# Understanding Co-translational Protein Targeting and Lithium Dendrite Formation through Free Energy Simulations and Coarse-Grained Models

Thesis by
Connie Wang

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy


CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2016
Defended May 20, 2016

Connie Wang
ORCID: [0000-0003-2971-3971]

# ACKNOWLEDGEMENTS

First I would like to thank Professor Thomas F. Miller, my advisor, who has encouraged and inspired me to become a more critical thinker and careful scholar during the course of my graduate career.

I would also like to thank the faculty of the Division of Chemistry and Chemical Engineering. In particular, I would like to thank my thesis committee, Professor Steven Mayo, Professor Shu-ou Shan, and Professor Zhen-Gang Wang for their support and guidance throughout my graduate career.

Thank you to the division administrative staff, in particular Priscilla Boon for keeping the Miller group running behind the scenes and for always being available for questions and conversations.

I appreciate the support and friendship of the Miller Group, past and present. Artur, Bin, Nandini, Jason, Josh, Taylor, Nick, Frank, Kuba, Matt, Mike, Michiel, Reid, Fran, Mark, Eric, Joonho, Lila, Brett, Ralph, Feizhi, Sebastian and Brooke, your support and willingness to provide comments for posters, talks and papers, and help with scientific questions has so greatly enriched my time at Caltech. Our walks to get coffee, jokes and group discussions made every day in the Miller group more enjoyable.

Thank you to the occupants of 151 South Michigan Ave, Helen, Matt, Eric, my friends and housemates who experienced all the ups and downs of classes, research and plumbing issues with me for the past five years. Thank you to all my friends and teammates who have always supported me and helped me enjoy life outside of Caltech.

Thank you to my family, my parents and my sister who have provided love and support and encouragement throughout my entire life.

# ABSTRACT

We describe the application of alchemical free energy methods and coarse-grained models to study two key problems: (i) co-translational protein targeting and insertion to direct membrane proteins to the endoplasmic reticulum for proper localization and folding, (ii) lithium dendrite formation during recharging of lithium metal batteries. We show that conformational changes in the signal recognition particle, a central component of the protein targeting machinery, confer additional specificity during the the recognition of signal sequences. We then develop a three-dimensional coarse-grained model to study the long-timescale dynamics of membrane protein integration at the translocon and a framework for the calculation of binding free energies between the ribosome and translocon. Finally, we develop a coarse-grained model to capture the dynamics of lithium deposition and dissolution at the electrode interface with time-dependent voltages to show that pulse plating and reverse pulse plating methods can mitigate dendrite growth.

# PUBLISHED CONTENT AND CONTRIBUTIONS

(1) Wang, C. Y.; Miller, T. F. *J. Biol. Chem.* **2014**, *289*, 30868–30879.

# TABLE OF CONTENTS

# LIST OF ILLUSTRATIONS

# LIST OF TABLES

*Chapter 1*

# INTRODUCTION

Computational simulations have a long history in studying the properties and dynamics of complex chemical systems in both biology and materials [1–3]. In the last 40 years, advances in computing power have pushed the capabilities of these simulation methods into the millisecond timescale which has allowed for the direct observation of critical events such as protein folding [4]. However, for many complex chemical systems, the reactions and dynamics of interest occur on time scales that are not yet accessible or are prohibitively expensive to simulate [5]. Furthermore, observation of a single trajectory is often insufficient to provide the statistics necessary to enable prediction and comparison to experiments. In the following dissertation, we describe the application of free energy methods that enable the direct calculation of experimentally measurable quantities and the development of coarse-grained models that allow for the simulation of much longer timescales. These methods are applied to study two key problems: (i) co-translational protein targeting which directs membrane proteins to the endoplasmic reticulum for proper localization and folding and (ii) lithium dendrite formation in lithium metal batteries.

