

# Double mutant of chymotrypsin inhibitor 2 stabilized through increased conformational entropy

*Yulian Gavrilov<sup>†</sup>, Felix Kümmerer<sup>†</sup>, Simone Orioli<sup>†,‡</sup>, Andreas Prestel<sup>†</sup>, Kresten Lindorff-Larsen<sup>†,\*</sup>, and Kaare Teilum<sup>†,\*</sup>*

<sup>†</sup>Structural Biology and NMR Laboratory and the Linderstrøm-Lang Centre for Protein Science,  
Department of Biology, University of Copenhagen

<sup>‡</sup>Structural Biophysics, Niels Bohr Institute, University of Copenhagen

## ABSTRACT

The conformational heterogeneity of a folded protein can affect both its function but also stability and folding. We recently discovered and characterized a stabilized double mutant (L49I/I57V) of the protein CI2 and showed that state-of-the-art prediction methods could not predict the increased stability relative to the wild-type protein. Here we have examined whether changed native state dynamics, and resulting entropy changes, can explain the stability changes in the double mutant protein, as well as the two single mutant forms. We have combined NMR relaxation measurements of the ps-ns dynamics of amide groups in the backbone and the methyl groups in the side-chains with molecular dynamics simulations to quantify the native state dynamics. The NMR experiments reveal that the mutations have different effects on the conformational flexibility of CI2: A reduction in conformational dynamics (and entropy) of the native state of L49I variant correlates with its decreased stability, while increased dynamics of the I57V and L49I/I57V variants correlates with their increased stability. These findings suggest that explicitly accounting for changes in native state entropy might be needed to improve the predictions of the effect of mutations on protein stability.

## Introduction

Controlling protein stability is an important challenge in biotechnology and biopharmaceutical applications. The ability to predict how protein mutations decrease stability can be used in clinical genetics to interpret gene or genome sequencing data.<sup>1</sup> Similarly, improving the stability of enzymes<sup>2</sup> or protein pharmaceuticals<sup>3</sup> by mutations may improve shelf life and tolerance to heat. The stability of a protein depends on interactions and dynamics of both the native and denatured states of the protein.<sup>4</sup>

The conformational dynamics of the native state of a protein contributes to its conformational stability. If the dynamics of the native state is increased without significantly affecting the intermolecular interactions it will lead to a more stable protein.<sup>5,6</sup> Typically, however, the increased dynamics is counteracted by fewer stabilizing interactions through entropy-enthalpy compensation in the more dynamic molecule.<sup>7,8</sup> In some cases evolution has circumvented this problem. Thus, in hyperthermophilic proteins Lys is preferred over Arg compared to the mesophilic homologues. This is attributed to high residual entropy of the Lys side-chain relative to Arg in folded protein structures.<sup>9</sup> However, even small local changes in a protein structure induced by point mutations can affect protein conformational equilibria, dynamics and its interactions.<sup>10,11</sup> NMR spectroscopy, in combination with computational methods such as molecular dynamics simulations, has successfully been applied to analyze the effect of mutations on conformational dynamics of various protein systems.<sup>8,12–17</sup> Here we apply these complementary techniques to study the small 7.3 kDa protein, chymotrypsin inhibitor 2 (CI2) from barley seeds,<sup>18</sup> which has been extensively used in both experimental and computational studies.<sup>19–24</sup> The conformational dynamics of the polypeptide backbone and the side-chains of CI2 on the ps-ns time-scales have been studied by NMR spectroscopy.<sup>25</sup> Previously, it was demonstrated that five single point-mutations in the hydrophobic

core of CI2 (L49V, L49A, I57V, V9A, and V47A) all lead to a similar yet small increase in backbone and side-chain flexibility throughout the structure. These (and other) core mutations only have a minor effect on the inhibitory function of CI2.<sup>26</sup>

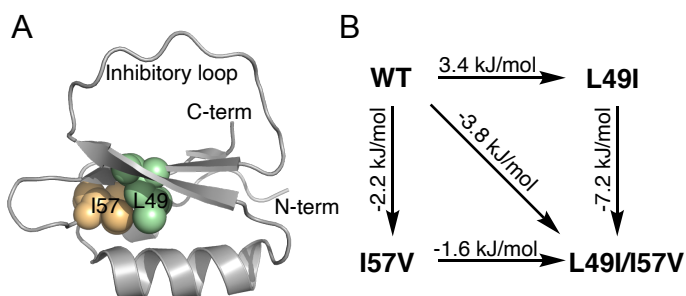
According to experimental<sup>19</sup> and computational studies<sup>27,28</sup> residues L49 and I57 are both important for formation of the CI2 folding nucleus and for the stability of the folded state as they participate in a central network of interactions.<sup>25-31</sup> L49 (together with R48) is a key energetic linchpin connecting CI2's hydrophobic core to its inhibitory loop.<sup>27,30</sup> Furthermore, L49 and I57 are involved in a network of interactions that dissipate the structural strain occurring when CI2 binds to its substrate.<sup>31</sup> This network includes residue I37 in the inhibitory loop. The neighboring residues, T36 and V38, point their side chains toward the highly conserved residues of the inhibitor loop, R48 and F50. Between these residues L49 forms part of the hydrophobic core together with A16 and I57.<sup>31</sup>

We recently showed that L49I destabilizes the folded state of CI2 ( $\Delta\Delta G_f = 3.4 \pm 0.1$  kJ/mol), while I57V stabilizes it slightly ( $\Delta\Delta G_f = -2.2 \pm 0.1$  kJ/mol).<sup>32</sup> Notably, the double mutant (L49I/I57V) is significantly more stable than either of the single mutants ( $\Delta\Delta G_f = -3.8 \pm 0.1$  kJ/mol; Figure 1). This synergistic stabilizing effect of the double mutant is unexpected and could not be predicted using commonly applied methods such as FoldX<sup>33</sup> and Rosetta.<sup>34</sup> Failure to properly capture long range structural changes and changes in the dynamic properties of the protein may contribute to the unsuccessful attempt to predicting the effect of the double mutation.

Here, we asked if the change in stability of the L49I/I57V double mutant can, at least in part, be explained by changes in the dynamics of CI2. We used NMR spectroscopy to experimentally assess the changes in fast dynamic of the three CI2 variants L49I, I57V and L49I/I57V relative to the wild type. To complement the experimental data, we also performed molecular dynamics (MD)



simulations and reweighted the obtained computational dynamic ensembles using experimental data. We found that I57V and L49I/I57V increases the dynamics of CI2 whereas L49I decreases the dynamics. The corresponding change in conformational entropy scales linearly with the change in conformational stability. Our findings suggest that explicitly accounting for changes in native state entropy might be needed to improve the predictions of the effect of mutations on protein stability, which may have implications for the interpretation of protein engineering and sequencing data.



**Figure 1.** (A) Overview of the positions of the mutated sites in CI2 (spacefilling). (B) Mutation cycle labelled with experimental  $\Delta\Delta G_f$  values at 25 °C.

## Materials and methods

**Protein production and sample preparation.** We expressed wild-type CI2 (UniProt: P01053, residues 22–84 with an additional N-terminal Met) and the three variants L49I, I57V, and L49I/I57V using previously described DNA constructs.<sup>32</sup> *E. coli* BL21(DE3) carrying the expression plasmid was grown in M9 minimal medium at 37 °C to  $OD_{600}=0.6$  before induction with 0.4 mM isopropyl b-D-1-thiogalactopyranoside (IPTG). Four different isotope compositions of the M9 media were used. i)  $^{15}\text{NH}_4\text{Cl}$  as the sole nitrogen source, ii)  $^{13}\text{C}_6\text{-Glc}$  as the sole carbon source and  $^{15}\text{NH}_4\text{Cl}$  as the sole nitrogen source, iii)  $^{13}\text{C}_6\text{-Glc}$  as the sole carbon source and  $^{15}\text{NH}_4\text{Cl}$

as the sole nitrogen source in 50% D<sub>2</sub>O and iv) 10% <sup>13</sup>C<sub>6</sub>-Glc / 90% unlabeled Glc as carbon source and <sup>15</sup>NH<sub>4</sub>Cl as the sole nitrogen source. The growth was continued overnight at 20 °C. Harvested cells were resuspended in 25mM Tris HCl, pH 8 with 1 mM of EDTA. Cell lysis was performed with two freeze-thaw cycles. Insoluble material was removed by centrifugation at 30000 g for 20 minutes at 5 °C.

Nucleic acids were precipitated with 2% polyethylene imine and removed by centrifugation at 20000 g for 15 min at 5 °C. CI2 was precipitated by adding ammonium sulfate to 70% saturation and centrifuging at 20000 g for 15 min at 5 °C. The ammonium sulfate precipitate was resuspended in 10 mM ammonium bicarbonate and pure CI2 was obtained by SEC on a Superdex 75 26/600 in 10 mM ammonium bicarbonate.

