

Non-local effects of point mutations on the stability of a protein module

Mateusz Chwastyk, Andrés M. Vera, Albert Galera-Prat, Melissabye Gunnoo, Damien Thompson, Mariano Carrión-Vázquez, and Marek Cieplak

Citation: *The Journal of Chemical Physics* **147**, 105101 (2017); doi: 10.1063/1.4999703

View online: <http://dx.doi.org/10.1063/1.4999703>

View Table of Contents: <http://aip.scitation.org/toc/jcp/147/10>

Published by the [American Institute of Physics](#)

Articles you may be interested in

[Communication: Hypothetical ultralow-density ice polymorphs](#)

The Journal of Chemical Physics **147**, 091101 (2017); 10.1063/1.4994757

[On the relationship between the local segmental dynamics and the tagged monomer dynamics in lamellar phases of diblock copolymers](#)

The Journal of Chemical Physics **147**, 104901 (2017); 10.1063/1.5001022



**COMPLETELY
REDESIGNED!**



**PHYSICS
TODAY**

Physics Today Buyer's Guide
Search with a purpose.

Non-local effects of point mutations on the stability of a protein module

Mateusz Chwastyk,^{1,a)} Andrés M. Vera,² Albert Galera-Prat,² Melissabye Gunnoo,³ Damien Thompson,³ Mariano Carrión-Vázquez,² and Marek Cieplak^{1,b)}

¹*Institute of Physics, Polish Academy of Sciences, Al. Lotników 32/46, PL-02668 Warsaw, Poland*

²*Instituto Cajal, Consejo Superior de Investigaciones Científicas, (CSIC), Ave. Doctor Arce, 37, 28002 Madrid, Spain*

³*Department of Physics, Bernal Institute, University of Limerick, Limerick V94 T9PX, Ireland*

(Received 15 April 2017; accepted 30 July 2017; published online 8 September 2017)

We combine experimental and theoretical methods to assess the effect of a set of point mutations on c7A, a highly mechanostable type I cohesin module from scaffoldin CipA from *Clostridium thermocellum*. We propose a novel robust and computationally expedient theoretical method to determine the effects of point mutations on protein structure and stability. We use all-atom simulations to predict structural shifts with respect to the native protein and then analyze the mutants using a coarse-grained model. We examine transitions in contacts between residues and find that changes in the contact map usually involve a non-local component that can extend up to 50 Å. We have identified mutations that may lead to a substantial increase in mechanical and thermodynamic stabilities by making systematic substitutions into alanine and phenylalanine in c7A. Experimental measurements of the mechanical stability and circular dichroism data agree qualitatively with the predictions provided the thermal stability is calculated using only the contacts within the secondary structures. *Published by AIP Publishing.* [<http://dx.doi.org/10.1063/1.4999703>]

I. INTRODUCTION

Cellulosomes¹⁻³ are multi-protein complexes that convert lignocellulosic biomass, consisting primarily of cellulose and hemicellulose, into simple fermentable sugars. In order to make the production of biofuels economically feasible, the efficiency of this process must be improved. One possible route to enhancement is the use of so-called designer cellulosomes,⁴ with protein domains (and in some cases the geometry of the multi-protein assembly) re-engineered to improve their mechanical and/or thermodynamic stability. The basic structural element of a cellulosome is a long protein known as scaffoldin that consists of a number of cohesins and several other protein domains (cell surface and cellulose binding units, cohesins of other types, and dockerins of other types). The type-I cohesins form pairs with type-I dockerins, and the dockerins may host enzymatic cellulases. The functioning of cellulosomes requires the cohesins to be mechanically strong⁵ and the cellulases to be effective enzymes.⁶⁻⁸ Thus one way to design a better cellulosome is to improve the mechanical stability of the type I cohesins combined with re-engineering of the cellulase units.⁹

Here, we consider point mutations in cohesin domain c7A of scaffoldin from CipA in *Clostridium thermocellum*. This domain is in the “connecting region” between two key attachment points, to the cell surface and to the substrate, and it appears to be subjected to a more intense mechanical stress than other cohesins. Its particularly large mechanical stability

(around 480 pN) has been measured by Valbuena *et al.*⁵ using single-molecule force spectroscopy (SMFS).

We perform an exhaustive computational “alanine screening” via numerical analysis of the structural role of mutations in c7A by systematically replacing each amino acid residue in the protein, one at a time, first by ALA (alanine) and then by PHE (phenylalanine), and measuring the associated change in the structure with respect to the native wild type (WT) protein. Ideally, one would be interested in considering single-site mutations into all twenty residues, but this task is beyond our computational means. Our focus on ALA is motivated by the fact that this choice is commonly adopted in studies of folding pathways through the so-called ϕ -value analysis¹⁰⁻¹³ as ALA is the smallest hydrophobic residue. PHE is still hydrophobic but is much larger in size and contains an aromatic ring. Thus PHE is expected to provide a contrast.

The effect of the transformation is sometimes local (L) as usually obtained by using empirical protein design algorithms such as FoldX^{14,15} and PROTS-RF¹⁶ that rely on compensating adjustments of just the first shell of surrounding side chains. However, we find that in most situations the effect is non-local (NL). Such situations arise, for instance, when the mutation results in a coordinated shift of a relatively ordered region that triggers shifts in the neighbouring region(s) giving significant NL effects well beyond the first shell of the mutation site. In the FoldX approach, the method involves finding a minimal model free energy across a library of rotamers and can be used iteratively for a set of neighboring sites. In contrast, our calculation is dynamical in nature, and it depends on the force-field used. Some mutation-related algorithms, such as Eris,¹⁷ seek to account for mutation-induced NL strain by incorporating backbone pre-relaxation, but these

^{a)}Current address: Department of Physics, Arizona State University, Tempe, Arizona 85287, USA.

^{b)}Electronic mail: mc@ifpan.edu.pl

techniques do not characterise the dynamical changes in the geometry.

Here, we obtain the mutated structures through short atomic-resolution simulations and characterise them using a structure-based coarse-grained (CG) model.^{18,19} This approach is similar to that taken to analyze the properties of proteins in extremophile organisms²⁰ and transient structures of polyglutamines.²¹ The geometry of a structure is described not only by several parameters referring to its overall shape but also by analyzing the list of contacts between the residues. The contacts are determined based on the placement of non-hydrogen atoms in the structure. Upon mutation, some contacts may disappear and some new contacts may arise. For mutations into ALA, only 11% of the structures give purely L changes in which no native contacts break and no new contacts form beyond the immediate mutation site. Purely NL effects arise in 49% of mutations and the rest show a mixture of L and NL changes. For mutations into PHE, we obtain 13% and 52% purely L and purely NL changes, respectively. A purely NL change must involve a local transformation in the geometry, but this transformation does not change the local contact map; the contact lengths (and strengths) may become different or the shifts in the residue positions may affect the number of contacts in the second shell.

The protein mechanostability is assessed by simulating stretching at constant speed, v_p , and calculating the height, F_{max} , of the tallest force peak. F_{max} is a measure of the ease with which the protein gets deformed. We have performed several spot checks and find that the sense of the shifts in F_{max} are consistent both with the SMFS measurements (below) and with all-atom (AA) simulations. An increase/decrease in F_{max} is usually due to the emergence/disappearance of some contacts within the mechanical clamp corresponding to the maximal force peak (see, e.g., Ref. 18). In the case of cohesin, each force peak is due to the shear between two β -strands.⁵

The thermodynamic stability can be captured by estimating the temperature, T_0 , at which the probability, P_0 , of all native contacts being present is $\frac{1}{2}$. It should be noted that various experimental techniques that are used to determine the thermal stability, such as circular dichroism (CD),²² differential scanning calorimetry (DSC),²³ and thermal shift assays,²⁴ typically are not sensitive to all changes in the structure. For instance, CD indicates the type and content of the secondary structure. For this reason, we also study P'_0 , which is defined as the probability of all contacts in the secondary structures being established simultaneously. In other words, P_0 counts conformations in which all contacts are present, whereas P'_0 counts conformations in which the contacts in the secondary structures are established regardless of the state of the other contacts. In the case of c7A, the only secondary structures are the β -sheets, and they contain about 50% of all contacts. We show that the thermal stability assessed through the shifts in T'_0 , the temperature at which P'_0 crosses $\frac{1}{2}$, is distinct from that assessed by shifts in T_0 and is consistent with our experimental CD data.

There are conceptual and practical reasons to use the CG model for the characterisation provided in this study. The primary conceptual reason is to obtain useful insights by focussing on the contact maps and thus defining the notion of

locality in a precise manner. In particular, the perspective of the contact map allows us to develop a method to provide an equilibrium measure of thermal stability—a measure that is well established in studies of lattice models of proteins.^{25–27} Assessing the thermal stability in AA models is difficult to achieve in a quantitative way. In an equilibrium run, one may study the temperature (T) dependence of some global parameter such as the radius of gyration, R_g , or the specific heat, C_v . However, the relation of such parameters to the folding temperature is either unclear (R_g) or doubtful (C_v).²⁸ Performing AA equilibrium simulations at a string of T s is very demanding. Therefore, one typically resorts to qualitative non-equilibrium measures such as obtained through submitting the system to one elevated T and then comparing the temporal evolution of, for instance, root-mean-square deviation (RMSD) in unfolding trajectories corresponding to WT and mutated systems. An example of this approach can be found in Ref. 29.

