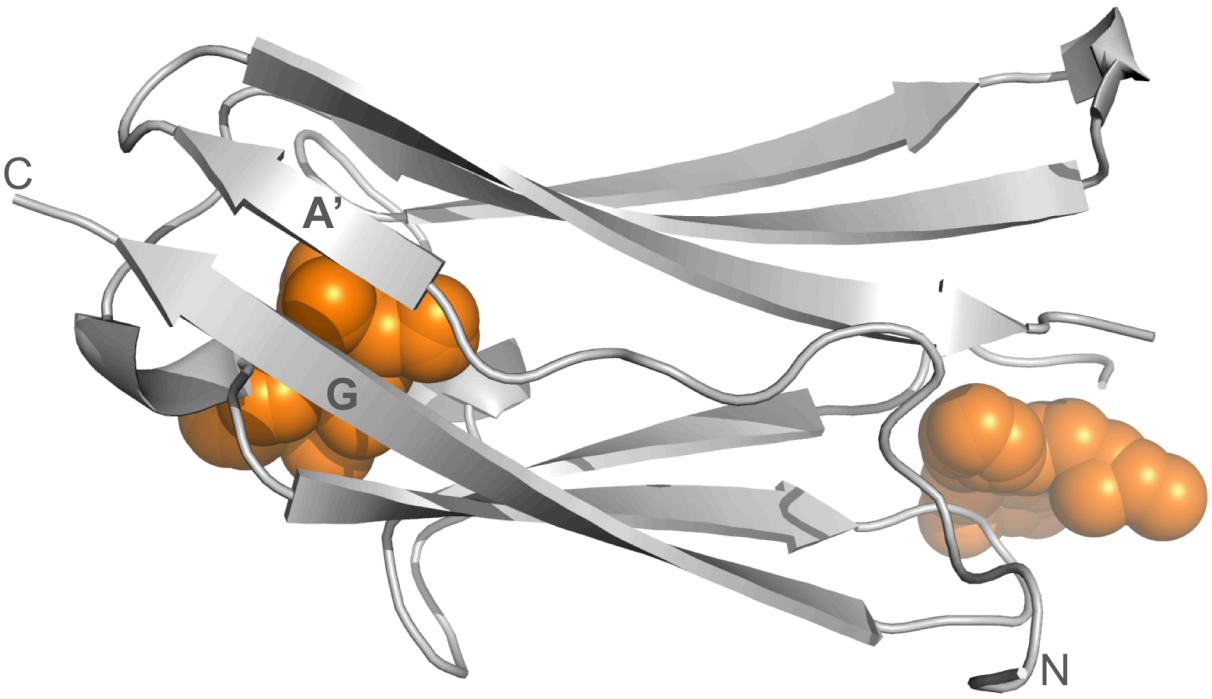
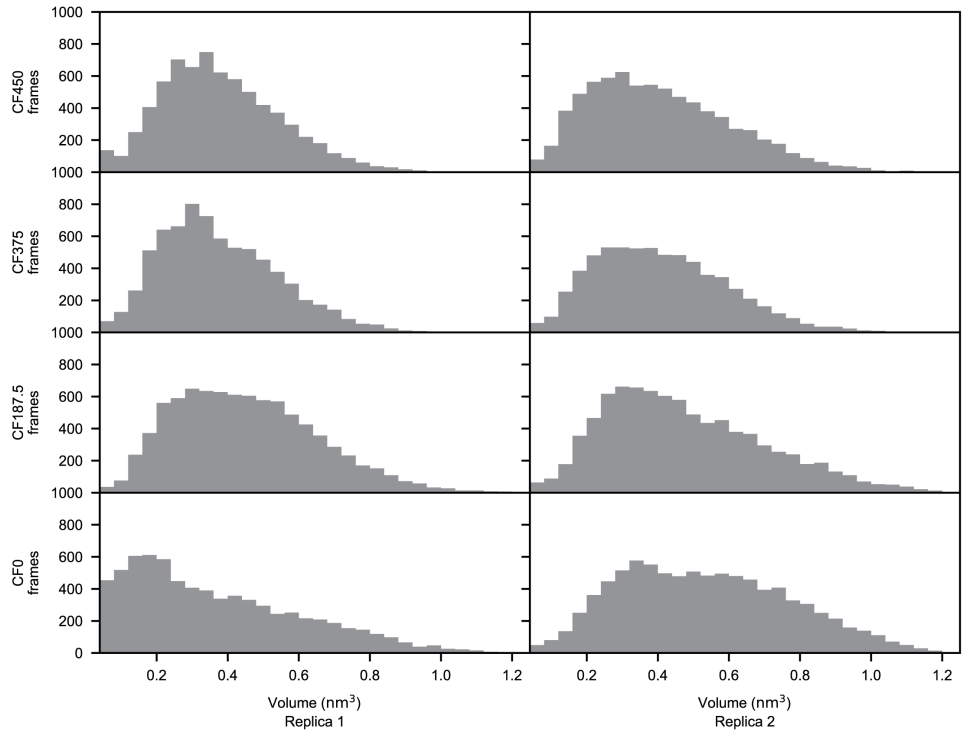
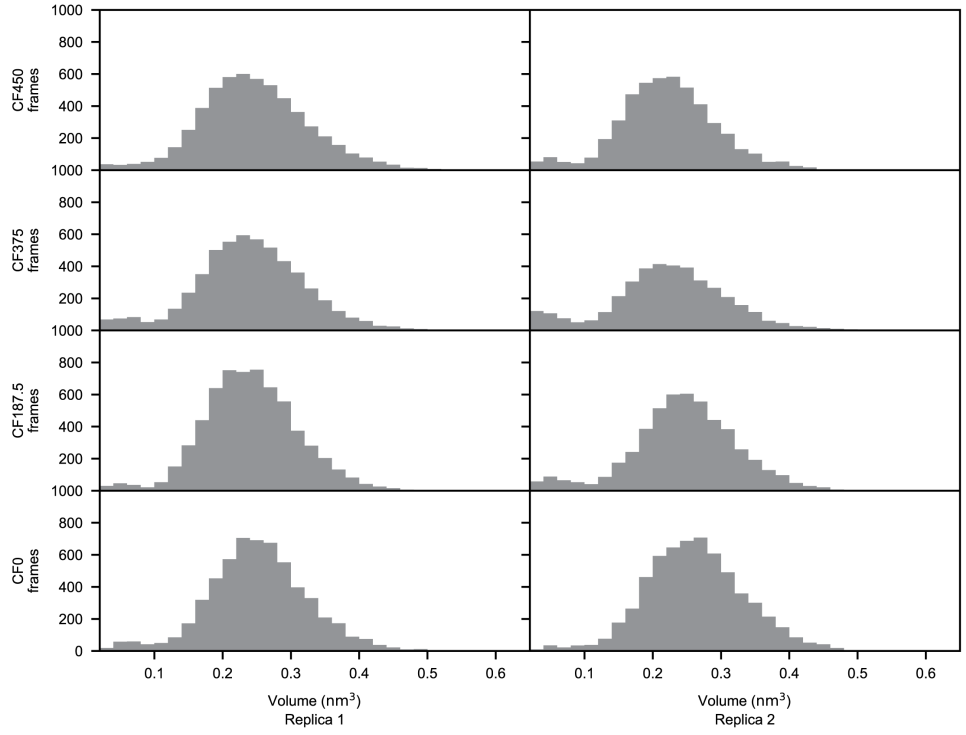