**NMR Experiments.** All NMR experiments were performed at constant temperature (25 °C) and buffer conditions (50 mM MES, pH 6.26, 5 % D<sub>2</sub>O and 0.02 % NaN<sub>3</sub>). <sup>15</sup>N-HSQC, HNC(O), HN(CA)CO, HNCA, HNCACB, CBCA(CO)NH spectra were acquired and used for backbone chemical shift (CS) assignments. The methyl chemical shifts were assigned using <sup>13</sup>C-HSQC, C(CO)NH and HC(CO)NH spectra. These spectra were acquired for all studied CI2 variants on a Varian DD2 800 MHz spectrometer equipped with a room temperature probe. The triple resonance spectra were recorded with non-uniform sampling at 25% and reconstructed by the compressed sensing method iterative soft thresholding using the software qMDD.<sup>35</sup> We stereospecifically assigned the methyl groups of the valine and leucine side chain from constant time <sup>13</sup>C-HSQC recorded on the sample expressed in the presence of 10% <sup>13</sup>C<sub>6</sub>-Glc.<sup>36</sup>

Backbone dynamics was probed with a series of <sup>15</sup>N *R*<sub>1</sub>, *R*<sub>2</sub> and heteronuclear <sup>1</sup>H-<sup>15</sup>N NOE relaxation experiments recorded on Bruker Avance III HD 600 and 750 MHz spectrometers cryoprobes using ~1.5 mM <sup>15</sup>N-labeled protein samples. Relaxation delays of 10, 200, 400, 600,

800, 1000 ms were chosen for  $R_1$  and 16.96, 84.8, 169.6, 254.4, 339.2, 424.8 ms for  $R_2$ , respectively. The  $^1\text{H}$ - $^{15}\text{N}$  NOE experiments were recorded with a 5 s interscan delay. All experiments were repeated twice, processed with nmrPipe<sup>37</sup> and analyzed in CcpNmr.<sup>38</sup> We performed model-free analysis<sup>39,40</sup> for the calculation of backbone order parameters  $S^2_{NH}$  using the software *relax*.<sup>41</sup> Values of  $R_1$ ,  $R_2$  and  $^1\text{H}$ - $^{15}\text{N}$  NOE for each amide group can be fitted to model-free models with up to six parameters.<sup>42</sup> Our initial analysis showed that fitting to models with more than three parameters resulted in unrealistic values of the order parameters. In the final analysis of the backbone NMR data, we therefore excluded models with more than three parameters.

Side chain  $^2\text{H}$  relaxation experiments were performed on  $\sim 0.7$  mM  $^{15}\text{N}/^{13}\text{C}/^2\text{H}_{50\%}$ -labeled protein samples using pulse sequences described elsewhere.<sup>43–45</sup> The relaxation data were acquired on Bruker Avance III HD 600 and 750 MHz spectrometers using following delays for  $R_{D_1}(D_z)$  and  $R_{D_3}(3D_z^2 - 2)$ : 1.4, 3.6, 7.6, 16.6, 33.8, and 50 ms; for  $R_{D_2}(D_+)$ : 0.2, 4.7, 9.7, 12.5, 16.5, and 20 ms; for  $R_{D_4}(D_+^2)$ : 3, 6, 12, 22, 34, and 50 ms. All side chain relaxation experiments were done in duplicates. The relaxation data were processed using python scripts, developed based on the scripts used by Hoffmann et al.<sup>46</sup> Similar to that study, only the original model-free equation (LS2)<sup>39,40</sup> and LS3 equation (with global tumbling treated as a fitting parameter) were used to fit methyl relaxation data. The global rotational diffusion times of the studied proteins were initially estimated based on backbone  $R_2/R_1$  ratio with the programs R2R1\_TM and QUADRIC ([www.palmer.hs.columbia.edu](http://www.palmer.hs.columbia.edu)).<sup>47</sup> Global tumbling of all CI2 variants was analyzed with an isotropic diffusion tensor, since  $D_{\parallel}/D_{\perp} < 1.3$  for all the variants.

**Chemical shift analysis.** We calculated secondary chemical shifts for C $\alpha$  relative to random coil chemical shifts using the method of Kjaergaard and Poulsen, which corrects for the effect of the neighboring residues ( $i-2, i-1, i+1, i+2$ , where  $i$  is the residue of interest) and temperature.<sup>48</sup>

**Entropy calculations using experimental order parameters.**  $\Delta S_{NH}^2$  and  $\Delta S_{axis}^2$  values obtained in this study were used to estimate the change in conformational entropy upon the mutations by applying the “entropy meter” approach:<sup>49</sup>

$$\Delta S/R = 0.88N_{res} \langle \delta \ln(1 - S_{NH}^2) \rangle - 0.56N_{\chi} \langle \delta S_{axis}^2 \rangle \quad (1)$$

where  $\Delta S$  is the change in conformational entropy,  $R$  is the gas constant,  $N_{res}$  is the total number of residues,  $N_{\chi}$  is the total number of side chain chi angles,  $\langle \delta S_{axis}^2 \rangle$  is the measured average change in methyl axis order parameters, and finally  $\langle \delta \ln(1 - S_{NH}^2) \rangle$  is the corresponding quantity for measured changes in the amide order parameters. The factors of 0.88 and 0.56 account for the reduction in entropy due to correlations of side chain and backbone torsional motions and for the dependence of entropy on order parameters.

**MD simulations.** The equations of motion were integrated using the leap-frog integrator implemented in GROMACS v2018.1, using a timestep of 2 fs. We performed the simulations using the Amber a99SB-disp force field with the TIP4P-D-based a99SB-disp water model.<sup>50</sup> Crystal structures of the CI2 WT (PDB: 7A1H) and its mutated variants L49I (PDB: 7AOK), I57V (PDB: 7A3M) and L49I/I57V (PDB: 7AON) were used as the starting conformations for the simulations.<sup>32</sup> The proteins were placed in a cubic box of 30.66 nm<sup>3</sup> (resulting in 0.5 nm distance from the protein to all box edges) and solvated with 6728 to 6836 water molecules. The net charge of CI2 is zero, so no ions were needed to neutralize the system. The system was minimized with 50000 steps of steepest descent and equilibrated for 150 ps in the NPT and NVT ensembles with harmonic position restraints (force constants of 1000 kJ mol<sup>-1</sup> nm<sup>2</sup>) on the heavy atoms of the

protein. Equilibration was performed using velocity rescaling thermostat<sup>51</sup> with temperature of 298 K, thermostat timescale of 0.1 ps and the Parrinello-Rahman barostat (NPT) with a reference pressure of 1 bar, a barostat time constant 2 ps and an isothermal compressibility of  $4.5 \times 10^{-5}$  bar<sup>-1</sup>. A cut-off range of 1.0 nm was chosen for the Van der Waals and Coulomb interactions and the Particle Mesh Ewald method<sup>52</sup> was applied to treat long-range electrostatics (4th order cubic interpolation, 0.16 nm grid spacing). The production simulations were run in the NVT ensemble, using velocity rescaling thermostat<sup>51</sup> with the time constant of 1 ps and the reference temperature of 298 K. We applied harmonic constrains to all bonds involving hydrogens using the LINCS algorithm. All production runs were 1  $\mu$ s long and the protein coordinates were saved every 1 ps and 3 replicates were used for each simulation, summing to a total of 3  $\mu$ s per structure. Since the initial structural fluctuations may be related to the equilibration process, all analyses were performed for the last 800 ns of each trajectory, discarding the first 200 ns.

**Calculation of order parameters from MD trajectories.** The calculation of the order parameters  $S^2_{NH}$  and  $S^2_{axis}$  was done following the procedures described by Hoffmann et al.<sup>46</sup> Briefly, we calculated the backbone time correlation function (TCF) of each NH bond vector up to a maximum lag-time of 400 ns. The resulting TCF was then fitted to two different Lipari-Szabo models:

$$C(t) = \frac{1}{5} S^2 e^{-t/\tau_R} \quad (2)$$

with  $S^2$  used as fitting parameter and

$$C(t) = \frac{1}{5} S^2 e^{-\frac{t}{\tau_R}} + \frac{1}{5} (1 - S^2) e^{-\frac{t}{\tau_e}} \quad (3)$$

where  $\tau^{-1} = \tau_R^{-1} + \tau_e^{-1}$  and with  $S^2, \tau_e$  as fitting parameters.  $C(t)$  was further multiplied by  $\xi = (1.02/1.041)^6$ , to include the effect of vibrational averaging on dipolar couplings.<sup>53</sup> An isotropic diffusion model was applied using the experimentally derived global rotational diffusion times  $\tau_R$

(see NMR methods). The obtained order parameters  $S_{NH}^2$  were then averaged over the three independent trajectories. Model selection for each residue was done using the Akaike Information Criterion (AIC):

$$AIC = \chi^2 + 2N_{params} \quad (4)$$

where  $\chi^2$  estimates the difference between the fitted and experimental values of the relaxation rates and  $N_{params}$  is the number of the fitting parameters in the model. The model with lowest  $AIC$  was chosen.