The practical reason is that one can perform the time evolution from the native structure but with a modified contact map, instead of trying to pick one typical conformation from an AA run. In this way, the data are inherently biased towards the experimental structure so our results are less dependent on the fine details of the molecular dynamics (MD) force field used. Furthermore, in AA stretching simulations, the values of v_p that are computationally feasible are orders of magnitude larger than those used in SMFS experiments^{30–34} and further orders of magnitude away from those encountered *in vivo*. By contrast, our CG model operates under near-experimental conditions and provides a well-benchmarked method to compute average values of F_{max} to assess the effects of mutations.³⁵ It should also be noted that the physics of stretching depends on the speed. At all-atom speeds, non-native contacts may be relevant for the dynamics, as evidenced, for instance, in Ref. 36 in the context of *Dictyostelium discoideum* filamin. It is likely, however, that the formation of such contacts is of no consequence at the experimental or physiological speeds.

The mechanical clamp that is ruptured at F_{max} in c7A corresponds to the terminal β -sheet, which is formed from segments 1–16 and 136–147 in two antiparallel β -strands (Fig. 1). Except for the G142F mutant, we find that point mutations into ALA or PHE within the clamp have only a minor effect on F_{max} and T_0 . We identify other sites that do provide a substantial increase in F_{max} , but, as expected, some of these mutations increase T_0 and some decrease it. The mutations that are identified as leading to substantial simultaneous improvements in both F_{max} and T_0 are substitution of VAL for ALA at residue site number 100, henceforth denoted as V100A, together with G126A, K78A, and T90F. All of these involve purely NL rearrangements. Notable examples of situations in which an increase in F_{max} is associated with a decrease in T_0 include I52A and N47F—both involve only L changes. Our experimental results do not come with a resolution that would allow us to probe the rearrangements. However, they are found to be consistent with our theoretical predictions if T'_0 is used to assess the thermodynamic stability.

We also investigated the effects of introducing disulphide bonds (DBs) into the protein. We find that double-site mutations that generate covalent DBs within the mechanical clamp make the protein extremely resistant to stretching. There is a

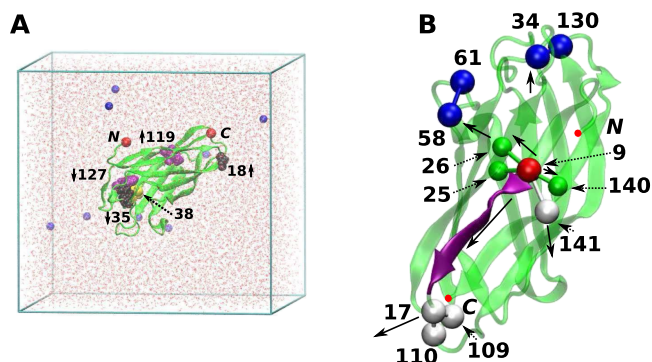


FIG. 1. (a) Simulation box showing the solvated cohesin protein prior to mutation (the green structure). The red spheres denote the terminal residues. The dark-gray spheres highlight areas of large calculated changes in the protein mechanostability on making point mutations into ALA. The larger changes are indicated by a solid arrow; an up-pointing arrow means an increase and a down-pointing one means a decrease. The purple spheres correspond to sites leading to substantial calculated changes in T_0 . The yellow sphere at site 38 (the dotted arrow) indicates the location of the CYS residue. Na^+ ions are represented by the blue spheres, and the small spots around the protein are water molecules. (b) Structural changes upon mutation K9A in c7A. The mutated residue is shown as a large red sphere. The small red spheres correspond to the termini N and C. Contacts at site 9 that persist between the native and mutant states are shown as green sticks connected to green spheres. These are (9–26, 5.11, +0.03), (9–25, 5.91, +0.01), and (9–140, 4.79, +0.14), where the first entry in each bracket indicates sites connected by the contact, the second—the native contact length in Å, and the third—the change in this length on mutation in Å. All conserved contacts slightly modify their average length. The contacts that disappear are shown as gray sticks connected to at least one gray sphere. These are (9–141, 6.26, +0.02), (17–110, 4.80, +1.85), and (17–109, 5.61, +1.59). The created contacts are shown as blue sticks connected to blue spheres. These are (34–130, 6.16, –0.03) and (58–61, 7.02, –0.02). The solid arrows indicate movements of the residues upon mutation. The dotted arrows point to some of the residues that are labelled. The β -strand in purple extends from site 9 to 17. It gets elongated from 23.95 Å to 24.25 Å. The distance between sites 26 and 58, natively equal to 10.08 Å, extends by 0.02 Å allowing for the formation of a contact between 58 and 61.

smaller enhancement of the mechanical stability if the DBs are introduced in other places in the protein.

II. METHODS

A. Theoretical

1. Determination of the mutated structures

The Protein Data Bank (PDB) structure code of c7A is 1AOH.³⁷ The sequence length of 1AOH is 147. The resolved structure of chain A of 1AOH has no α -C atoms in the first four residues and no coordinates of the side-chain atoms beyond β -C at nine sites: LYS9, LYS16, ASN47, GLU50, GLU53, ARG77, SER110, LYS136, and GLN138. For the identical chain B sequence, no α -C atoms are missing, but there are nine defective side groups whose list partially overlaps with that for chain A. We reconstructed the full structure of 1AOH through a combination of methods (statistical and small scale AA MD simulations) as described in Ref. 38. We use this repaired structure of the WT native protein.

Our AA simulations were conducted using the NAMD code version CVS-2013-11-07 for Linux-x86_64-MPI³⁹ with the CHARMM22 force field used to describe the potential energy surface of the solvated protein.^{40,41} Starting mutant structures were obtained by replacing a single residue in the native protein with a new residue using Mutator Plugin,

Version 1.3, in the VMD package.⁴² The simulations were performed using a water box with dimensions of 85 Å in length, 80 Å in width, and 80 Å in as shown in Fig. 1. The protein was rotated to give a 20 Å margin between protein and the boundaries of the box. The box contained ~16 680 TIP3P molecules of water.⁴³ The charge neutrality was achieved by adding about seven Na^+ ions, depending on the charge of the residue(s) being mutated. The total number of atoms in the system was ~52 000. The cutoff radius for non-bonded interactions was set to 1.2 nm. We used the smooth particle mesh Ewald procedure⁴⁴ with a grid spacing of 0.16 nm to treat the long range electrostatics. This is a variant of the particle mesh Ewald method⁴⁵ implemented in NAMD. Periodic boundary conditions were applied in all directions.

The first part of the simulations involved energy minimization of the system for 50 000 conjugate gradient steps. In the second part, the system was heated up to T of 310 K. The heating process was done in three stages, each lasting for approximately 100 ps, first at $T = 110$ K, then 210 K. The last stage was conducted at 310 K and the ensemble switched from constant volume to constant pressure to model the cell at atmospheric pressure. The temperature was controlled by the standard Langevin algorithm and the pressure by the Langevin piston pressure algorithm. The time step was 1 fs.

The conformations of the structure were then captured every 0.002 ns to generate 62 frames for each mutant. Each frame was analyzed to determine the instantaneous contact map, which was derived by checking for the existence of atomic overlaps. Specifically, each non-hydrogen heavy atom is represented by a sphere of radius equal to the van der Waals radius of the atom enlarged by a factor of 1.24 to account for attraction.⁴⁶ A contact is said to exist if at least two such spheres from two amino acids overlap. In the terminology of Ref. 34, this procedure is denoted as OV.

Our results are reproducible (provided the parallelisation is sufficiently extended—of order 100 cores) as tested by redoing selected calculations with different random numbers. However, we also performed runs of 1 ns (instead of 0.125 ns) for I52A, P63A, I80A, V118A, and G143A. We have observed that the longer simulations resulted in the emergence of about 8% additional contacts, but they were all very weak (with a typical well-depth of 0.02 ϵ). Some of the native contacts got weakened but by not more than 5% (2.5% on average). Neither of these changes had any substantial effect on the calculated mechanical and thermal properties. In particular, the force vs. displacement curves were modified in a way that were within the thermal fluctuations of the shorter runs.

2. The CG model

Comparison of the mechanical and thermodynamical properties of the mutants was done within the CG model in which the contacts between amino acids i and j were represented by the Lennard-Jones potential $V(r_{ij}) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$, where r_{ij} is the distance between the α -C atoms in i and j , and σ_{ij} are chosen so that the r_{ij} 's in the native contacts are in the minima of the contact potentials. This model has been tested to be consistent with the experimental data on stretching of 38 proteins.¹⁹

Many contacts persist through all of the 62 frames. In this situation, ϵ_{ij} is set to a constant parameter ϵ , which has been calibrated to be about 106–110 pN Å.^{19,34} The parameters σ_{ij} are calculated from the average WT distance r_{ij} . Other contacts exist only in a fraction, f_{ij} , of the frames. The corresponding values of ϵ_{ij} are taken to be equal to $f_{ij}\epsilon$. To determine the values of σ_{ij} in the partially occupied frames, we consider two models: model I in which the σ_{ij} s are determined from the native distances; model II in which one determines the average r_{ij} in the occupied frames and calculates σ_{ij} from these average values. The α -C atoms are connected harmonically along the chain. All other interactions between the α -C atoms are softly repulsive and set to a repulsive distance of 4 Å. The backbone stiffness is modelled by a chirality term.

