The Dynamic Structure of Thrombin in Solution

Brian Fuglestad,[†] Paul M. Gasper,[†] Marco Tonelli,[‡] J. Andrew McCammon,[†] Phineus R. L. Markwick,[†] and Elizabeth A. Komives[†*]

[†]Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, California; and [‡]National Magnetic Resonance Facility at Madison, University of Wisconsin, Madison, Wisconsin

ABSTRACT The backbone dynamics of human α-thrombin inhibited at the active site serine were analyzed using R_1, R_2, and heteronuclear NOE experiments, variable temperature TROSY 2D [^{1}H-^{15}N] correlation spectra, and R_{ex} measurements. The N-terminus of the heavy chain, which is formed upon zymogen activation and inserts into the protein core, is highly ordered, as is much of the double beta-barrel core. Some of the surface loops, by contrast, remain very dynamic with order parameters as low as 0.5 indicating significant motions on the ps-ns timescale. Regions of the protein that were thought to be dynamic in the zymogen and to become rigid upon activation, in particular the γ-loop, the 180s loop, and the Na^+ binding site have order parameters below 0.8. Significant R_{ex} was observed in most of the γ-loop, in regions proximal to the light chain, and in the β-sheet core. Accelerated molecular dynamics simulations yielded a molecular ensemble consistent with measured residual dipolar couplings that revealed dynamic motions up to milliseconds. Several regions, including the light chain and two proximal loops, did not appear highly dynamic on the ps-ns timescale, but had significant motions on slower timescales.

INTRODUCTION

In blood coagulation, a cascade of proteolytic events converts zymogens to active proteases. Ultimately, prothrombin is cleaved twice removing the lipid-binding domain, and activating it for catalysis (1,2). The resulting α-thrombin cleaves 12 known substrates, including fibrinogen and factor (f) XIII (responsible for forming and cross-linking the fibrin clot) and the protease-activated receptors (responsible for platelet activation and many other cell signaling events). Thrombin upregulates its own formation by activating essential cofactors fV and fVIII and the protease fXI. Thrombin also downregulates its own formation when, in complex with thrombomodulin, it activates protein C, which is responsible for inactivating fVa and fVIIIa. The allosteric regulation of thrombin specificity toward multiple substrates is not yet understood (3,4).

The coagulation proteases all resemble trypsin, but whereas trypsin has a broad substrate specificity cleaving after any exposed Lys or Arg, the coagulation proteases have exquisite substrate specificity. Sequence alignment analysis revealed that the coagulation proteases have nonconserved surface insertions at several loops (5), and it has been suggested that only the trypsin-like catalytic core is conserved (6).

The structure of thrombin has mainly been analyzed by x-ray crystallography, and the many available structures reveal thrombin surface loops trapped in different conformations under crystallographic conditions. Di Cera's group recently captured the same mutant form of thrombin in two different conformational states in the same crystals (7). Hydrogen-deuterium exchange followed by mass spectrometry suggested that most of the surface loops are dynamic based on their rapid amide exchange behavior (8). In the first report of NMR resonance assignments of variously liganded forms the resonances for exosite 1 (both 70s and 30s loops) were missing unless a ligand was bound directly to these loops and resonances for the γ-loop were not observed under any conditions (9). Some resonances for the light chain were also missing. Although the authors of this work asserted that the absence of resonances is indicative of μs-ms dynamics, many other reasons including proteolysis of surface loops, or weak association of the proteins at high NMR concentrations could also account for such observations.

In this study, we provide a detailed quantitative analysis of the dynamic properties of thrombin across a broad range of timescales using a combination of NMR experiments and state-of-the-art molecular dynamics (MD) simulation. In the past two decades, NMR has emerged as the method of choice for investigating protein dynamics across a broad spectrum of timescales (10,11). In this study, we have employed nuclear spin relaxation (R_1, R_2, heteronuclear NOE), residual dipolar coupling (RDC) experiments, and relaxation due to chemical exchange (R_{ex}) measurements to study the dynamic properties of D Phe-Pro-Arg chloromethylketone (PPACK)-inhibited thrombin on multiple timescales. Although both nuclear spin relaxation and RDCs are associated with magnetic dipolar coupling, they probe

Submitted March 19, 2012, and accepted for publication May 30, 2012.

*Correspondence: ekomives@ucsd.edu

To accommodate readers who use one of several different numbering schemes for thrombin, we report residues in the chymotrypsin numbering, in which loops are denoted 60A, 60B, etc., followed by the sequential numbering used in the NMR data plots given in parentheses.

Editor: Patrick Loria.

0006-3495/12/07/0079/10

doi: 10.1016/j.bpj.2012.05.047

protein dynamics on very different timescales. Nuclear spin relaxation results from the time-dependent stochastic modulation of physical interactions between spin-active nuclei, inducing random field fluctuations, which relax the excited spin state back to equilibrium. These experiments are now routinely employed to probe local dynamic fluctuations up to the characteristic rotational correlation time of the system, which is nearly 20 ns for thrombin. In the framework of the well-known Lipari-Szabo model-free analysis, the temporal reorientation fluctuation of an internuclear vector is described using three parameters: the internal and total correlation times and an order parameter, S^2, which characterizes the spatial restriction of the internuclear vector.

RDCs have long been recognized as very powerful tools for structure determination, particularly when combined with other distance and dihedral angle restraints derived from NOEs and scalar J-couplings, respectively (12). However, it is now well accepted that RDCs can also be applied to the study of protein dynamics (13). RDCs are averaged over all orientations of the magnetic dipolar interaction vector sampled up to a timescale defined by the inverse of the alignment-induced coupling, and therefore report on time-averaged motions up to the millisecond range. This early work focused on rigid body domain motions (14,15). More recently attention has focused on the study of slow backbone conformational dynamics primarily in ubiquitin using both motional models and MD simulation strategies: The methods include the self-consistent RDC-based model free approach (16), the direct interpretation of dipolar couplings method (17), the structure-free Gaussian axial fluctuation model approach (18), ensemble refinement with orientational restraints (19), and the accelerated molecular dynamics (AMD) approach (20) employed in this work. For ubiquitin, the agreement between the RDCs back calculated from the AMD ensemble was statistically significantly better than the match to any crystal structure. Even more dramatic improvements were observed when the same approach was applied to the first four ankyrin repeats of IκBα (21).

Measurements of R_{ex} are an indicator of structural fluctuations on the 100 μs to ms timescales (11). These measurements are a result of the nuclear probe (backbone nitrogens in this study) undergoing exchange between states of distinct chemical shifts on the NMR timescale. The effect of temperature on line broadening was also measured. With increasing temperature, peak intensities are expected to increase as the tumbling of the molecule becomes faster. Deviations from this expected behavior are indicative of exchange processes and may give insight into the characteristics of these exchange processes.