For the side chain order parameters  $S_{axis}^2$ , each trajectory was first cut into 10 ns long blocks. After that, internal TCFs (without global tumbling) were calculated for each  $C_{methyl}-H_{methyl}^i$  ( $i = 1,2,3$ ) bond vector of methyl-bearing side chains up to a maximum lag-time of 5 ns and averaged over the three  $C_{methyl}-H_{methyl}^i$  ( $i = 1,2,3$ ) bonds. Next, the internal TCFs were fitted to six exponentials plus an offset,

$$C_{int}(t) = \sum_{i=1}^6 A_i e^{-t/\tau_i} + S_{long}^2 \quad (5)$$

where  $\sum_{i=1}^6 A_i + S_{long}^2 = 1$ ,  $0 \leq A_i \leq 1$ ,  $0 \leq S_{long}^2 \leq 1$ ,  $0 \leq \tau_1 \leq 10 ps$ ,  $0 \leq \tau_2 \leq 50 ps$ ,  $0 \leq \tau_3 \leq 200 ps$ ,  $0 \leq \tau_4 \leq 500 ps$ ,  $0 \leq \tau_5 \leq 1000 ps$ ,  $\tau_6 \geq 0 ps$  and  $S_{long}^2$  is the long-time limit order parameter. The internal TCF was then multiplied by a single exponential using the experimentally derived global rotational diffusion times  $\tau_R$ :

$$C(t) = C_{int}(t)e^{-t/\tau_R} \quad (6)$$

The resulting total TCF was then fitted to

$$C(t) = S^2 e^{-t/\tau_R} + (1 - S^2) e^{-t/\tau_{red}} \quad (7)$$

with  $S^2 = 1/9 S_{axis}^2$ ,  $\tau_{red} = \frac{\tau_R \tau_f}{\tau_R + \tau_f}$  and  $S^2, \tau_f$  as fitting parameters. Finally, all  $S^2$  were averaged over the 3x80 blocks.

In order to assess the convergence of the calculations we focused on the side chain  $S^2_{axis,sim}$  values, since they, due to the higher flexibility of the side chains, may converge slower than the backbone order parameters.<sup>54-56</sup> We averaged the 240 original blocks of 10 ns into 10 blocks of 240 ns and performed a leave- $N$ -out cross-validation with  $N=1\dots5$  to evaluate the convergence of the side chain order parameters.<sup>57</sup> Specifically, for each  $N$  we split the 10 blocks in a reference set of  $10-N$  and a test set of  $N$  blocks (the left-out blocks), averaged the  $S^2_{axis,sim}$  values within the two sets and computed the RMSD between the resulting average values of the order parameters. For the same  $N$ , this procedure was repeated for all the possible combinations of the left out blocks and the corresponding RMSDs are averaged to obtain  $\langle \text{RMSD} \rangle$ . As shown in **Figure S1** the  $\langle \text{RMSD} \rangle$  curves for  $N=5$  converge to values around 0.04, which is smaller than the typical difference we computed between the  $S^2_{axis,sim}$  of the different protein variants. For this reason, we can consider our estimations of  $S^2_{axis,sim}$  converged.

**Reweighting simulations using  $S^2_{NH}$  and  $S^2_{axis}$ .** The simulated trajectories were analyzed in a block-wise fashion using 100 non-overlapping blocks of 10 ns each. This resulted in 300 blocks for the set of 3x 1  $\mu$ s-long simulations of each CI2 variant. The backbone  $S^2_{NH}$  and side chain  $S^2_{axis}$  order parameters were calculated for each 10 ns block in a similar way as described above but with a maximum lag-time of 5 ns. The resulting TCF was then fitted only to the classical Lipari-Szabo model (equation (5)). Further we performed the weighted average of the order parameters over the blocks, where the weights,  $w$ , are obtained through the optimization of a functional which determines the agreement between simulated and experimental order parameters. The approach is similar to previously described methods for fitting ensembles using order parameters,<sup>54,58</sup> but using a reweighting approach similar to our recently described ABSURDer method.<sup>57</sup> The functional form for the target function that we here used is:

$$\mathcal{J}(\mathbf{w}) = \sum_{n=1}^N \left[ \frac{S_{exp,n}^2 - \sum_{i=1}^N w_i S_{r,n}^2}{\sqrt{(\sigma_n^{exp})^2 + (\sigma_n^{calc})^2}} \right]^2 + \theta \sum_{i=1}^N w_i \log \left( \frac{w_i}{w_i^0} \right) = \chi^2(\mathbf{w}) - \theta S(\mathbf{w}) \quad (8)$$

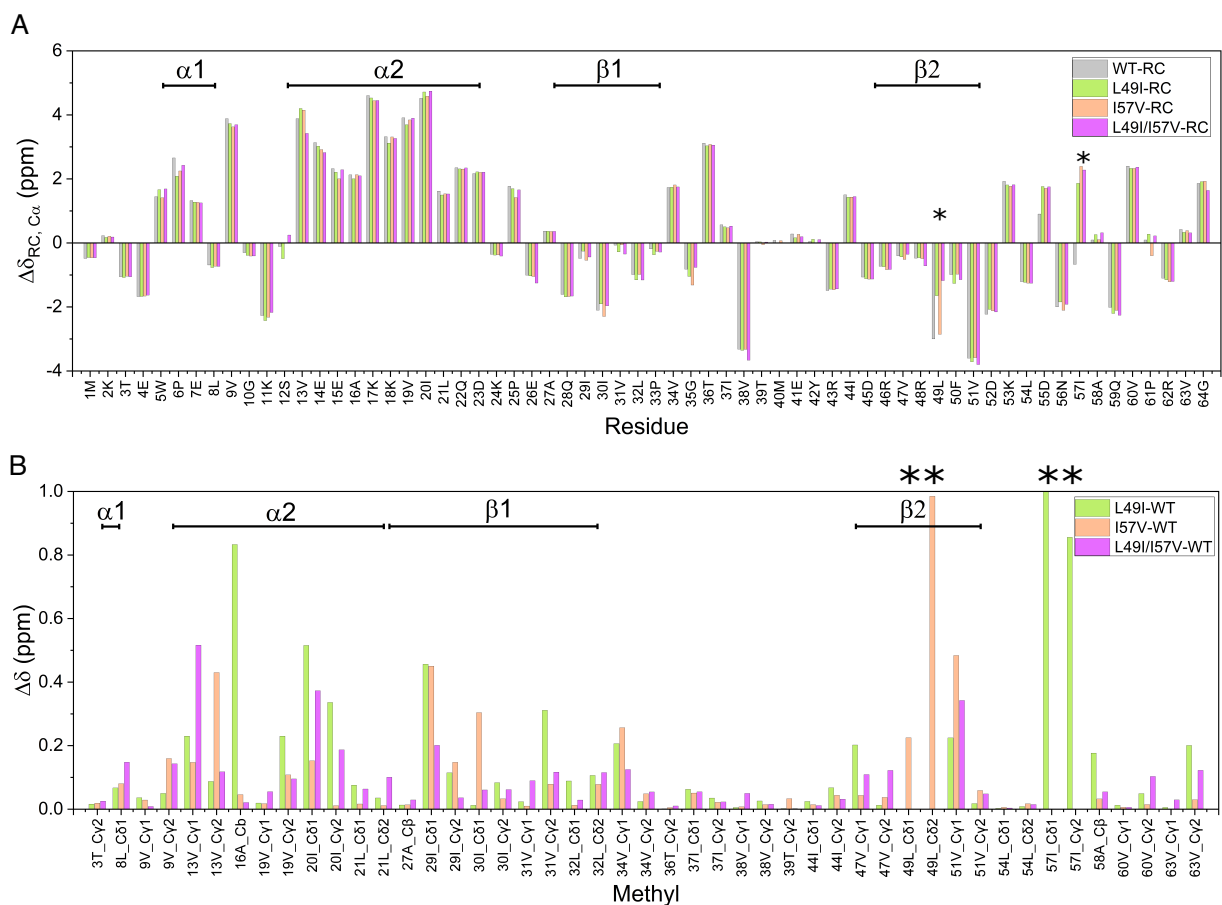
where  $n$  ranges over the overall number of experimentally defined order parameters and  $N$  represents the number of blocks which the trajectory has been cut into,  $\sigma_n^{exp}$  is the experimental error of the  $n$ -th order parameter,  $\sigma_n^{calc}$  is the standard error of the simulated order parameters, estimated over the different blocks. The optimal set of weights is obtained as the corresponding minimum in weight space,  $\mathbf{w}^* = \min_{\mathbf{w}} S^2(\mathbf{w})$ . Each weight is associated with a block from the trajectory and it encodes the relevance of each block to the dynamical ensemble. As reference weights we use  $w_i^0 = 1/N$  to represent the initial, uniform weights provided by the prior (original simulated ensemble). The parameter  $\theta$  is used to set the balance between our trust in the prior and in the experimental data. See Kümmerer et al.<sup>57</sup> and Orioli et al.<sup>59</sup> for the further details.

## Results

Initially, we assessed the overall structural differences in the four variants of CI2 by comparing the chemical shift differences of C $\alpha$  atoms, in the form of secondary chemical shifts, and of the side chain methyl groups (**Figure 2**). Overall, the changes in chemical shifts reflect only minor and local structural rearrangements upon residue substitutions as expected from the crystal structures of the variants.<sup>32</sup>

**Backbone dynamics.** We assessed the conformational dynamics on the ps-ns timescale of the four CI2 variants using NMR relaxation experiments and interpreted it with the model-free (MF) formalism. We analyzed both the dynamics of the backbone amide groups and of the side-chain methyl groups for Ala, Ile, Leu, Thr, and Val. The side-chain methyls of the two Met residues were excluded from the analysis due to low data quality.