Stretching of each structure, mutated or WT, was performed at $k_B T = 0.3\epsilon$ and $v_p = 0.005 \text{ \AA}/\tau$ (k_B is the Boltzmann constant), where τ is of order 1 ns (the time scale results from the overdamped motion of the effective atoms). It is accomplished by attaching elastic springs to the termini and anchoring one of them.¹⁸ The T used is close to the optimal folding conditions and is of the order of room/body temperature. The values of F_{max} used in comparisons are averaged over 25 trajectories for each mutant and 300 trajectories for the WT. For the WT, $\langle F_{max} \rangle$ is 3.67 $\epsilon/\text{\AA}$ in model I and 3.87 $\epsilon/\text{\AA}$ in model II. It should be noted that both models are distinct from a “canonical” model, in which the contact map is determined based on atomic overlaps in the native state instead of allowing for a fractional occupation of a contact arising in an AA evolution around the native state. The canonical model yields F_{max} of 4.2 $\epsilon/\text{\AA}$ ³⁸ but is of little utility when determining changes induced by mutations. Probability P_0 is calculated from two equilibrium trajectories of 30 000 τ . The trajectories start from the native state and the first 1000 τ are used for equilibration. We did spot checks using ten trajectories to make sure that the results are robust, yielding an error of ± 0.001 on P_0 .

The local geometry of the mutated structures is described by the contact maps. The global geometry is characterised by three parameters: root-mean square deviation (RMSD), R_g , and the distortion parameter w . In the CG model, RMSD involves the α -C atoms and is calculated with reference to the PDB structure and then averaged over 30 000 τ . This procedure also applies to the WT situation and it yields RMSD_{WT} of 7.8 Å. It is convenient to present results in terms of d_s defined as $(\text{RMSD} - \text{RMSD}_{WT})/\text{RMSD}_{WT}$. Parameter w is dimensionless and is defined in terms of the three main radii of gyration derived from the tensor of inertia:⁴⁷ $w = \frac{\Delta R}{\bar{R}}$ where $\bar{R} = \frac{1}{2}(R_1 + R_3)$ and $\Delta R = R_2 - \bar{R}$. Here, $R_\alpha = \sqrt{D_\alpha/M}$, where D_1 is the smallest eigenvalue of the tensor of inertia and D_3 is the largest. $w = 0$ corresponds to a spherical symmetry, $w > 0$ to elongated shapes and $w < 0$ to flattened shapes.

3. All-atom stretching

c7A was modelled with the CHARMM36^{40,48} force field and solvated with the TIP3P water model.⁴⁹ The water box of size 500 Å length and 60 Å width and height, was large enough to ensure that the protein is solvated at all times throughout the stretching. The model was neutralised with NaCl and periodic boundary conditions were applied in all directions. The whole

protein-water system contained $\approx 280\,000$ atoms. Prior to MD simulations, all systems were submitted to an energy minimization and equilibration. The protocol involved increasing T in three steps from 110 K to 210 K and 310 K. All simulations were performed using the NAMD package.^{39,50} The point-mutant structures were generated using the Mutator Plugin Version 1.3, from VMD software.⁴² In the case of cysteine mutants, the DB between the sulfur atoms was patched using VMD.

Steered molecular MD simulations (SMD)^{39,51} of 1AOH and its mutants were performed by anchoring one terminus and attaching the other to a spring of force constant 8 (kcal/mol)/Å². The spring was moving with v_p of 50 Å/ns. Prior to pulling, a 20 ns simulation of 1AOH in a smaller water box ($\approx 60\,000$ molecules) was executed in the NPT ensemble (fixed pressure, P , temperature, and the number of atoms, N) with T maintained at 310 K using Langevin dynamics for pressure kept at 1 bar and temperature coupling. Particle mesh Ewald (PME)⁴⁵ method was used to treat long-range electrostatic interactions. The SHAKE algorithm⁵² was applied to constrain all bond lengths involving hydrogen atoms. For each system, seven simulations were conducted, using the same starting configuration and trajectories were sampled every 10 ps for analysis. Unfolding forces correspond to the maximum force peak in the force-extension profile. Structural stability of 1AOH was assessed through RMSD (based on the heavy atoms), showing backbone and sidechain values of $1.11 \pm 0.13 \text{ \AA}$ and $1.82 \pm 0.19 \text{ \AA}$, respectively, consistent with the literature values for similar sized proteins.^{53,54} In the CG model, the α -C-based RMSD evaluated over 20 ns used in the AA simulations is about 1.34 Å, i.e., about 20% of the estimate obtained after averaging over 30 000 τ , which is 1500 times longer. The final structure of the AA preparatory evolution was used as the starting point in the SMD pulling simulations.

B. Experimental

pQE31-c7A⁵ was used as a template to generate the selected mutants. Mutations were introduced in the oligonucleotides used for PCR (for Q138F, G142F, and F127A) or by using overlapping PCR including the mutation (Y44F and I119A). Thermal stability mutants were cloned into pET-24 d (Novagen) using the NcoI XhoI sites, resulting in constructions where a cohesin module preceded a hexahistidine tag. G142F c7A mutant was cloned into KpnI XbaI restriction sites of the pAFM vector.⁵⁵ Plasmids were transformed into BL21 star (DE3) and cells were grown in LB media until an $\text{OD}_{595 \text{ nm}}$ of 0.5–0.8 was reached. Then expression was induced by the addition of 300–500 μM IPTG and incubation at 16 °C for 4–20 h. Cells were harvested by centrifugation at 6000 g for 10 min and lysed in 50 mM phosphate, 500 mM NaCl, 50 mM Imidazole, pH 7.4 buffer using 1 mg/ml lysozyme, 5 $\mu\text{g}/\text{ml}$ RNase A, 5 $\mu\text{g}/\text{ml}$ DNase, and 1% Triton X-100. After centrifugation at 39 000 g for 30 min, samples were purified by Ni²⁺ affinity in HisTrap HP columns (GE Healthcare, Uppsala, Sweden) in an FPLC (ÄKTA Purifier, GE Healthcare) and eluted in a buffer containing 500 mM imidazole. Protein samples were concentrated in 50 mM phosphate, 100 mM NaCl, 1 mM DTT, 0.5 mM EDTA, pH 7.6 buffer and further purified by size

TABLE I. The summary of the results pertaining to stability and geometry for point mutations into ALA in c7A. The first column shows the site of the mutation and the second column shows the residue type being mutated. The third and fourth columns show the changes in F_{max} as calculated within model II and I, respectively. All results are rank-ordered by the change in F_{max} obtained using model II, and the black horizontal line separates mutations that give predicted increase and decrease in F_{max} . The fifth and sixth columns show the changes in T_0 in model II and I, respectively. The errors for T_0 are of order $0.0002 \text{ } \epsilon/k_B$. The seventh column gives d_s , which is a normalized deviation from the average WT RMSD. The next two columns provide the average values of w and R_g . The penultimate column gives the number of contacts formed as a result of the mutation (they appear in at least 80% of the simulation frames). The last column gives the number of contacts that disappeared (they show in less than 20% of the frames). The first line provides the same data for the WT system.