The experimental NMR studies were complemented by a theoretical investigation using both conventional molecular dynamics (CMD) simulations and AMD, an enhanced conformational space sampling algorithm. We report here the first, to our knowledge, direct measurements of the backbone dynamics of PPACK-inactivated human α-thrombin. The results reveal an ensemble in which much of the surface of thrombin remains surprisingly dynamic even in the active-site liganded form.

MATERIALS AND METHODS

Protein purification and sample preparation

Isotopically labeled wild-type human thrombin (^{2}H-^{15}N or ^{2}H-^{15}N-^{13}C) was produced by a procedure modified from the previously published method (22). The procedure involves expression of prethrombin-2 in *Escherichia coli* from a modified pET23a+ vector containing the sequence of prethrombin-2 plus residues 310–327 of prothrombin at the N-terminus, which were critical for proper refolding. Details of the protein preparation are found in the Supporting Material.

The fully active α-thrombin was confirmed by fibrinogen clotting times (activity of >3500 units/mg), chromogenic substrate assays, and MALDI-TOF mass spectrometry of reduced and oxidized samples (revealing the correct molecular mass of the oxidized form of 34 kD and of the heavy chain upon reduction of 30.5 kD (including PPACK and seven acetamido groups on the cysteines).

Covalent inhibition of thrombin was achieved by incubation with a 10× molar excess of PPACK (Haematologic Technologies, Essex Junction, VT) at room temperature for 1 h. To separate excess PPACK from thrombin and exchange the protein into NMR buffer, the protein solution was concentrated and loaded onto a Superdex S75 size exclusion column (GE Healthcare Life Sciences, Piscataway, NJ) eluted with 25 mM sodium phosphate pH 6.5, 150 mM NaCl, 0.05% sodium azide. A final PPACK-thrombin concentration of 150 μM and 10% (v/v) D_2O was used for the NMR experiments at a final volume of 300 μL in a Shigemi tube (Shigemi, Allison Park, PA, catalog No. BMS-005).

NMR experiments and data analysis

NMR resonance assignment experiments were performed at 298°K unless otherwise specified on either a Bruker Avance III 600 MHz (UCSD Pharmacology), Varian VS 800 MHz (UCSD Chemistry), or a Varian VNS 800 MHz (National Magnetic Resonance Facility at Madison) all equipped with cryoprobes. Triple resonance assignment experiments performed were HNCO, HN(CA)CO, HN(CO)CA, HN(COCA)CB at 600 MHz as well as nuclear Overhauser effect spectroscopy (NOESY)-^{1}H-^{15}N-HSQC (τ_{mix} = 200 ms), transverse relaxation-optimized spectroscopy (TROSY)-HNCA, and TROSY-HN(CA)CB at 800 MHz (23–26). The details of the NMR experiments used here are found in the Supporting Material.

Accelerated MD calculations

The details of the AMD protocol have been discussed previously in the literature (27,28) and specifics of the protocol used here are found in the Supporting Material. In previous studies (20), we assumed that a single alignment tensor was sufficient to describe the orientation of the molecule in the alignment medium. In this work, we employed a multiple alignment tensor model because the PPACK-thrombin system exhibits considerable flexibility, particularly on slower timescales. Whereas for the previous study of ubiquitin the backbone root mean-square deviation (RMSD) across the AMD trajectory with respect to a randomly chosen structure for this system varied up to 1.4 Å; the backbone RMSD varied up to 2.9 Å for PPACK-thrombin. Large backbone RMSD values may represent a significant change in the preferential orientation of the molecule in the alignment medium. As the exchange between conformational states occurs on timescales considerably slower than the rotation diffusion coefficient (~20 ns according to Stokes theory), the system has ample time to adopt a different

preferential orientation as the system evolves from one conformational state to the next. The introduction of a multiple alignment tensor analysis presented in this work essentially couples the slow intramolecular conformational dynamics of the system to the preferential orientation of the molecule, although still maintaining the appropriate time and ensemble averaging properties of the RDC observables. The agreement between the theoretical and experimental RDCs was monitored using the *R*-factor:

$$R_{\text{factor}} = \sqrt{\frac{\sum_{i=1}^{N}\left(X(i)_{\text{theory}} - X(i)_{\text{exp}}\right)^2}{2\sum_{i=1}^{N}\left(X(i)_{\text{exp}}\right)^2}},$$

where $X(i)_{\text{theory}}$ and $X(i)_{\text{exp}}$ are the theoretically determined and experimental observables, respectively. The optimal torsional acceleration level (and hence the optimal conformational space sampling) for the reproduction of the experimental RDCs was found to be at [E_b(dih)-V_0(dih) = 960 kcal/mol, α(dih) = 240 kcal/mol]. The *R*-factors for the agreement of the experimental RDCs were 0.37 to the x-ray structure, 0.32 to the CMD ensemble, and 0.20 to the optimal AMD ensemble. Without the use of the multiple alignment tensor approach, the agreement was 0.23.

The internal dynamics present in the different CMD and AMD simulations of PPACK-thrombin were assessed by calculating the S^2 values relevant for the Lipari-Szabo type analysis (29) of ^{15}N autorelaxation data and for the S^2 values derived from the AMD ensemble that best fits the RDC data, respectively. In all cases, molecular ensembles generated from the standard CMD simulations and the free energy-weighted molecular ensembles generated from the AMD trajectories were superposed onto the backbone atoms (N, C^{α}, C′) of all heavy chain residues for the appropriate average structure. Order parameters, S^2 were calculated as

$$S^2 = \frac{1}{2}\left[3\sum_{i=1}^{3}\sum_{j=1}^{3}\langle \mu_i\, \mu_j\rangle^2 - 1\right],$$

where μ_i are the Cartesian coordinates of the normalized internuclear vector of interest. Others have shown that S^2 values calculated from standard MD simulations in this way were in excellent agreement with experimental S^2 values calculated using the Lipari-Szabo autocorrelation function approach (30).

Residual local frustration analysis

To analyze the crystal structure of PPACK-inhibited thrombin, we used an algorithm for determining residual local frustration, i.e., whether a contact between amino acid residues is energetically favorable or not in the folded state (31). This algorithm assesses residue-residue interactions by systematically perturbing the identity of individual residues and evaluating the resulting total energy change. The amino acids forming a particular contact are changed to other amino acids generating a set of decoys for which the total energy of the protein is recomputed. After constructing a histogram of the energy of the decoys and comparing the distribution to the native energy, cutoffs are implemented to identify minimally frustrated or highly frustrated contacts. Energetically favorable contacts between residues are depicted by green lines and highly frustrated or energetically unfavorable contacts are red.