Chapter 2 describes the characterization of the conformational dynamics and substrate selectivity of the signal recognition particle (SRP) using a thermodynamic free energy cycle approach and microsecond timescale molecular dynamics simulations. The SRP is a central component of the co-translational protein targeting machinery that binds to the N-terminal signal peptide (SP) of nascent proteins and selects them for targeting to the translocon for membrane integration or translocation. The SRP is unique for the large conformational changes that are associated with targeting. We determine the role of these conformational changes in determining substrate specificity by calculating the shift in relative conformational stability of the SRP upon substrate binding to quantify allosteric coupling between SRP domains. Microsecond timescale trajectories are also performed to study the intrinsic flexibility of the SRP conformational landscape; illustrating that 10-nm lengthscale changes occur via the rigid-body movement of SRP domains connected by the flexible linker region. This work has been published as "Allosteric response and substrate sensitivity in peptide binding of the signal recognition particle" C. Y. Wang

and T. F. Miller III, *J. Biol. Chem.*, 289, 30868 (2014).

Chapter 3 describes development of a coarse-grained simulation model to study the co-translational integration of the nascent protein at the translocon, focusing on the establishment of topology during early stage membrane protein integration. Development of a new three-dimensional coarse-grained model of membrane protein integration enables access to the long timescales of protein translocation needed to understand this process, while preserving the needed sequence-level detail. The development of the 3D model used in this work was done in close collaboration with Michiel Niesen and Reid Van Lehn and is in preparation for submission. A discrete state free energy approach is then developed to study the energetics of RNC-translocon binding as a function of stalled nascent chain lengths. Calculation of binding free energies allows a direct comparison with experimental results. Furthermore, these simulations enable a comparison between equilibrium conformational preferences and the intermediates of the nonequilibrium protein integration process.

Chapter 4 describes the study of lithium dendrite formation during recharging of lithium metal batteries. Lithium metal batteries are a promising electrode technology that offers a very high theoretical energy density, but have long been deemed unsuitable for rechargeable battery applications because of the instability of the electrode surface—which leads to dendritic growth during recharging. Although this system is very different from co-translational membrane protein integration, the issues of long timescales and large lengthscales are similar, and we again apply a coarse-graining approach to achieve the relevant timescales. We extend a recently developed coarse-grained reaction-diffusion model that captures several of the essential features of voltage-dependent lithium deposition and allows us to reach the timescales necessary to study various time-dependent charging protocols. This work incorporates the effect of voltage-dependent lithium dissolution and allows for investigation of time-dependent reverse pulse charging protocols which show promise in mitigating dendrite growth. This work additionally incorporates the effect of mechanical barriers and electrolyte inhomogenieties in the SEI to allow for simulation of novel strategies to mitigate dendrite growth through artificial thin films.

## References

(1) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.

(2) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585–90.

(3) Cowley, E. R.; Jacucci, G.; Klein, M. L.; McDonald, I. R. *Phys. Rev. B* **1976**, *14*, 1758–1769.

(4) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. *Science* **2011**, *334*, 517–520.

(5) Elber, R. *J. Chem. Phys.* **2016**, *144*, 060901.

*Chapter 2*

# ALLOSTERIC RESPONSE AND SUBSTRATE SENSITIVITY IN PEPTIDE BINDING OF THE SIGNAL RECOGNITION PARTICLE

## 2.1 Introduction

Protein targeting pathways govern the delivery of both secretory and integral membrane proteins to their appropriate cellular destinations [1]. Co-translational protein targeting relies on the recognition of the N-terminal signal peptide (SP) of nascent-protein sequences. This process requires the signal recognition particle (SRP), a protein-RNA complex and universally conserved component of the co-translational protein targeting machinery [2]. The SRP binds to the ribosome/nascent-protein complex (RNC) and targets nascent proteins for either translocation or membrane integration via the Sec translocon [3–9].

Nascent proteins that undergo targeting via the co-translational pathway have SPs that exhibit an 8-12 residue hydrophobic core and a propensity for alpha-helical secondary structure [10, 11]. However, the detailed mechanism by which SP binding triggers selection for co-translational targeting remains unclear. In this work, we use free-energy (FE) calculations and microsecond-timescale trajectories to investigate the coupling between SP binding and SRP conformational dynamics and to provide new insight into the role of SP binding in co-translational protein targeting.