Figure S3. **Pockets identified by fpocket on the X-ray structure of MyBP-C C1.** The X-ray structure (PDB ID: 3CX2) is shown as white cartoon, while the fpocket alpha spheres are shown in orange. The location of these two 'static' pockets correspond approximately to the one of the p3 (left) and p1 (right) pockets observed during the MD trajectories (Figure S1).

Pocket 1



Pocket 2



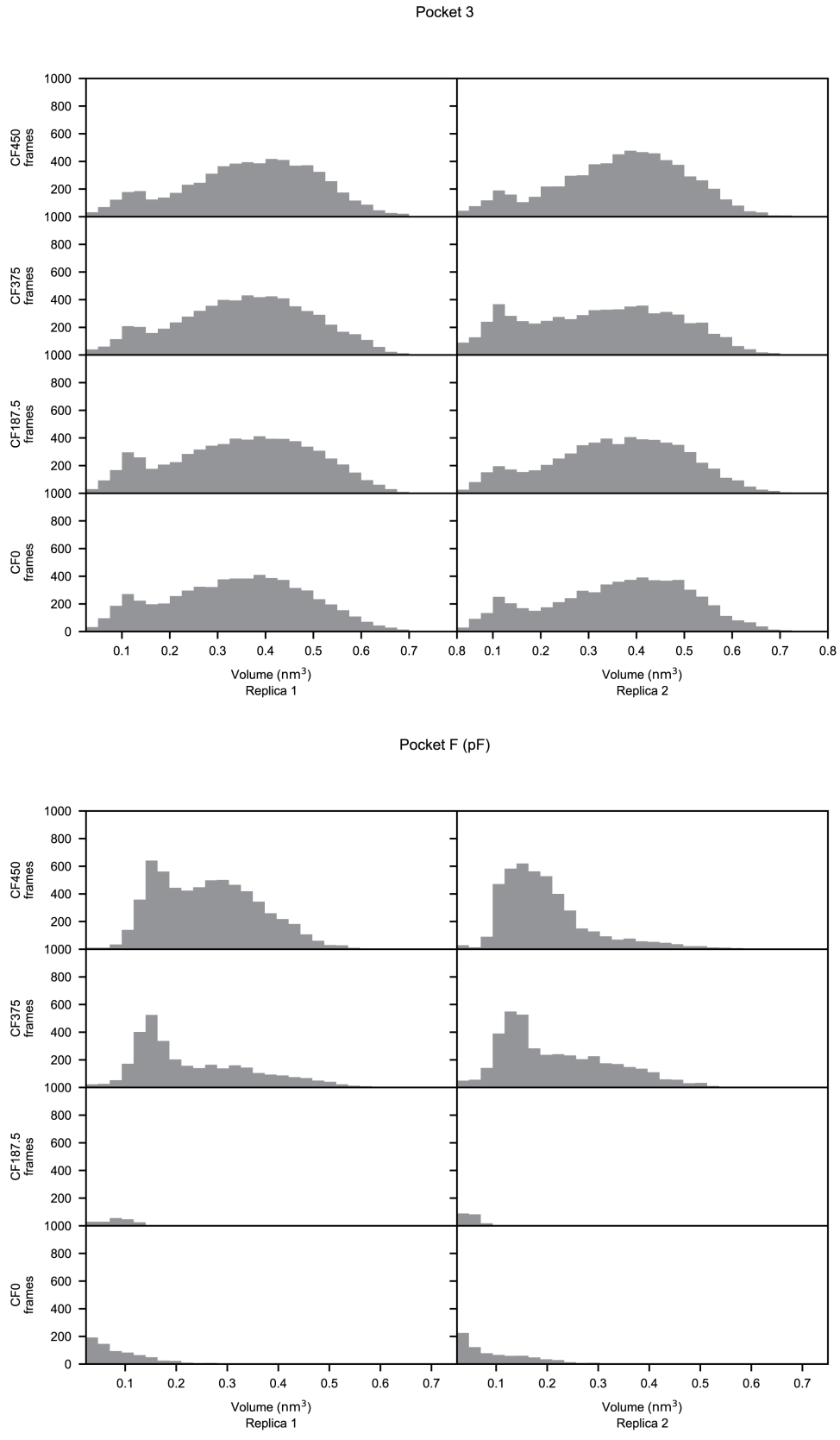


Figure S4. **Probability distribution of pocket volumes.** Distributions of pocket volumes observed during the simulations are reported for the p1-p3 and pF pockets. Volume values are calculated with MDpocket (Supplementary Methods). Only values different from zero are shown.