No.	AA	$\Delta F_{max}^{\epsilon, \sigma} (\text{ } \epsilon/\text{Å})$	$\Delta F_{max}^{\epsilon} (\text{ } \epsilon/\text{Å})$	$\Delta T_0^{\epsilon \sigma} (\text{ } \epsilon/k_B)$	$\Delta T_0^{\epsilon} (\text{ } \epsilon/k_B)$	d_s	w	$R_g (\text{Å})$	>80%	<20%	
0	WT	0.00 ± 0.00	0.00 ± 0.00	0.000	0.000	0.000	0.72	15.48	0	0	
18	GLY	0.30 ± 0.06	0.18 ± 0.06	0.000	0.000	-0.038	0.71	15.48	7	-2	
52	ILE	0.28 ± 0.07	0.16 ± 0.06	-0.003	-0.005	-0.051	0.72	15.48	-2	4	
100	VAL	0.23 ± 0.08	0.17 ± 0.06	+0.003	0.000	-0.013	0.72	15.49	2	4	
91	GLY	0.19 ± 0.06	0.10 ± 0.06	0.000	-0.003	-0.051	0.71	15.49	7	4	
55	GLU	0.17 ± 0.07	0.10 ± 0.06	-0.002	-0.005	-0.026	0.72	15.48	1	2	
141	ASP	0.13 ± 0.06	0.14 ± 0.07	+0.006	0.001	-0.064	0.72	15.47	3	3	
96	THR	0.13 ± 0.06	0.02 ± 0.07	-0.006	-0.009	0.128	0.71	15.56	-2	8	
126	GLY	0.11 ± 0.08	0.12 ± 0.08	+0.010	+0.006	0.090	0.70	15.47	-4	6	
97	GLU	0.11 ± 0.07	0.04 ± 0.07	+0.008	+0.006	-0.051	0.71	15.46	7	2	
17	PRO	0.10 ± 0.07	0.11 ± 0.06	0.000	0.000	-0.077	0.71	15.45	5	2	
28	SER	0.08 ± 0.09	0.16 ± 0.07	0.000	+0.005	0.103	0.72	15.48	3	4	
37	ASN	0.06 ± 0.08	0.12 ± 0.06	+0.005	+0.003	0.000	0.72	15.45	1	5	
74	TYR	0.05 ± 0.09	-0.01 ± 0.06	-0.005	-0.006	0.038	0.71	15.48	-2	5	
78	LYS	0.05 ± 0.06	0.07 ± 0.06	+0.009	+0.008	0.051	0.71	15.51	5	2	
44	TYR	0.04 ± 0.06	-0.02 ± 0.07	+0.009	+0.006	0.141	0.72	15.46	-11	14	
86	GLU	0.04 ± 0.06	0.03 ± 0.06	-0.007	-0.008	0.026	0.72	15.49	-7	10	
93	TYR	0.04 ± 0.07	0.06 ± 0.06	-0.010	-0.006	0.090	0.72	15.50	2	5	
10	VAL	0.03 ± 0.07	0.01 ± 0.05	-0.006	-0.006	-0.038	0.71	15.49	3	2	
59	LEU	0.02 ± 0.09	0.06 ± 0.08	+0.009	+0.006	-0.064	0.72	15.45	-4	5	
60	ILE	0.02 ± 0.08	0.01 ± 0.05	-0.010	-0.009	0.000	0.71	15.49	-2	9	
89	GLY	0.01 ± 0.07	-0.03 ± 0.06	-0.001	-0.006	0.128	0.73	15.49	-3	7	
69	PHE	0.00 ± 0.07	0.06 ± 0.05	-0.005	-0.006	0.064	0.71	15.50	-2	10	
82	PHE	0.00 ± 0.06	-0.04 ± 0.06	-0.004	-0.006	0.103	0.71	15.47	4	11	
98	ASP	-0.02 ± 0.08	-0.04 ± 0.06	+0.005	+0.009	0.115	0.73	15.48	0	10	
32	SER	-0.06 ± 0.07	-0.04 ± 0.06	+0.010	+0.009	0.051	0.71	15.48	-4	2	
62	ASP	-0.06 ± 0.06	0.01 ± 0.07	-0.006	-0.006	0.103	0.71	15.52	6	4	
145	ASN	-0.08 ± 0.06	-0.04 ± 0.06	+0.006	+0.006	-0.038	0.72	15.46	3	7	
140	PHE	-0.11 ± 0.06	-0.07 ± 0.06	-0.010	-0.008	0.013	0.71	15.49	9	4	
143	GLY	-0.12 ± 0.07	-0.11 ± 0.06	0.000	0.000	-0.013	0.71	15.47	-7	4	
24	PRO	-0.12 ± 0.08	-0.10 ± 0.07	-0.006	-0.006	0.115	0.69	15.47	-7	5	
108	VAL	-0.13 ± 0.08	-0.12 ± 0.07	-0.009	-0.006	0.000	0.70	15.46	0	4	
127	PHE	-0.13 ± 0.07	-0.08 ± 0.06	-0.011	-0.010	0.064	0.71	15.47	-6	11	
76	ASP	-0.13 ± 0.08	-0.05 ± 0.06	+0.005	+0.006	-0.038	0.72	15.48	-1	11	
6	VAL	-0.14 ± 0.06	-0.01 ± 0.06	-0.006	-0.009	0.013	0.72	15.50	2	4	
56	PRO	-0.18 ± 0.07	-0.09 ± 0.06	0.000	0.000	0.103	0.71	15.53	-6	1	
61	VAL	-0.18 ± 0.07	-0.08 ± 0.06	-0.010	-0.006	0.013	0.72	15.51	2	4	
99	GLY	-0.19 ± 0.07	-0.06 ± 0.06	0.000	0.003	0.077	0.72	15.51	-4	4	
129	ASN	-0.20 ± 0.04	-0.11 ± 0.06	+0.003	0.000	-0.115	0.73	15.47	-5	0	
47	ASN	-0.21 ± 0.06	-0.12 ± 0.06	-0.002	-0.006	0.103	0.71	15.51	-4	7	
119	ILE	-0.22 ± 0.06	-0.09 ± 0.06	+0.014	+0.014	0.051	0.71	15.50	-8	7	
45	ASP	-0.22 ± 0.05	-0.13 ± 0.06	-0.010	-0.003	0.128	0.72	15.50	3	2	
116	LEU	-0.23 ± 0.06	-0.17 ± 0.05	-0.002	0.000	0.000	0.71	15.51	-4	10	
8	ILE	-0.26 ± 0.07	-0.16 ± 0.05	0.000	0.000	-0.038	0.72	15.46	-1	8	
38	CYS	-0.27 ± 0.06	-0.15 ± 0.05	-0.005	0.000	-0.051	0.73	15.49	3	3	
63	PRO	-0.28 ± 0.06	-0.21 ± 0.05	-0.004	-0.001	-0.026	0.71	15.49	2	5	
101	PHE	-0.30 ± 0.06	-0.15 ± 0.06	0.000	0.000	0.064	0.72	15.45	-4	12	
142	GLY	-0.30 ± 0.06	-0.19 ± 0.06	0.000	0.000	0.064	0.72	15.51	-10	5	
9	LYS	-0.30 ± 0.06	-0.30 ± 0.05	-0.006	-0.002	0.115	0.72	15.51	-10	14	
121	PHE	-0.31 ± 0.07	-0.23 ± 0.06	0.000	-0.001	0.013	0.71	15.49	-2	6	
139	PHE	-0.31 ± 0.06	-0.27 ± 0.06	-0.009	-0.006	0.141	0.73	15.54	-6	8	
35	ILE	-0.32 ± 0.05	-0.27 ± 0.05	-0.001	0.000	-0.038	0.72	15.49	-4	9	
Average:							0.020 ± 0.006	0.716 ± 0.001	15.480 ± 0.002		

TABLE II. Similar to Table I but for mutations into PHE.

No.	AA	$\Delta F_{max}^{\epsilon,\sigma}$ ($\epsilon/\text{\AA}$)	$\Delta F_{max}^{\epsilon}$ ($\epsilon/\text{\AA}$)	$\Delta T_0^{\epsilon\sigma}$ (ϵ/k_B)	ΔT_0^{ϵ} (ϵ/k_B)	d_s	w	R_g (\AA)	>80%	<20%
47	ASN	0.39 ± 0.06	0.21 ± 0.06	-0.006	-0.006	0.000	0.72	15.50	6	0
114	ASN	0.31 ± 0.06	0.23 ± 0.06	0.000	-0.003	0.077	0.70	15.53	4	8
125	GLY	0.28 ± 0.08	0.15 ± 0.07	+0.003	0.000	-0.026	0.72	15.47	2	1
96	THR	0.26 ± 0.06	0.07 ± 0.06	+0.004	+0.005	-0.026	0.72	15.42	3	7
86	GLU	0.24 ± 0.08	0.20 ± 0.06	-0.006	-0.006	0.038	0.72	15.48	-1	10
90	THR	0.23 ± 0.06	0.15 ± 0.08	+0.006	+0.006	0.038	0.72	15.45	4	8
128	ALA	0.20 ± 0.06	0.10 ± 0.06	+0.003	+0.009	0.000	0.72	15.45	10	1
58	GLU	0.18 ± 0.06	0.22 ± 0.08	+0.005	+0.003	0.051	0.72	15.44	7	7
12	THR	0.17 ± 0.08	0.09 ± 0.07	-0.006	-0.003	0.077	0.72	15.50	-1	3
110	SER	0.16 ± 0.07	0.04 ± 0.06	+0.005	0.000	0.013	0.71	15.51	-2	7
137	THR	0.16 ± 0.06	0.09 ± 0.08	+0.005	0.000	0.064	0.72	15.47	8	7
45	ASP	0.14 ± 0.08	0.02 ± 0.07	-0.005	-0.005	0.026	0.71	15.51	-4	7
19	ASP	0.12 ± 0.06	0.12 ± 0.06	-0.006	-0.003	0.077	0.71	15.46	3	4
2	ASP	0.09 ± 0.06	0.04 ± 0.07	+0.006	+0.003	0.013	0.73	15.43	4	8
34	GLY	0.09 ± 0.07	0.03 ± 0.07	-0.006	-0.003	0.526	0.71	15.57	-22	18
52	ILE	0.09 ± 0.07	0.02 ± 0.05	-0.005	-0.003	0.026	0.72	15.46	8	9
17	PRO	0.07 ± 0.06	0.20 ± 0.07	-0.009	-0.006	0.218	0.72	15.60	-5	8
20	THR	0.07 ± 0.07	0.01 ± 0.07	+0.006	0.000	0.013	0.72	15.48	1	3
83	LEU	0.06 ± 0.07	0.02 ± 0.05	0.000	+0.006	0.064	0.73	15.46	6	2
130	ASN	0.04 ± 0.07	0.02 ± 0.06	+0.005	+0.006	-0.026	0.71	15.39	2	9
131	ASP	0.03 ± 0.06	0.01 ± 0.07	+0.003	+0.006	0.038	0.71	15.43	4	6
138	GLN	0.03 ± 0.07	0.02 ± 0.05	+0.014	+0.009	0.013	0.73	15.51	-2	6
23	ILE	0.03 ± 0.06	0.08 ± 0.05	+0.009	+0.006	0.064	0.72	15.48	7	10
80	ILE	0.02 ± 0.07	0.03 ± 0.07	-0.011	-0.005	0.128	0.70	15.48	-8	9
68	SER	0.00 ± 0.07	0.01 ± 0.05	0.000	-0.006	0.231	0.72	15.56	-9	11
120	LYS	-0.04 ± 0.06	-0.03 ± 0.07	-0.005	-0.005	0.090	0.72	15.54	-1	4
129	ASN	-0.04 ± 0.06	-0.01 ± 0.05	-0.009	-0.005	0.026	0.71	15.49	-1	9
21	VAL	-0.04 ± 0.07	0.00 ± 0.07	-0.005	0.000	0.013	0.72	15.48	2	5
33	LYS	-0.05 ± 0.06	-0.03 ± 0.06	-0.009	0.000	-0.026	0.72	15.49	2	2
24	PRO	-0.06 ± 0.07	0.01 ± 0.07	-0.006	-0.006	-0.026	0.72	15.51	5	2
74	TYR	-0.06 ± 0.08	-0.04 ± 0.06	+0.006	+0.003	0.026	0.71	15.47	1	5
104	ILE	-0.07 ± 0.07	-0.05 ± 0.06	+0.003	+0.006	0.128	0.72	15.51	3	9
141	ASP	-0.07 ± 0.07	0.02 ± 0.05	-0.007	-0.003	0.038	0.71	15.50	-5	2
73	VAL	-0.07 ± 0.08	-0.00 ± 0.07	-0.011	-0.003	0.179	0.71	15.55	-5	9
76	ASP	-0.08 ± 0.08	-0.02 ± 0.07	-0.006	0.000	0.013	0.73	15.50	1	5
26	ARG	-0.09 ± 0.06	-0.11 ± 0.05	+0.006	0.000	0.064	0.72	15.47	1	9
70	ASP	-0.14 ± 0.08	-0.07 ± 0.06	-0.006	-0.003	0.051	0.71	15.49	0	4
88	SER	-0.14 ± 0.06	-0.01 ± 0.05	-0.007	-0.006	0.051	0.72	15.51	-1	5
44	TYR	-0.15 ± 0.07	-0.10 ± 0.06	-0.011	-0.010	0.128	0.72	15.45	-9	12
75	PRO	-0.16 ± 0.07	-0.08 ± 0.06	+0.006	+0.006	0.103	0.72	15.51	-9	2
50	GLU	-0.17 ± 0.05	-0.09 ± 0.06	-0.006	-0.005	-0.026	0.72	15.50	0	7
108	VAL	-0.20 ± 0.05	-0.00 ± 0.05	-0.007	-0.006	0.141	0.72	15.54	-13	9
93	TYR	-0.20 ± 0.06	-0.04 ± 0.06	-0.003	-0.006	0.385	0.72	15.54	-13	13
15	ALA	-0.23 ± 0.07	-0.18 ± 0.06	0.000	0.000	0.077	0.72	15.50	-3	7
31	PRO	-0.24 ± 0.06	-0.13 ± 0.05	0.000	+0.001	0.103	0.72	15.51	0	6
43	SER	-0.24 ± 0.06	-0.13 ± 0.05	0.000	0.000	-0.064	0.72	15.47	5	0
49	LEU	-0.24 ± 0.06	-0.15 ± 0.05	-0.003	-0.004	0.064	0.72	15.54	-3	7
42	TYR	-0.25 ± 0.06	-0.15 ± 0.06	0.000	0.000	0.013	0.70	15.51	-11	4
9	LYS	-0.26 ± 0.06	-0.16 ± 0.05	0.000	0.000	0.192	0.72	15.55	-6	10
144	VAL	-0.31 ± 0.06	-0.26 ± 0.06	0.000	0.000	0.090	0.72	15.49	-8	11
143	GLY	-0.35 ± 0.05	-0.18 ± 0.05	-0.009	-0.009	0.154	0.72	15.46	-3	6
142	GLY	-1.22 ± 0.06	-0.99 ± 0.05	-0.003	0.000	0.449	0.71	15.52	-23	18
Average:						0.041 ± 0.008	0.716 ± 0.001	15.487 ± 0.003		

