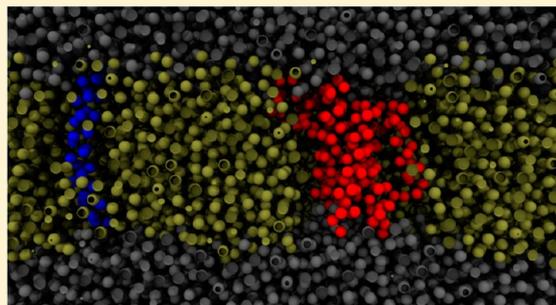


## Multifaceted Substrate Capture Scheme of a Rhomboid Protease

Tyler Reddy<sup>†,§</sup> and Jan K. Rainey<sup>\*,†,‡</sup><sup>†</sup>Department of Biochemistry & Molecular Biology and <sup>‡</sup>Department of Chemistry, Dalhousie University, Halifax, Nova Scotia B3H 4R2, Canada

## Supporting Information

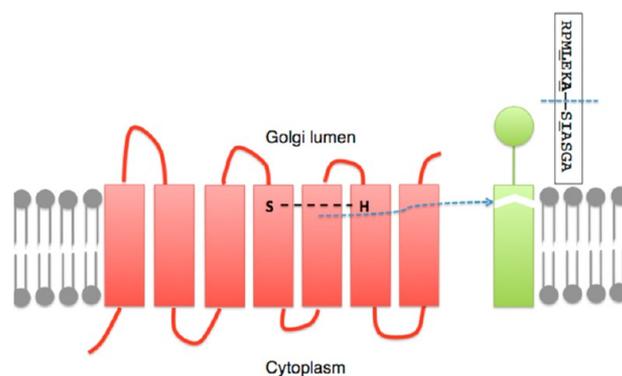
**ABSTRACT:** Rhomboid proteases are integral membrane serine proteases that catalyze peptide bond hydrolysis in biological membranes. Little is currently known about the interaction of enzyme and substrate. Coarse-grained molecular dynamics simulations in hydrated lipid bilayers are employed herein to study the interaction of the *E. coli* rhomboid protease GlpG (ecGlpG) with the transmembrane domain (TMD) of the substrate Spitz. Spitz does not associate with ecGlpG exclusively at the putative substrate gate near TMD 5. Instead, there are six prominent and stable interaction sites, including one between TMDs 1 and 3, with the closest enzyme–substrate proximity occurring at the ends of helical TMDs or in loops. Bilayer thinning is observed proximal to ecGlpG, but there is no evidence of additional thinning of the bilayer upon interaction with substrate. We suggest that the initial interaction between enzyme and substrate, or substrate capture event, is not limited to a single site on the enzyme, and may be driven by juxtamembrane electrostatic interactions. The findings are of additional interest because catalytically inactive rhomboids (iRhoms) are now known to interact with the substrates of their catalytically active counterparts and to antagonize the enzyme-driven pathways.



## INTRODUCTION

The structure and function of water-soluble serine proteases has been studied for decades.<sup>1</sup> Classical examples, such as trypsin and chymotrypsin, are synthesized as inactive precursors called zymogens, which are then cleaved into proteolytically active forms. The irreversible peptide bond hydrolysis catalyzed by serine proteases is therefore carefully regulated, as might be expected. More recently, a new class of serine proteases has been discovered in the form of polytopic membrane proteins conserved in all kingdoms of life.<sup>2</sup> These were named rhomboid proteases, as rhomboid protease mutant *Drosophila* embryos exhibited rhomboid-shaped head skeletons.<sup>3</sup> Generally, rhomboid proteases cleave single-pass transmembrane proteins at or near the top (i.e., juxtamembrane portion) of the transmembrane domain (TMD; Figure 1), although there are known exceptions. Rhomboid proteases, unlike their soluble counterparts and some other integral membrane proteases, are constitutively active after synthesis. For those cases where rhomboid substrates have been identified, the primary mechanism of regulation appears to be physical segregation (e.g., compartmentalization to different organelles) of the enzyme and its substrate.

In the past decade, a great breadth and depth of both biological and medical relevance for rhomboid proteases has become clear. *Drosophila* Rhomboid-1 cleaves the class 1 membrane protein substrate Spitz within its TMD, releasing the N-terminus as an epidermal growth factor receptor (EGFR) ligand.<sup>4</sup> This process takes place at the Golgi apparatus, leading to eventual release from the cell via the secretory pathway during embryonic development.<sup>4</sup> Since the discovery of the



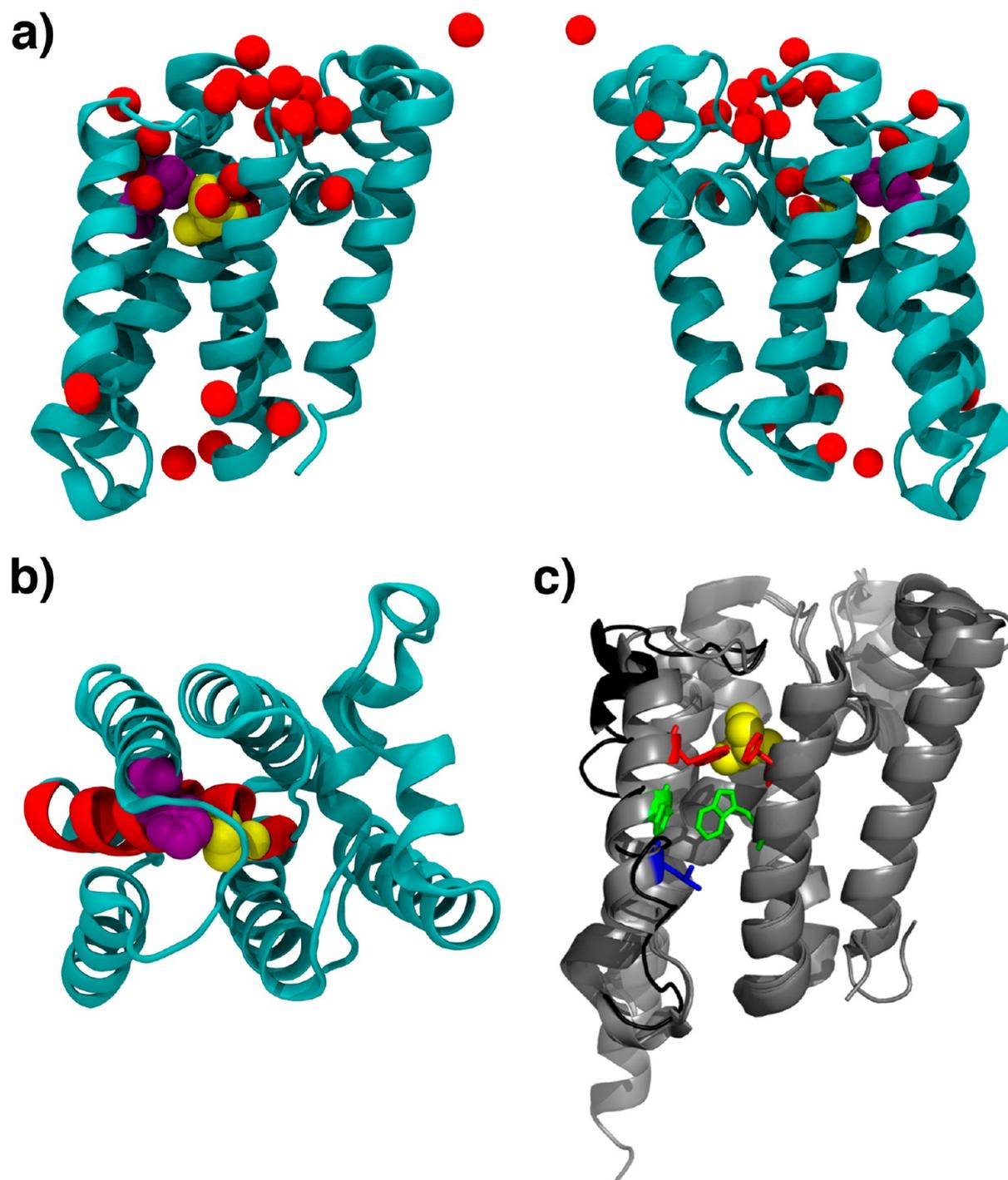
**Figure 1.** A schematic representation of *Drosophila* Rhomboid-1 (red) and its substrate Spitz (green) in the phospholipid bilayer (gray) of the Golgi apparatus. The membrane-buried Ser and His residues of the catalytic dyad are shown along with the intramembrane cleavage of the substrate. The primary sequence in the inset shows the cleavage point in the substrate (blue dotted line) flanked by characteristic hydrophobic residues at the P4 and P2' positions, and a small residue at the P1 position,<sup>24</sup> all highlighted. The cleavage occurs within the ASIASGA substrate sequence.