RESULTS AND DISCUSSION

Backbone assignments of active site-inhibited human thrombin at 25°C

In an attempt to access loop regions of thrombin that appeared highly dynamic, we collected a suite of three-dimensional experiments for resonance assignments of human α-thrombin with PPACK bound at the active site. Of the 278 possible backbone amides, 245 were assigned for 88% coverage of the backbone. Backbone resonances (H^N, N, C′, C^{α}, C^{β}) matched well with previously published assignments of the protein at 37°C (9), with some differences (Fig. 1). At 25°C, the C-terminus of the light chain including resonances G1f through D1a were visible as was N60g, R126, D178, and K236. Remarkably, the resonances for the entire γ-loop, T147 through Q151, which were missing from the previously published assignments were observed and assigned in our spectra. Resonances V31 through K36, R67 through G69, I79, and E80, I88, W141, and Q156, which were assigned at 37°C, were not observed at 25°C. The temperature dependence of the line broadening is most likely due to chemical exchange in these residues.

Residual dipolar couplings

RDCs were measured using spin-state selected TROSY 2D [^{1}H-^{15}N] correlation experiments to measure chemical shift differences between the TROSY and ^{1}H anti-TROSY peaks in both isotropic and Pf1 phage aligned samples (32). Surprisingly, PPACK-thrombin aligned in only 3 mg/ml Pf1 phage and RDC values under these alignment conditions ranged from −39.07 to 48.68 (Fig. 2 *a*). Molecular dipole calculations indicate that the thrombin molecule is highly polarized (33), and it is likely that the alignment is electrostatically driven. RDCs could be measured for 209 of the 245 assigned residues. The RDCs calculated from the x-ray crystal structure (PDB ID: 1PPB (34),) using the program, PALES (35), matched poorly with the experimental values (R^2 = 0.72, Fig. 2 *b*). Specifically, of 209 measured RDCs, 55 were >10 Hz different from those calculated for the best-fit alignment tensor for the crystal structure. The agreement was substantially worse than was observed for the more rigid protein, ubiquitin (20) and somewhat worse than was observed for IκBα (21). The largest discrepancies were found in the loop regions.

Backbone dynamics experiments

R_1 and R_2 relaxation rates and ^{15}N-{^{1}H}NOEs were measured for PPACK-[^{2}H, ^{15}N] thrombin at 600 MHz (Fig. 3). Much of the protein appeared rigid as indicated by high ^{15}N-{^{1}H}NOEs and consistent R_2/R_1 ratios, but some regions had remarkably lower ^{15}N-{^{1}H}NOE values (Fig. 3 *c*). In particular, the C-terminus of the light chain (residues 14G–14M; 29–36) the γ-loop (residues 149–152; 185–193), and the 180s loop (residues 186A–186D; 229–233) had ^{15}N-{^{1}H}NOE values lower than 0.4, similar to the disordered termini of the protein. Such low ^{15}N-{^{1}H}NOE values strongly indicate motions on ps-ns timescales.

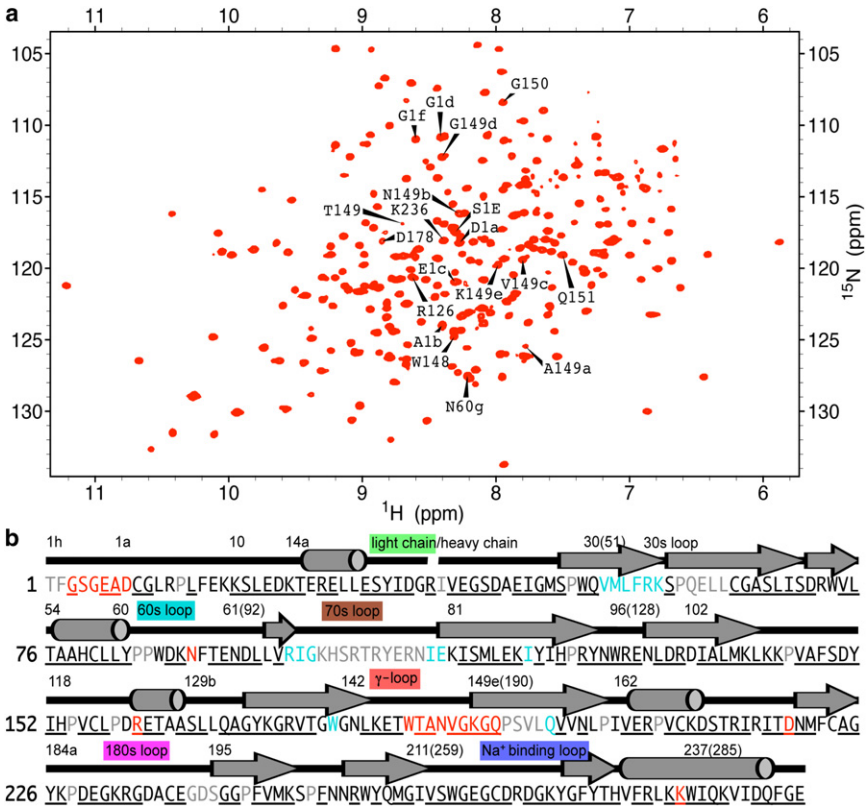

FIGURE 1 (*a*) TROSY 2D [^{1}H-^{15}N] correlation spectrum of PPACK-inhibited ^{2}H, ^{15}N-labeled human α-thrombin at 800 MHz and 25°C. Newly assigned resonances are labeled. (*b*) Sequence of human α-thrombin showing both commonly used numbering schemes (*chymotrypsin numbering above the text, sequential numbering at left and in parentheses*). Resonances that were assigned previously and also by us are in black, remaining unassigned resonances are in gray, those that are newly assigned in this work are in red, and resonances assigned only at 37°C (9) are in cyan. Important loops are labeled above the sequence in the same color scheme used in Fig. 6 *c* and Movie S1. Residues for which backbone dynamics data were obtained are underlined.

These regions also had lower than expected R_2 values (Fig. 3 *b*) indicative of heterogeneous dynamics.

The relaxation data were interpreted within the framework of the Lipari-Szabo model-free analysis using the program, TENSOR2. Initially, the R_1 and R_2 relaxation rates were used to calculate the overall rotational correlation time. A comparative analysis of isotropic, axially symmetric, and fully anisotropic diffusion tensors, revealed that the axially symmetric model provided the best representation of the rotational diffusion as identified by a statistical Monte Carlo analysis (36,37). The rotational correlation time was 16.3 ns, consistent with the molecular mass of thrombin, and $D_{\|}/D_{\perp}$ of 0.868. Order parameters (S^2) were calculated for 192 of the thrombin resonances (Fig. 3 *d*). The program TENSOR2 offers a variety of motional models of increasing complexity to fit the relaxation data (38). Over half of the residues (a total of 101 residues) could be fit with the simplest motional model (model 1) and nearly all of these had S^2 values above 0.8 and were in the core of the structure. Many residues (a total of 40) required fitting with a more complex motional model (model 5) and most of these had S^2 values below 0.8. These were located in the N-terminus and C-terminus of the light chain, the C-terminus of the heavy chain, the 30s loop, residues 30 and 36A (51 and 58) and the γ-loop, residues 149A–150 (186–191). Residues in the 60s loop, residues 60–60A and