exclusion chromatography in a Superdex 200 Increase 10/300 GL (GE Healthcare). Clean fractions were concentrated in 10 mM formate, 0.2 mM EDTA buffer at pH 4.0 for thermal stability measurements (measured at room T), and in phosphate

buffered saline (PBS) for AFM-SMFS experiments. All measurements (both for CD and for AFM) have been done under the same buffer conditions, allowing for comparisons within each group.

The experimental methods pertaining to CD and SMFS are described in the [supplementary material](#).

III. RESULTS AND DISCUSSION

A. Geometry of the mutated structures

The calculated global geometry-related parameters are listed in Tables I and II as well as in Tables 1–4 in the [supplementary material](#). As expected, parameters w and R_g are not very sensitive to point mutations. w is between 0.69 and 0.73 indicating cigar-shaped structures in the simulations. R_g changes between 15.39 and 15.57 Å (the range is smaller for mutations into ALA). The changes in the RMSD are usually minor; the values of d_s are mostly close to 0.

We now focus on changes in the contact map. In this section, we consider a new contact to be present if it exists in more than 80% of the frames in the AA simulations. If a native contact shows in less than 20% of the frames—it is considered to be absent. The changes are assessed with respect to the WT runs and not with respect to the native PDB structure. We did spot checks by repeating the simulations with a different thermal noise to make sure that the structural transformations derived are robust.

The results of AA simulations generally depend on the length of the computation, but long simulations are costly. We have observed that parallel simulations on at least 100 cores yield reproducibility even for short trajectories considered here. This is because parallel simulations involve cells with buffers, and many buffers provide effective self-averaging. We have performed spot checks using trajectories that are an order of magnitude longer, and they yielded similar values for the mechanical and thermal stabilities.

Figure 1 for K9A shows an example of the situation in which the changes in the contact map are NL. The average distance between the α -C's in residues 9 and 141 increases from 6.3 to 6.7 Å and the corresponding contact disappears. The contact between 9 and 140 persists but the distance between the α -C's increases from 4.6 to 4.8 Å. The contacts between 9–25 and 9–26 stay. The corresponding distances get slightly shorter for the former (5.9 Å) and a bit longer for the latter (5.1 Å). Furthermore, the β -strand between sites 9 and 17 (shown in purple) gets longer, which destroys the contacts 17–109 and 17–110. Since the 9–26 contact elongates, it pushes residue 58 closer to 61 and 34 to 130, creating two new contacts. Overall, three contacts are destroyed—only one of them locally—and two new non-locally induced contacts are formed, with a net reduction in the thermodynamic and mechanical stabilities.

Figure 2 shows five other examples of mutations (F127A, I119A, Y44F, Q138F, and G142F) in which formation of at least one new contact or annihilation of one native contact occurs far away from the mutated site. The effect is particularly dramatic for Q138F in which the disappearance of the N-terminal contact 1–138 adjusts the structure so that four new distant contacts appear and the thermodynamic and mechanical stabilities both increase (see Table II).

The same figure also shows an example of a structural change that is borderline L, created by the mutation G18A. This mutation does not change T_0 but yields the largest increase in F_{max} in the set of all mutations into ALA (Table I), even though site 18 does not belong to the mechanical clamp. No contact disappears. One new contact, 49–109, is formed. Its formation is due to the following mechanism. The length of contact 18–110 increases significantly allowing site 109 to come closer to 49, from the distance of 6.27 to 5.87 Å. At the

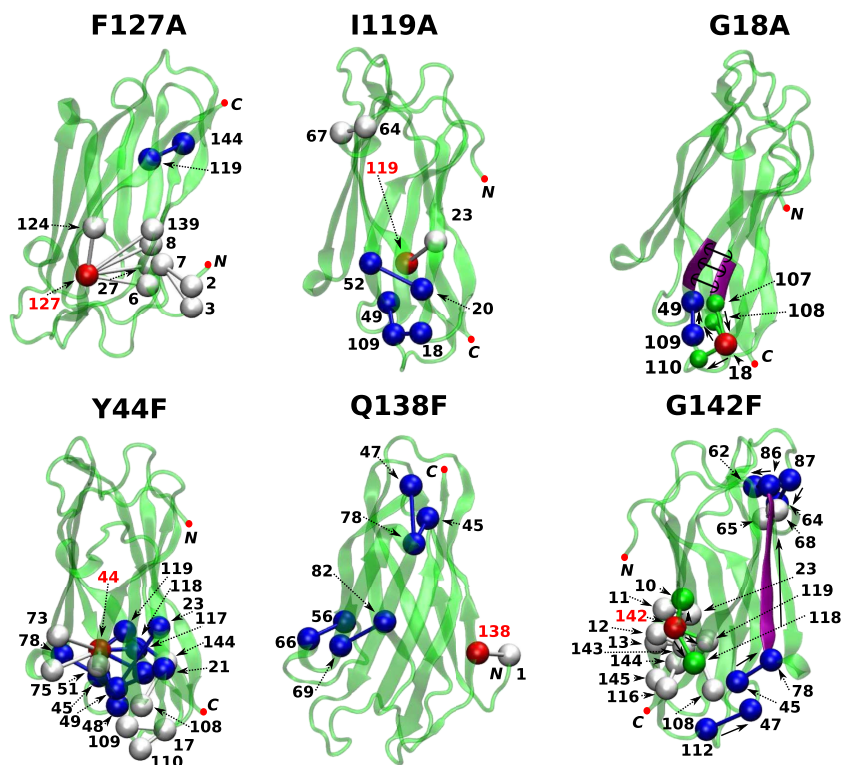


FIG. 2. Examples of changes in the contact map upon mutation. The symbols used are as in panel (a) of Fig. 1. The top two panels are for mutations into ALA. The bottom two panels are for mutations into PHE. The panels on the left correspond to the situations in which the thermodynamic stability decreases the most, and the panels in the middle correspond to the situations in which thermodynamic stability increases. The mechanical stability increases for Q138F, but it becomes smaller in the other cases. In the top-right panel for mutation G18A, the structure shown in purple is a β -sheet that shifts (indicated by the double-headed arrows) when contact distances vary as a result of the mutation. The contact distance in 18–107 decreases from 6.74 Å to 6.55 Å. On the other hand, the distances in 18–108 and 18–110 increase from 5.34 Å to 6.25 Å and from 4.34 Å to 6.60 Å, respectively. The panel corresponding to G142F is discussed in the in the text.