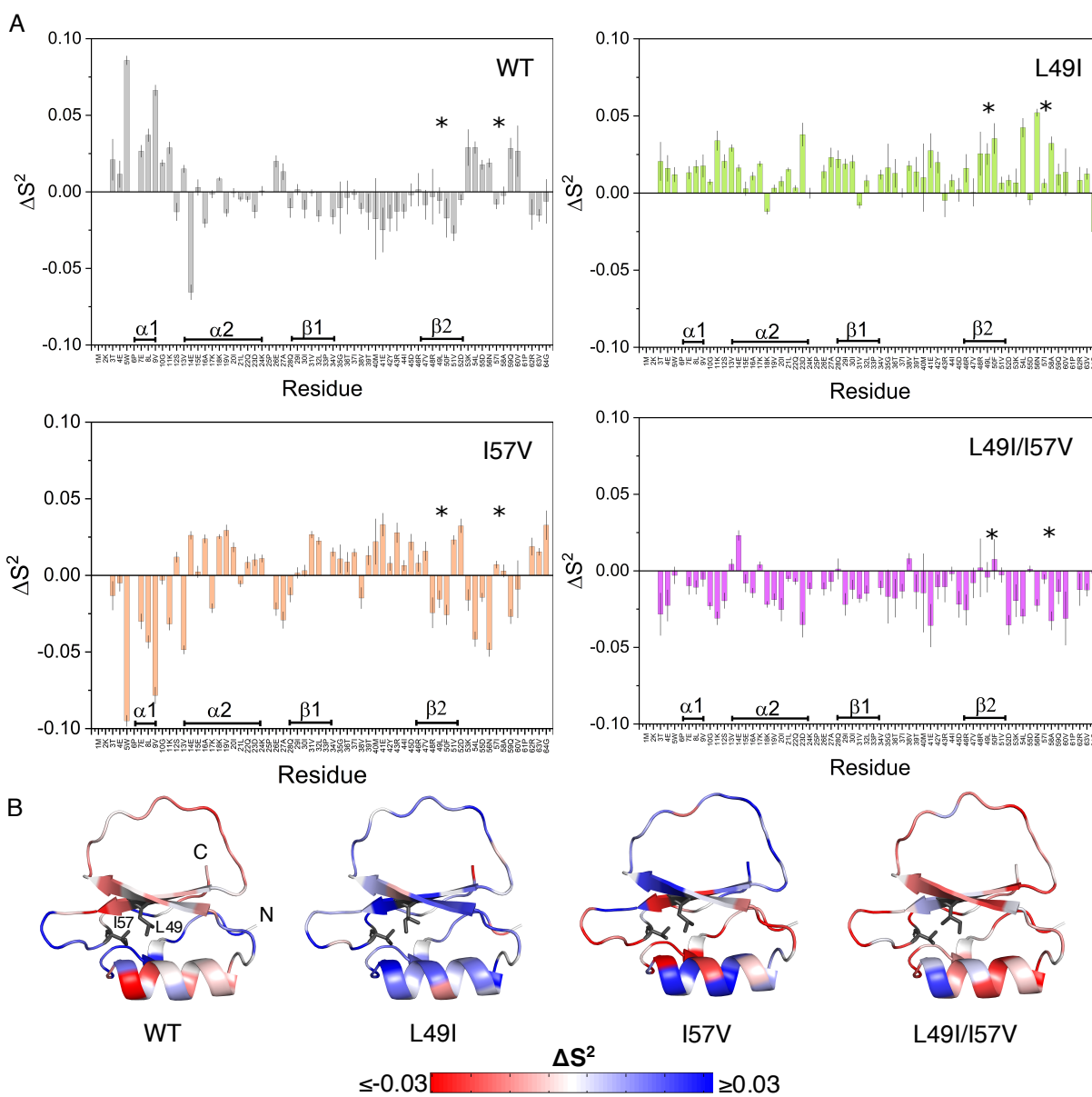


**Figure 2.** Chemical shift analysis. (A) Secondary  $C\alpha$  chemical shifts. (B) Absolute difference in methyl  $^{13}C$  CSs of the wild type (WT) and mutated variants of CI2. Mutation sites are highlighted with the asterisks. Secondary structure elements are depicted at the top of each panel.

For the backbone  $^{15}N$  amides, we acquired  $R_1$ ,  $R_2$  and  $\{^1H\}$ - $^{15}N$  NOE relaxation parameters at two magnetic fields (600 and 750 MHz) for each residue. The six experimental data points allowed us to sample the spectral densities,  $J(\omega)$ , at five different frequencies for each residue. We analyzed the data with the model-free formalism and to avoid overfitting we excluded MF models containing more than three parameters. Since the initial analysis in *relax* showed that the ratio between parallel and perpendicular components ( $D_{||}/D_{\perp}$ ) of the rotational diffusion tensor was less than 1.3 for all four CI2 variants, we used an isotropic model for the global molecular tumbling.

For each CI2 variant, we were able to analyze the data from all non-proline backbone amides, except for K2. The consistency of the relaxation data for the remaining residues were assessed by comparing the values of the spectral densities at zero frequency,  $J(0)$ , calculated from data collected at either of two different fields. In the absence of  $\mu\text{s}$ – $\text{ms}$  motions, the  $J(0)$  values calculated from the two datasets are expected to be the same.<sup>60</sup> For all four CI2 variants  $J(0)$  values from the two fields are highly correlated (**Figure S2**). Most of the deviating amide groups are also characterized by elevated  $R_1R_2$  values. Unlike  $R_2/R_1$  ratio, the  $R_1R_2$  product includes the effects of chemical exchange but not the effects of motional anisotropy.<sup>61</sup> Comparison of the  $R_1R_2$  values between the CI2 variants suggests that L49I may experience a decrease in  $\mu\text{s}$ - $\text{ms}$  conformational dynamics whereas I57V and L49I/I57V, oppositely, may experience an increase compared to the wild type (**Figure S3**). We note that a high correlation between the  $R_2/R_1$  ratio and the  $R_1R_2$  product suggests little motional anisotropy and thus justifies an isotropic model for the global molecular tumbling of CI2 variants.

Overall, the order parameters,  $S^2_{NH}$ , for the backbone amides are quite similar for all variants with a value around 0.8 in the structured regions of the protein. which drops to  $\sim 0.7$  in the more flexible inhibitory loop around residues I37-I44 (**Figures S4A**). The lowest value, indicating greatest flexibility, is found at residue M40 that plays an important functional role in chymotrypsin inhibition. We note that several residues around this region, according to their elevated  $R_1R_2$ , values experience slower dynamics.



**Figure 3.** The difference in NMR-derived order parameters (for backbone N-H groups) between one of the CI2 variants and the average value of three other CI2 variants:  $\Delta S^2_{NH} = S^2_{NH}(\text{variant}) - S^2_{NH}(\text{average-of-three-other-variants})$ . Secondary structure elements are shown at the bottom of each panel. Mutation sites are highlighted by asterisks. (A)  $\Delta S^2_{NH}$  plots (B) The difference in order parameters is shown as a color scale on the structure of CI2.

We compared the  $S^2_{NH}$  values between the variants of CI2 to assess the effects of the mutations on the dynamics. To minimize the effects of outliers, we calculated the  $\Delta S^2_{NH}$  values as the difference between the  $S^2_{NH}$  values for each variant and the average values of the other three variants, as recently suggested.<sup>62</sup> For example, in the case of WT:  $\Delta S^2_{WT} = S^2_{WT} - \langle S^2 \rangle_{L49I, I57V, L49I/I57V}$ . The analysis demonstrates that the differences in ps-ns conformational dynamics overall is small, however, we do see some sequence dependent variation (**Figure 3**).  $\Delta S^2_{WT}$  is positive for the first eleven residues, residues 26E, 27A, and in the region close to the mutation site 57 (53K-56N, and 59Q-60V) suggesting smaller than average amplitude of the backbone dynamics of WT for these residues. For most of the other residues  $\Delta S^2_{WT}$  are around 0.01-0.02 units lower than the average values of the other three variants. The destabilizing L49I mutation results in a decrease in the dynamics of almost all amide groups in CI2: most of  $\Delta S^2_{L49I}$  values are in between 0.01-0.2 units. The stabilizing I57V mutation results in a distribution of  $\Delta S^2_{I57V}$  values which almost mirrors