*Drosophila* rhomboid, a number of other functions have been ascribed to this class of protease. It should be noted that the *E. coli* rhomboid protease, ecGlpG, has been employed in the

Received: May 24, 2012

Revised: July 5, 2012

Published: July 6, 2012

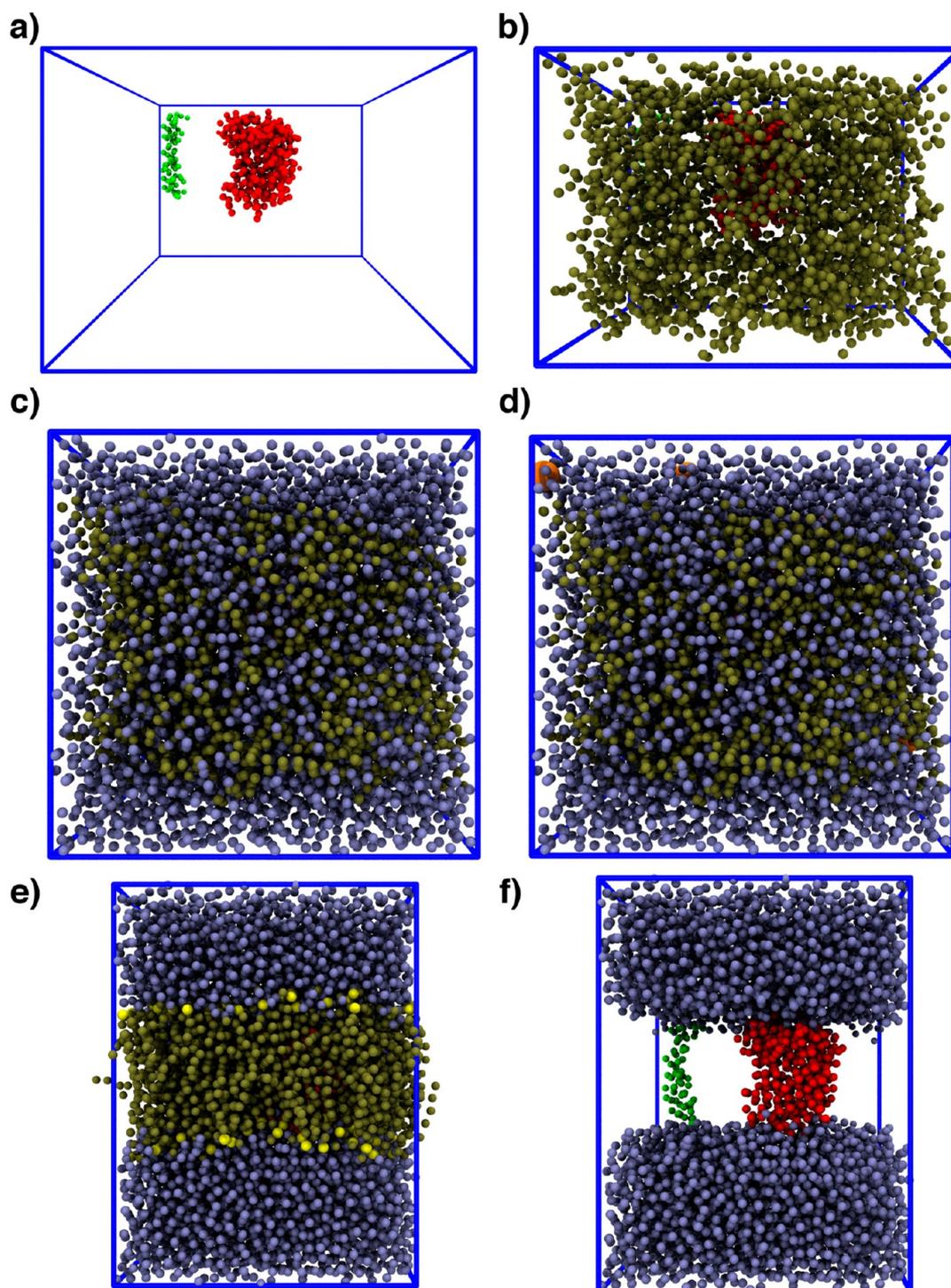


**Figure 2.** High-resolution X-ray crystal structures of rhomboid proteases providing mechanistic insight for substrate entry and catalysis. The residues of the catalytic dyad, Ser (yellow) and His (purple), are highlighted. (a) Side-on mirror views of the *E. coli* GlpG (ecGlpG) structure (PDB ID: 2IC8)<sup>14</sup> used as the basis of the simulations herein demonstrating accessibility of the active site to water molecules (red). (b) Top-down view (i.e., from periplasm) of the same structure with ecGlpG TMD4, which contains the catalytic Ser, highlighted in red. This catalytically essential TMD is protected laterally from the bilayer by a ring of TMDs. (c) Side-on superposed view of ecGlpG structures (PDB IDs: 2IC8, 2O7L;<sup>14,18</sup> gray) along with a GlpG structure from *Haemophilus influenzae* (PDB ID: 2NR9;<sup>17</sup> gray, with TMD5 shown in black). Enzymatic activity is increased 4-fold by mutation of the three TMD5 residues to Val, and either 7-fold (green) or 10-fold (red) for mutation of highlighted bulky hydrophobic residue pairs on TMDs 2 and 5 to Ala.<sup>23</sup>

majority of rhomboid studies to date, including the structural study used as the starting point for the simulations described herein, but its substrate and native role have not been identified.

In mice, knockout of the mitochondrial rhomboid PARL (presenilin-associated rhomboid-like) leads to susceptibility to

innate cellular apoptosis, and eventually death from systemic atrophy.<sup>5</sup> The soluble form of the PARL substrate (OPA1) appears to be required to maintain the form of the mitochondrial cristae, thus attenuating the rate of cytochrome *c* release and apoptosis. The PARL rhomboid homologue in yeast (Pcp1) is also involved in maintenance of mitochondrial



**Figure 3.** Representative coarse-grained simulation system (10 768 CG particles total; Table 1) setup procedure. (a) The substrate (green; 72 CG particles) and enzyme (red; 401 CG particles) are initially placed in a simulation container with dimensions  $12 \times 12 \times 9 \text{ nm}^3$ . (b) 268 POPE molecules (brown; 3484 CG particles) are added to the  $z$  constrained container to encourage bilayer formation in  $xy$  plane. (c) The box is expanded vertically to a cube with 12 nm sides and the system solvated with 6803 CG waters (blue). (d) Eight water particles are replaced with chloride ion particles (orange; relative radii exaggerated for clarity, i.e., particle visible in top left). (e) The POPE bilayer (phosphates colored yellow) self-assembles during a 200 ns equilibration with the proteins restrained. (f) Equilibrated system in (e) excluding the POPE bilayer for clarity.

morphology.<sup>6</sup> There is a correlation between reduced levels of PARL and diabetes in humans,<sup>7</sup> though causality remains controversial. Furthermore, mutation of the OPA1 rhomboid substrate in humans has been associated with dominant optic atrophy causing early onset blindness,<sup>8</sup> with apparent optic nerve fiber loss from mitochondrial dysfunction.

Rhomboids are also involved in the final host cell invasion step of the Apicomplexan parasite *Toxoplasma gondii*.<sup>9</sup> Estimates indicate that a third of all humans have toxoplasmosis.<sup>10</sup> Although it is not normally dangerous, it can cause serious complications for pregnant women and immunocompromised individuals. Similarly, the Apicomplexan parasite that causes

malaria in humans, *Plasmodium falciparum*, also encodes a number of rhomboid proteases which may likewise cleave TMD substrates as part of the invasion process.<sup>11</sup> Beyond use by parasites in host invasion, rhomboid activity has been described in prokaryotes, with the opportunistic human pathogen *P. stuartii* requiring its rhomboid homologue (AarA) in order to release a quorum sensing signal involved in antibiotic resistance.<sup>12</sup> The AarA substrate, TatA, oligomerizes to form part of the twin-arginine translocase complex, which may directly or indirectly allow the release of the quorum sensing signal.

Beyond these diverse functions and the identification of rhomboids across phylogeny, a strong incentive for the study of rhomboids is the fact that they provide models of the presenilin enzyme implicated in Alzheimer's disease that are more tractable to study. Presenilin, although it is also a polytopic intramembrane protease, must undergo endoproteolytic cleavage, requires three other proteins to form the active  $\gamma$ -secretase complex, requires prior proteolytic cleavage of the substrate by  $\beta$ -secretase, and cleaves the substrate in two different TMD regions.<sup>13</sup> In comparison, rhomboids require no other proteins, cofactors, or prior proteolytic cleavage of either enzyme or substrate in order to catalyze hydrolysis of their substrates.

There is, therefore, substantial motivation to study rhomboid proteases. The first major question about these enzymes was how they could catalyze a proteolytic reaction within the hydrophobic confines of the phospholipid bilayer. The determination of a number of X-ray crystal structures of *E. coli* and *Haemophilus influenzae* GlpG rhomboid proteases revealed that water could access the active site cleft from the periplasm (Figure 2a).<sup>14–22</sup> The second major question, the answer to which remains controversial, is how a bitopic and putatively predominantly  $\alpha$ -helical TMD substrate partitions from the hydrophobic bilayer into the protected active site in the core of the rhomboid 6-helix bundle (Figure 2b). This is highlighted by the fact that the catalytic Ser is protected at the center of the enzyme by the surrounding ring of TMDs. Superposition of several recent crystal structures suggests that TMD5 is flexible and may act as a substrate gate to the active site (e.g., Figure 2c). Indeed, mutation of large hydrophobic residues in TMDs 2/5 to smaller residues can increase rhomboid activity 10-fold relative to WT.<sup>23</sup>

The structural detail now available for rhomboid proteases is contrasted by the relative paucity of information available about the substrates. In fact, the endogenous substrates for many rhomboids have not even been identified. Identification of rhomboid substrates from species other than *Drosophila* led to a proposed cleavage consensus sequence in the substrate (ASIASGA; Figure 1), but the requirements for cleavage are now known to be less specific.<sup>24</sup> There is currently no high-resolution structure of a rhomboid substrate available, and the only structural information about the rhomboid–substrate interaction comes from a crystal structure of a mechanism-based isocoumarin inhibitor bound in the ecGlpG active site,<sup>20</sup> and a more recent structure of ecGlpG complexed with a phosphonofluoridate inhibitor.<sup>25</sup> The latter structure is consistent with ligand entry between TMDs 2 and 5, but the inhibitors in both studies are much smaller than a peptide substrate. There is, therefore, a substantial need to perform studies that include both enzyme and substrate.

