**Supporting Information for**

Dynamic Consequences of Mutation of Tryptophan 215 in Thrombin

Riley Peacock[1], Jessie Davis[1], Phineus R. L. Markwick[2] and Elizabeth A. Komives[1]

[1]Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0378 USA.

[2] San Diego Supercomputer Center University of California, San Diego 10100 Hopkins Dr, La Jolla, CA 92093 USA.

Corresponding author email ekomives@ucsd.edu

**Supplementary Information**

**Detailed Description of Accelerated Molecular Dynamics (AMD)**

Accelerated Molecular Dynamics (AMD) (*1, 2*) is an extended biased potential molecular dynamics approach that allows for the efficient study of bio-molecular systems up to time-scales several orders of magnitude greater than those accessible using standard classical molecular dynamics (CMD) methods, while still maintaining a fully atomistic representation of the system. Compared to many other approaches, AMD affords efficient enhanced conformation space sampling without any *a priori* understanding of the underlying free energy surface, not does it require the specific prior definition of a reaction coordinate or a set of collective variables.

In the standard AMD approach, a reference, or 'boost energy', Eb, is defined, which lies above the minimum of the potential energy surface (PES). At each step in the simulation, if the potential energy V(r), lies below this boost energy, a continuous, non-negative bias potential, $\Delta$V(r), is added to the actual potential. If the potential energy is greater than the boost energy, it remains unaltered. The extent to which the potential energy surface is modified depends on the energy difference between the boost energy and the actual potential. This essentially raises the low energy valleys on the potential energy landscape, decreasing the magnitude of the energy barriers and thereby accelerating the exchange between low energy conformational states, while still maintaining the essential details of the potential energy landscape. Explicitly, the modified potential, V*(r), on which the system evolves during an AMD simulation is given by [1]:

$$V^*(r) = V(r) \quad ; V(r) \geq E_b$$

$$V^*(r) = V(r) + \Delta V(r) \quad ; V(r) < E_b$$

and the bias potential, $\Delta$V(r), is defined as:

$$\Delta V(r) = \frac{(E_b - V(r))^2}{E_b - V(r) + \alpha}$$

The extent of acceleration (*i.e* how aggressively one samples the conformational space) is determined

by the choice of the boost energy, Eb, and the acceleration parameter, α. Conformational space sampling can be enhanced by either increasing the boost energy, or decreasing α. Fig. S1 shows a schematic representation of the AMD method for a 1dimensional potential energy surface, where the boost energy is kept constant and the modified potential is depicted for a variety of acceleration parameters, α. During the course of the AMD simulation, if the potential energy surface is modified, the forces on the atoms are re-calculated for the modified potential.

In the present work, we implemented a "dual boost" AMD approach (*3*), in which two acceleration potentials are simultaneously applied to the system: The first acceleration potential is applied to the torsion terms only, and a second, weaker acceleration is applied across the the entire potential. For both WT thrombin and the W215 mutant, the specific torsional acceleration parameters were defined as [Eb(dih) - <V0(dih)>] = [4 kcal/mol * No residues], and the acceleration parameter, α(dih), was set to one-fifth of this value. The total background acceleration parameters were fixed at [Eb(tot) - <V0(tot)>] = α(tot) = [0.16 kcal/mol * No. atoms in simulation cell]. The average minimum energy potentials, <V0(dih)> and <V0(tot)>, for each system were obtained from 20ns CMD simulations performed as part of the initial equilibration procedure.