**Table 1.** The total average difference in order parameters for each CI2 variant.

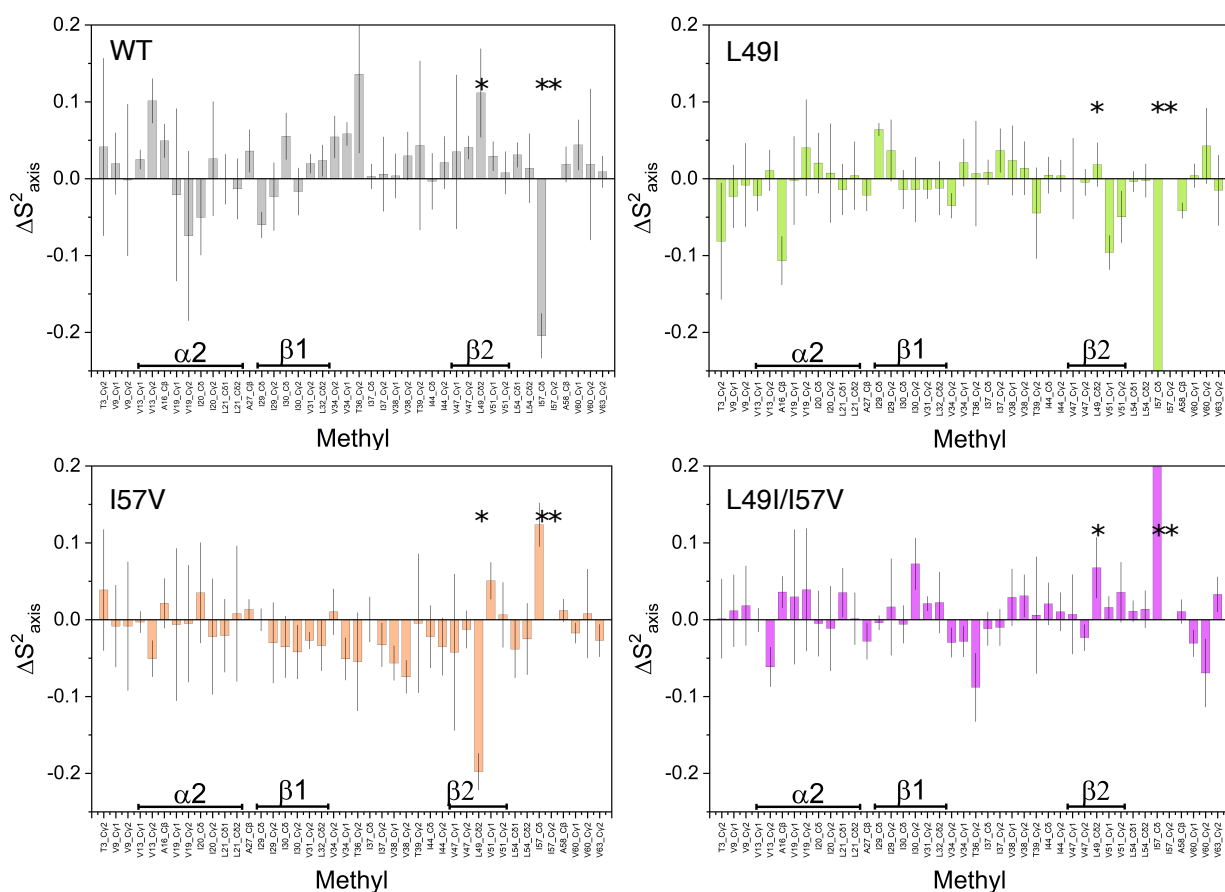
Variant	$\Delta S^2_{NH, total}$ (backbone) <sup>a</sup>			$\Delta S^2_{axis, total}$ (side chains) <sup>a</sup>		
	experimental	simulated		experimental	simulated	
		original	reweighted		original	reweighted
WT	0.1±0.1	0.2±0.1	-0.03±0.05	0.7±0.4	-0.7±0.1	0.3±0.2
L49I	0.8±0.1	-0.1±0.1	0.4±0.1	-0.3±0.2	-0.3±0.1	-0.3±0.2
I57V	-0.11±0.04	0.1±0.1	0.14±0.04	-0.6±0.3	0.8±0.2	0.4±0.2
L49I/I57V	-0.8±0.1	-0.03±0.09	-0.61±0.05	0.1±0.2	0.2±0.1	-0.7±0.2

<sup>a</sup>The values are calculated as a difference between a variant and the average of three other variants. Simulated data are presented before and after the reweighting procedure. The uncertainty is presented as the standard error of the mean.

that for WT, namely positive  $\Delta S_{I57V}^2$  for the first eleven residues and for residues 26E, 27A, 53K-56N, and 59Q-60V. The I57V/L49I variant is characterized by the increased dynamics for almost all backbone amide groups. The summed (excluding the mutation sites) differences in the backbone order parameters,  $\Delta S_{total}^2 = \sum_{n=1}^N \Delta S_i^2$  (where  $i$  is an amide group and  $N$  is the total number of amide groups), are listed in **Table 1**.

**Side chain dynamics.** We assigned the chemical shifts for all side chain methyl groups of WT CI2. However, some were excluded from the relaxation analysis. For L8  $^2\text{H-C}^{\delta 2}$ , V31  $^2\text{H-C}^{\gamma 1}$ , and V63  $^2\text{H-C}^{\gamma 1}$  it was not possible to estimate the volume or the maximum height due to significant peak overlap. Moreover, the chemical shifts for L8- $\text{C}^{\delta 1}$ , L32- $\text{C}^{\delta 1}$ , and L49- $\text{C}^{\delta 1}$  (in case of WT and I57V) were greater than 24 ppm. This may result in strong  $^{13}\text{C}$ - $^{13}\text{C}$  coupling between  $\text{C}^{\gamma}$  and  $\text{C}^{\delta}$  in leucine residues and the  $^2\text{H}$  relaxation data can be compromised.<sup>46</sup> In total, we included 42 peaks in the analysis of side chain ps-ns dynamics.

Fast dynamics of side chain  $\text{CH}_2\text{D}$  isotopomers can be tracked based on two to five relaxation rates. In this study we acquired four side-chain relaxation rates at two magnetic fields (600 and 750 MHz):<sup>44,45</sup>  $R_{D1}(\text{D}_z)$  – longitudinal magnetization;  $R_{D2}(\text{D}_+)$  – in-phase transverse magnetization;  $R_{D3}(3\text{Dz}^2 - 2)$  – quadrupolar order magnetization; and  $R_{D4}(\text{D}_+^2)$  – double quantum magnetization. Collection of these four relaxation rates results in an overdetermination since they depend on the spectral densities at only three different frequencies (zero frequency  $J(0)$ , the deuterium frequency at the given field strength  $J(\omega_D)$ , and double frequency of deuterium  $J(2\omega_D)$ ).<sup>44</sup> Therefore, according to the relationship  $R_{D4} = 1/2R_{D1} + 1/6R_{D3}$  we found that the consistency of the datasets at both magnetic fields is good ( $R^2 \geq 0.97$ , **Figure S5**). All acquired side chain deuterium relaxation rates were fitted to LS2 and LS3 model-free models.

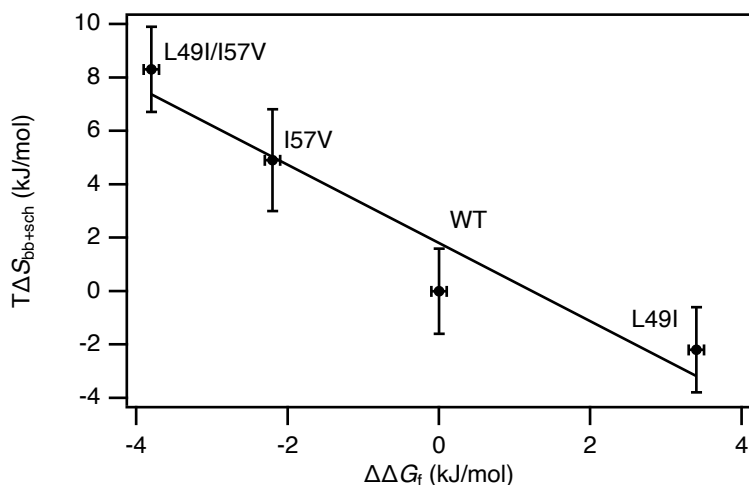


**Figure 4.** The difference in NMR-derived squared generalized order parameters (for sidechain methyl groups) between one of the CI2 variants and the average value of three other CI2 variants:  $\Delta S^2_{\text{axis}} = S^2_{\text{axis}}(\text{variant}) - S^2_{\text{axis}}(\text{average-of-three-other-variants})$ . The borders of the structural elements are depicted at the bottom of each panel, mutation sites are highlighted by asterisks.

As in case of backbone amide order parameters, side chain methyl  $S^2_{\text{axis}}$  values are quite similar for all the variants (**Figure S4B**). Similar to  $S^2_{\text{NH}}$ , we calculated  $\Delta S^2_{\text{axis}}$  values as the difference between  $S^2_{\text{axis}}$  values of each of the variants and the average values of other three variants (**Figure 4**). A direct comparison suggests that there may be subtle changes in the ps-ns dynamics of side-chain methyl groups. However, the errors are close to the value of  $\Delta S^2_{\text{axis}}$ . In general, according to

the total summed  $\Delta S^2_{\text{axis}}$ , side chain methyl groups of the three mutants are more flexible than the wild type (**Table 1**).

**Entropy calculations.** To obtain a quantitative measure of the effects of the mutations we used the  $\Delta S^2_{\text{NH}}$  and  $\Delta S^2_{\text{axis}}$  values to estimate the change in the conformational entropy by applying the “entropy meter” approach.<sup>63,49</sup> The entropy contribution to the free energy calculated in this way correlates well with experimental  $\Delta\Delta G_f$  values (Pearson’s  $r = -0.96$ ; **Figure 5**). The individual contributions from the backbone and the side-chains are in the same order of magnitude although most of the error in the total value originate from the side-chain data (**Table S1**). We note that the variation in the calculated  $T\Delta S$  is larger than the variation in  $\Delta\Delta G_f$ , which could either result from entropy-enthalpy compensation or because the “entropy meter” overestimates the entropy in our case. Still, the conclusion from the strong correlation is that changes in conformational entropy are important for the changes in conformational stability we observe for CI2.



**Figure 5.** Correlation between changes in thermodynamic stability of CI2 variants values from unfolding experiments ( $\Delta\Delta G_f$ ) and changes in the entropy contribution from ps-ns dynamics to the free energy ( $T\Delta S_{\text{bb+sch}}$ ).