We have probed the interaction between the *E. coli* rhomboid (ecGlpG) and the region of Spitz (a known substrate of

ecGlpG)<sup>26</sup> comprising the predicted TMD residues on biologically relevant time and size scales in a bilayer setting, which has not been possible with previously reported atomistic simulations.<sup>27,28</sup> The coarse-grained molecular dynamics (CG-MD) methodology we employ has previously been used to study the dimerization of glycophorin A (GpA) in a bilayer setting, with helix crossing angles and helix–helix interface residues that matched experimental results from nuclear magnetic resonance (NMR) spectroscopy and mutagenesis.<sup>29</sup> The authors also accurately estimate the  $\Delta\Delta G$  of dimerization for GpA TMD mutant peptides relative to atomistic MD and experimental approaches. Likewise, the same force field was more recently used to reproduce key aspects of the more complex talin heterodimer NMR structure.<sup>30</sup> The authors observe a right-handed helix packing with a crossing angle of  $-30^\circ$ , in agreement with the packing of the  $\alpha$ Ib $\beta$ 3 TM dimer seen in NMR structures determined in lipid bicelles and in nonaqueous ( $\text{CH}_3\text{CN}/\text{H}_2\text{O}$ ) solution (both  $-27^\circ$ ). A similar fidelity with experimental results has been reported using the related MARTINI CG-MD force field.<sup>31,32</sup>

It must be emphasized that the study of protein–protein interactions by classical MD simulation techniques, whether atomistic or coarse-grained, is not sufficient to fully grasp the mechanistic details of enzyme function. Namely, quantum mechanical bond breaking and formation is required to elucidate a catalytic process. However, as catalytically inactive rhomboids (iRhoms) have recently been shown to bind to the same substrates and regulate the same pathways as their catalytically active counterparts,<sup>33</sup> there is an additional justification for performing classical MD to study protein–protein interactions of rhomboid proteases in the absence of catalysis.

Thus, there are precedents for the use of CG-MD in probing experimentally verified protein–protein interactions in the membrane and for the biological relevance of studying said interactions in the absence of catalytic activity. Therefore, we employ the CG-MD methodology to perform 20 5  $\mu\text{s}$  replicate trajectories of the encounter between ecGlpG and an  $\alpha$ -helical Spitz-TMD in a hydrated 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine (POPE) bilayer (Figure 3).

## ■ COMPUTATIONAL METHODS

**Software and Force Field.** GROMACS<sup>34–36</sup> 4.0.7 was used for simulation setup, energy minimization, equilibration simulations, and 5  $\mu\text{s}$  replicate simulations suitable for analysis. For all simulations, we employed a previously described coarse-grained molecular dynamics force field<sup>37,38</sup> (similar to MARTINI<sup>39,40</sup>) with an approximate 4:1 mapping of heavy atoms to representative particles. This force field previously exhibited retention of critical details of interprotein interactions observed in detailed atomistic simulations and experimental studies.<sup>29</sup> We also employ a Gaussian network model with a spring constant (1000 kJ/mol/nm<sup>2</sup>) that has been shown to reliably reproduce protein flexibilities from atomistic simulation and experiment.<sup>41–43</sup> The phase partitioning properties of amino acid side chains in the force field we employ have been thoroughly validated.<sup>44</sup> Molecular representations were visualized in VMD<sup>45</sup> or PyMOL,<sup>46</sup> and trajectory analysis was performed using Python code dependent on the MDAnalysis toolkit (see section Trajectory Analysis Algorithms).<sup>47</sup>

**Simulation Setup Procedure.** The stepwise simulation setup procedure is demonstrated in Figure 3 and is similar to previously described CG-MD bilayer simulation protocols.<sup>29,30</sup>

Table 1. Summary of CG-MD Simulations Performed<sup>a</sup>

simulation system	equilibrated simulation box dimensions (nm <sup>3</sup> )	no. of ecGlpG binding partner CG particles	no. of POPE molecules	no. of POPE CG particles	no. of water CG particles
Spitz TMD starting near ecGlpG TMDs 1/3	9.8 × 9.8 × 13.1	72	268	3484	6803
Spitz TMD starting near ecGlpG TMD 5	9.8 × 9.8 × 13.2	72	270	3510	6792
FGFR3 TMD negative control starting near ecGlpG TMD 5	9.9 × 9.9 × 13.0	67	277	3601	6770
ecGlpG with no binding partner	9.7 × 9.7 × 13.3	0	271	3523	6819

<sup>a</sup>The *E. coli* GlpG (ecGlpG) enzyme consists of 401 coarse-grained (CG) particles in each simulation system and 10 replicate 5  $\mu$ s simulations were performed for each system.

System compositions are summarized in Table 1. The atomistic starting structure for the rhomboid protease was from a 2.1 Å ecGlpG structure (PDB ID 2IC8),<sup>14</sup> while the 34-residue Spitz TMD construct (sequence KRPRPMLKASIASGAM-CALVFMFLVCLAFYLRLK) was built as an ideal  $\alpha$ -helix in PyMOL ( $\phi = -54.2^\circ \pm 10.9$ ;  $\psi = -47.0^\circ \pm 0.03$ ). The negative control peptide was a 33-residue ideal  $\alpha$ -helical construct of the fibroblast growth factor receptor 3 (FGFR3) TMD (sequence RRAGSVYAGILSYGVGFFLFILVVAAVTLCLRLR), chosen for its similar size to the Spitz construct and its tractable behavior in CG-MD simulations (Reddy and Sansom, manuscript in preparation).

Coarse-grained representations of the enzyme (401 CG particles) and substrate (72 CG particles) or negative control peptide (67 CG particles) were initially placed in a Z-restrained (12 × 12 × 9 nm<sup>3</sup>) rectangular container (Figure 3a). Either of two starting configurations were employed to avoid bias of substrate association with enzyme (Figure 4a), and a 70 Å center-to-center enzyme–substrate starting separation places the substrate well outside the long-range electrostatic and Lennard-Jones potential interaction radius (12 Å) of the enzyme in this force field (Figure 4b,c). The 70 Å interprotein starting separation is also greater than previous values used to study the association of GpA (55 Å)<sup>29</sup> and the talin heterodimer (60 Å),<sup>30</sup> both using the same force field in a membranous environment. The system was then solvated with 268 POPE molecules (3484 CG particles; Figure 3b). The system container was expanded vertically to 12 × 12 × 12 nm<sup>3</sup> and solvated with water (6803 CG particles) to encourage bilayer formation in the *xy* plane (Figure 3c). Water particles were replaced with chloride ion CG particles to produce a neutral system (Figure 3d) and a POPE phospholipid bilayer self-assembled around the restrained proteins during a 200 ns equilibration (Figure 3e,f). A total of 40 5  $\mu$ s simulations were carried out: 10 replicates for each of the two substrate starting configurations; 10 replicates for the negative control configuration (where FGFR3 was always placed in corner nearest ecGlpG TMD5, as in Figure 4a right panel); and 10 replicates for ecGlpG in the absence of a substrate or negative control peptide (Table 1). The simulations were conducted with 40 fs time steps at 323 K using unique starting velocity seeds for each of the replicate simulations, and frames were written every 400 ps. Protein, lipid, and solvent were separately temperature coupled to an external bath with a temperature of 323 K using the Berendsen thermostat<sup>48</sup> and a 1 ps time constant. Semi-isotropic exponential relaxation pressure coupling was performed with a time constant of 1 ps.

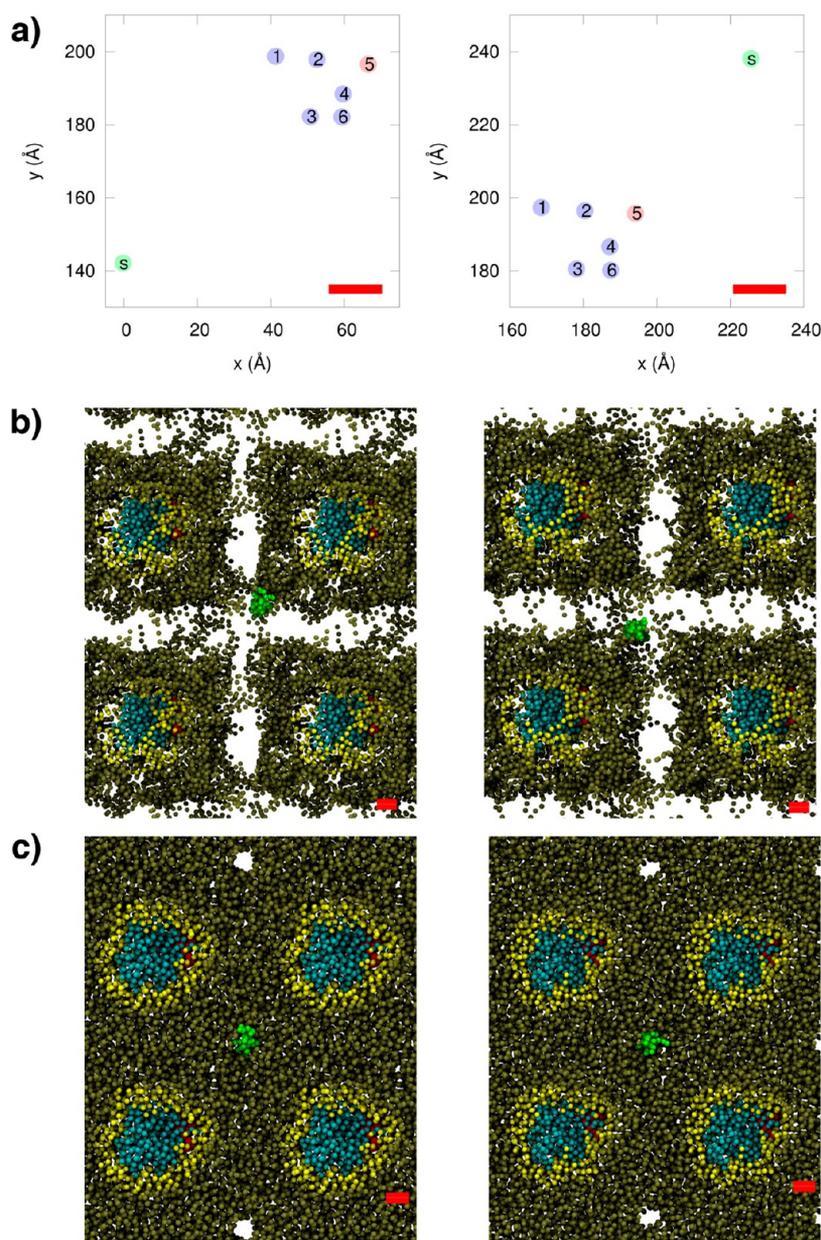
**Trajectory Analysis Algorithms.** The following algorithm descriptions were implemented in Python code for the analysis of MD trajectories exposed by the MDAnalysis toolkit.<sup>47</sup>