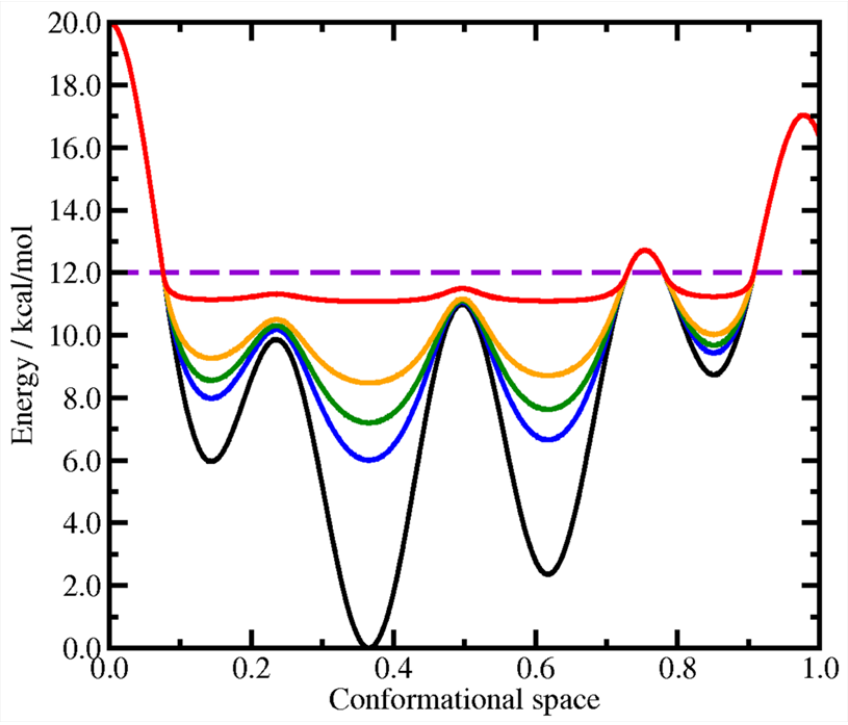



*Fig. 1: Schematic representation of the AMD method. The hypothetical true potential energy function, V(r), is shown by the black line. For a fixed boost energy, Eb=12.0 kcal/mol (dashed violet line), a series of modified potential energy functions, V*(r) are depicted for a variety of acceleration parameters [α = 12.0 kcal/mol (blue), 8 kcal/mol (green), 5 kcal/mol (orange), and 1kcal/mol (red)]. A similar raising and flattening of the potential energy surface can be achieved by holding the acceleration parameter, α fixed, and increasing the boost energy.*

**Obtaining Accurate Free Energy Statistics**

The application of the bias potential destabilizes low energy regions of the conformational space on the potential energy landscape and therefore the population statistics on the modified potential are inherently incorrect. However, as it is known at each step in the simulation exactly how much the system is destabilized, one can, in principle, retrieve the correct population statistics by re-weighting each point in the configuration space on the modified potential by the strength of the Boltzmann factor of the bias energy, *exp*[$\beta\Delta V(r(ti))$], at that particular point. Despite the fact that this Boltzmann canonical re-weighting factor is theoretically, thermodynamically exact, the statistical noise error prohibits the practical acquisition of accurate free energy statistics using this simple re-weighting procedure: The statistical noise error results from the presence of small local molecular distortions directly induced by the application of the bias potential. Although the energy variation for each individual term is very small, when summed over all degrees of freedom (torsional and total in “dual boost” AMD), the resulting fluctuation in the total energies can become comparable to, or larger than the underlying variation in the true free energy surface, and the exponential free energy re-weighting protocol results in the fact that only a very small fraction (<5%) of the molecular conformers sampled across the entire AMD trajectory contribute to the population statistics.

A simple, but robust approach to obtaining accurate free energy statistics is the *bias potential block averaging* method (*2*). In this procedure, the bias potential obtained at each step in the AMD trajectory is averaged over a given number of steps, or 'a block', and the free energy statistics for each member (conformer) in that block are obtained by exponentially re-weighting the block-averaged bias potentials. The correct free energy-weighted relative population, pj, of each block, j, is given as:

$$p_j = p_j^* \frac{< e^{\beta \Delta V} >}{\sum_{j=1}^{M} < e^{\beta \Delta V} >} \quad ; j = 1, M$$

The size of the block (ie. the number of MD-steps, or conformers over which the averaging procedure is performed) is determined by calculating the auto-correlation function of the bias potential, $<\Delta V(0)\cdot\Delta V(t)>$. This correlation function can be fit to a multi-exponential decay: The fast initial time exponent is associated with the temporal oscillation of the statistical noise function, whilst the slower

time exponents refer to the underlying average temporal variations in the bias potential as the system evolves from one conformational state to the next on the modified potential. The most suitable integration period over which the block average is performed lies between the first and second exponents. Following the work of Miao *et al (4),* the exponential averages, <*exp*[βΔV]> in the above block-averaging protocol are approximated using a cumulant expansion to the second order:

$$< e^{\beta \Delta V} > = exp\left\{\sum_{k=1}^{2} \frac{\beta^k}{k!} C_k\right\}$$

Where

$$C_1 = \langle \Delta V \rangle$$

$$C_2 = \langle \Delta V^2 \rangle - \langle \Delta V \rangle^2$$

**Estimating Conformational Space Sampling Time-Scales in AMD Simulations**

In the standard AMD approach, the system evolves on a modified potential at an accelerated rate with a non-linear time-scale of Δt*, given as (*1*):

$$\Delta t_i^* = \Delta t * e^{\beta \Delta V[r(t_i)]}$$

where Δt is the actual time-step of the simulation on the unmodified potential. In principle therefore, it is also possible to estimate the time-scale of events during an AMD trajectory as:

$$t^* = \sum_{1}^{N} \Delta t_i^* = t\langle e^{\beta \Delta V[r(t_i)]} \rangle$$