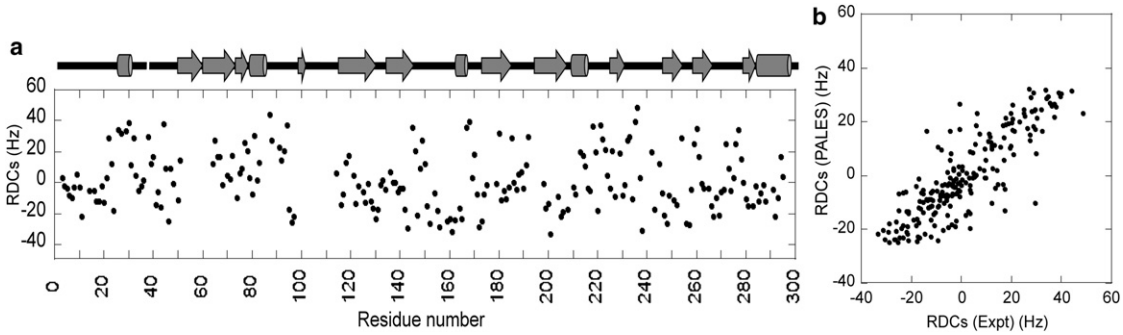

FIGURE 2 (*a*) RDC values measured by TROSY-anti-TROSY experiment collected at 800 MHz and 25°C using 3 mg/ml Pf1 phage for alignment. A schematic of the secondary structure is depicted above the plot. A break in the secondary structure diagram indicates where the light chain ends and the heavy chain starts. (*b*) Plot of measured RDCs versus those calculated from the x-ray structure (PDB code 1PPB) using PALES (35).

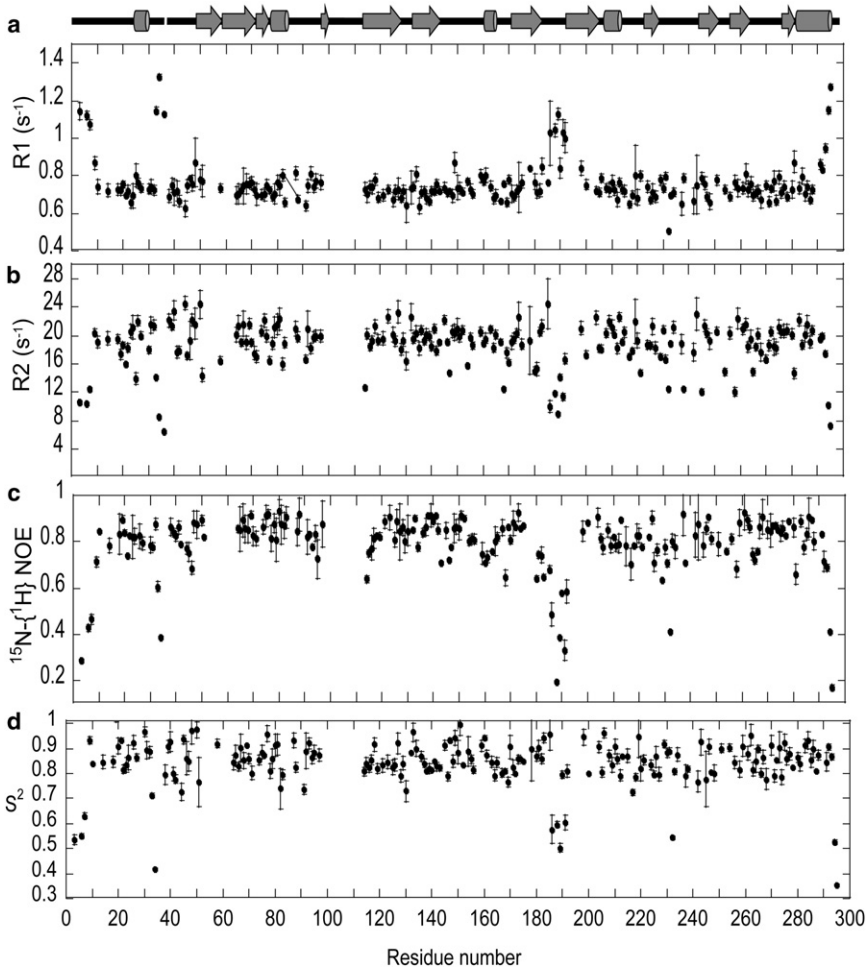

FIGURE 3 Backbone relaxation data ((*a*) R_1, (*b*) R_2, and (*c*) heteronuclear NOE) collected at 600 MHz, 25°C for [^{2}H, ^{15}N] PPACK-inhibited human α-thrombin as described in the Methods section in the Supporting Material. (*d*) S^2 values derived from the model-free analysis of the relaxation data using TENSOR2. The schematic of the secondary structure is depicted above the plot.

60I (82–83 and 91), the 90s loop residues 96 and 97A (128 and 130), and the 180s loop residues 184, 185, 186D, and 191 (225, 227, 232, and 237) also had S^2 values below 0.8 (see Fig. 6 *a*). The observation of so many resonances requiring model 5 fitting was strong evidence of heterogeneous fast timescale dynamics. To extract the contribution of chemical exchange to spin-spin relaxation, TROSY Hahn-echo experiments were performed (39). Of the 225 residues, only 15 had significant line broadening due to chemical exchange ($R_{ex} > 6\ s^{-1}$, see Fig. 5 *b*). These residues were: 1F and 3 (3,11) of the light chain; 23 and 27 (27 and 48); 66 (97) leading up to the 70s loop; 87 (119) in the surface strand that connects the 70s loop to the 90s loop; 144, 147, 149a, 149c, and 151 (180, 183, 186, 188, and 192) of the γ-loop; 138, 157, and 199 (174, 198, and 245) in the core β-sheet; and 233 (281) that forms a kink in the C-terminal helix. Of these residues, only 23 (44) and 138 (174) required a R_{ex} term to properly fit a Lipari-Szabo model to the relaxation data. These direct measurements of R_{ex} are much more accurate than the estimates obtained from the TENSOR2 analysis of the R_1, R_2, and ^{15}N-{^{1}H} NOE data.

Variable temperature experiments

Resonances for most of the 30s loop and the 70s loop were neither observed in previous experiments (at 37°C) nor in our experiments (at 25°C) (cf. Fig. 1). To further probe the temperature dependence of the thrombin backbone resonance intensities, we carried out variable temperature ^{1}H-^{15}N TROSY-HSQC experiments. These experiments revealed complex motions for several of the thrombin surface loops. For example, resonances in the γ-loop were not observed at 37°C (9), but are observed at lower temperatures (see Fig. 5 *c*). Resonances at the C-terminus of the light chain show similar behavior. In both cases, the decreased signal intensity at higher temperature indicates the region is undergoing exchange processes that are contributing to relaxation (40).