**Parsing Bilayer Thickness.** The thickness of the POPE bilayer was assessed proximal and distal to ecGlpG and Spitz over the course of each trajectory. POPE phosphate particles were assigned to one of the two bilayer leaflets only if they were within 12.5 Å of another POPE phosphate particle in that leaflet. Assignment to a leaflet was made in the first frame to avoid the complication of bilayer deformation during the trajectory, which could otherwise confound leaflet assignment with this algorithm. In each frame of a given trajectory, POPE phosphates were classified as protein-local (within 16 Å of Spitz or ecGlpG) or -distal (>16 Å from Spitz or ecGlpG). The protein-local and -distal bilayer thicknesses were then calculated as the *z* coordinate separation between the centers of geometry for the full set of POPE phosphates in the two leaflets that fall within or outside of these 16 Å shells. The number of lipids in the protein-local and -distal shells was also tracked to ensure more than a single POPE molecule was used in the measurements (not shown). As phosphate particles were rigidly assigned to one of two leaflets in the first frame of a given replicate, POPE flip-flop between leaflets would confound bilayer thickness calculations, so each trajectory was also parsed to ensure that POPE particles did not flip in the bilayer (see Flip-Flop Analysis below).

The per-frame measured bilayer thicknesses were then classified according to the association of enzyme (E) and substrate (S). ES complex formation was defined as an interprotein *C<sub>α</sub>* separation  $\leq 6$  Å (see Closest Approach Analysis below). The full set of categorized bilayer thickness values were averaged with uncertainty measured as one standard deviation from the mean.

**ecGlpG Root-Mean-Square Fluctuation.** ecGlpG root-mean-square fluctuation (RMSF) values were calculated in two different ways for comparison. For each trajectory, the values were either based on a reference configuration that was the average configuration of ecGlpG taken over all frames of the trajectory (canonical), or the first frame of the trajectory (noncanonical). In each frame of a trajectory, ecGlpG *C<sub>α</sub>* coordinates were translated and rotated such that an root-mean-square deviation (rmsd)-minimized superposition with the reference configuration was achieved. Per-residue *C<sub>α</sub>* RMSF values were calculated over 12 501 frames in each trajectory and then averaged across all trajectories. When using an average-based reference, there was an initial pass through the trajectory to calculate the arithmetic mean coordinates for each ecGlpG *C<sub>α</sub>* particle. RMSF values were also estimated from X-ray crystal structure *B*-factors using the following relationship:<sup>49</sup>

$$B = \left( \frac{8\pi^2}{3} \right) (\text{RMSF})^2 \quad (1)$$



**Figure 4.** Starting substrate configurations are not biased to association at a particular site on the enzyme. The red scale bar in each panel represents the 12 Å long-range cutoff for both electrostatic and Lennard-Jones interactions. The substrate was placed either closer to ecGlpG TMDs 1/3 (left panels) or to TMD5 (right panels). (a) The geometric center of the substrate (s, green) was placed 70 Å from the geometric center of ecGlpG for each starting configuration. The geometric centers of the ecGlpG TMDs are shown as purple circles, except for the putative substrate gate (TMD5), which is colored red. (b) A top-down (periplasmic) view of the two starting configurations prior to POPE bilayer self-assembly (e.g., stage shown in Figure 3(d)). Multiple copies of the enzyme (blue; TMD5 in red) result from demonstration of the effect of the periodic boundary cells around the substrate. POPE molecules that fall within the 12 Å long-range interaction cutoff of ecGlpG are shown in yellow, while the remainder of the bilayer is shown in brown. (c) The same systems following 200 ns of protein-restrained equilibration (e.g., following stage shown in Figure 3e). These represent starting configurations for 5  $\mu$ s replicate simulations. Solvent has been excluded for clarity in all panels.

**Flip-Flop Analysis.** In the first frame of each trajectory, POPE phosphate particles were again assigned to a leaflet based on the formation of a network of connections with other POPE phosphates using a 12.5 Å maximum distance cutoff. The absolute assignments to leaflets were tested by tracking minimum and maximum phosphate  $z$  coordinates for all POPE molecules in each leaflet. The centroid of all phosphates in the system was tracked as the midpoint of the bilayer, through which transitions would be consistent with POPE flip-flop activity.

The N- and C-terminal Spitz TMD  $C_{\alpha}$  particle  $z$  coordinates were similarly tracked relative to the overall POPE phosphate centroid as well as the minimum and maximum phosphate  $z$  coordinates in the system. Transitions of the Spitz termini through the phosphate centroid would be consistent with substrate flip-flop activity.

**Closest Approach Analysis.** The initial and sustained closest approach distances between Spitz and ecGlpG were tracked for each replicate simulation. In each frame, ecGlpG  $C_{\alpha}$  particles were categorized into TM segments based on the crystal structure topology assignments.<sup>14</sup> A matrix of distances was

generated between all possible pairs of  $C_\alpha$  particles in each ecGlpG TM segment and each  $C_\alpha$  particle in Spitz. The minimum value from each matrix corresponds to the closest approach between Spitz and a particular ecGlpG TM segment.

**Enzyme–Peptide Positional Contour Maps and Polar Angle Categorization.** Contour maps were produced to assess the preferred enzyme–peptide (Spitz or FGFR3 TMD) interaction sites. The ecGlpG starting configuration for 5  $\mu$ s simulations, where Spitz starts nearest TMDs 1/3, was used as a reference configuration. The reference coordinates were adjusted such that the ecGlpG centroid was at the origin of the system, and only  $C_\alpha$  particles were employed. The instantaneous ecGlpG  $C_\alpha$  coordinates in each frame of a trajectory were similarly adjusted to place the ecGlpG centroid at the origin. For consistency, the same translation and rotation was applied to Spitz. The rotation matrix necessary to superpose the (centered) instantaneous ecGlpG to the reference with minimal rmsd was determined and used to transform the Spitz coordinates. Overall, Spitz is thus adjusted to a reference frame where ecGlpG remains relatively fixed (within the bounds of small rmsd fluctuations). The Spitz centroid coordinates were tracked in this reference frame for each frame of a given trajectory. The merged data across all trajectories was divided into bins ( $x$  and  $y$  coordinates with 0.4  $\text{\AA}$  widths) and the probability of finding the substrate centroid in each coordinate bin was plotted. Tracking the Spitz centroid coordinates in this reference frame for a given trajectory as a function of simulation progression was also used for analysis.

The overall contour map was categorized by polar angle boundaries based on inspection of high-probability interaction regions. As the ecGlpG reference configuration was centered at the origin, the polar angles were conveniently defined as arcs from the positive  $x$  axis. The position of Spitz in the above ecGlpG reference frame was categorized according to polar angle for each frame of each trajectory, for subsequent interface-categorized analyses.

**Predominant Interacting Residues.** The CG  $C_\alpha$  particles most likely to participate in close interactions between ecGlpG and Spitz were assessed. In each trajectory, those frames where Spitz and ecGlpG were more than 7  $\text{\AA}$  apart were excluded from the analysis. In each frame, a matrix of all possible interprotein  $C_\alpha$  interaction distances was produced, similar to the description provided in the Closest Approach Analysis above. The five smallest distances (and the identities of residues involved in these interactions) in each frame were appended to an overall list aggregated across all replicate trajectories. The frequency of occurrence for a given  $C_\alpha$  particle in the aggregated set of closest interactions was then calculated for both ecGlpG and Spitz.

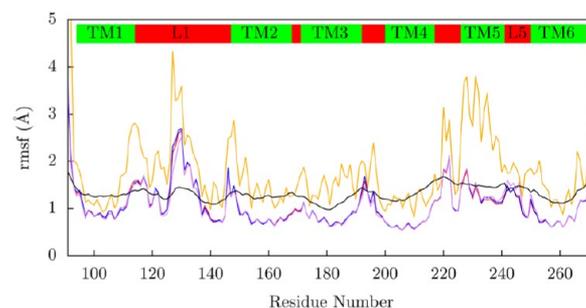
**Spitz Helix Tilt Analysis.** The helical tilt angles of Spitz in a fixed ecGlpG reference frame were visualized. For each replicate trajectory, the ecGlpG reference frame and the transformation of Spitz coordinates in each frame were as described above for positional contour maps. The centroid and minimum/maximum  $z$  coordinates for Spitz  $C_\alpha$  particles in each frame were calculated. Spitz was then centered at the origin and the first eigenvector of its  $C_\alpha$  particles calculated. The resulting eigenvector was then translated back to the appropriate reference frame and scaled from a unit vector to the length of the helix estimated from the recorded  $z$  coordinate boundaries of Spitz  $C_\alpha$  particles. The in-house code automatically generated the necessary input (i.e., cylinder primitives) for visualization of eigenvectors in VMD.