**Molecular dynamics simulations.** We performed MD simulations of the CI2 variants using a99SB-disp with the modified TIP4PD water model,<sup>50</sup> a force field that substantially improves the accuracy for simulations of the disordered elements without sacrificing accuracy for structured regions. This force field was chosen due to the possible higher flexibility of several residues in the inhibitory CI2 loop. For all CI2 variants we conducted 3 simulations of 1  $\mu$ s each, obtaining a total of 3  $\mu$ s of sampling per variant.

As two rotamers of I49 are observed in the crystal structure of the L49I/I57V variant (PDB: 7AON) we ran two independent simulations starting from each of the rotamers. In the first rotamer structure, I49I/I57V(A), the I49  $\chi^1$  angle (N-C $\alpha$ -C $\beta$ -C $\gamma$ ) is in gauche- conformation and in the second structure, I49I/I57V(B), the I49  $\chi^1$  angle is in gauche+ conformation. For L49I/I57V, the I49  $\chi^1$  angle populates the gauche- rotamer for 95% of the time and the gauche+ for the remaining 5%. This behavior is independent of the initial rotameric state. I49  $\chi^1$  flips from gauche+ to gauche- within the first 15 ns of the simulations, which indicated that the force field prefers the gauche- conformation of I49  $\chi^1$ . We therefore only consider I49I/I57V(A) in the following. Initial analysis of the simulated ensembles revealed no changes in the overall structure of either of the variants: the average all-atom RMSD is in the range of 0.26-0.29 nm and per-residue RMSF does not exceed 0.6 nm for the most flexible residue (M40) in all five systems.

**Simulation-derived order parameters.** For a more detailed comparison of NMR and MD data we analyzed the acquired trajectories in terms of backbone amide and side chain methyl order parameters. As for the experimental data we used the model-free formalism. For consistency with experimental data the extended model-free models were omitted. The impact of the amount of sampling on the values of the order parameters was checked by employing a leave-N-out cross-

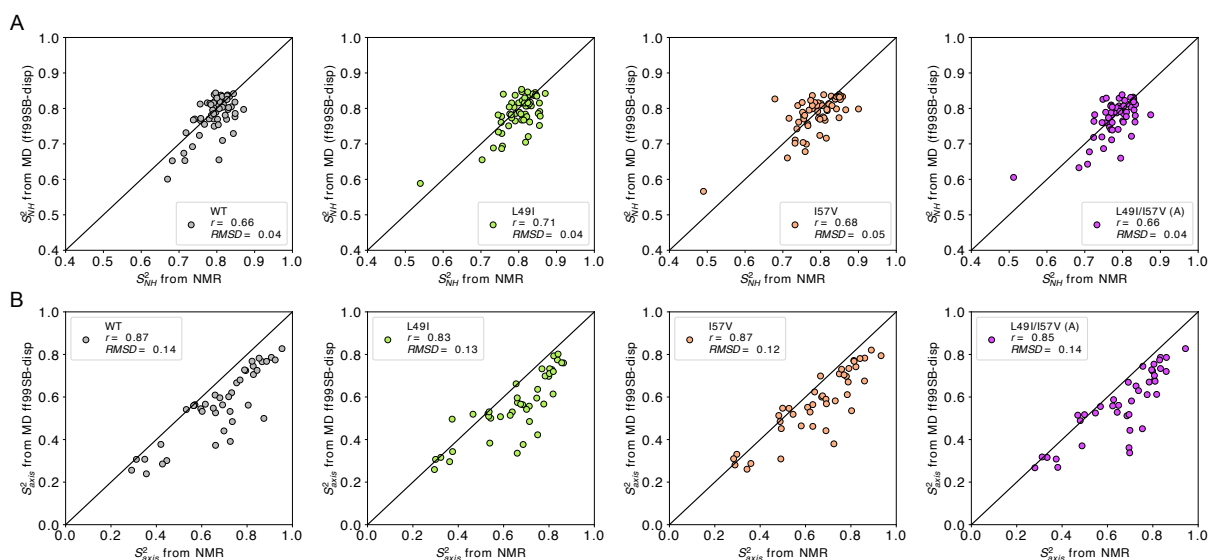


validation. The analysis suggests that our estimations of the order parameters are converged (**Figure S1**).

$S^2_{NH, sim}$  values are quite similar for all CI2 variants and reflect the absence of any major effect of the mutations. However, the correlations between  $S^2_{NH, sim}$  and  $S^2_{NH}$  for the four variants are not strong (Pearson's correlation coefficient,  $r = 0.55-0.62$ ; **Figures S6A**). This may be related to several issues. First, the simulations may not be converged, though our analysis described above suggests this to be a minor effect. Second, the  $^{15}\text{N}$  transverse relaxation rates can be significantly affected by  $\mu\text{s}$ - $\text{ms}$  timescale dynamics itself ( $R_{ex}$  contribution, **Figure S3**) and it is not always easy to accurately correct for these effects.<sup>64,65</sup> Third, despite the application of a state-of-the-art force field, imperfections can still reduce the match between simulations and experiments. These limitations of the computational approach may significantly affect not only the absolute values of the order parameters but also the difference in the dynamics of CI2 variants. As in case of the backbone, limited agreement between NMR and MD sidechain order parameters ( $S^2_{axis, sim}$ ) may be related to the force field imperfections and was observed in other studies (**Figure S6B**).<sup>46,66</sup>

**Reweighting MD trajectories.** It was shown previously that the direct integration of experimental data with molecular simulations can significantly improve a detailed description of the structural and dynamic properties of biomolecules. Such integration can be performed *a priori* using MD simulations biased with experimental data or by reweighting unbiased simulations using experimental data *a posteriori*.<sup>67</sup> The second approach allows us to reweight the simulated trajectories also using our NMR relaxation data.<sup>59</sup> The relaxation rates are used in the fitting procedure to obtain  $S^2_{NH}$  and  $S^2_{axis}$  and can be considered as a source of experimental data.<sup>46,66</sup> Due to slow  $\mu\text{s}$ - $\text{ms}$  time scale motions for some residues in CI2, the  $R_2$  rate values can be affected by exchange (**Figure S3**). This contribution is not captured in MD simulations, which can result in

overfitting during the reweighting procedure if we directly target the relaxation rates. Consequently, instead we reweighted our simulated trajectories using  $S^2_{NH}$  and  $S^2_{axis}$  values. We used an approach inspired by our recently developed ABSURDer method<sup>57</sup> but targeting the order parameters as also done previously.<sup>54,58</sup> For each CI2 variant  $S^2_{NH,sim}$  and  $S^2_{axis,sim}$  values were calculated separately for each of the 300 blocks of 10 ns and further reweighted against the experimental  $S^2_{NH}$  and  $S^2_{axis}$  values.



**Figure 6.** Experimental backbone  $S^2_{NH}$  (A) and side chain  $S^2_{axis}$  (B) order parameters of CI2 compared to their *reweighted* computational counterparts extracted from MD simulations. Pearson correlations ( $r$ ) and RMSDs were calculated for each variant separately.

The reweighting procedure resulted in a noticeable improvement of the agreement between the experimental and computational data (**Figure 6**). Pearson's correlation coefficient ( $r$ ) and RMSD for the backbone  $S^2_{sim}$  and  $S^2_{NH}$  improves for all CI2 variants (from  $r = 0.45$ - $0.62$ , RMSD =  $0.05$ - $0.08$  to  $r = 0.59$ - $0.71$ , RMSD =  $0.04$ - $0.05$ ; **Figures 6A**). Similar to the experimental data, we compared  $\Delta S^2_{NH,sim}$  values calculated as a difference between  $S^2_{NH,sim}$  values of each of the variants and the average values of other three variants.<sup>62</sup> The  $\Delta S^2_{NH,sim,total}$  values get closer to the

experimental  $\Delta S^2_{NH, total}$  values and better reflect experimentally observed changes in the order parameters (with an exception of I57V; **Table 1**). In case of the side chain methyl order parameters ( $S^2_{axis, sim}$ ) after the reweighting  $r$  and RMSD also improve (from  $r = 0.55-0.69$ , RMSD = 0.27-0.30 to  $r = 0.62-0.87$ , RMSD = 0.12-0.27; **Figure 6B**).

The trajectory reweighting results in a significant improvement of the correlation between experimentally measured and MD-derived backbone and/or side chain order parameters for all CI2 variants, except I57V. However, even after the reweighting the MD simulations do not capture all small variations in the dynamics of the four CI2 variants. Consequently, we omitted calculations of conformational entropy from the simulations. Similar to the analysis of the experimental data (**Figures 3 and 4**), the reweighted computational ensembles did not reveal any clear changes in the structural and dynamic properties of CI2. This suggests that the changes in stability and conformational entropy are the combined effect of several subtle contributions. In accordance with the previous study of CI2 dynamics,<sup>26</sup> the mutations thus affect the conformational dynamics of all 64 amino acid residues as a whole, probably by introducing many small changes in the contact network of the protein.