Relation between dynamics and local residual frustration

Over half of the residues in thrombin have high S^2 values, and form a rigid core. This is consistent with analysis

of the results of a residual local frustration analysis (31), which shows a strong network of minimally frustrated contacts throughout the core of the protein (see Fig. 6 *b*). It should be noted that this part of the molecule is also the part that is evolutionally conserved (5). Adjacent to the active site is Ile-16 (36), which makes a very large number of minimally frustrated contacts, and this corresponds to the new N-terminus generated by proteolytic activation that inserts into the core of the protein. Crosspeaks for the oxyanion hole, residues 193–196 (239–242) are not observed, and it is interesting that these residues are engaged in highly frustrated contacts. This is the only region of thrombin with frustrated contacts located in the core of the protein. Conversely, the dynamic surface loops have many highly frustrated contacts (see Fig. 6 *b*).

MD simulations

CMD simulations were carried out for 20 ns to reach the same timescale as the rotational diffusion time of the thrombin molecule (11,41). The S^2 values calculated from the CMD simulations matched well with those determined from the R_1, R_2, and ^{15}N-$\{^1H\}$NOE experiments (see Fig. 5 *a*). This result is a testament to the quality of the contemporary force field used in this study, ff99SB, and is particularly remarkable in light of the many assumptions that have to be made when interpreting spin relaxation data (42,43).

The excellent agreement between experimental and simulated S^2 values encouraged us to extend our theoretical study to investigate slower timescale motions using an enhanced conformational space sampling algorithm; AMD (27). AMD is an extended biased potential MD approach that allows for the efficient study of biomolecular systems up to timescales several orders of magnitude greater than accessible using standard CMD approaches, yet maintains a fully atomistic representation of the system. AMD has already been employed with great success to study the dynamics and conformational behavior of a variety of biomolecular systems including polypeptides, folded and natively unstructured proteins (28).

A series of five dual-boost AMD simulations were performed at increasing levels of acceleration (see the Supporting Material, Methods). For each AMD simulation a corrected canonical ensemble was obtained by performing a Boltzmann free energy reweighting protocol. Successively larger acceleration levels sample greater conformational space and the optimal conformational space sampling for the reproduction of the RDC data were obtained using singular value decomposition, following a protocol similar to that previously applied to ubiquitin and IκBα (20,21). A multiple alignment tensor model was implemented because the PPACK-thrombin system exhibited considerable flexibility, particularly on slower timescales (see details in Methods). The slower motions of the surface loops resulted in slight changes in the molecular alignment tensor such that the use of the multiple alignment tensors improved the agreement of the RDCs calculated from the AMD ensemble with the measured RDCs. This should not be confused with the molecular tumbling analysis implemented in TENSOR2 that determines whether the residues with relaxation parameters within the trimmed mean fit a particular tumbling model. Using this procedure, the optimal acceleration level, and hence the optimal conformational space sampling for the reproduction of the experimental RDCs was found to be [E_b(dih)-V_0(dih) = 960 kcal/mol, α(dih) = 240 kcal/mol] (see Methods). This optimal AMD ensemble recapitulated the measured RDC values much better (R^2 = 0.92) (Fig. 4 *b*) than both the x-ray crystal structure (R^2 = 0.72) (Fig. 2 *b*) and the CMD ensemble (R^2 = 0.80) (Fig. 4 *a*). The most dramatic improvement in the agreement with the experimental RDCs was found for residues in the loop regions (Fig. 4 *c*). S^2 values were calculated from the standard MD simulation (see Methods), and the results from the CMD ensemble were in excellent agreement with the experimental values calculated from the Lipari-Szabo autocorrelation function analysis, with a few small deviations for the most flexible residues (Fig. 5 *a*). A temporally heterogeneous distribution of dynamics is revealed in the AMD ensemble (as well as in the experiments), particularly in the light chain residues 14G–14M (29–36), the γ-loop, residues 149A–150 (186–191), the 180s loop, residues 184,185, 186D, and 191 (225, 227, 232, and 237), and the Na^+ binding site, residues 216–223 (264–271). The loops do not appear to hinge, but rather they appear to be appendages that extend from the rigid core in such a way that within the span of a few residues the S^2 values transition from completely ordered to highly disordered with motions on multiple timescales (Fig. 6 *a* and 6 *c*, Movie S1 in the Supporting Material).

The γ-loop remains remarkably dynamic in the inhibited form. The resonances of this loop were missing from all liganded states in previous studies that were carried out at 37°C (9), and at first we were convinced that the previous studies had been inadvertently carried out with proteolyzed protein. The entire γ-loop was visible in our studies carried out at 25°C, but it showed remarkably low S^2 values, some of the lowest reported for a region other than the terminus of a protein this size. These results help explain the proteolytic sensitivity of the γ-loop, which is also called the autolysis loop, because it can be cleaved by thrombin itself. The fact that the γ-loop does not become ordered even in the active site-inhibited form is remarkable.

Resonances in the 70s loop in anion-binding exosite 1 were only observed in previous work when a ligand was bound at this exosite and it was thus thought to be dynamic (9). We also did not observe resonances for this loop at 25°C. Surprisingly, however, the S^2_{RDC} values calculated from the AMD ensemble (which represents dynamics out to milliseconds) revealed that the 70s loop is no more