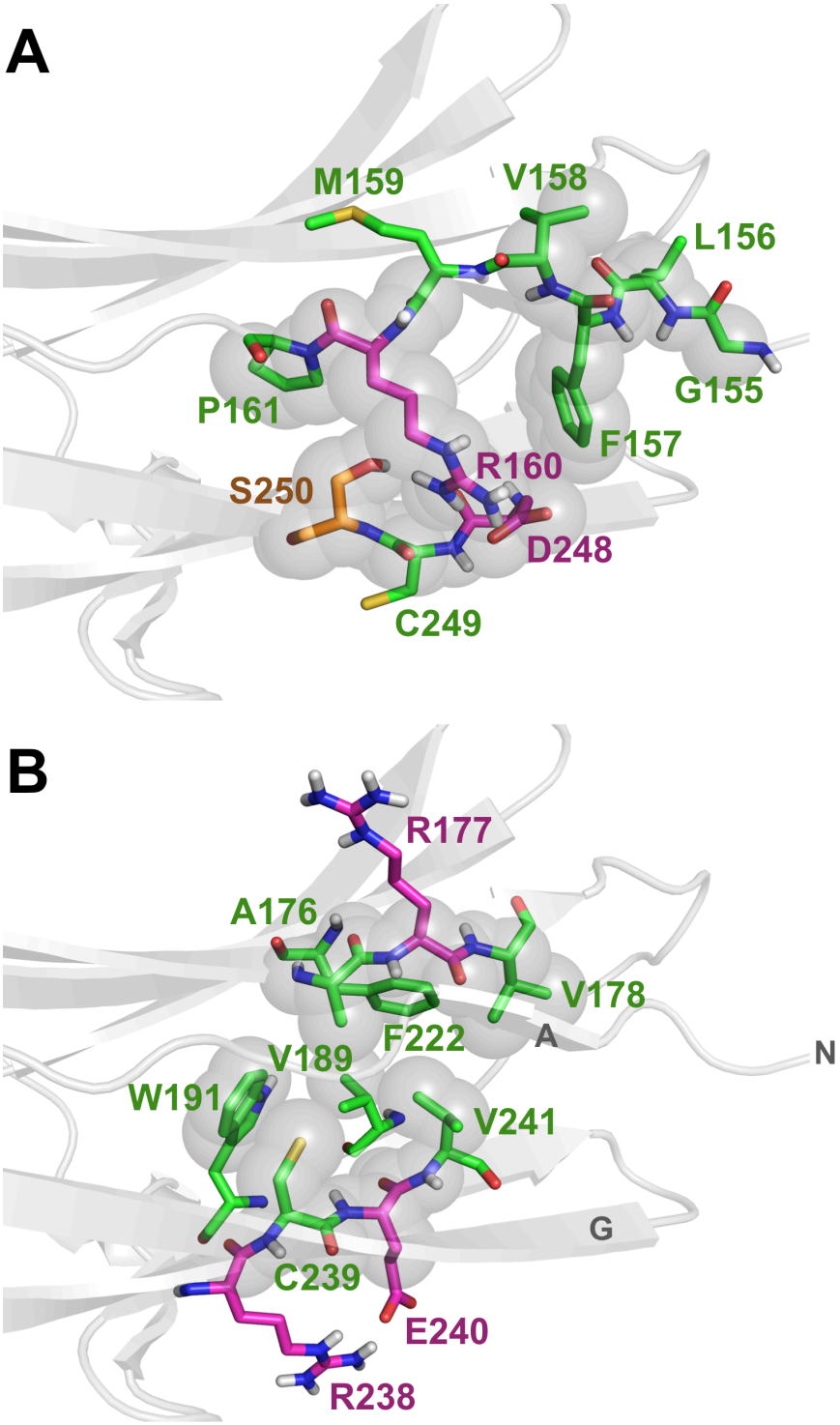


Figure S5. Residues lining the pF pocket. Residues in the rim (A) and bottom (B) regions of the pF pocket are shown as licorice and coloured according to their polarity (hydrophobic residues in green, charged residues in magenta and polar non-charged residues in orange), with the atoms lining the pocket highlighted as spheres.