## RESULTS AND DISCUSSION

### Consistency with Previous Atomistic Simulations.

Despite previous validation of the CG force field for the reproduction of crucial experimental and atomistic computational details,<sup>29,30</sup> we felt it prudent to directly compare key observations in the previously reported atomistic results of Bondar et al.<sup>27</sup> for ecGlpG in POPE with our CG simulations. First, we observe an approximate 4  $\text{\AA}$  thinning of the POPE bilayer proximal to the enzyme consistent with the atomistic results. Specifically, the bilayer thickness  $>16$   $\text{\AA}$  from both ecGlpG and Spitz was  $44.4 \pm 0.3$   $\text{\AA}$  while the thickness within 16  $\text{\AA}$  of ecGlpG was  $40.3 \pm 0.8$   $\text{\AA}$  both before and after association with Spitz (10 replicates). Although Bondar et al. speculate that an additional perturbation of the bilayer may occur upon interaction of enzyme and substrate, we observe that the substrate causes some bilayer-thinning on its own before association with ecGlpG ( $41.6 \pm 1.7$   $\text{\AA}$  within a 16  $\text{\AA}$  shell of Spitz), but not synergistically with the enzyme ( $40.4 \pm 1.5$   $\text{\AA}$  within a 16  $\text{\AA}$  shell of Spitz) compared with ecGlpG-local thickness prior to substrate association ( $40.3 \pm 0.8$   $\text{\AA}$ ).

Likewise, the general flexibility trends reported by Bondar et al. were reproduced in our CG-MD simulations. The CG ecGlpG exhibits elevated root-mean-square fluctuation (RMSF) near the L1 loop and TMD5 (Figure 5). The magnitude of

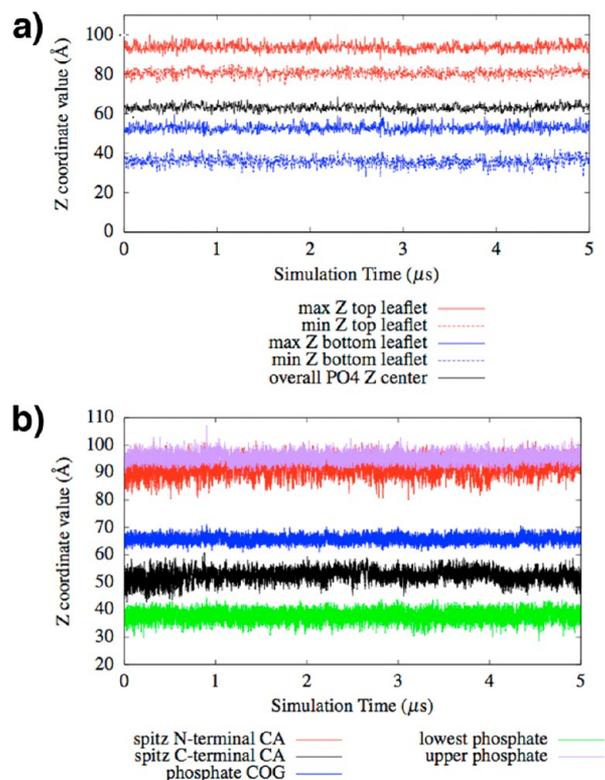


**Figure 5.**  $C_\alpha$  root-mean-square fluctuation (RMSF) for ecGlpG across all replicate simulations. Standard RMSF values, measuring per-residue fluctuations relative to an average structure, were aggregated for the 20 replicate simulations involving the substrate (red; both starting configurations included), the 10 replicates involving the FGFR3 TMD negative control (blue), and the 10 replicate simulations involving ecGlpG in the absence of any other peptide (purple). The RMSF calculation for the 20 (ecGlpG + substrate) replicates was also performed using the first frame of each replicate as a reference configuration (orange). Additionally, RMSF values were back-calculated (black) from  $C_\alpha$  B-factors, as described in Computational Methods, averaged from crystal structures (PDB ID: 2IC8),<sup>14</sup> (PDB ID: 2IRV),<sup>16</sup> (PDB ID: 2NRF),<sup>15</sup> (PDB ID: 2O7L),<sup>18</sup> (PDB ID: 2XOV),<sup>20</sup> and (PDB ID: 3B45).<sup>19</sup>

observed RMSF values for the CG enzyme is similar to the approximate 0–2  $\text{\AA}$  range reported from atomistic simulations. The much longer time scale accessible in CG-MD and large number of replicate simulations may be a possible explanation for any discrepancies. Furthermore, ecGlpG flexibility trends remained largely unchanged in the presence of the FGFR3 TMD negative control or in the absence of any other peptide in the simulation. As expected, calculated RMSF values were larger when the first frame of a trajectory was used as a reference configuration instead of the standard approach that employs the average coordinates as a reference, but the trends were still consistent. We observe considerably more flexibility in certain parts of ecGlpG when compared with RMSF values

back-calculated from available X-ray crystallography temperature factors.<sup>49</sup> This was also the case in the previous all-atom simulation study.

**Appropriateness of Bilayer Behavior and Substrate Orientation.** Although the consistency with the atomistic results of Bondar et al. provides confidence in the behavior of the CG-MD simulations, we also performed a set of analyses to ensure that the system was behaving as expected. We checked for phospholipid “flip-flop” activity between POPE leaflets by tracking the minimum, maximum, and average phosphate vertical coordinates in each CG leaflet. The results from a representative simulation (Figure 6a) are consistent with the

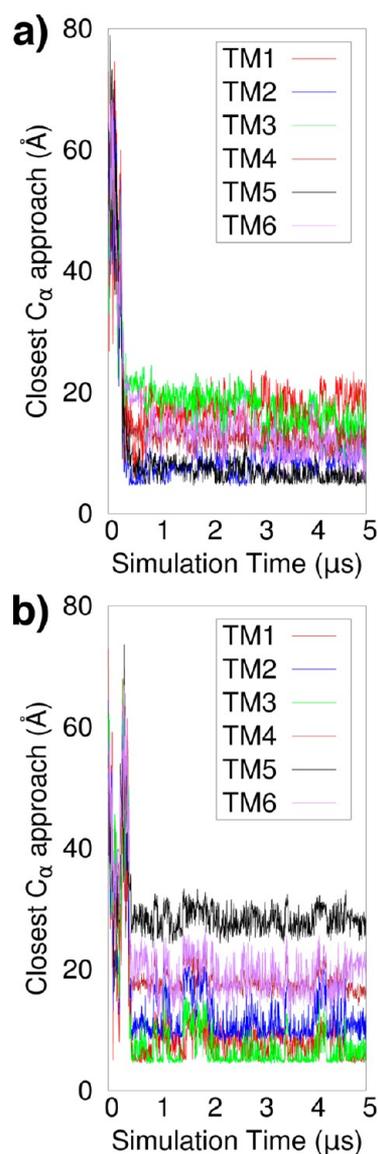


**Figure 6.** Testing for phospholipid and peptide “flip-flop” activity within the phospholipid bilayer of representative CG-MD simulations. The bilayer always formed in the  $xy$  plane such that  $z$  coordinates are vertical (along the bilayer normal). (a) Minimum and maximum POPE headgroup phosphate coordinates were tracked in each leaflet. A transition across the overall phosphate center of geometry coordinates (black) would be an indication of possible flip-flop activity. (b) A similar tracking analysis was performed for the N-terminal  $C_\alpha$  (red) and C-terminal  $C_\alpha$  (black) of Spitz relative to the overall phosphate center of geometry (COG; blue). The upper (purple) and lower (green) POPE phosphate boundaries are also highlighted.

trend over all replicates of a lack of spontaneous phospholipid flip-flop. Similar analysis performed on the termini of the peptide substrate construct indicated that Spitz did not flip-flop in the bilayer either (Figure 6b). This lack of flip-flop behavior by either lipids or TMD peptides is fully expected, but is critical to confirm. With confidence in the behavior of the system, we proceeded with an in-depth analysis of the trajectories.

**Initial and Persistent Interaction between Enzyme and Substrate Not Exclusive to Putative Gate.** We performed separate sets of CG-MD simulation replicates with Spitz starting near either ecGlpG TMD5 or TMDs 1/3 (Figure

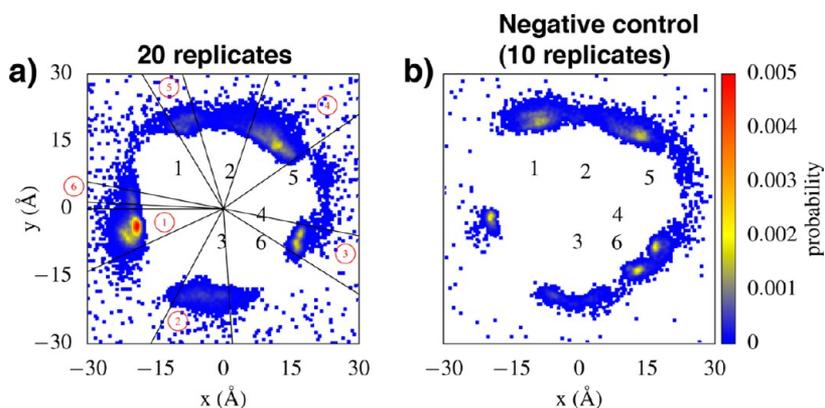
4a) to avoid introduction of any bias in the association of the substrate with the putative substrate gate. We tracked the closest approach between the  $C_\alpha$  particles of Spitz and each TM segment of ecGlpG for each replicate trajectory (e.g., Figure 7)



**Figure 7.** Monitoring the initial and sustained closest interactions between Spitz and ecGlpG TMD segments. The closest approach between  $C_\alpha$  particles in both the enzyme and substrate was calculated in each frame of the example replicate simulations shown in parts (a) and (b). Although in some cases Spitz initially contacted ecGlpG near the putative substrate gate near TMDs 2/5 (a), in the majority of cases another enzyme–substrate interaction site was preferred (b).

to assess both initial and persistent close interactions between enzyme and substrate. We also calculated the positional probability of the geometric center of Spitz in a fixed-enzyme reference frame to allow classification of enzyme–substrate interaction regions into a set of polar angle boundaries (Figure 8), summarizing the major enzyme–substrate interaction “hotspots” (or exosites) in a consistent coordinate system.