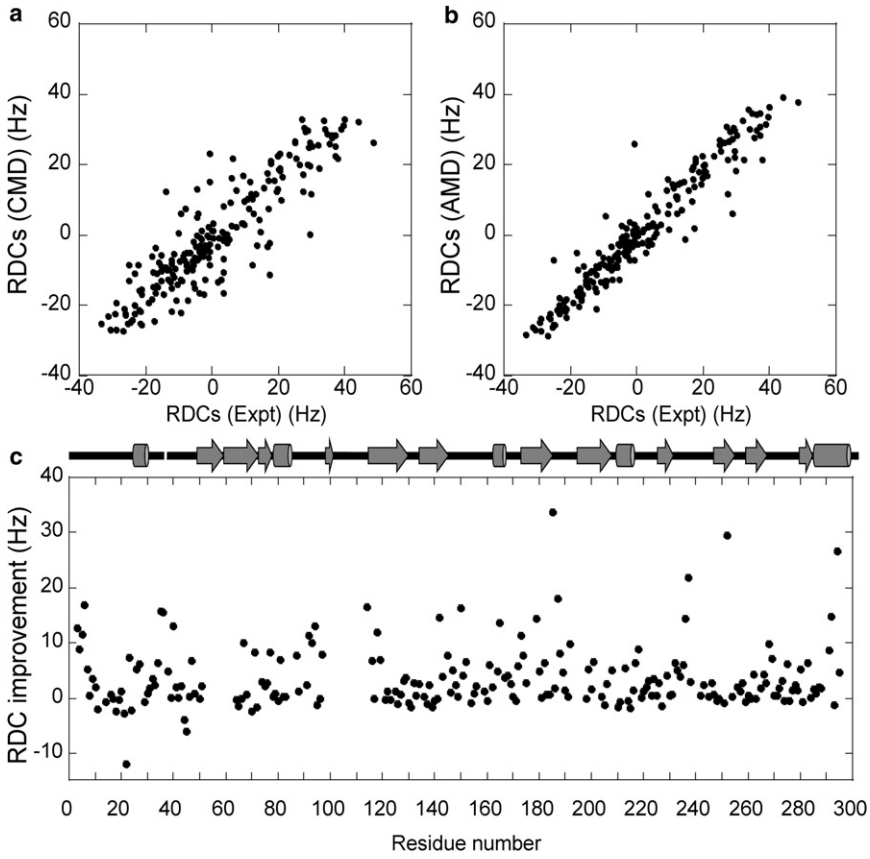

FIGURE 4 (*a*) Plot of measured RDCs versus those calculated from the CMD ensemble (R^2 = 0.80). (*b*) Plot of measured RDCs versus the RDCs calculated from the best-fit AMD ensemble (R^2 = 0.92). (*c*) Plot of the RDC improvement as defined by $|(RDC_{expt}\text{-}RDC_{x\text{-}ray})|\text{-}|(RDC_{expt}\text{-}RDC_{AMD})|$. The schematic of the secondary structure is depicted above the plot.

dynamic than other visible dynamic regions of the protein. Typically, the reason that resonances are not observed is that they are transitioning between conformational states with different chemical shifts on an intermediate timescale in the NMR experiment. It is interesting that the AMD simulations do not show any evidence of large conformational fluctuations in this loop, in fact, it appears uniformly structured (Fig. 6 *c*, 70s loop shown in *brown*). The results of the AMD simulations show that the absence of peaks in the 70s loop is not due wholly to dynamics as previously suggested (9) and illustrates the necessity of quantitative study of dynamics in thrombin.

CONCLUSION

Due to its size and complex folding pathway, direct measurements of backbone dynamics in thrombin have not been reported previously. In this first report, we wanted to explore the solution structure of human α-thrombin with a small inhibitor bound at the active site. Parts of the light chain, the entire γ-loop, and tips of the 60s and 90s loops and the Na^+-binding site had order parameters < 0.8 indicative of substantial dynamics on the ps-ns timescale. TROSY Hahn-echo experiments revealed a substantial number of residues undergoing relaxation due to chemical exchange, including residues in the light chain, the beginning of the heavy chain, one residue in the 90s loop, one residue in the strand that connects the 70s to the 90s loop, much of the γ-loop, two residues in regions proximal to the light chain, and three residues in the β-sheet core. The emerging picture was thus of a very dynamic molecule.

To obtain a more complete quantitative description of the dynamic properties of the system across multiple timescales, MD simulations were employed. AMD simulations were calibrated using RDC data and these extended MD simulations provided a more complete picture of temporally heterogeneous dynamic behavior of observed residues. Together the experiments and simulations yield an unprecedented molecular ensemble view of thrombin that truly represents the solution structure, revealing that PPACK-thrombin possesses a largely rigid core corresponding to the evolutionarily conserved protease with highly dynamic and less conserved loop appendages. The order parameters obtained from the AMD simulations (S^2_{RDC} values) were often much lower than the S^2 values calculated from the relaxation data revealing the true extent of thrombin's motions on longer timescales. Many residues had high S^2 values and much lower S^2_{RDC} values, but did not appear to be undergoing R_{ex} (Fig. 6 *d*). These residues may be undergoing motions slower than the slow limit of the relaxation calculated S^2 values (τ_c, ~20 ns) but faster than the fast limit of the chemical exchange regime (~100 μs). This is

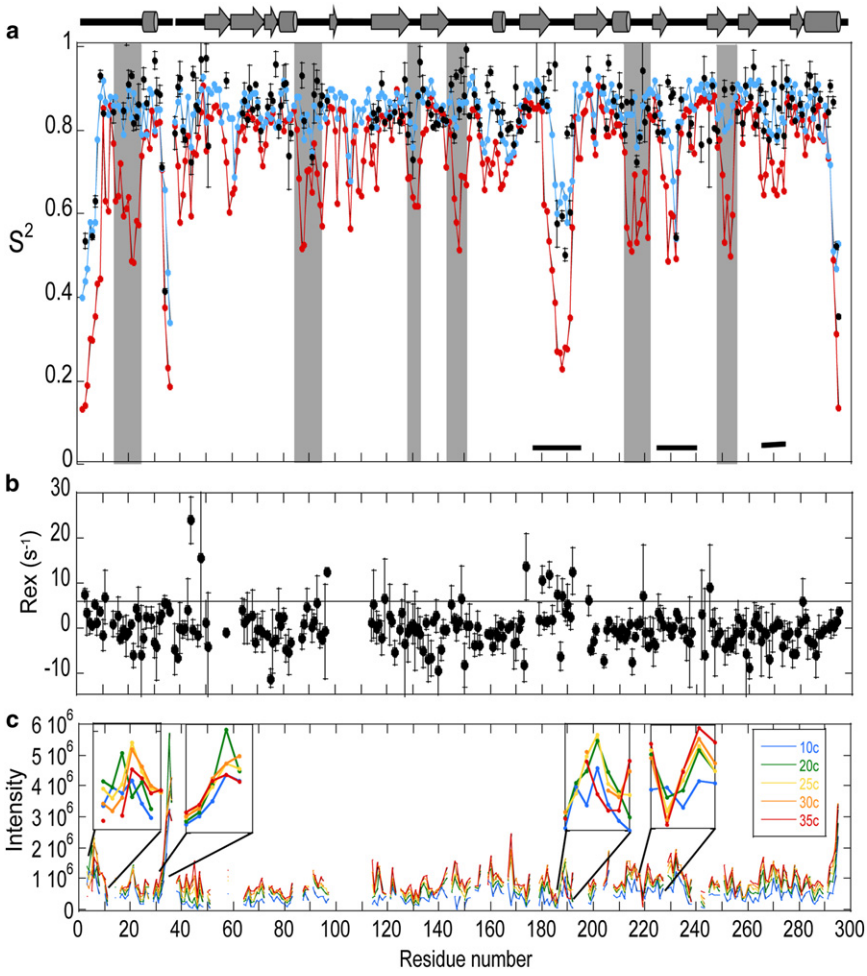

FIGURE 5 (*a*) Plot of the S^2 values calculated from the relaxation data using TENSOR2 (*black*), compared to the S^2 values calculated from the CMD ensemble (*blue*), and compared to the S^2_{RDC} values calculated from the RDC-optimized AMD ensemble (*red*). Regions of the heavy chain comprising the activation domain that were thought to become less dynamic in α-thrombin as compared to prethrombin-2 are marked with horizontal bars. Regions with significant S^2_{RDC} values that show little motional contribution from relaxation calculated S^2 values and measured R_{ex} are boxed in gray. (*b*) R_{ex} values calculated from the TROSY Hahn-echo experiment. Horizontal line at $R_{ex} = 6\ s^{-1}$ is the threshold for significant contributions to relaxation from chemical exchange. (*c*) Plot of the intensity of the TROSY peak for each resonance at different temperatures from 35°C down to 10°C. For most of the protein, the resonance intensity increases with temperature, however for some regions (*boxed and enlarged*) the relationship between intensity and temperature is indicative of complex slower timescale dynamics.

strong evidence of motions in the supra-τ_c regime (19). In addition, the AMD simulations revealed the dynamic behavior of the residues that were invisible in the NMR experiments.