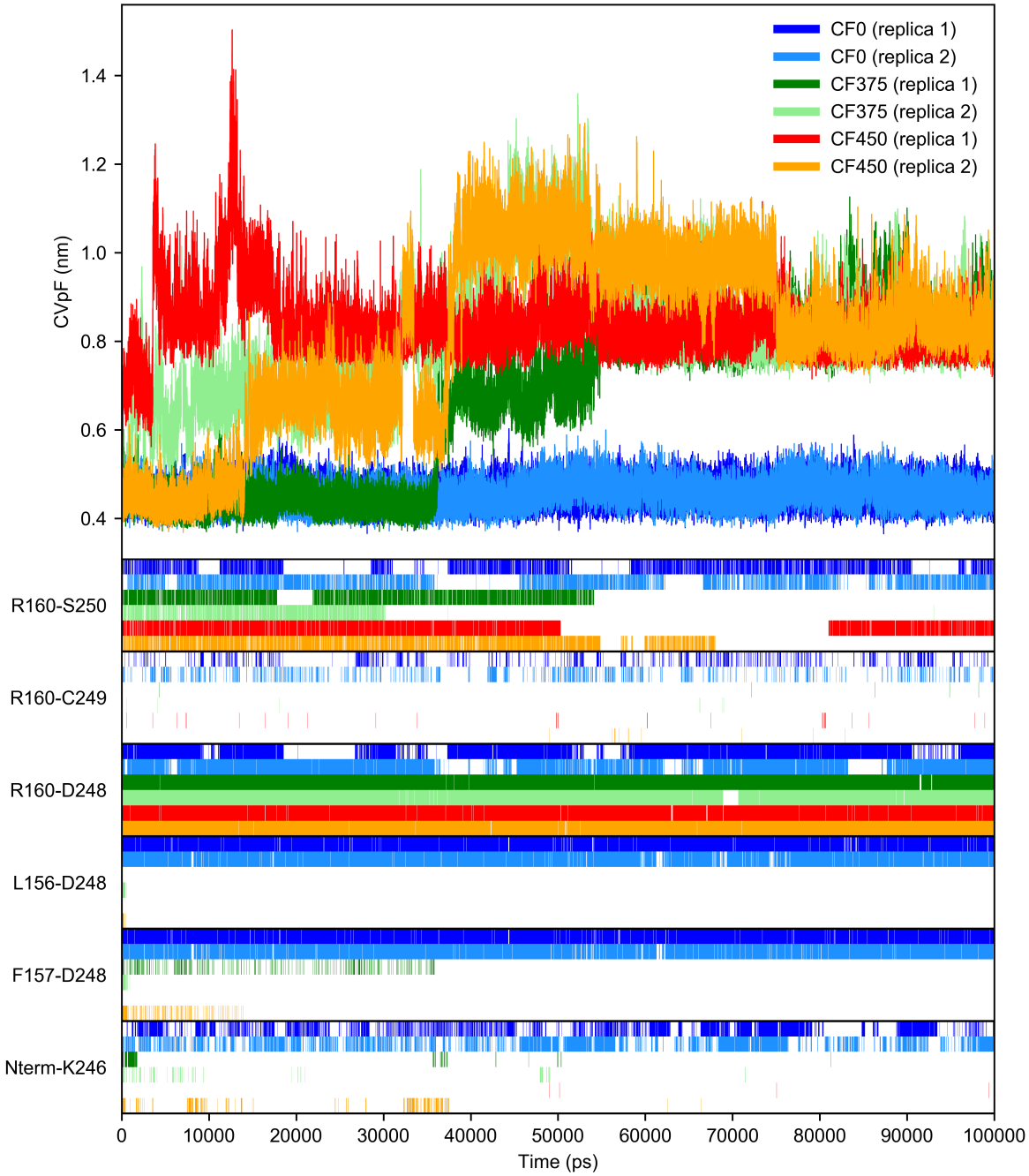


Figure S6. Opening of the pF pocket in the CF simulations. The time evolution of the collective variable CV_{pF} is shown in the top panel for the CF0 (blue hues), CF375 (green hues) and CF450 (orange hues) simulations. Hydrogen-bond existence maps are reported in the bottom panels for selected residue pairs.

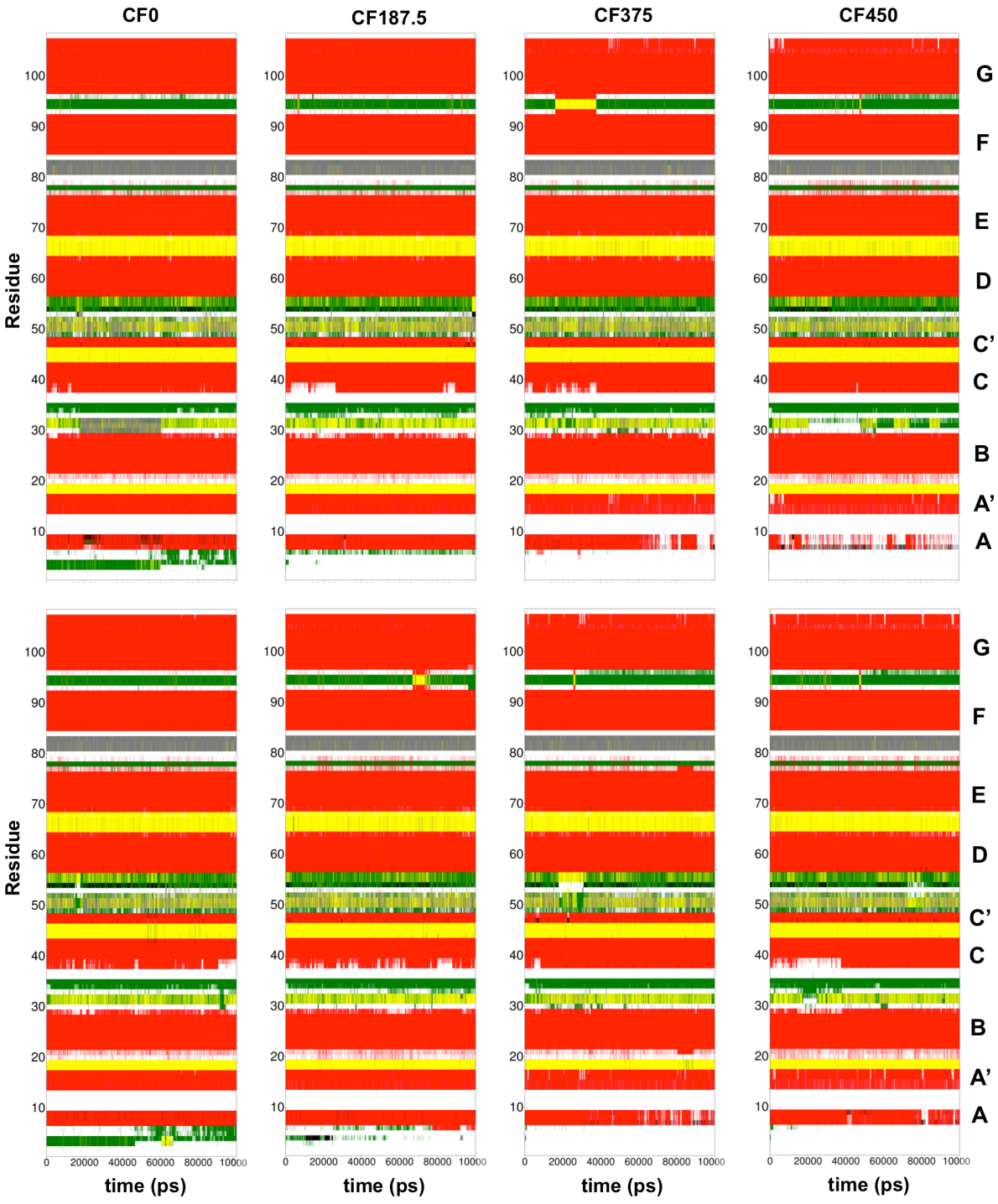


Figure S7. Time evolution of secondary structure in the CFX simulations. The plot shows the DSSP secondary structure for each residue in replica 1 (upper panels) and replica 2 (lower panels) of CFX simulations. A colour code is used for β-strands (red), beta bridges (black), turns (yellow) and bends (green). The names of the β-strands are reported on the right.