In some simulation replicates, the substrate did indeed initially and persistently interact with ecGlpG TMD 2/5 (the proposed gate) (e.g., the replicate shown in Figure 7a). In most cases, however, the substrate associated with another region of



**Figure 8.** (a) Positional probability of the geometric center of the Spitz TMD construct monitored in an rmsd-fixed reference frame (that of ecGlpG in the first frame of the first replicate simulation). The ecGlpG TM segment geometric centers are indicated by black numbers. To avoid bias, the starting configuration of Spitz was placed nearer either TMDs 1 and 3 or TMD5 for half of the replicates (Figure 4), and the results were merged (20 replicate 5  $\mu$ s simulations). Six highly populated Spitz–ecGlpG interaction sites are apparent (red circled numbers). Arcs used to classify enzyme–substrate interaction regions are conveniently defined from the positive  $x$  axis since the reference configuration is centered at the origin. The interfacial regions are described as follows: region 1,  $-2.70 \geq \theta \geq \pi$  rad; region 2,  $-1.50 \geq \theta \geq -2.06$  rad; region 3,  $-0.20 \geq \theta \geq -0.56$  rad; region 4,  $1.25 \geq \theta \geq 0.61$  rad; region 5,  $2.11 \geq \theta \geq 1.86$  rad; and region 6,  $3.09 \geq \theta \geq 2.94$  rad. (b) Corresponding probability of observing negative control fibroblast growth factor receptor 3 (FGFR3) TMD in proximity to ecGlpG (10 replicate 5  $\mu$ s simulations with ecGlpG in hydrated POPE bilayer under the same conditions).

the enzyme, with preference for an area near TMDs 1/3 (e.g., the replicate shown in Figure 7b). An assessment of the overall positional probability of the substrate in a fixed enzyme reference frame over all available replicates clearly indicates a preference for enzyme–substrate interaction at the region near TMDs 1/3 (Figure 8a; region 1), though there is still a highly populated interaction region near the proposed gate at TMDs 2/5 (Figure 8a; region 4). Although similar trends were observed for some interfaces in the negative control replicates (Figure 8b), there was a lower specificity, especially near TMDs 1/3.

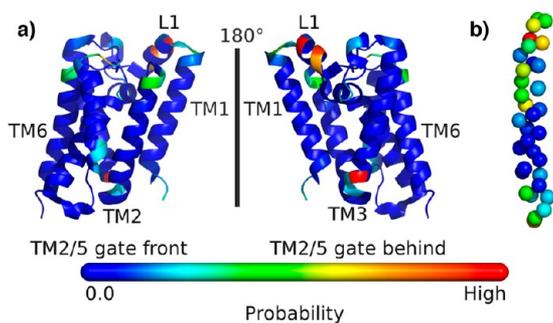
**Enzyme–Substrate Interactions Occur in Juxtamembrane Regions.** The interfacial probability maps described above are useful for classifying the system into a set of distinct interaction regions but do not address the topology of the interaction between the enzyme and substrate. Therefore, we compared the frequencies with which each residue in the enzyme and substrate were observed to be involved in the closest protein–protein interactions (Figure 9). The most frequent interactions occur in loops or at the ends of TMD helices in ecGlpG, while the Spitz substrate exhibits a matching

trend near its termini rather than in the hydrophobic bilayer core region.

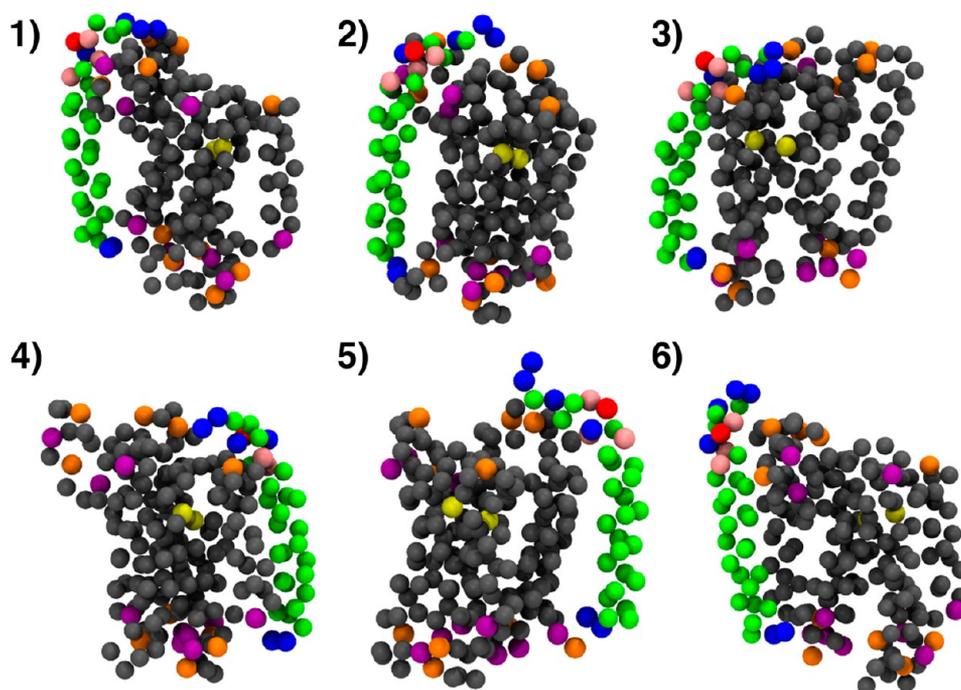
This probabilistic analysis does not answer the question of whether juxtamembrane enzyme–substrate interactions were predominant at each of the six polar interfaces identified for this system and detailed in Figure 8a. This is clarified by examination of the enzyme–substrate complexes formed at each interface site (Figure 10) using the frame of closest enzyme–substrate approach to demonstrate the most favorable interactions. Notably, the substrate TMD curves toward the enzyme at its termini forming favorable charge–charge interactions, while the central hydrophobic portion of the substrate typically bends away from the enzyme. It therefore appears that substrate–enzyme electrostatic interactions are the driving force for the preferential presentation of substrate termini to enzyme exosites. Naturally, charged residues are preferentially located at juxtamembrane regions rather than within the hydrophobic core of the bilayer, providing a straightforward explanation for the observed preferential juxtamembrane interactions (Figure 9).

Finally, the preferred interaction sites for each interfacial region were classified according to polar angle and as a function of primary amino acid sequence so that the specific interacting residues could be precisely identified on the enzyme and substrate at each interface. This exhaustive analysis is provided in the Supporting Information, and the results at each interface for enzyme and substrate are consistent with preferential juxtamembrane interactions.

**Enzyme–Substrate Interactions Are Persistent but Interface Transitions Are Possible.** Given the above trends for enzyme–substrate interaction, the question arises as to whether or not each enzyme–substrate interface is stable or whether there are frequent transitions between these interaction sites on the enzyme. The geometric center of the substrate in the fixed-enzyme reference frame was tracked as a function of simulation progress (e.g., Figure 11a,b). In 11/20 replicates the substrate does not appear to be very mobile in the enzyme-fixed reference frame after complex formation (Figure



**Figure 9.** Map over ecGlpG (a) and Spitz (b, CG representation of  $C_{\alpha}$  particles) illustrating probability of involvement in closest approach (over 10 replicate 5  $\mu$ s simulations). Water and lipid bilayer are excluded for clarity. The cytosolic side is at the bottom and the periplasmic side at the top. The highest probabilities are 0.092 and 0.11 for ecGlpG and Spitz, respectively.



**Figure 10.** Configuration of closest enzyme–substrate approach for each of the six most populated interfaces (Figure 8a). The simulation frame from each of the interfaces (numbered as in Figure 8a) where the enzyme and substrate have the closest center-to-center distance is illustrated. To highlight possible electrostatic interactions, basic residues are shown for the substrate (blue) and enzyme (purple) alongside acidic residues for the substrate (red) and enzyme (orange). The ecGlpG catalytic dyad is shown in yellow. The active site of the enzyme and periplasm are located at the top of each representation, and the N-terminal P4, P1, and P2' sites<sup>24</sup> are shown in pink. The phospholipid bilayer and solvent have been excluded for clarity, and the enzyme–substrate complex is rotated for clarity.