where N is the total number of molecular dynamics steps performed over the whole simulation, and <*exp*[βΔV(r(ti))]> is the so-called “boost factor”. However, according to Transition State Theory, the exchange rate between two states depends on both the magnitude of the energy barrier on the free energy surface and the transmission coefficient. Following the well-known Kramers' theory, the transmission coefficient is a function of the curvature of the free energy surface on approaching the transition state and the internal friction coefficient. The primary source of error in estimating observed transition rates (and therefore also estimating the effective time-scale of an AMD simulation) arises from

the fact that the application of the bias potential perturbs both these parameters and therefore the transmission coefficient is ill-defined. The question therefore remains: At the acceleration level used in this work, how can one estimate the effective time-scale of the AMD simulations?

During the early stages of the development of AMD, comparative analysis of successful AMD studies applied to a variety of systems (many of which were performed by P. Markwick, one of the authors of this paper) revealed that for torsional acceleration, the optimal value of [Eb(dih) - <V0(dih)>] is equal to 3-5 kcal/mol times the number of solute residues in the system, and the associated acceleration parameter, α(dih), should be set to one fifth of this value (*2*). These acceleration parameters afford efficient conformational space sampling, without generating instabilities in the trajectory and avoiding a random walk, which occurs when the modified potential surface becomes iso-energetic, causing the system to spend a large proportion of time sampling energetically unfavourable regions of the PES. Similarly, for the background, total acceleration, applied in "dual boost" AMD simulations the optimal acceleration parameters are [Eb(tot) - <V0(tot)>] = α(tot) = [0.16 kcal/mol * No. atoms in simulation cell] (*3*).These are the AMD parameters that have been employed in this work, and that are presently used across the entire AMD community. Recently, two independent groups have compared AMD simulations across a variety of proteins to available long, brute-force CMD simulations (*5, 6*). In all cases, it was found that the "optimal" acceleration parameters defined above afforded an effective speed up in conformational space sampling of 3-4 orders of magnitude. Specifically, both independent studies found that AMD simulations performed over the number of MD steps equivalent to a CMD simulation of 100s of nanoseconds, provided configurational space sampling, and torsional entropy values (a thermodynamic measure of the extent of torsional fluctuation) ostensibly identical to that observed in 1-ms brute-force CMD simulations for the same protein systems. This observation of an effective speed up of 3-4 orders of magnitude is in very good agreement with an earlier AMD/NMR study of WT thrombin (*7*). In that study, we showed that AMD simulations of WT thrombin performed over the equivalent of 10s of nanoseconds of CMD, using the "optimal" acceleration parameters described above, afforded an excellent description of experimental NMR-based Residual Dipolar Coupling (RDC) data which reports on an ensemble and time average in the microsecond regime [10s-100s of microseconds], up to the

chemical shift coalescence limit [1-ms]. These results concurred with previous studies on ubiquitin in which it was shown that AMD simulations using the same “optimal” acceleration parameters performed over the equivalent of 10s of nanoseconds of CMD accurately predicted scalar J-couplings and chemical shifts, both of which, like RDCs, report on an ensemble and time-average up to the chemical shift coalescence limit (*2*).

In the present work, for each system (WT thrombin and the W215A mutant thrombin) two AMD simulations were performed for 750,000,000 steps with a (real time) time-step of 2-fs. This is computationally equivalent to performing a 1.5μs CMD simulation. Assuming an effective speed up in the rate of conformational space-sampling by 3 to 4 orders of magnitude due to the application of the bias potential, we anticipate that the configurational space sampling afforded in each of the AMD simulations is associated with dynamics occurring on time-scales of milliseconds to tens of milliseconds, thereby identifying slow motions, including rare local unfolding/refolding events which can be experimentally probed by HDXMS in the fast limit.