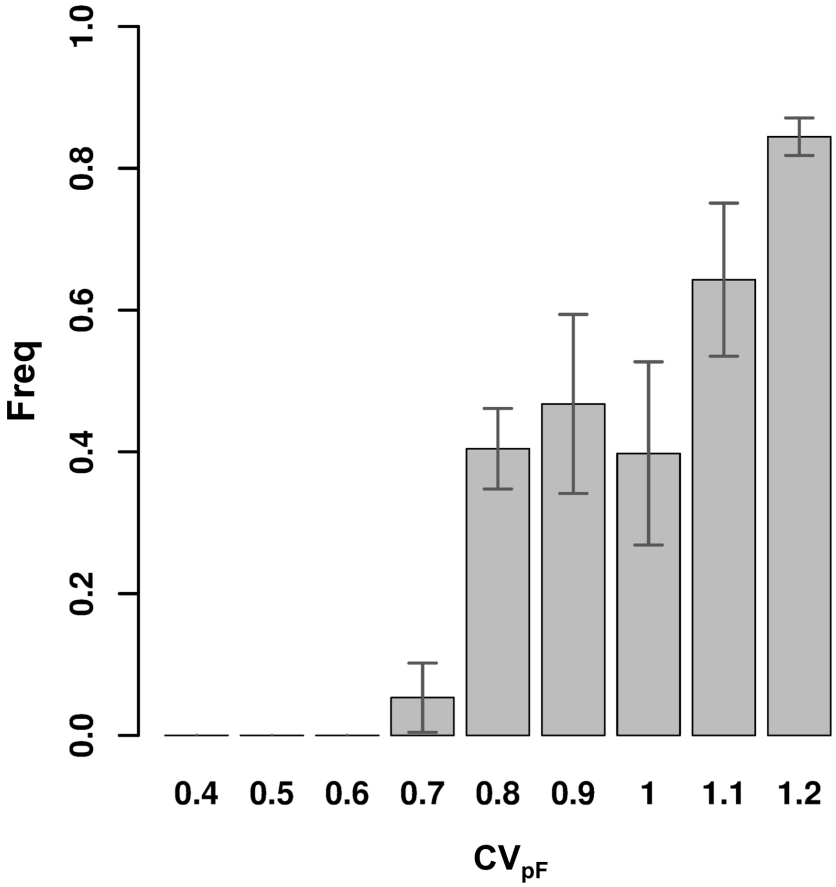


Figure S8. **Relationship between pF pocket volumes and CV_{pF} values.** The structures sampled during the CF450 trajectories were binned according to their CV_{pF} value (0.1-nm intervals) and the fraction of open-pF states (pF volume > 0.17 nm^3) was plotted for each CV_{pF} bin. The grey bars show the average fraction calculated over the two replicas, while the error bars indicate the fraction observed for each replica as minimum and maximum.

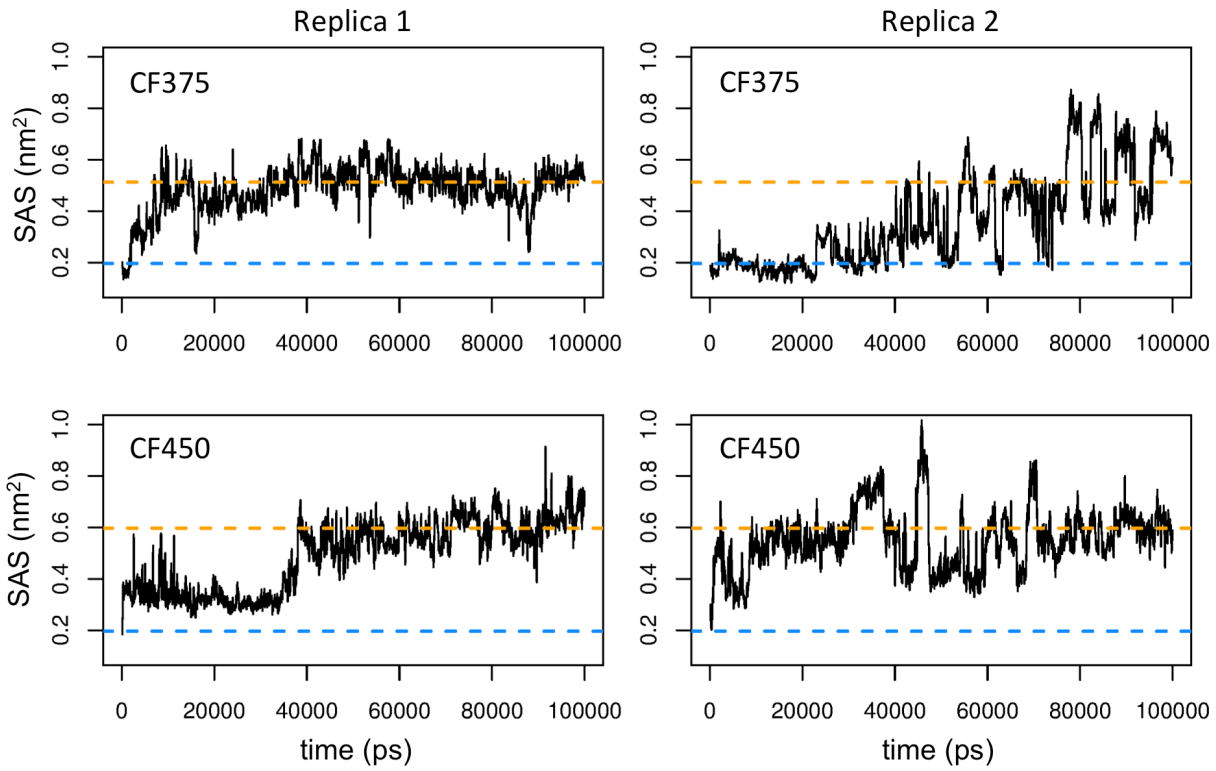


Figure S9. **Solvent accessible surface (SAS) of representative pF-pocket residues in the CF375 and CF450 simulations.** Residues R177, V241 and D248 were selected since they have the largest SAS change between CF0 and CF450 simulations. SAS values were calculated with GROMACS (g_sas tool). The dashed lines indicate the reference values calculated for the closed-pocket (blue) and open-pocket (orange) states as averages over the CF0 (replica 1) and CF375/CF450 (replica 1, last 50 ns) simulations, respectively.