## Discussion

We have demonstrated that ps-ns conformational dynamics of CI2 can be affected by mutations in the hydrophobic core and find that the level of the conformational entropy correlates with  $\Delta\Delta G_f$ . Based on the NMR measurements, we conclude that the entropy and in particular the backbone entropy provides a major contribution to the stability of the CI2 variants analyzed in this work. From the entropy meter estimates the increase in entropy is even larger than the gain in conformational stability suggesting that the mutations could be associated with an unfavorable loss

of intermolecular interactions. We note, however, complications arising from entropic contributions arising from dynamics on time scales not probed by the relaxation measurements,<sup>68</sup> and which may not be accurately captured by the entropy meter approach. Our analysis of the crystal structures of the four CI2 variants showed that the structures are very similar and only subtle structural changes are observed throughout the protein.<sup>32</sup> This picture is mirrored by our current NMR analysis. The very small changes in dynamics between the variants are thus scattered throughout the protein. The largest difference is near the N-terminal for residues 5-9 that form a helical turn. This region appears most rigid in the wild type and most dynamic in the I57V variant. We speculate that the packing between the residue at position 57 and V9 is slightly weakened by the presence of a Val position 59 and that this effect is alleviated by having an Ile at position 49. However, we do not see any structural rearrangements around V9 or any other residues in the helical turn.

Our MD simulations of the CI2 variants support our findings from the experimental data, i.e. that differences in the dynamics are subtle and distributed throughout the protein, although the simulations cannot fully reproduce our experimental data. The difference between NMR and MD derived dynamic properties may be related to the presence of fluctuations in the native state of CI2 and its variants that can affect fast protein dynamics but cannot be intensively sampled by relatively short simulations. This assumption may be supported by a previous computational study of CI2.<sup>69</sup> In that work MD simulations, restrained with protection factors for local H-exchange, revealed a native state ensemble of CI2 including rare, large fluctuations from the native state, which may take place on a timescale of milliseconds or more and cannot be related to the folding pathway of CI2.<sup>69</sup> This is consistent with our experimental  $R_1R_2$  values suggesting that mutations of CI2 may also affect the dynamics on slower ( $\mu$ s-ms) time scales.

Our best explanation for increased stability is thus that the mutations releases the previously proposed unfavorable strain in the hydrophobic mini-core of A16, L49 and I57,<sup>70</sup> and that this occurs without any actual rearrangement of the structure and thus of the enthalpic interactions that stabilize the folded state. This is much in line with the conclusions based on a structural analysis of the I57V mutant by Fersht and co-workers.<sup>71</sup>

In summary, we have demonstrated that in case of CI2 and its mutated variants, changes in fast ps-ns dynamics may significantly contribute to changes in protein stability. Our analysis suggests that also slower  $\mu$ s-ms timescale dynamics of CI2 may be affected by these mutations.

## AUTHOR INFORMATION

### **Corresponding Author**

\*E-mail: [kaare.teilum@bio.ku.dk](mailto:kaare.teilum@bio.ku.dk). Phone: +45 35 32 20 29, or E-mail: [lindorff@bio.ku.dk](mailto:lindorff@bio.ku.dk)

Phone: +45 35 32 20 27.

### **Funding Sources**

This work was supported by the following grants: Carlsberg Foundation (postdoc grant no. CF17-0491 to Y.G.); the Lundbeck Foundation to the BRAINSTRUC structural biology initiative (155-2015-2666, to K.L.-L.); the NordForsk Nordic Neutron Science Programme (to K.L.-L.); Novo Nordisk Foundation to the NMR infrastructure facility, cOpenNMR, (grant no. NNF18OC0032996) and to the ROBUST Resource for Biomolecular Simulations (grant no. NNF18OC0032608). We also acknowledge access to the Biocomputing Core Facility at the Department of Biology, University of Copenhagen.

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENT

The authors would like to thank Falk Hoffmann for the help with the calculation of order parameters from MD simulations and Yong Wang for the assistance with MD simulations setup.

## REFERENCES

- (1) Stein, A., Fowler, D. M., Hartmann-Petersen, R., and Lindorff-Larsen, K. (2019) Biophysical and Mechanistic Models for Disease-Causing Protein Variants. *Trends Biochem. Sci.* *44*, 575–588.
- (2) Morley, K. L., and Kazlauskas, R. J. (2005) Improving enzyme properties: when are closer mutations better? *Trends Biotechnol.* *23*, 231–237.
- (3) Manning, M. C., Chou, D. K., Murphy, B. M., Payne, R. W., and Katayama, D. S. (2010) Stability of Protein Pharmaceuticals: An Update. *Pharm. Res.* *27*, 544–575.
- (4) Karshikoff, A., Nilsson, L., and Ladenstein, R. (2015) Rigidity versus flexibility: the dilemma of understanding protein thermal stability. *FEBS J.* *282*, 3899–3917.
- (5) Gavrilov, Y., Dagan, S., Reich, Z., Scherf, T., and Levy, Y. (2018) An NMR Confirmation for Increased Folded State Entropy Following Loop Truncation. *J. Phys. Chem. B* *122*, 10855–10860.
- (6) Bigman, L. S., and Levy, Y. (2020) Entropic Contributions to Protein Stability. *Isr. J. Chem.* *60*, 705–712.

- (7) Teilum, K., Olsen, J. G., and Kragelund, B. B. (2011) Protein stability, flexibility and function. *Biochim. Biophys. Acta* 1814, 969–976.
- (8) Mittermaier, A., and Kay, L. E. (2004) The response of internal dynamics to hydrophobic core mutations in the SH3 domain from the Fyn tyrosine kinase. *Protein Sci.* 13, 1088–1099.
- (9) Berezovsky, I. N., Chen, W. W., Choi, P. J., and Shakhnovich, E. I. (2005) Entropic Stabilization of Proteins and Its Proteomic Consequences. *PLoS Comput. Biol.* 1, 11.
- (10) Solomentsev, G., Diehl, C., and Akke, M. (2018) Conformational Entropy of FK506 Binding to FKBP12 Determined by Nuclear Magnetic Resonance Relaxation and Molecular Dynamics Simulations. *Biochemistry* 57, 1451–1461.
- (11) Verteramo, M. L., Stenström, O., Ignjatović, M. M., Caldararu, O., Olsson, M. A., Manzoni, F., Leffler, H., Oksanen, E., Logan, D. T., Nilsson, U. J., Ryde, U., and Akke, M. (2019) Interplay between Conformational Entropy and Solvation Entropy in Protein–Ligand Binding. *J. Am. Chem. Soc.* 141, 2012–2026.
- (12) Evenäs, J., Malmendal, A., and Akke, M. (2001) Dynamics of the Transition between Open and Closed Conformations in a Calmodulin C-Terminal Domain Mutant. *Structure* 9, 185–195.
- (13) Millet, O., Mittermaier, A., Baker, D., and Kay, L. E. (2003) The Effects of Mutations on Motions of Side-chains in Protein L Studied by 2H NMR Dynamics and Scalar Couplings. *J. Mol. Biol.* 329, 551–563.
- (14) Siggers, K., Soto, C., and Palmer, A. G. (2007) Conformational Dynamics in Loop Swap Mutants of Homologous Fibronectin Type III Domains. *Biophys. J.* 93, 2447–2456.

- (15) Kleckner, I. R., and Foster, M. P. (2011) An introduction to NMR-based approaches for measuring protein dynamics. *Biochim. Biophys. Acta Proteins Proteom.* 1814, 942–968.
- (16) Stafford, K. A., Robustelli, P., and Palmer, A. G. (2013) Thermal Adaptation of Conformational Dynamics in Ribonuclease H. *PLoS Comput. Biol.* 9, e1003218.
- (17) Kovermann, M., Rogne, P., and Wolf-Watz, M. (2016) Protein dynamics and function from solution state NMR spectroscopy. *Q. Rev. Biophys.* 49, e6.
- (18) McPhalen, C. A., and James, M. N. (1987) Crystal and molecular structure of the serine proteinase inhibitor CI-2 from barley seeds. *Biochemistry* 26, 261–269.
- (19) Itzhaki, L. S., Otzen, D. E., and Fersht, A. R. (1995) The Structure of the Transition State for Folding of Chymotrypsin Inhibitor 2 Analysed by Protein Engineering Methods: Evidence for a Nucleation-condensation Mechanism for Protein Folding. *J. Mol. Biol.* 254, 260–288.
- (20) Ladurner, A. G., Itzhaki, L. S., Daggett, V., and Fersht, A. R. (1998) Synergy between simulation and experiment in describing the energy landscape of protein folding. *Proc. Natl. Acad. Sci. U. S. A.* 95, 8473–8478.
- (21) Elcock, A. H. (2006) Molecular Simulations of Cotranslational Protein Folding: Fragment Stabilities, Folding Cooperativity, and Trapping in the Ribosome. *PLoS Comput. Biol.* 2, e98.
- (22) Matouschek, A., Otzen, D. E., Itzhaki, L. S., Jackson, S. E., and Fersht, A. R. (1995) Movement of the position of the transition state in protein folding. *Biochemistry* 34, 13656–13662.
- (23) Neira, J. L., Itzhaki, L. S., Otzen, D. E., Davis, B., and Fersht, A. R. (1997) Hydrogen exchange in chymotrypsin inhibitor 2 probed by mutagenesis. *J. Mol. Biol.* 270, 99–110.



(24) Fersht, A. R. (2000) Transition-state structure as a unifying basis in protein-folding mechanisms: Contact order, chain topology, stability, and the extended nucleus mechanism. *Proc. Natl. Acad. Sci. U. S. A.* 97, 1525–1529.

(25) Whitley, M. J., Zhang, J., and Lee, A. L. (2008) Hydrophobic Core Mutations in CI2 Globally Perturb Fast Side-Chain Dynamics Similarly without Regard to Position. *Biochemistry* 47, 8566–8576.