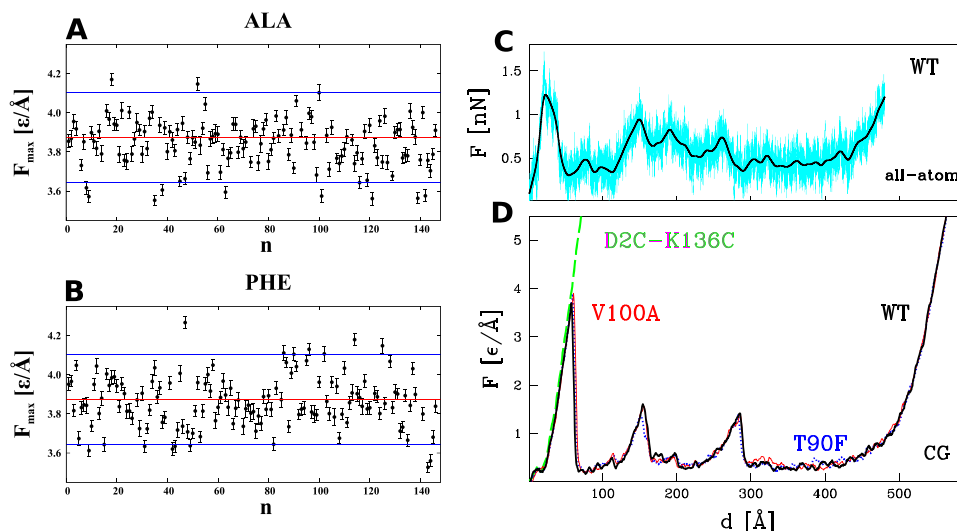


FIG. 5. (a) and (b) Values of F_{max} obtained using model II. Top and bottom: after mutating to ALA and PHE, respectively. The red central horizontal lines correspond to the WT value. The blue horizontal lines delineate the region in which the deviation in F_{max} away from the WT value is less than $0.17 \text{ } \epsilon/\text{\AA}$. Larger deviations indicate values that are significantly different from WT. Mutation G142F results in such a large lowering in F_{max} that it falls beyond the scale of the figure. (c) An example of the dominant stretching trajectory (5 cases out of 7) obtained through AA SMD for the WT cohesin. The average F_{max} , before smoothing, is $1.46 \pm 0.11 \text{ nN}$. In the minority set of trajectories, the force peak at $\approx 180 \text{ \AA}$ is the highest (with F_{max} of $1.26 \pm 0.1 \text{ nN}$); such trajectories do not arise in the CG model at lower v_p values. (d) Single-trajectory examples of the $F-d$ plots for the WT cohesin (black solid line) and the mutated structures: V100A (solid red), T90F (dotted blue), and the doubly mutated D2C-K136C (dashed green). There is no force peak when the DB is made between sites 2 and 136. For the other mutations, the first peak is maximal, and the values of F_{max} are nearly identical. They are both higher than for the WT. The simulations were done using model II.

changes in F_{max} and T_0 . These panels refer to mutations into ALA. Similar plots for mutations into PHE are shown in Fig. 1 in the [supplementary material](#). The insignificant changes are contained within the rectangles in the centers of the figures. The upper-right quadrangles correspond to mutations that yield significant increase both in the mechanical and thermodynamic stabilities. The corresponding mutation sites that are identified both in model I and in model II are marked by the blue digits. These are V100A, G126A, K78A, and T90F. A large increase in F_{max} with no accompanying change in T_0 is observed for G18A (NL changes) and N114F (L changes). Examples of situations in which an increase in F_{max} is associated with a decrease in T_0 include I52A and N47F.

The systematic substitutions into particular amino acids allow one to consider mutations whose impact on the protein stabilities is not obvious. Another approach is to select residues that could be suspected of being important based on them being related to the region of the mechanical clamp. One would expect that making mutations that are likely to increase the number of contacts within the clamp should result in the improved stability. Thus we have considered mutations to Q and W: N14Q, N14W, A15W, K16W, Q135W, K136R, K136W, T137Q, and T137W. In addition, we studied K136R to check the role of the change in the polarity. Stretching implemented within model II showed that only mutation A15W noticeably increases F_{max} , by $0.21 \text{ } \epsilon/\text{\AA}$. Mutations T137W and Q135W are found to decrease the mechanostability by 0.55 and $0.19 \text{ } \epsilon/\text{\AA}$, respectively. All of these mutations do not change T_0 much, less than $0.01 \text{ } \epsilon/\text{\AA}$.

We also considered residues that do not belong to the mechanical clamp but may stabilize it by additional contacts. TRP (tryptophane) is expected to introduce more local

contacts than other residues so we considered mutations V108W, A112W, P113W, N114W, I30W, P31W and the double mutation A15W-K136W. Of these, only P31W and V108W were found to increase F_{max} significantly—by $0.20 \text{ } \epsilon/\text{\AA}$ and $0.23 \text{ } \epsilon/\text{\AA}$, respectively. These results are listed in Table 5 in the [supplementary material](#).

C. Mechanical stability assessed through all-atom simulations

Figure 5(c) shows a typical trajectory for the WT c7A obtained through AA SMD. Even though these simulations are performed at much higher speed than in the CG model, they provide the atomic-scale details. The $F-d$ pattern is overall similar to that in the CG model except that the first force peak arises earlier and its height is much higher. The first effect is primarily due to the stiffer spring constant and the second is related to the different v_p s that are related to the different spatial resolutions of the models.

The mutants tested are G126A, I52A, K78A, V100A, G18A, N114F, N47F, and G142F. The shifts in F_{max} (in the dominant pathway) were found to vary between -0.1 nN and 0.3 nN , the smallest for G142F and the largest for I52A, with the errors of order 0.1 nN , approximately $1 \text{ } \epsilon/\text{\AA}$. The sign of the shift is generally consistent with the data shown in Tables I and II, except for N47F, for which the shift is opposite, and V100A for which there is no shift. It should be noted, again, that mechanostability has no absolute meaning—it depends on the v_p . It is possible that the relevant mechanical clamp in N47F adjusts its sequential location on varying the v_p .

D. Disulphide bonds between cysteines

We now consider the possibility of introducing DBs into the cohesin to boost its mechanostability. Protein c7A contains

one cysteine, at site 38 (see Fig. 1). In WT, this residue forms contacts with L83, F84, F101, G126, and F127. Mutating these residues into CYS could form DBs. Our CG model (the DBs are represented by the harmonic springs) confirms that none of these mutations would have an impact on the thermodynamic stability. However, F_{max} does increase by $0.20 \text{ e}/\text{\AA}$ for mutations L83C, F84C, G126C, and F127C and by $0.26 \text{ e}/\text{\AA}$ for F101C. These constitute substantial increases. Notice that these results were obtained without checking for the modifications in the whole structure through the AA simulations. Thus all non-bonded contacts come with the identical value of ϵ .

AA SMD simulations for F127C yield an increase in F_{max} of about 0.15 nN and hold the average α -C- α -C distance, $d_{\alpha,\alpha}$, between sites 38 and 127 at 4.11 \AA , indicating that this coupling is not likely to form a viable DB. The reason is that the effective range for $d_{\alpha,\alpha}$ in DBs falls between 4.4 and 6.8 \AA .⁵⁶ Experimentally, this range is still narrower between 5.4 and 5.8 \AA .⁵⁷

An alternative approach to increase F_{max} is through double mutations. We have found that introducing DBs within the mechanical clamp leads to a theoretically infinite stability (with no isolated force peaks) as illustrated in Fig. 5(d) for D2C-K136C. A similar behavior is found for A5C-Q138C, V10C-G142C, and A15C-G147C. On the other hand, isolated force peaks arise when the mutations do not affect the mechanical clamp directly. For instance, for D4C-G29C and T12C-L116C, we get an increase in F_{max} by $0.13 \text{ e}/\text{\AA}$ and $1.68 \text{ e}/\text{\AA}$, respectively. However, AA SMD for T12C-L116C yields $d_{\alpha,\alpha}$ of 7.32 \AA , which suggests that this bridge cannot actually be formed. The derived $d_{\alpha,\alpha}$ for A15C-G147C, V10C-G142C are 5.31 \AA and 4.89 \AA , respectively, making formation of these bridges more likely.

IV. EXPERIMENTAL VALIDATION

In order to validate our theoretical results, we performed experimental measurements of representative mutants displaying large variation in thermal and mechanical stability. We first consider the mechanical stability, as measured by AFM-based SMFS. One concern in AFM studies of highly mechanostable proteins is the possibility of protein detachment prior to module unfolding. This effect could potentially generate a detection biasing toward less stable modules and lead to a poor accuracy in the determination of F_{max} . For this reason, we have decided to limit the AFM part of our experimental study to only one cohesin mutant G142F that has been predicted by us to yield the largest change in F_{max} resulting from the point mutation into ALA or PHE.