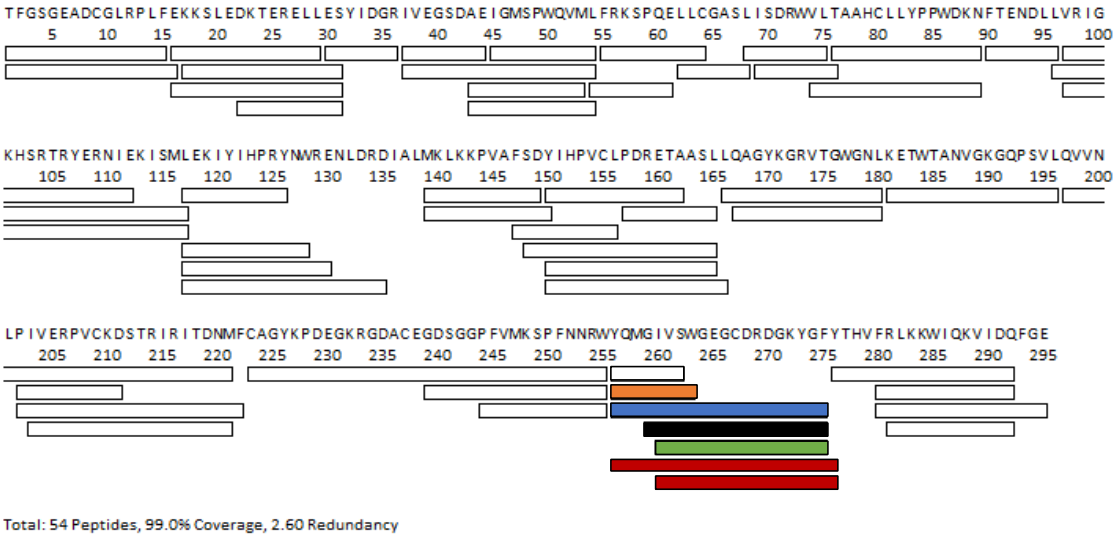
REFERENCES


1. Hamelberg, D., Mongan, J., and McCammon, J. A. (2004) Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules., *J Chem Phys 120*, 11919-11929.
2. Markwick, P. R., and McCammon, J. A. (2011) Studying functional dynamics in bio-molecules using accelerated molecular dynamics., *Phys Chem Chem Phys 13*, 20053-20065.
3. Hamelberg, D., de Oliveira, C. A., and McCammon, J. A. (2007) Sampling of slow diffusive conformational transitions with accelerated molecular dynamics., *J Chem Phys 127*, 155102.
4. Miao, Y., Sinko, W., Pierce, L., Bucher, D., Walker, R. C., and McCammon, J. A. (2014) Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation., *J Chem Theory Comput 10*, 2677-2689.
5. Kamenik, A. S., Kahler, U., Fuchs, J. E., and Liedl, K. R. (2016) Localization of millisecond dynamics: Dihedral entropy from accelerated MD., *J Chem Theory Comput 12*, 3449-3455.
6. Pierce, L. C., Salomon-Ferrer, R., Augusto, F., de Oliveira, C., McCammon, J. A., and Walker, R. C. (2012) Routine Access to Millisecond Time Scale Events with Accelerated Molecular

Dynamics., *J Chem Theory Comput 8*, 2997-3002.

7. Fuglestad, B., Gasper, P. M., Tonelli, M., McCammon, J. A., Markwick, P. R., and Komives, E. A. (2012) The dynamic structure of thrombin in solution, *Biophys J 103*, 79-88.

## Figure S1




**Supplemental Figure 1:** HDX-MS coverage map of all peptides analyzed for WT thrombin, W215A, W215I, F227A, and F227V. All white peptides were identified for all enzymes tested. Black peptides were identified in WT and W215A datasets only. Orange peptides were Identified in WT, F227A, and F227V datasets only. Blue peptides were identified in WT, F227A, and W215A datasets only. Green peptides were identified in WT, W215I, and W215A datasets only. Red peptides were identified in the F227V dataset only.