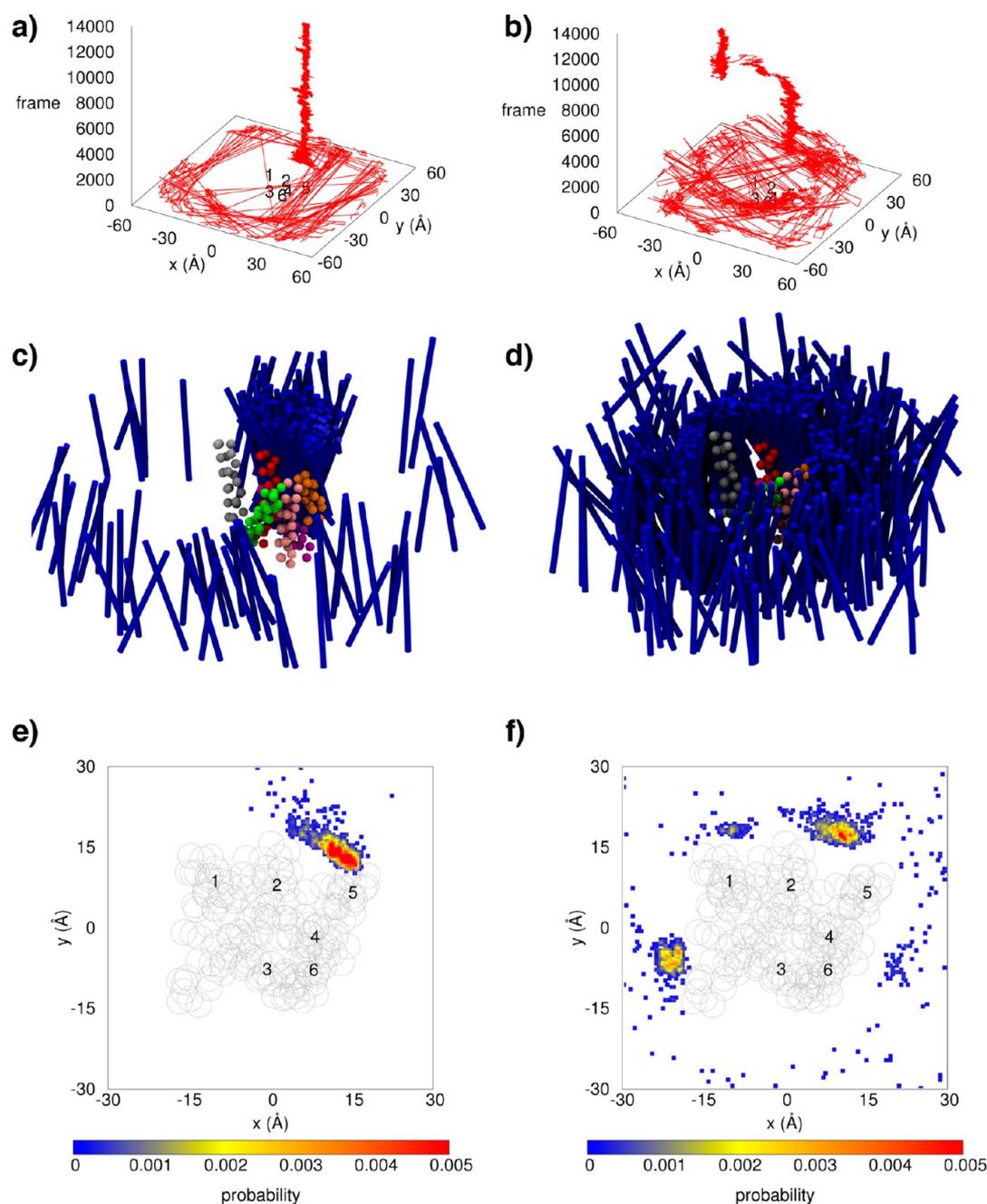
11a). However, in some cases transitions between different enzyme–substrate interfaces clearly occur (Figure 11b).

Tracking of the geometric center of the substrate in this manner is not very sensitive to the tilt angle of the helix in the bilayer, as the tilt angle could theoretically change substantially with little effect on the coordinates of the center of the TMD. Thus, the first eigenvector of the substrate (see Computational Methods) was used to approximate its main helical axis and these eigenvectors were visualized in the fixed-enzyme reference frame (e.g., Figure 11c,d). Clearly, the substrate is able to sample a number of different tilt angles after association with the enzyme, revealing less rigid interfaces than might be implied by the center-tracking analysis (Figure 11a,b) or by positional probability (Figure 11e,f) alone.

**Functional Relevance: Rhomboids, iRhoms, and Limitations.** Substrate specificity for a number of enzymes is directed by exosite binding, where an exosite is defined as a secondary binding site remote from the active site.<sup>50</sup> The presence of multiple substrate exosites on ecGlpG suggests that the substrate is “captured” by the enzyme at a secondary site prior to entry to the active site proper, as has been previously speculated for rhomboid proteases. Specifically, functional experiments on rhomboid substrates suggest that the substrate TMD segment binds to an exosite on rhomboid prior to active site binding, since the TMD is required even when the substrate cleavage site is experimentally adjusted to a position outside of the membrane bilayer.<sup>24</sup> This may be a common feature of the rhomboid family, as the presence of multiple exosites was recently suggested by structural biology and mutagenesis on the *Haemophilus influenzae* rhomboid, hiGlpG.<sup>21</sup>

The exosite scheme is not necessarily exclusive to the proposal that ecGlpG TMD5 acts as a substrate gate, as the initial capture event may be independent of the entry of the substrate to the active site at another location. Indeed, our observation of “hopping” between exosites in some replicates on the  $\sim$ microsecond time scale (e.g., Figure 11b) implies the feasibility of initial substrate capture followed by sampling of exosites until catalysis is able to take place. Interestingly, matrix metalloproteinase (MMP) exosites have been proposed to be involved in “substrate preparation” and “unwinding” of native collagen, the so-called triple-helicase activity, prior to collagen cleavage.<sup>51</sup> Following from this, rhomboid exosites might have a role to play in the helical destabilization of substrates that has been suggested as an important feature for cleavage competency in previous works.<sup>52,53</sup> The identification of preferential enzyme–substrate interactions near charged sites in juxtamembrane regions may also provide some inspiration for future mutagenesis experiments and inhibitor design. This is particularly exciting in light of the fact that there are currently no rhomboid-specific protease inhibitors available despite the breadth of biologically and medically relevant roles presented in the Introduction. Indeed, inhibition of substrate binding at an exosite has been proposed for anti-MMP therapies to combat connective tissue degradation.<sup>51</sup> A similar strategy would provide the capability to distinctly differentiate rhomboid inhibitors from other serine proteases.

There is also significant value in studying the enzyme–substrate interaction independent of catalysis. A natural criticism of the use of classical MD methodologies, whether CG or atomistic, to study this (or any other) enzyme is that bond breaking and formation cannot occur without a quantum mechanical component to the simulation system. In the case of



**Figure 11.** Tracking Spitz in the fixed reference frame of ecGlpG for representative cases where the substrate remains relatively fixed (left) or transitions between exosites (right), respectively. (a,b) The geometric centers of the ecGlpG TMDs are shown in black numbers at the centers of the plots. The geometric center of Spitz (colored red) is tracked as a function of the progression of a given simulation (by frame number). (c,d) Spitz helical tilt angles relative to ecGlpG fixed reference frame. Angled periplasmic views of ecGlpG surrounded by the first eigenvectors (dark blue) of the Spitz  $C_{\alpha}$  particles are shown. The rhomboid TM helices ( $C_{\alpha}$  particles only; loops excluded) are colored in silver (TM 1), red (TM 2), green (TM 3), purple (TM4), orange (TM5), and pink (TM6). (e,f) Corresponding contour plots for the representative replicates. The positional probability of the geometric center of Spitz from the same replicate simulations is shown with ecGlpG TM segment geometric center positions labeled at the centers of the plots along with an outline of the van der Waals radii of ecGlpG  $C_{\alpha}$  particles in the force field (gray circles).

the rhomboid proteases, there is a phylogenetic branch of catalytically inactive rhomboids, termed iRhoms, which are usually lacking the Ser and/or His of the catalytic dyad, or have a proline introduced near the active-site Ser.<sup>4</sup> Strikingly, a *Drosophila* iRhom was shown to associate with natural rhomboid substrates in the ER, to retain these substrates in the ER, and to promote ER associated proteasomal degradation (ERAD) of the substrates.<sup>33</sup> The authors confirmed that iRhoms were directly antagonizing the EGFR-signaling activity

caused by their catalytically active counterparts in the Golgi, leading to the suggestion that the regulation of enzyme-driven pathways by their catalytically dead counterparts could be a common evolutionary strategy. If so, insights into enzyme–substrate interaction from classical MD simulations may still have a large role to play in studying the interactions of other catalytically dead, but biologically relevant, proteins despite inaccessibility of the simulations to the catalytic process proper.

The capacity for bending of the substrate N-terminus toward the enzyme in each of the six major interfaces (Figure 10) is consistent with presentation of the scissile peptide bond laterally to the active site. Unfortunately, the secondary structure restraints imposed in our force field prevent an informed testing of the helix-unwinding hypothesis proposed as being an essential cleavage prerequisite<sup>52,53</sup> and allowing exposure of the peptide backbone to nucleophilic attack. We felt it would be premature to repeat the simulations in the absence of the substrate structural restraints, at least until a high-resolution substrate structure becomes available and warrants a detailed all-atom investigation of the substrate protein-fold stability at various ecGlpG exosites.

An additional potential limitation of our simulation setup is that the 12 Å long-range electrostatic cutoff (Figure 4) may not appropriately mimic the electrostatics driving the substrate capture event *in vivo*, given that charge–charge interactions appear to be driving the interfacial interaction. However, the fact that this particular electrostatics simulation scheme has previously reproduced crucial details from atomistic simulations and structural biology results<sup>29,30</sup> is a good indicator that the results of our simulations would not be significantly hindered by the length scale of this interaction.

Beyond the observance of multiple highly populated enzyme–substrate interaction sites between Spitz and ecGlpG, the retention of some of these interaction sites between ecGlpG and the negative control FGFR3 TMD peptide (Figure 8b), albeit with less specificity, suggests a degree of promiscuity to the initial enzyme–substrate interaction event. The promiscuity of rhomboids has recently been uncovered in their capacity to cleave polytopic TMD constructs of varying sizes,<sup>54</sup> and is consistent with the capacity of rhomboid proteases from a wide range of species to cleave substrates derived not only from different rhomboid–substrate pairs but also from different kingdoms of life.<sup>52,55</sup> Indeed, if the initial enzyme–substrate interaction is electrostatically driven, the presence of charged residues in the juxtamembrane region of a TMD substrate may suffice to allow an initial capture of a broad range of substrates. For example, the FGFR3 TMD negative control construct used in this study contains two Arg residues at the N-terminus and two Arg residues near the C-terminus, and interacts with enzyme at a number of regions favored by the substrate. It is, however, known that normally the substrate P4 and P2' positions require hydrophobic residues and the P1 position must contain a small residue in order for cleavage proper to occur.<sup>24</sup>

## CONCLUSIONS

Rhomboids may be the most well-conserved class of membrane proteins known,<sup>2,56</sup> and we have performed the first set of simulations involving both a rhomboid protease and its substrate. Our results suggest that there is an initial electrostatically driven enzyme–substrate capture event in the juxtamembrane region of the bilayer, with minimal interaction between hydrophobic cores of the rhomboid and substrate transmembrane domains. Upon substrate capture at an interface site, dynamics is inherent both in the orientation of the substrate and in transitions of the substrate between the highly populated enzyme–substrate interaction sites. The extensive substrate dynamics and sampling of multiple enzyme–substrate interaction sites would likely facilitate substrate entry via an enzyme gate, even if this were located distal to the initial capture site of the enzyme. Further insight

would be obtained both by mutagenesis studies on charged juxtamembrane residues in the enzyme and substrate and by additional simulations of rhomboid–substrate interactions using high-resolution substrate structure(s) once they become available.