The CG simulations suggest that the mutation results in a reduction in F_{max} of 1.22 or $0.99 \text{ e}/\text{\AA}$, depending on whether we consider model I or II. The average of the two is about $1.1 \text{ e}/\text{\AA}$ or about 120 pN with at least 25% error. The experimental results, on the other hand, are shown in Fig. 6—the substantial reduction is indeed observed, from $480 \pm 14 \text{ pN}$ ⁵ in WT to $229 \pm 3 \text{ pN}$, but the change is larger than predicted. Note, however, that the values of F_{max} depend on v_p . In the theoretical calculations, v_p is of order $5 \times 10^{-4} \text{ nm/ns}$, which is 1250 times faster than the experimental one. It is not clear how lowering the speed would affect F_{max} , but the very fact of a substantial

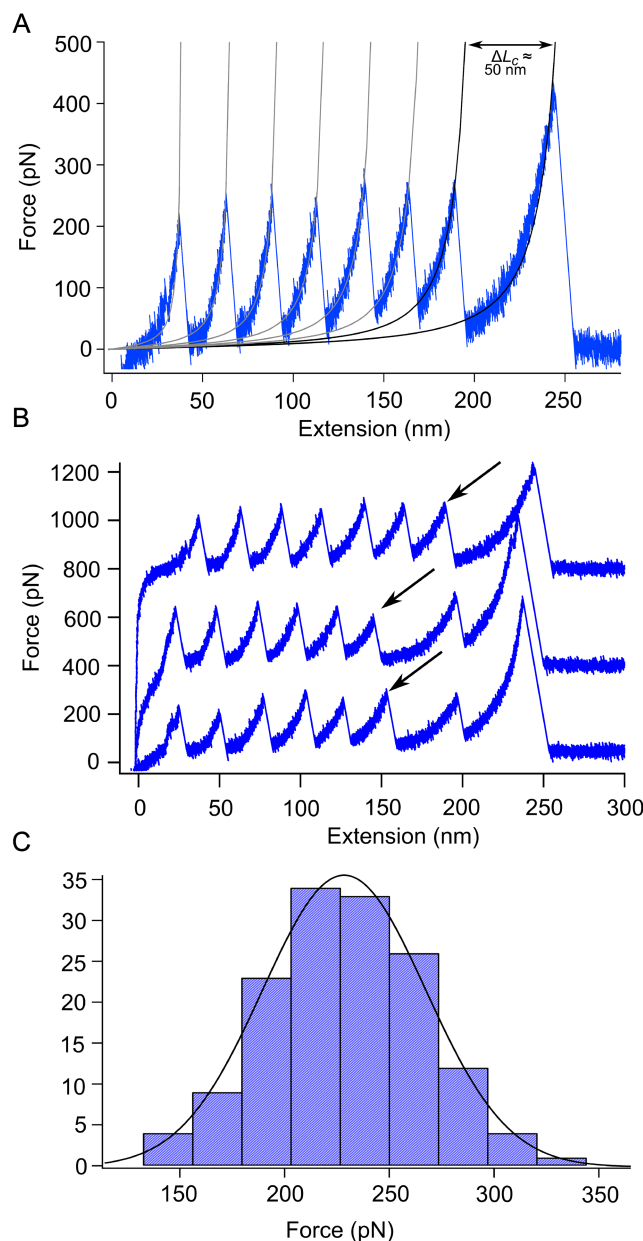


FIG. 6. AFM-SMFS results. (a) SMFS recording showing the unfolding of six I27 modules and one *c7A* G142F mutant, the latter clearly identifiable based on its larger increase in contour length ($\approx 50 \text{ nm}$). Solid lines represent worm-like chain fittings to the force peaks. (b) *c7A* G142F mutant (arrows), displays a mechanical stability comparable to the I27 single molecule marker. Two of the three traces shown are shifted upwards for clarity. (c) Force unfolding histogram of *c7A* G142F. The average value is $\langle F_{max} \rangle = 229 \pm 3 \text{ pN}$.

reduction in F_{max} on mutation can be considered as a positive validation of the prediction.

In the experiment, we have actually used a heteropolyprotein consisting of one cohesin 7 mutant module and six I27 modules acting as single molecule marker. G142F unfolding was clearly distinguishable from the I27 signal due to its larger increase in contour length upon unfolding [$\approx 50 \text{ nm}$ for G142F vs. $\approx 28 \text{ nm}$ for I27, see panel (a) in Fig. 6]. A visual inspection of the recordings indicated that G142F displayed a mechanostability comparable to that of the I27 marker [panels (b) and (c)].

We now consider the thermal stability, for which we synthesized and tested four mutants. Two of them, I119A and

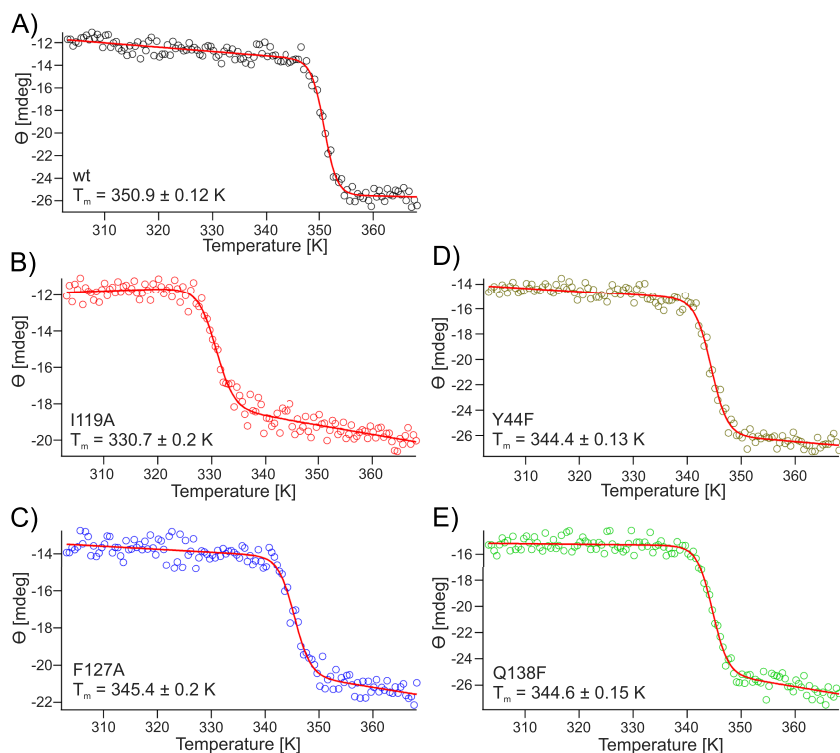


FIG. 7. Thermal unfolding of *c7A* and 4 point mutants. CD signal as a function of T for *c7A* WT (a), I119A (b), I127A (c), Y44F (d), and Q138F (e). The red line corresponds to the fitting of experimental data. The calculated T_m was 77.8 ± 0.1 °C for *c7A* WT, 57.5 ± 0.2 °C for *c7A* I119A, 72.2 ± 0.2 °C for *c7A* F127A, 71.3 ± 0.13 °C for *c7A* Y44F, and 71.5 ± 0.15 °C for *c7A* Q138F.

Q138F, have been predicted to enhance T_0 and two, F127A and Y44F, to lower it [Fig. 4(a)], the top panel. The CD measurements shown in Fig. 7, on the other hand, indicate a reduction in stability in all four cases. The largest reduction is observed for I119A.

As mentioned in the Introduction, the CD technique does not measure the “true” stability but only the part that is contained in the secondary structures. When we study the shifts in T'_0 as determined from P'_0 , then all four cases yield a reduction in T'_0 . Actually, we observe a reduction or no change for all mutations except for a few, such as N130A, as shown in Fig. 4(a) (the bottom panel), for which there is an enhancement (even though the consideration of all contacts, through P_0 , is predicted not to induce any change in T_0).

The downward shifts in T'_0 for I119A, Q138F, F127A, and Y44F are -0.004 , -0.006 , 0.001 , -0.006 ϵ/k_B , respectively, which is in the range of several °C. However, if one looks at P'_0 at higher T s than T'_0 then one observes a more noticeable decrease in P'_0 for I119A. This is consistent with the experimental observation that the largest decrease in the thermodynamic stability occurs for I119A, in the set of proteins studied.

V. CONCLUSIONS

We have proposed a theoretical method to study the effects of mutations in a protein. The method combines AA simulations with contact-based assessment of the structural changes. The contacts are derived during the time evolution and not just taken from the last structure obtained. The perspective of the contacts also illuminates the spatial extent of the resulting changes and leads to the conclusion that the effects are usually not local—they may extend up to 50 Å.

We have applied the method to cohesin *c7A* by making systematic replacements of all residues into ALA and PHE, one at a time. We have analyzed the impact of the mutations on the mechanical and thermodynamic stabilities by using a structure-based CG model. The approximations involved in using the CG model allow the simulations to access the experimentally relevant time scales, but the detailed results come with substantial errors. Thus one should treat the method as a comparison tool and to identify the most profound variations. We have found that the changes in F_{max} do not necessarily come from mutations in the vicinity of the mechanical clamp.

We have identified mutations that would lead to a substantial increase in mechanical and thermodynamic stabilities of *c7A*. The largest possible increases in F_{max} found are of order 0.2 $\epsilon/\text{Å}$, i.e., of order 20 pN. Much larger increases in mechanostability can be obtained by introducing double mutations to cysteines within the region of the mechanical clamp.

Our predictions are qualitatively consistent with the experimental findings. However, there is an issue of which quantity should be used to assess the thermodynamic stability. For enzymatic proteins, the criterion should be based on the T shift related to the catalytic activity and hence to the occupation of the contacts in the catalytic region. For structural proteins, such as *c7A*, we think that all contacts matter and not merely those that form the secondary structures. This poses the following question. Since the CD-based methods couple only to a subset of all contacts (of order half of them in the case of *c7A*), should the selection of mutants of a particular level of thermodynamic stability be based on theory or experiment? Perhaps on theory. In conclusion, our theoretical method seems to be a valid approximation for a rapid screening of the effects of mutations in the mechanical and thermal stabilities of proteins.