**Figure S2 pg. 1**

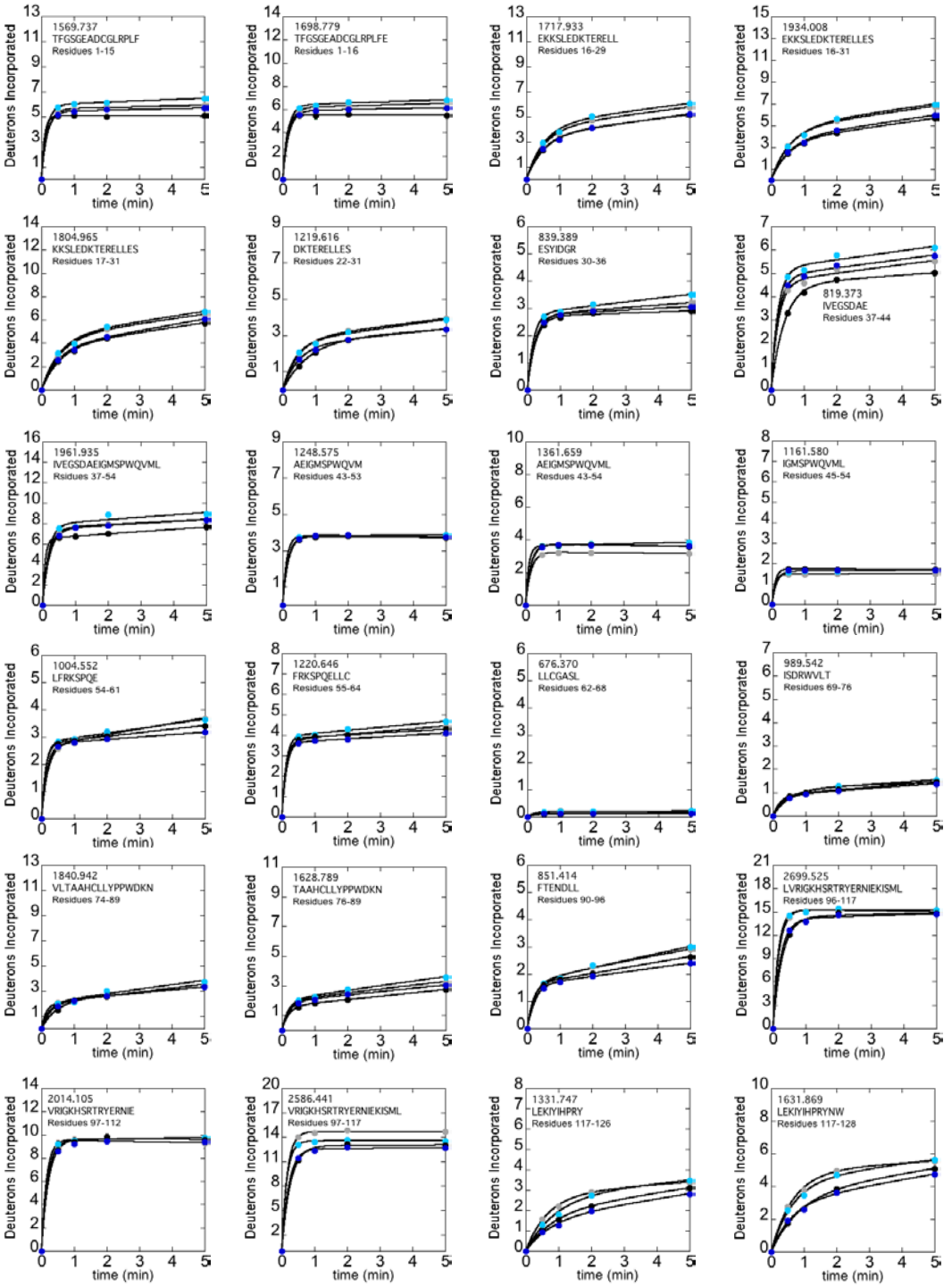