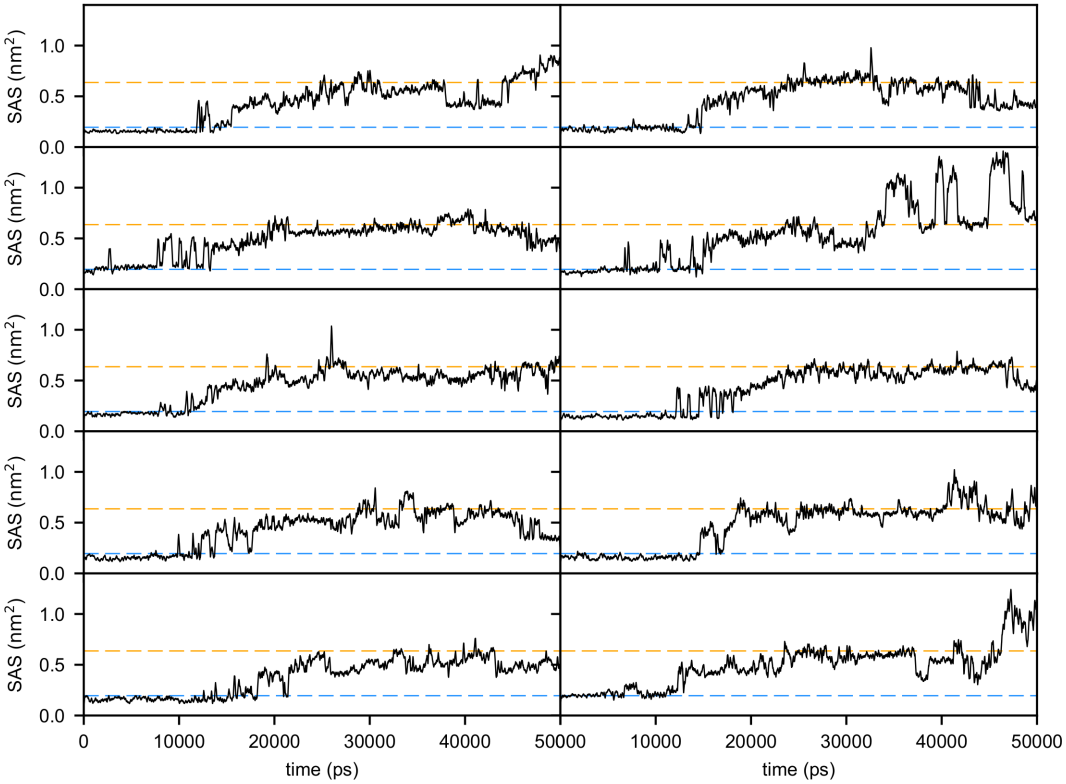


Figure S10. Solvent accessible surface (SAS) of representative pF-pocket residues in the SCV-CF0 simulations. Residues R177, V241 and D248 were selected since they have the largest SAS change between CF0 and CF450 simulations. SAS values were calculated with GROMACS (g_sas tool). The dashed lines indicate the reference values calculated for the closed-pocket (blue, 0.195 nm^2) and open-pocket (orange, 0.632 nm^2) states as averages over the CF0 (replica 1) and CF450 (replica 1) simulations, respectively.