Previous dynamic studies of enzymes have mostly observed that loops surrounding the active site display a hinging motion on the millisecond timescale between an open and closed state. Our combined relaxation experiments and simulations revealed multiple timescale dynamics of the thrombin loops, thus differing from this paradigm. The complex temperature dependence of resonance peak intensities in these regions is further evidence of the temporal

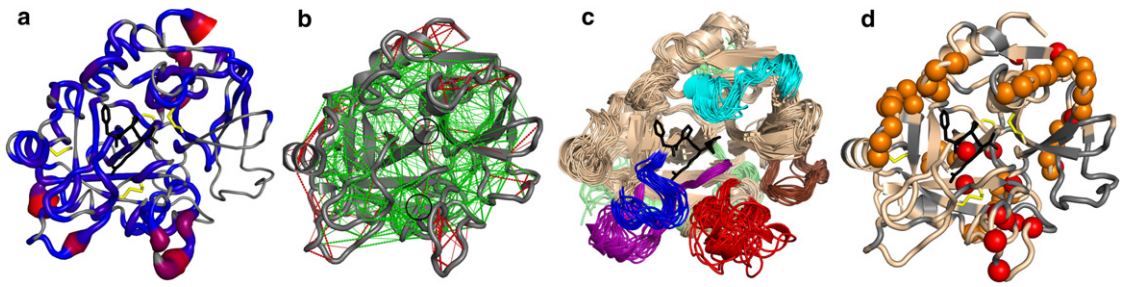

FIGURE 6 (*a*) Structure of PPACK-inhibited human α-thrombin depicting the ps-ns motions as determined from the backbone relaxation experiments. S^2 values are depicted as a backbone sausage model where the scale was red ($S^2 \leq 0.5$) to blue ($S^2 \geq 0.8$) with unobserved residues in gray. (*b*) Structure of PPACK-inhibited human α-thrombin showing the minimally frustrated contacts (*green*) and the highly frustrated contacts (*red*); the highly frustrated oxyanion hole and the minimally frustrated N-terminus of the heavy chain are circled. (*c*) Structural ensemble of PPACK-inhibited human α-thrombin calculated using AMD that best matched the RDC data. The dynamic regions are colored: light chain (*light green*), 60s loop (*cyan*), 70s loop (*brown*), γ-loop (*red*), 180s loop (*purple*), Na^+-binding site (*blue*). (*d*) Structure of PPACK-thrombin depicting residues with measured Rex $> 6\ s^{-1}$ (*red*), those with predicted supra-τ_c motions (*orange*), and unobservable residues (*gray*).

heterogeneity of the loop motions. The loops do not appear to hinge, but rather they appear to be flexible appendages that extend from the rigid core. One could say that thrombin resembles an anemone in the sense that its active site is surrounded by loops that exhibit both fast and long time scale dynamics.

It has long been proposed that one feature of zymogen activation is the further ordering of four structural segments that were observed to be deformed and flexible in crystal structures of the zymogen form: residues light chain up to residue 19, the γ-loop (residues 142–152; 178–195), the 180s loop (residues 184–193; 225–239), and the Na^+ binding site (residues 216–223; 264–271), which together form the activation domain (44). These regions also display temporally heterogeneous dynamics. Our results show that, in fact, zymogen activation does not result in ordering of these regions (marked with horizontal bars in Fig. 5 *a*) because they remain highly dynamic even in the active form albeit at longer timescales. The residues undergoing motions on the supra-τ_c timescale appear to surround a track of residues undergoing R_{ex} (Fig. 6 *d*). These apparent pathways of dynamic residues may be particularly important in the allosteric communication between the anion binding exosites and the active site (8,45).

SUPPORTING MATERIAL

Details of the thrombin preparation and purification, the NMR experimental methods, and the AMD protocol, and references and movie are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(12)00627-3.

We thank Tracy Handel and Michael Massiah for helpful discussions.

Financial support was from R01 HL070999. Some NMR experiments were performed at National Magnetic Resonance Facility at Madison. B.F. acknowledges training support from the Molecular Biophysics training grant (T32 GM08326) and from the American Heart Association. P.G. acknowledges training support from the Cell and Molecular Genetics training program. The computational work was funded in part by the National Institutes of Health, National Science Foundation, and Howard Hughes Medical Institute, the Center for Theoretical Biological Physics, the National Biomolecular Computing Resource, and the San Diego Supercomputer Center.