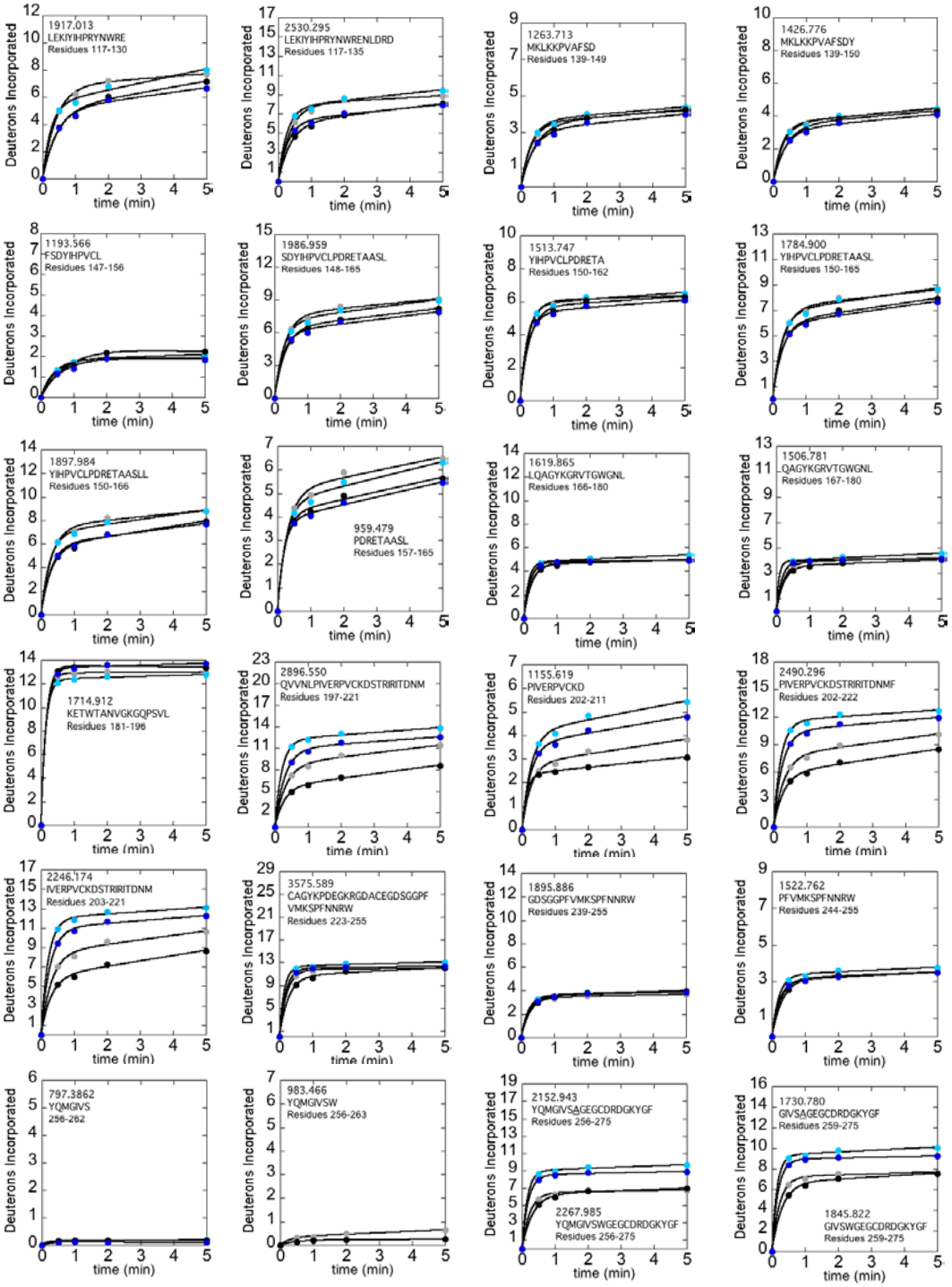
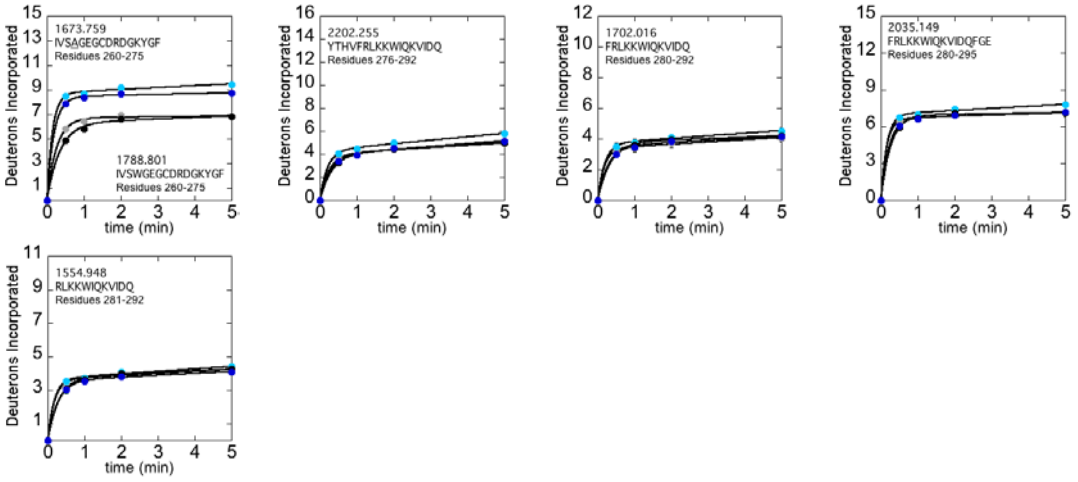