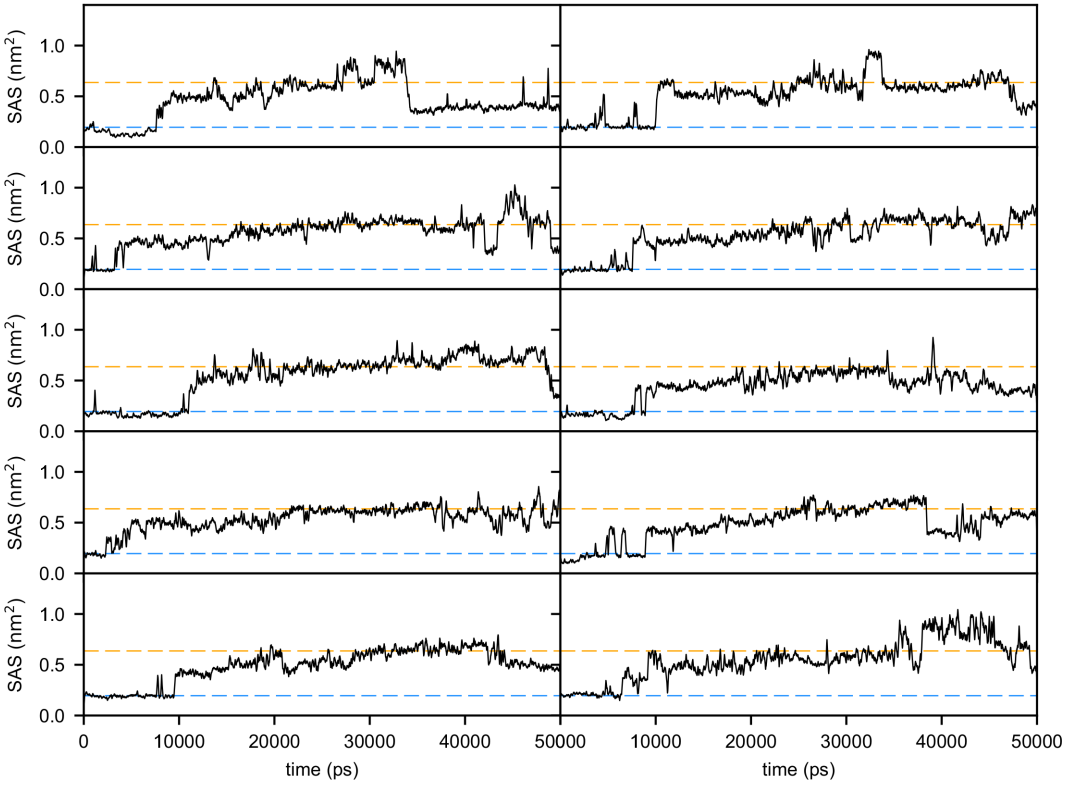


Figure S11. Solvent accessible surface (SAS) of representative pF-pocket residues in the SCV-CF187.5 simulations. See Figure S10 caption.

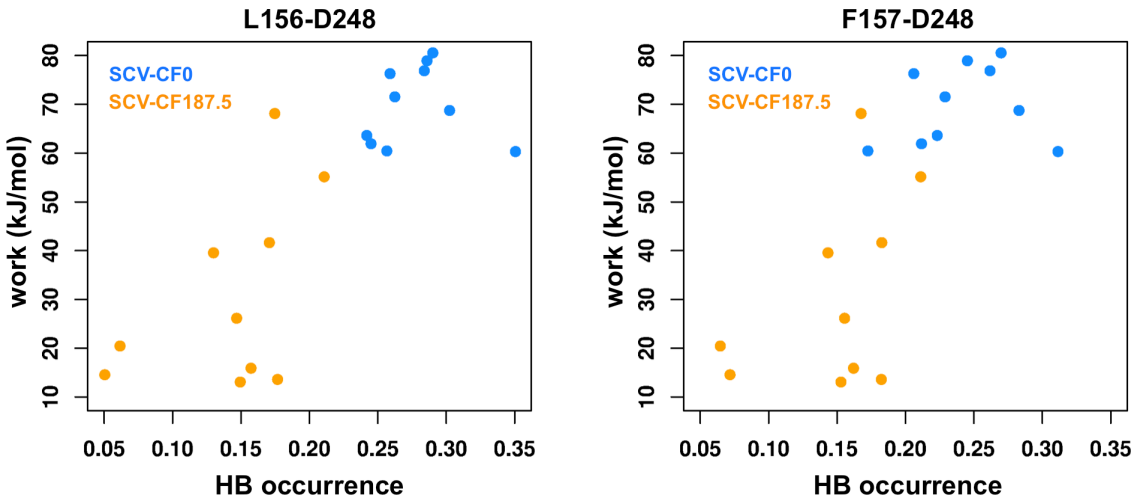


Figure S12. **Correlation between work values and hydrogen bond occurrence in SCV-CF simulations.** The plot shows the work performed to open the pF pocket against the fraction of frames in which a hydrogen bond was observed (HB occurrence) between L156 and D248 (left) and between F157 and D248 (right) in SCV-CF0 (blue) and SCV-CF187.5 (orange) simulations. The Pearson correlation coefficients between work values and HB occurrence are 0.81 for L156-D248 and 0.74 for F157-D248. A larger spread was found in the SCV-CF187.5 simulations compared to the SCV-CF0 ones both for work and HB occurrence values, as also indicated by the standard deviations of the six sets of values (Table S3).