REFERENCES

1. MacFarlane, R. G. 1964. An enzyme cascade in the blood clotting mechanism, and its function as a biochemical amplifier. *Nature.* 202:498–499.
2. Davie, E. W., and O. D. Ratnoff. 1964. Waterfall sequence for intrinsic blood clotting. *Science.* 145:1310–1312.
3. Huntington, J. A. 2005. Molecular recognition mechanisms of thrombin. *J. Thromb. Haemost.* 3:1861–1872.
4. Davie, E. W., and J. D. Kulman. 2006. An overview of the structure and function of thrombin. *Semin. Thromb. Hemost.* 32 (*Suppl 1*):3–15.
5. Süel, G. M., S. W. Lockless, …, R. Ranganathan. 2003. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat. Struct. Biol.* 10:59–69.
6. Page, M. J., and E. Di Cera. 2010. Combinatorial enzyme design probes allostery and cooperativity in the trypsin fold. *J. Mol. Biol.* 399: 306–319.
7. Niu, W., Z. Chen, …, E. Di Cera. 2011. Crystallographic and kinetic evidence of allostery in a trypsin-like protease. *Biochemistry.* 50:6301–6307.
8. Koeppe, J. R., A. Seitova, …, E. A. Komives. 2005. Thrombomodulin tightens the thrombin active site loops to promote protein C activation. *Biochemistry.* 44:14784–14791.
9. Lechtenberg, B. C., D. J. Johnson, …, J. A. Huntington. 2010. NMR resonance assignments of thrombin reveal the conformational and dynamic effects of ligation. *Proc. Natl. Acad. Sci. USA.* 107:14087–14092.
10. Markwick, P. R. L., and M. Nilges. 2011. Computational approaches to the interpretation of NMR data for studying protein dynamics. *Chem. Phys.* 396:124–134.
11. Salmon, L., G. Bouvignies, …, M. Blackledge. 2011. Nuclear magnetic resonance provides a quantitative description of protein conformational flexibility on physiologically important time scales. *Biochemistry.* 50:2735–2747.
12. Tjandra, N., and A. Bax. 1997. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science.* 278:1111–1114.
13. Prestegard, J. H., H. M. al-Hashimi, and J. R. Tolman. 2000. NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Q. Rev. Biophys.* 33:371–424.
14. Tolman, J. R., J. M. Flanagan, …, J. H. Prestegard. 1997. NMR evidence for slow collective motions in cyanometmyoglobin. *Nat. Struct. Biol.* 4:292–297.
15. Braddock, D. T., M. Cai, …, G. M. Clore. 2001. Rapid identification of medium- to large-scale interdomain motion in modular proteins using dipolar couplings. *J. Am. Chem. Soc.* 123:8634–8635.
16. Lakomek, N. A., T. Carlomagno, …, J. Meiler. 2006. A thorough dynamic interpretation of residual dipolar couplings in ubiquitin. *J. Biomol. NMR.* 34:101–115.
17. Tolman, J. R. 2002. A novel approach to the retrieval of structural and dynamic information from residual dipolar couplings using several oriented media in biomolecular NMR spectroscopy. *J. Am. Chem. Soc.* 124:12020–12030.
18. Salmon, L., G. Bouvignies, …, M. Blackledge. 2009. Protein conformational flexibility from structure-free analysis of NMR dipolar couplings: quantitative and absolute determination of backbone motion in ubiquitin. *Angew. Chem. Int. Ed. Engl.* 48:4154–4157.
19. Lange, O. F., N. A. Lakomek, …, B. L. de Groot. 2008. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science.* 320:1471–1475.
20. Markwick, P. R., G. Bouvignies, …, M. Blackledge. 2009. Toward a unified representation of protein structural dynamics in solution. *J. Am. Chem. Soc.* 131:16968–16975.
21. Cervantes, C. F., P. R. L. Markwick, …, E. A. Komives. 2009. Functional dynamics of the folded ankyrin repeats of I kappa B alpha revealed by nuclear magnetic resonance. *Biochemistry.* 48:8023–8031.
22. Johnson, D. J., T. E. Adams, …, J. A. Huntington. 2005. Crystal structure of wild-type human thrombin in the Na+-free state. *Biochem. J.* 392:21–28.
23. Talluri, S., and G. Wagner. 1996. An optimized 3D NOESY-HSQC. *J. Magn. Reson. B.* 112:200–205.
24. Kay, L. E., M. Ikura, …, A. D. Bax. 1990. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J. Magn. Reson.* 89:496–514.
25. Yamazaki, T., W. Lee, …, L. E. Kay. 1994. A suite of triple resonance NMR experiments for the backbone assignment of 15N, 13C, 2H labeled proteins with high sensitivity. *J. Am. Chem. Soc.* 116:11655–11666.
26. Pervushin, K., R. Riek, …, K. Wüthrich. 1997. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical

shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. USA.* 94:12366–12371.

27. Hamelberg, D., J. Mongan, and J. A. McCammon. 2004. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* 120:11919–11929.
28. Markwick, P. R. L., and J. A. McCammon. 2011. Studying functional dynamics in bio-molecules using accelerated molecular dynamics. *Phys. Chem. Chem. Phys.* 13:20053–20065.
29. Lipari, G., and A. Szabo. 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.
30. Genheden, S., C. Diehl, …, U. Ryde. 2010. Starting-condition dependence of order parameters derived from molecular dynamics simulations. *J. Chem. Theory Comput.* 6:2176–2190.
31. Ferreiro, D. U., J. A. Hegler, …, P. G. Wolynes. 2007. Localizing frustration in native proteins and protein assemblies. *Proc. Natl. Acad. Sci. USA.* 104:19819–19824.
32. Ding, K., and A. M. Gronenborn. 2003. Sensitivity-enhanced 2D IPAP, TROSY-anti-TROSY, and E.COSY experiments: alternatives for measuring dipolar ^{15}N-^{1}HN couplings. *J. Magn. Reson.* 163:208–214.
33. Baerga-Ortiz, A., A. R. Rezaie, and E. A. Komives. 2000. Electrostatic dependence of the thrombin-thrombomodulin interaction. *J. Mol. Biol.* 296:651–658.
34. Bode, W., D. Turk, and A. Karshikov. 1992. The refined 1.9-A x-ray crystal structure of D-Phe-Pro-Arg chloromethylketone-inhibited human alpha-thrombin: structure analysis, overall structure, electrostatic properties, detailed active-site geometry, and structure-function relationships. *Protein Sci.* 1:426–471.
35. Zweckstetter, M. 2008. NMR: prediction of molecular alignment from structure using the PALES software. *Nat. Protoc.* 3:679–690.
36. Schurr, J. M., H. P. Babcock, and B. S. Fujimoto. 1994. A test of the model-free formulas. Effects of anisotropic rotational diffusion and dimerization. *J. Magn. Reson. B.* 105:211–224.
37. Dosset, P., J. C. Hus, …, D. Marion. 2000. Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR.* 16:23–28.
38. Mandel, A. M., M. Akke, and A. G. Palmer, 3rd. 1995. Backbone dynamics of *Escherichia coli* ribonuclease HI: correlations with structure and function in an active enzyme. *J. Mol. Biol.* 246:144–163.
39. Wang, C., M. Rance, and A. G. Palmer, 3rd. 2003. Mapping chemical exchange in proteins with MW $>$ 50 kD. *J. Am. Chem. Soc.* 125:8968–8969.
40. McConnell, H. M. 1958. Reaction rates by nuclear magnetic resonance. *J. Chem. Phys.* 28:430–431.
41. Jarymowycz, V. A., and M. J. Stone. 2006. Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem. Rev.* 106:1624–1671.
42. Case, D. A. 2002. Molecular dynamics and NMR spin relaxation in proteins. *Acc. Chem. Res.* 35:325–331.
43. Showalter, S. A., E. Johnson, …, R. Brüschweiler. 2007. Toward quantitative interpretation of methyl side-chain dynamics from NMR by molecular dynamics simulations. *J. Am. Chem. Soc.* 129:14146–14147.
44. Huntington, J. A. 2012. Thrombin plasticity. *Biochim. Biophys. Acta.* 1824:246–252.
45. Treuheit, N. A., M. A. Beach, and E. A. Komives. 2011. Thermodynamic compensation upon binding to exosite 1 and the active site of thrombin. *Biochemistry.* 50:4590–4596.