SUPPLEMENTARY MATERIAL

See [supplementary material](#) for more details about the experimental procedures (CD and SMFS) and the tables of the theoretical results for all single-site substitutions into alanine and phenylalanine. It also shows a crossplot of the predicted changes of the thermostability and mechanostability after making point mutations into PHE.

ACKNOWLEDGMENTS

We thank D. V. Laurents for his help with CD measurements and their analysis. This research has been supported by the ERA-NET grant ERA-IB (No. EIB.12.022) (FiberFuel) and the European Framework Programme VII NMP Grant No. 604530-2 (CellulosomePlus) and by PLGrid Infrastructure. It was also co-financed by the Polish Ministry of Science and Higher Education from the resources granted for the years 2014–2017 in support of international scientific projects. D.T. acknowledges Science Foundation Ireland (SFI) for financial support under Grant No. 15/CDA/3491 and M.Ch., M.G., and D.T. acknowledge computing resources at the SFI/Higher Education Authority Irish Centre for High-End Computing (ICHEC).

- ¹P. Beguin and M. Lemaire, *Crit. Rev. Biochem. Mol. Biol.* **31**, 201 (1996).
- ²E. A. Bayer, H. Chanzy, R. Lamed, and Y. Shoham, *Curr. Opin. Struct. Biol.* **8**, 548 (1998).
- ³E. A. Bayer, J. P. Belaich, Y. Shoham, and R. Lamed, *Annu. Rev. Microbiol.* **58**, 521 (2004).
- ⁴M. Gunnoo *et al.*, *Adv. Mater.* **28**, 5619 (2016).
- ⁵A. Valbuena, J. Oroz, R. Hervas, A. M. Vera, D. Rodriguez, M. Menendez, J. I. Sułkowska, M. Cieplak, and M. Carrion-Vazquez, *Proc. Natl. Acad. Sci. U. S. A.* **106**, 13791 (2009).
- ⁶M. Anbar, R. Lamed, and E. A. Bayer, *ChemCatChem* **2**, 997 (2010).
- ⁷M. Anbar, O. Gul, R. Lamed, U. O. Sezerman, and E. A. Bayer, *Appl. Environ. Microbiol.* **78**, 3458 (2012).
- ⁸M. Sandgren, P. J. Gualfetti, A. Shaw, L. S. Gross, M. Saldajeno, A. G. Day, T. A. Jones, and C. Mitchinson, *Protein Sci.* **12**, 848 (2003).
- ⁹C. Lehmann, F. Sibilla, Z. Maugeri, W. R. Streit, P. D. de María, R. Martinez, and U. Schwaneberg, *Green Chem.* **14**, 2719–2726 (2012).
- ¹⁰A. Matouschek, J. T. Kellis, Jr., L. Serrano, and A. R. Fersht, *Nature* **340**, 122–126 (1989).
- ¹¹S. E. Jackson and A. R. Fersht, *Biochemistry* **30**, 10428–10435 (1991).
- ¹²D. S. Otzen, N. ElMasry, S. E. Jackson, and A. R. Fersht, *Proc. Natl. Acad. Sci. U. S. A.* **91**, 10422–10425 (1994).
- ¹³L. S. Itzhaki, D. E. Otzen, and A. R. Fersht, *J. Mol. Biol.* **254**, 260–288 (1995).
- ¹⁴R. Guerois, J. E. Nielsen, and L. Serrano, *J. Mol. Biol.* **320**, 369–387 (2002).
- ¹⁵J. W. H. Schymkowitz, F. Rousseau, I. C. Martins, J. Ferkinghoff-Borg, F. Stricher, and L. Serrano, *Proc. Natl. Acad. Sci. U. S. A.* **102**, 10147–10152 (2005).
- ¹⁶Y. Li and J. Fang, *PLoS One* **7**, e47247 (2012).
- ¹⁷S. Yin, F. Ding, and N. V. Dokholyan, *Structure* **15**, 1567–1576 (2007).
- ¹⁸J. I. Sułkowska and M. Cieplak, *J. Phys.: Condens. Matter* **19**, 283201 (2007).
- ¹⁹M. Sikora, J. I. Sułkowska, and M. Cieplak, *PLoS Comput. Biol.* **5**, e1000547 (2009).
- ²⁰B. Rózycki and M. Cieplak, *J. Chem. Phys.* **141**, 235102 (2014).
- ²¹A. Gómez-Sicilia, M. Sikora, M. Cieplak, and M. Carrión-Vázquez, *PLoS Comput. Biol.* **11**, e1004541 (2015).
- ²²N. J. Greenfield, *Nat. Protoc.* **1**, 2876–2890 (2007).
- ²³M. J. O’Neil, *Anal. Chem.* **36**, 1238–1245 (1964).
- ²⁴K. U. Linderstrom-Lang and J. A. Schellman, *The Enzymes*, edited by B. D. Boyer (Academic Press, New York, 1959), Vol. 1, 2nd ed., pp. 443–510.
- ²⁵A. Šali, E. Shakhnovich, and M. Karplus, *Nature* **369**, 248–251 (1994).
- ²⁶N. D. Socci and J. N. Onuchic, *J. Chem. Phys.* **101**, 1519–1528 (1994).
- ²⁷M. Cieplak and J. R. Banavar, *Phys. Rev. E* **88**, 040702(R) (2013).
- ²⁸K. Wolek and M. Cieplak, *J. Chem. Phys.* **144**, 185102 (2016).
- ²⁹G. B. Akcapinar, A. Venturini, P. L. Martelli, R. Casadio, and U. O. Sezerman, *Protein Eng., Des. Sel.* **28**, 127–135 (2015).
- ³⁰M. Carrion-Vazquez, H. B. Li, H. Lu, P. E. Marszalek, A. F. Oberhauser, and J. M. Fernandez, *Nat. Struct. Biol.* **10**, 738–743 (2003).
- ³¹D. J. Brockwell, G. S. Beddard, E. Paci, D. K. West, P. D. Olmsted, D. A. Smith, and S. E. Radford, *Biophys. J.* **89**, 506–519 (2005).
- ³²A. Galera-Prat, A. Gomez-Sicilia, A. F. Oberhauser, M. Cieplak, and M. Carrion-Vazquez, *Curr. Opin. Struct. Biol.* **20**, 63–69 (2010).
- ³³H. C. Kotamarthi, R. Sharma, and S. R. K. Aivarapu, *Biophys. J.* **104**, 2273–2281 (2013).
- ³⁴K. Wolek, A. Gómez-Sicilia, and M. Cieplak, *J. Chem. Phys.* **143**, 243105 (2015).
- ³⁵M. Cieplak, in *Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes—From Bioinformatics to Molecular Quantum Mechanics*, edited by A. Liwo (Springer, Berlin, Heidelberg, 2014), pp. 295–315.
- ³⁶M. Kouza, C.-K. Hu, H. Zung, and M. S. Li, *J. Chem. Phys.* **131**, 215103 (2009).
- ³⁷G. A. Tavares, P. Beguin, and P. M. Alzari, *J. Mol. Biol.* **273**, 701–713 (1997).
- ³⁸M. Chwastyk, A. Poma Bernaola, and M. Cieplak, *Phys. Biol.* **12**, 046002 (2015).
- ³⁹J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, *J. Comput. Chem.* **26**, 1781–1802 (2005).
- ⁴⁰A. D. MacKerell *et al.*, *J. Phys. Chem. B* **102**, 3586–3616 (1998).
- ⁴¹A. D. MacKerell, Jr., M. Feig, and C. L. Brooks III, *J. Comput. Chem.* **25**, 1400 (2004).
- ⁴²W. Humphrey, A. Dalke, and K. Schulten, *J. Mol. Graphics* **14**, 33–38 (1996).
- ⁴³W. L. Jorgensen, J. Chandrasekhar, J. D. Madhura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.* **79**, 926–935 (1983).
- ⁴⁴U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, *J. Chem. Phys.* **103**, 8577–8592 (1995).
- ⁴⁵T. Darden, D. York, and L. G. Pedersen, *J. Chem. Phys.* **98**, 10089 (1993).
- ⁴⁶J. Tsai, R. Taylor, C. Chothia, and M. Gerstein, *J. Mol. Biol.* **290**, 253–266 (1999).
- ⁴⁷A. Starzyk and M. Cieplak, *J. Chem. Phys.* **139**, 045102 (2013).
- ⁴⁸R. B. Best *et al.*, *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
- ⁴⁹M. W. Mahoney and W. L. Jorgensen, *J. Chem. Phys.* **112**, 8910–8922 (2000).
- ⁵⁰L. Kale *et al.*, *J. Comput. Phys.* **151**, 283–312 (1999).
- ⁵¹H. Lu, and K. Schulten, *Proteins: Struct., Funct., Genet.* **35**, 453–463 (1999).
- ⁵²P. Gonnert, *J. Comput. Phys.* **220**, 740–750 (2007).
- ⁵³C. A. Laughton, *J. Mol. Biol.* **235**, 1088–1097 (1994).
- ⁵⁴C. L. Ross, X. Liang, Q. Liu, B. E. Murray, M. Hook, and V. K. Ganesh, *J. Biol. Chem.* **287**, 34856–34865 (2012).
- ⁵⁵A. Steward, J. L. Toca-Herrera, and J. Clarke, *Protein Sci.* **11**, 2179–2183 (2002).
- ⁵⁶J. S. Richardson, *Adv. Protein Chem.* **34**, 167–339 (1981).
- ⁵⁷M. T. Petersen, P. H. Johnson, and S. B. Petersen, *Protein Eng., Des. Sel.* **12**, 535–548 (1999).