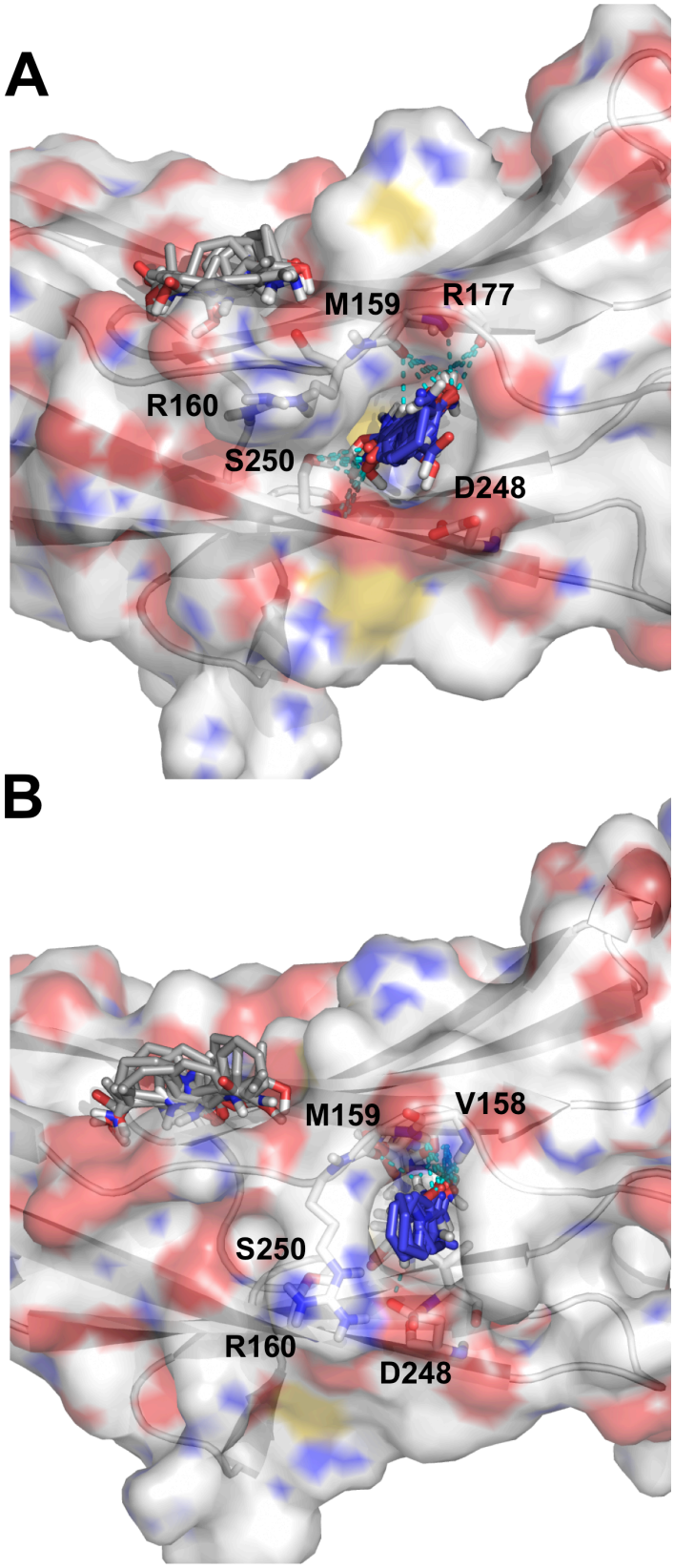


Figure S13. FTMap hotspots. The hotspots (consensus clusters) from the FTMap analysis of the CF375 - replica 2 (A) and CF450 - replica 2 (B) representative structures are shown as sticks. Only hotspots with at least 16 probe clusters are shown. The probes in the pF pocket are highlighted in blue, while probes in a nearby hotspot are visible in grey. Polar contacts between the pF pocket probes and the protein are shown as cyan dashed lines, while the residues involved in the contacts are shown as sticks and labelled. The fragment composition of the pF pocket hotspot is described in Table S4.

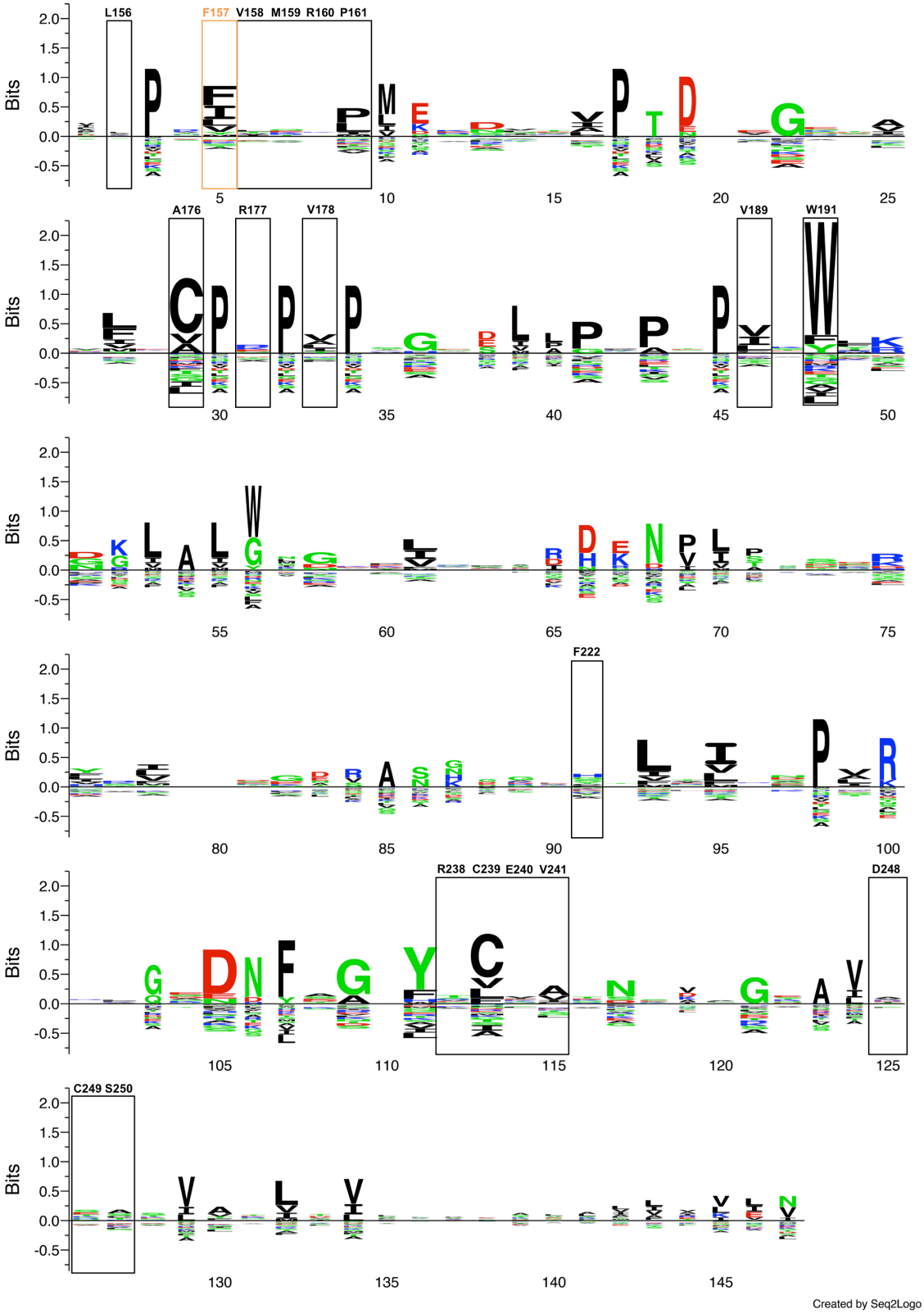


Figure S14. **Sequence logo for the I-set family (Pfam3D subset).** The plot was generated with Seq2Logo[18] and it shows the amino acids enriched (positive *y* axis) and depleted (negative *x* axis) at each position of the multiple sequence alignment of I-set domains with known 3D structure (see Supplementary Methods). The residues lining the pF pocket in the human cardiac My-BPC C1 domain are enclosed in boxes and labelled according to the My-BPC sequence numbering (UniProt ID: Q14896).