**Figure S2 pg. 2**



## Figure S2 pg. 3




**Supplemental Figure 2:** HDX-MS uptake plots showing deuterium uptake overtime for WT thrombin (grey) and W215A (cyan) at 100 mM NaCl, and for WT thrombin (black) and W215A (blue) at 300 mM NaCl. Mutations, if present, are underlined in the peptide sequence within the uptake plot. All Experiments were done in triplicate, and error are bars shown.

**Figure S3 pg. 1**

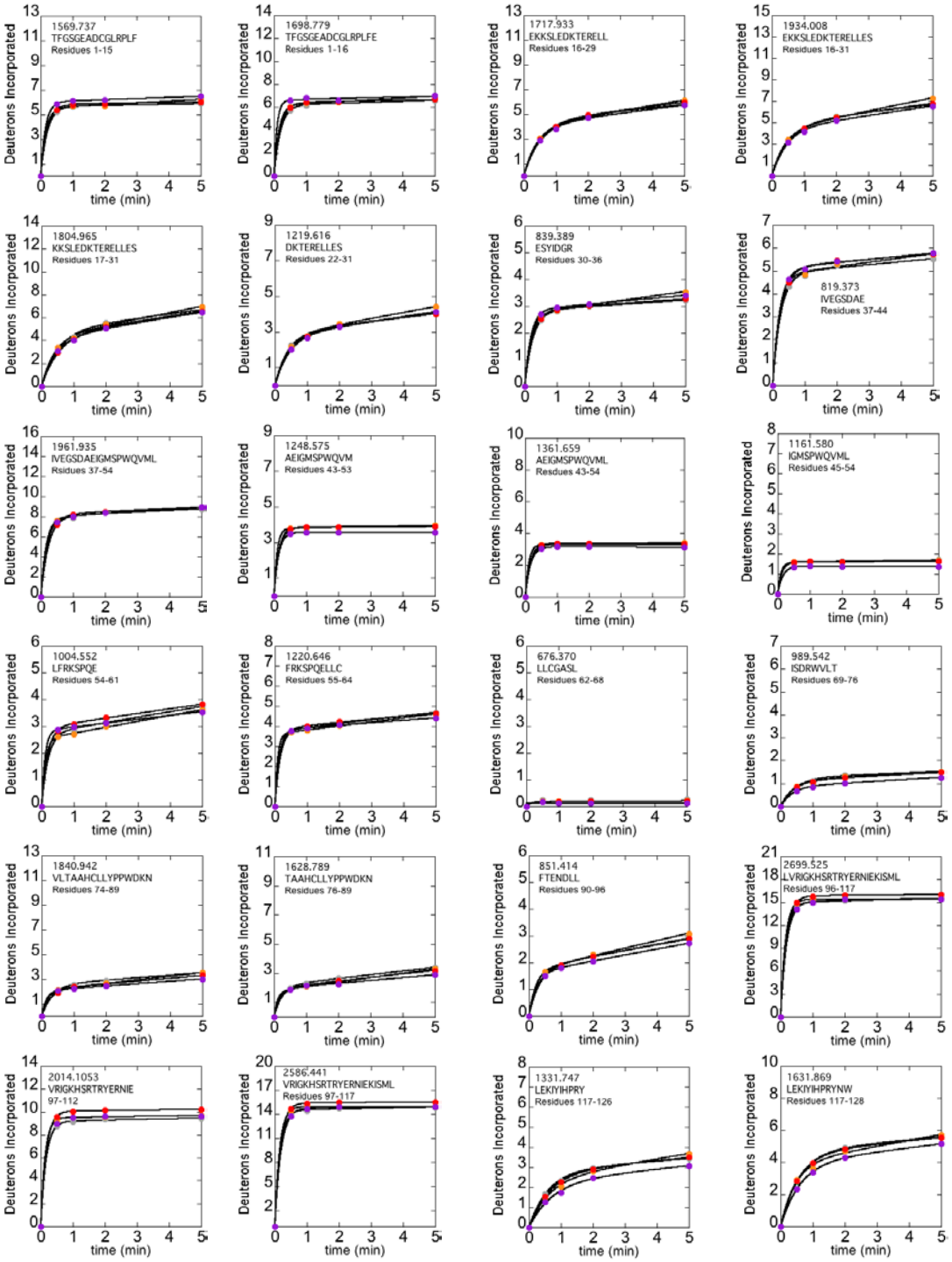