## ASSOCIATED CONTENT

### Supporting Information

Detailed plots of enzyme–substrate per-residue close interactions at each of the major interfaces are provided in a PDF file. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [jan.rainey@dal.ca](mailto:jan.rainey@dal.ca).

### Present Address

<sup>§</sup>Structural Bioinformatics & Computational Biochemistry Unit, Department of Biochemistry, University of Oxford, Oxford, UK.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank members of the Structural Bioinformatics and Computational Biochemistry Unit (SBCB) at the Department of Biochemistry, University of Oxford, for helpful advice on the project, Dr. Philip Fowler for critical reading of the manuscript, and Prof. Mark S. P. Sansom and Drs. Benjamin Hall and Phillip Stansfeld provided helpful advice during the peer review process. This work was supported by a doctoral level Natural Sciences and Engineering Research Council (NSERC) Canada Graduate Scholarship, a Killam predoctoral scholarship, and an NSERC Michael Smith Foreign Study Supplement to T.R. and a Nova Scotia Health Research Foundation Health Research Grant to J.K.R. Currently, J.K.R. is supported by a Canadian Institutes of Health Research New Investigator Award. Simulations were performed using resources at the SBCB and ACEnet, the regional high performance computing consortium for universities in Atlantic Canada. ACEnet is funded by the Canada Foundation for Innovation, the Atlantic Canada Opportunities Agency, and the provinces of Newfoundland and Labrador, Nova Scotia, and New Brunswick.

## REFERENCES

- (1) Kraut, J. *Annu. Rev. Biochem.* **1977**, *46*, 331.
- (2) Urban, S. *Genes Dev.* **2006**, *20*, 3054.
- (3) Mayer, U.; Nusslein-Volhard, C. *Genes Dev.* **1988**, *2*, 1496.
- (4) Freeman, M. *Annu. Rev. Genet.* **2008**, *42*, 191.
- (5) Cipolat, S.; Rudka, T.; Hartmann, D.; Costa, V.; Serneels, L.; Craessaerts, K.; Metzger, K.; Frezza, C.; Annaert, W.; D'Adamio, L.; et al. *Cell* **2006**, *126*, 163.
- (6) Herlan, M.; Vogel, F.; Bornhovd, C.; Neupert, W.; Reichert, A. S. *J. Biol. Chem.* **2003**, *278*, 27781.
- (7) Walder, K.; Kerr-Bayles, L.; Civitarese, A.; Jowett, J.; Curran, J.; Elliott, K.; Trevasakis, J.; Bishara, N.; Zimmet, P.; Mandarino, L.; et al. *Diabetologia* **2005**, *48*, 459.
- (8) Alexander, C.; Votruba, M.; Pesch, U. E. A.; Thiselton, D. L.; Mayer, S.; Moore, A.; Rodriguez, M.; Kellner, U.; Leo-Kottler, B.; Auburger, G.; et al. *Nat. Genet.* **2000**, *26*, 211.
- (9) Brossier, F.; Jewett, T. J.; Sibley, L. D.; Urban, S. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 4146.
- (10) Montoya, J. G.; Liesenfeld, O. *Lancet* **2004**, *363*, 1965.

- (11) Baker, R. P.; Wijetilaka, R.; Urban, S. *PLoS Pathog.* **2006**, *2*, e113.
- (12) Stevenson, L. G.; Strisovsky, K.; Clemmer, K. M.; Bhatt, S.; Freeman, M.; Rather, P. N. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1003.
- (13) Erez, E.; Fass, D.; Bibi, E. *Nature* **2009**, *459*, 371.
- (14) Wang, Y. C.; Zhang, Y. J.; Ha, Y. *Nature* **2006**, *444*, 179.
- (15) Wu, Z. R.; Yan, N.; Feng, L.; Oberstein, A.; Yan, H. C.; Baker, R. P.; Gu, L. C.; Jeffrey, P. D.; Urban, S.; Shi, Y. G. *Nat. Struct. Mol. Biol.* **2006**, *13*, 1084.
- (16) Ben-Shem, A.; Fass, D.; Bibi, E. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 462.
- (17) Lemieux, M. J.; Fischer, S. J.; Cherney, M. M.; Bateman, K. S.; James, M. N. G. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 750.
- (18) Wang, Y. C.; Ha, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2098.
- (19) Wang, Y. C.; Maegawa, S.; Akiyama, Y.; Ha, Y. *J. Mol. Biol.* **2007**, *374*, 1104.
- (20) Vinothkumar, K. R.; Strisovsky, K.; Andreeva, A.; Christova, Y.; Verhelst, S.; Freeman, M. *EMBO J.* **2010**, *29*, 3797.
- (21) Brooks, C. L.; Lazareno-Saez, C.; Lamoureux, J. S.; Mak, M. W.; Lemieux, M. J. *J. Mol. Biol.* **2011**, *407*, 687.
- (22) Vinothkumar, K. R. *J. Mol. Biol.* **2011**, *407*, 232.
- (23) Baker, R. P.; Young, K.; Feng, L.; Shi, Y. G.; Urban, S. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 8257.
- (24) Strisovsky, K.; Sharpe, H. J.; Freeman, M. *Mol. Cell* **2009**, *36*, 1048.
- (25) Xue, Y.; Chowdhury, S.; Liu, X.; Akiyama, Y.; Ellman, J.; Ha, Y. *Biochemistry* **2012**, *51*, 3723.
- (26) Urban, S.; Schlieper, D.; Freeman, M. *Curr. Biol.* **2002**, *12*, 1507.
- (27) Bondar, A. N.; del Val, C.; White, S. H. *Structure* **2009**, *17*, 395.
- (28) Zhou, Y.; Moin, S. M.; Urban, S.; Zhang, Y. *Structure* **2012**, *20*, 1255.
- (29) Psachoulia, E.; Fowler, P. W.; Bond, P. J.; Sansom, M. S. P. *Biochemistry* **2008**, *47*, 10503.
- (30) Kalli, A. C.; Hall, B. A.; Campbell, I. D.; Sansom, M. S. *Structure* **2011**, *19*, 1477.
- (31) Janosi, L.; Prakash, A.; Doxastakis, M. *Biophys. J.* **2010**, *99*, 284.
- (32) Sengupta, D.; Marrink, S. J. *Phys. Chem. Chem. Phys.* **2010**, *12*, 12987.
- (33) Zettl, M.; Adrain, C.; Strisovsky, K.; Lastun, V.; Freeman, M. *Cell* **2011**, *145*, 79.
- (34) Lindahl, E.; Hess, B.; Van Der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306.
- (35) Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701.
- (36) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435.
- (37) Bond, P. J.; Sansom, M. S. P. *J. Am. Chem. Soc.* **2006**, *128*, 2697.
- (38) Bond, P. J.; Holyoake, J.; Ivetac, A.; Khalid, S.; Sansom, M. S. P. *J. Struct. Biol.* **2007**, *157*, 593.
- (39) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812.
- (40) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. *J. Chem. Theory Comput.* **2008**, *4*, 819.
- (41) Periole, X.; Cavalli, M.; Marrink, S. J.; Ceruso, M. A. *J. Chem. Theory Comput.* **2009**, *5*, 2531.
- (42) Wang, Y. M.; Rader, A. J.; Bahar, I.; Jernigan, R. L. *J. Struct. Biol.* **2004**, *147*, 302.
- (43) Kundu, S.; Melton, J. S.; Sorensen, D. C.; Phillips, G. N. *Biophys. J.* **2002**, *83*, 723.
- (44) Bond, P. J.; Wee, C. L.; Sansom, M. S. P. *Biochemistry* **2008**, *47*, 11321.
- (45) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics Modell.* **1996**, *14*, 33.
- (46) Schrodinger, L. L. C. The PyMOL Molecular Graphics System, Version 1.3r1, 2010.
- (47) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. *J. Comput. Chem.* **2011**, *32*, 2319.
- (48) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
- (49) Case, D. A.; Darden, T. A. D.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; et al. AMBER 12 University of California, San Francisco, 2012.
- (50) Yegneswaran, S.; Tiefenbrunn, T. K.; Fernandez, J. A.; Dawson, P. E. *J. Thromb. Haemostasis* **2007**, *5*, 2062.
- (51) Overall, C. M. *Mol. Biotechnol.* **2002**, *22*, 51.
- (52) Urban, S.; Freeman, M. *Mol. Cell* **2003**, *11*, 1425.
- (53) Akiyama, Y.; Maegawa, S. *Mol. Microbiol.* **2007**, *64*, 1028.
- (54) Erez, E.; Bibi, E. *Biochemistry* **2009**, *48*, 12314.
- (55) Urban, S.; Wolfe, M. S. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 1883.
- (56) Koonin, E. V.; Makarova, K. S.; Rogozin, I. B.; Davidovic, L.; Letellier, M. C.; Pellegrini, L. *Genome Biol.* **2003**, *4*, R19.