(26) Whitley, M. J., and Lee, A. L. (2011) Exploring the role of structure and dynamics in the function of chymotrypsin inhibitor 2. *Proteins* 79, 916–924.

(27) Shakhnovich, E., Abkevich, V., and Ptitsyn, O. (1996) Conserved residues and the mechanism of protein folding. *Nature* 379, 96–98.

(28) Dokholyan, N. V., Li, L., Ding, F., and Shakhnovich, E. I. (2002) Topological determinants of protein folding. *Proc. Natl. Acad. Sci. U. S. A.* 99, 8637–8641.

(29) Jiao, X., Chang, S., Li, C., Chen, W., and Wang, C. (2007) Construction and application of the weighted amino acid network based on energy. *Phys. Rev. E* 75, 051903.

(30) Hilser, V. J., Dowdy, D., Oas, T. G., and Freire, E. (1998) The structural distribution of cooperative interactions in proteins: Analysis of the native state ensemble. *Proc. Natl. Acad. Sci. U. S. A.* 95, 9903–9908.

(31) Atilgan, A. R., Akan, P., and Baysal, C. (2004) Small-World Communication of Residues and Significance for Protein Dynamics. *Biophys. J.* 86, 85–91.

(32) Hamborg, L., Granata, D., Olsen, J. G., Roche, J. V., Pedersen, L. E., Nielsen, A. T., Lindorff-Larsen, K., and Teilum, K. (2021) Synergistic stabilization of a double mutant in chymotrypsin inhibitor 2 from a library screen in *E. coli*. *Commun. Biol.* *4*, 980.

(33) Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L. (2005) The FoldX web server: an online force field. *Nucleic Acids Res.* *33*, W382–W388.

(34) Das, R., and Baker, D. (2008) Macromolecular Modeling with Rosetta. *Annu. Rev. Biochem.* *77*, 363–382.

(35) Kazimierczuk, K., and Orekhov, V. Y. (2011) Accelerated NMR Spectroscopy by Using Compressed Sensing. *Angew. Chem. Int. Ed.* *50*, 5556–5559.

(36) Neri, D., Szyperski, T., Otting, G., Senn, H., and Wuethrich, K. (1989) Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional carbon-13 labeling. *Biochemistry* *28*, 7510–7516.

(37) Delaglio, F., Grzesiek, S., Vuister, Geerten W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* *6*, 277–293.

(38) Skinner, S. P., Fogh, R. H., Boucher, W., Ragan, T. J., Mureddu, L. G., and Vuister, G. W. (2016) CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis. *J. Biomol. NMR* *66*, 111–124.

(39) Lipari, G., and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* *104*, 4546–4559.

(40) Lipari, G., and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc.* *104*, 4559–4570.

(41) d’Auvergne, E. J., and Gooley, P. R. (2008) Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. *J. Biomol. NMR* *40*, 107–119.

(42) Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., and Gronenborn, A. M. (1990) Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.* *112*, 4989–4991.

(43) Muhandiram, D. R., Yamazaki, T., Sykes, B. D., and Kay, L. E. (1995) Measurement of  $^2\text{H}$   $T_1$  and  $T_{1\rho}$  Relaxation Times in Uniformly  $^{13}\text{C}$ -Labeled and Fractionally  $^2\text{H}$ -Labeled Proteins in Solution. *J. Am. Chem. Soc.* *117*, 11536–11544.

(44) Millet, O., Muhandiram, D. R., Skrynnikov, N. R., and Kay, L. E. (2002) Deuterium Spin Probes of Side-Chain Dynamics in Proteins. 1. Measurement of Five Relaxation Rates per Deuteron in  $^{13}\text{C}$ -Labeled and Fractionally  $^2\text{H}$ -Enriched Proteins in Solution. *J. Am. Chem. Soc.* *124*, 6439–6448.

(45) Skrynnikov, N. R., Millet, O., and Kay, L. E. (2002) Deuterium Spin Probes of Side-Chain Dynamics in Proteins. 2. Spectral Density Mapping and Identification of Nanosecond Time-Scale Side-Chain Motions. *J. Am. Chem. Soc.* *124*, 6449–6460.

(46) Hoffmann, F., Xue, M., Schäfer, L. V., and Mulder, F. A. A. (2018) Narrowing the gap between experimental and computational determination of methyl group dynamics in proteins. *Phys. Chem. Chem. Phys.* *20*, 24577–24590.

(47) Lee, L. K., Rance, M., Chazin, W. J., and Palmer, A. G. (1997) Rotational diffusion anisotropy of proteins from simultaneous analysis of  $^{15}\text{N}$  and  $^{13}\text{C}\alpha$  nuclear spin relaxation. *J. Biomol Nmr* *9*, 287–298.

(48) Kjaergaard, M., and Poulsen, F. M. (2011) Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. *J. Biomol. NMR* *50*, 157–165.

(49) Sharp, K. A., O'Brien, E., Kasinath, V., and Wand, A. J. (2015) On the relationship between NMR-derived amide order parameters and protein backbone entropy changes: Measurement of Protein Backbone Entropy by NMR. *Proteins* *83*, 922–930.

(50) Robustelli, P., Piana, S., and Shaw, D. E. (2018) Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U. S. A.* *115*, E4758–E4766.

(51) Bussi, G., Donadio, D., and Parrinello, M. (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.* *126*, 014101.

- (52) Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.* *103*, 8577–8593.
- (53) Case, D. A. (1999) Calculations of NMR dipolar coupling strengths in model peptides. *J. Biomol. NMR* *15*, 95–102.
- (54) Best, R. B., and Vendruscolo, M. (2004) Determination of Protein Structures Consistent with NMR Order Parameters. *J. Am. Chem. Soc.* *126*, 8090–8091.
- (55) Genheden, S., Akke, M., and Ryde, U. (2014) Conformational Entropies and Order Parameters: Convergence, Reproducibility, and Transferability. *J. Chem. Theory Comput.* *10*, 432–438.
- (56) Allnér, O., Foloppe, N., and Nilsson, L. (2015) Motions and Entropies in Proteins as Seen in NMR Relaxation Experiments and Molecular Dynamics Simulations. *J. Phys. Chem. B* *119*, 1114–1128.
- (57) Kümmerer, F., Orioli, S., Harding-Larsen, D., Hoffmann, F., Gavrillov, Y., Teilum, K., and Lindorff-Larsen, K. (2021) Fitting Side-Chain NMR Relaxation Data Using Molecular Simulations. *J. Chem. Theory Comput.* *17*, 5262–5275.
- (58) Lindorff-Larsen, K., Best, R. B., Depristo, M. A., Dobson, C. M., and Vendruscolo, M. (2005) Simultaneous determination of protein structure and dynamics. *Nature* *433*, 128–132.
- (59) Orioli, S., Larsen, A. H., Bottaro, S., and Lindorff-Larsen, K. (2020) Chapter Three - How to learn from inconsistencies: Integrating molecular simulations with experimental data, in *Progress in Molecular Biology and Translational Science* (Strodel, B., and Barz, B., Eds.), pp 123–176. Academic Press.

(60) Morin, S., and M. Gagné, S. (2009) Simple tests for the validation of multiple field spin relaxation data. *J. Biomol. NMR* 45, 361–372.

(61) Kneller, J. M., Lu, M., and Bracken, C. (2002) An Effective Method for the Discrimination of Motional Anisotropy and Chemical Exchange. *J. Am. Chem. Soc.* 124, 1852–1853.

(62) Wallerstein, J. (2019) Molecular recognition and dynamics in proteins studied by NMR. Ph.D. Dissertation, Lund University, Lund.

(63) Kasinath, V., Sharp, K. A., and Wand, A. J. (2013) Microscopic Insights into the NMR Relaxation-Based Protein Conformational Entropy Meter. *J. Am. Chem. Soc.* 135, 15092–15100.

(64) Palmer, A. G. (2004) NMR Characterization of the Dynamics of Biomacromolecules. *Chem. Rev.* 104, 3623–3640.

(65) Jarymowycz, V. A., and Stone, M. J. (2006) Fast Time Scale Dynamics of Protein Backbones: NMR Relaxation Methods, Applications, and Functional Consequences. *Chem. Rev.* 106, 1624–1671.

(66) Hoffmann, F., Mulder, F. A. A., and Schäfer, L. V. (2020) Predicting NMR relaxation of proteins from molecular dynamics simulations with accurate methyl rotation barriers. *J. Chem. Phys.* 152, 084102.

(67) Bottaro, S., and Lindorff-Larsen, K. (2018) Biophysical experiments and biomolecular simulations: A perfect match? *Science* 361, 355–360.

(68) Hoffmann, F., Mulder, F. A. A., and Schäfer, L. V. (2021) How much entropy is contained in NMR relaxation parameters? preprint, *ChemRxiv* doi:10.33774/chemrxiv-2021-1686w.

(69) Best, R. B., and Vendruscolo, M. (2006) Structural Interpretation of Hydrogen Exchange Protection Factors in Proteins: Characterization of the Native State Fluctuations of CI2. *Structure* 14, 97–106.

(70) Ladurner, A. G., Itzhaki, L. S., and Fersht, A. R. (1997) Strain in the folding nucleus of chymotrypsin inhibitor 2. *Fold. Des.* 2, 363–368.

(71) Jackson, S. E., Moracci, M., elMasry, N., Johnson, C. M., and Fersht, A. R. (1993) Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. *Biochemistry* 32, 11259–11269.