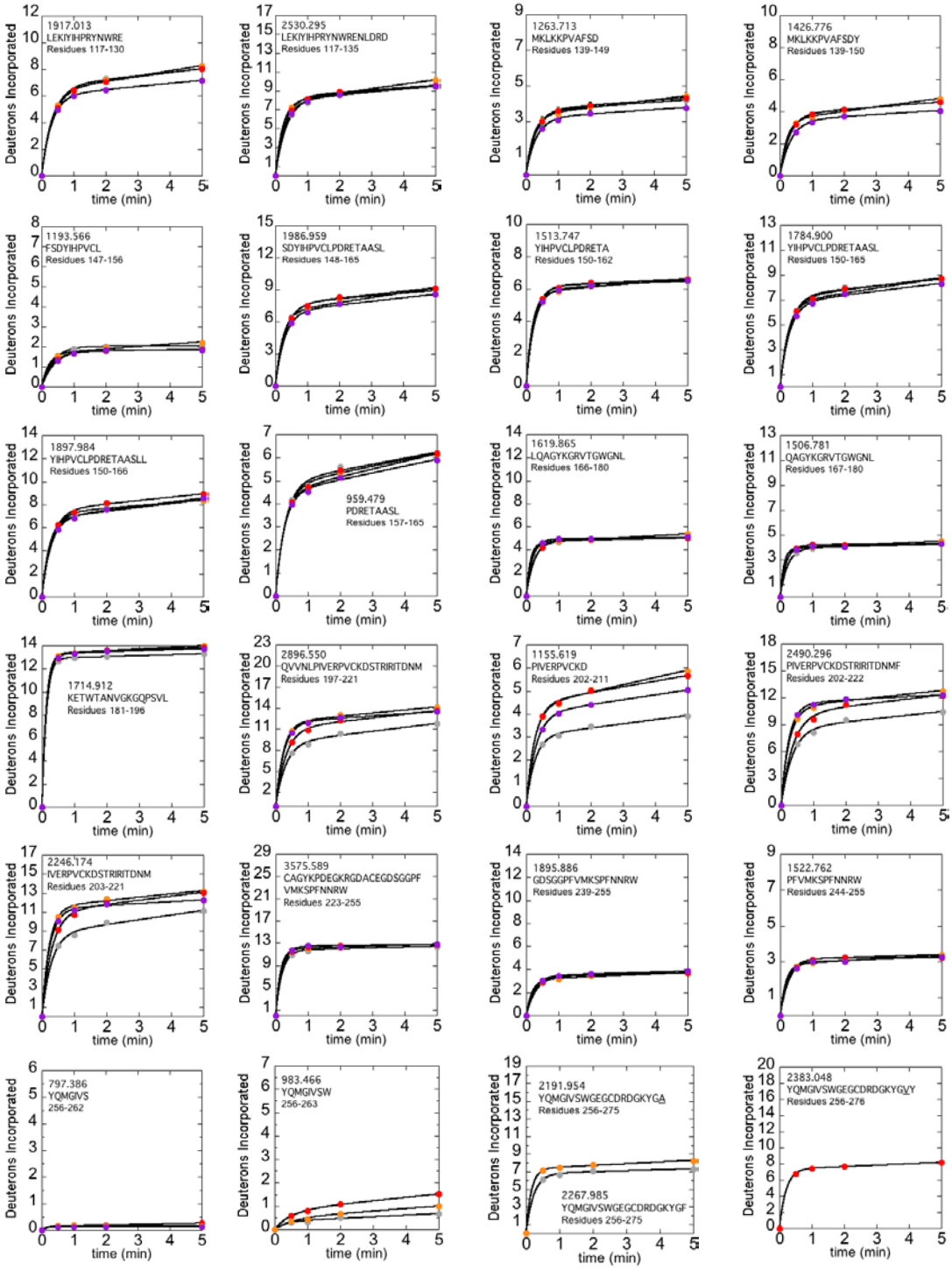
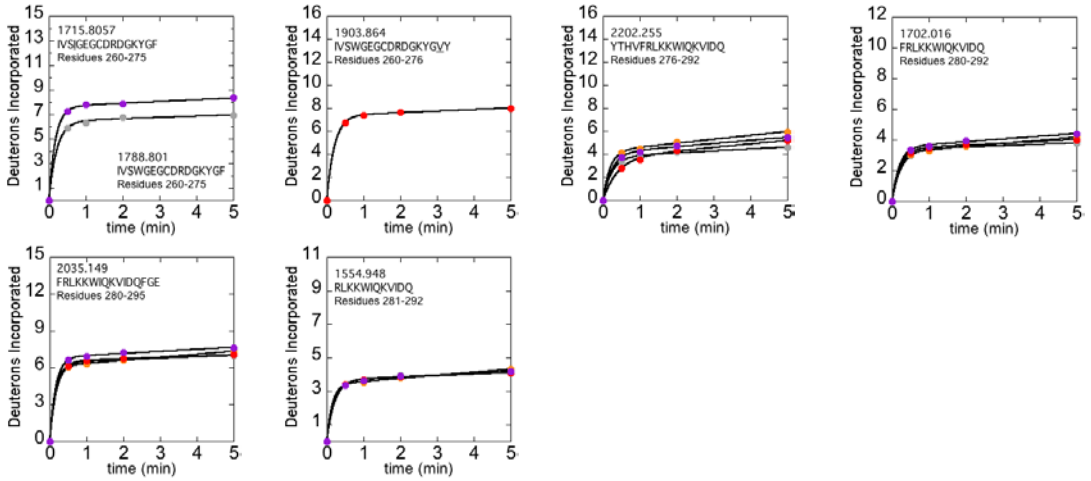


**Figure S3 pg. 2**



## Figure S3 pg. 1




**Supplemental Figure 3:** HDX-MS uptake plots showing deuterium uptake overtime for WT thrombin (grey), F227A (orange), F227V (red), and W215I (purple) at 100 mM NaCl. Mutations, if present, are underlined in the peptide sequence within the uptake plot. All Experiments were done in triplicate, and error are bars shown.