Both structural [12–21] and biochemical work [22–24] suggest that the SRP exhibits multiple stable conformations that are important for protein targeting. The conserved functional core of the SRP (Figure 2.1A) is comprised of an RNA component and a multidomain protein component [25, 26]. The protein component consists of a methionine-rich M domain that contains both the SP binding site and the RNA binding site, as well as an NG domain that contains a catalytic GTPase [27–29]. The M and NG domains are connected by a flexible 30-residue linker region [12]. Figure 2.1A,B represents two conformations of the SRP that have been proposed to play a central role in the initial step of SP binding to the SRP [13]. Recent crystal structures reveal that while the structure of the individual domains is quite similar, the relative arrangements of the domains differ substantially between the two conformations [12, 13]. Throughout this paper we refer to these two con-

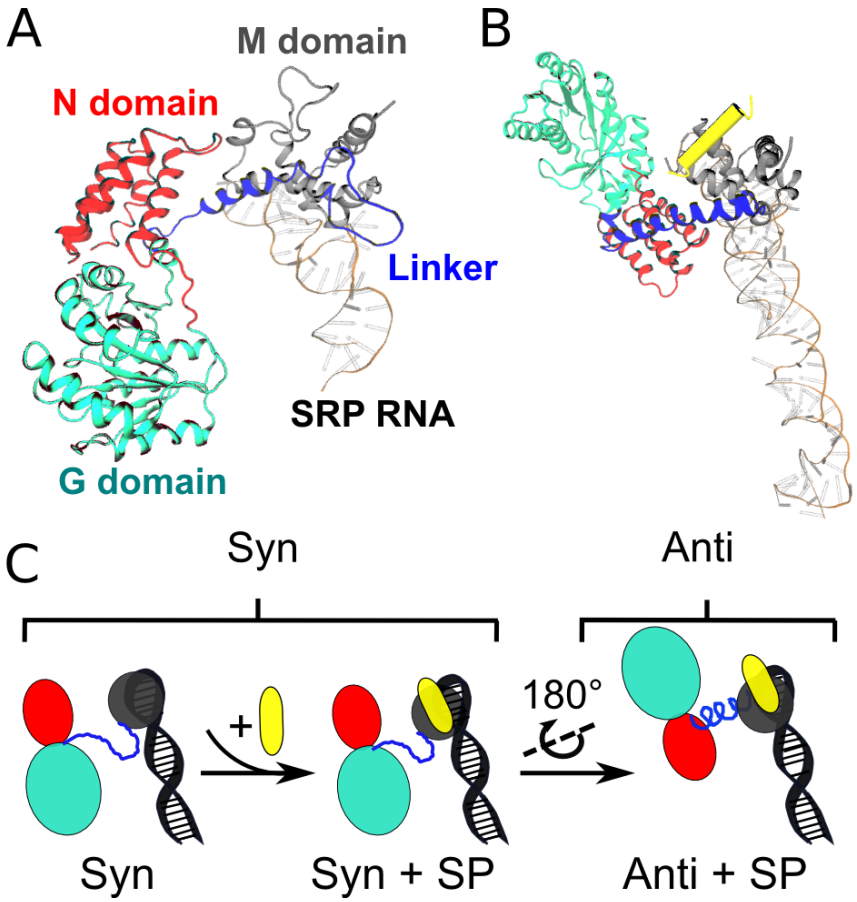



Figure 2.1: A. Crystal structure of the *S. Solfataricus* SRP in the Syn conformation [12]. A disordered linker (blue) region connects the N (red) and G (green) domains to the M (gray) domain which is bound to the SRP RNA. B. Crystal structure of the *M. jannaschi* SRP in the Anti conformation with bound SP (yellow) [13]. C. The proposed mechanism in which SP binding leads to a shift in the SRP conformational distribution from the Syn conformation to the Anti conformation.

formations as Syn and Anti, indicating the relative orientation of the SRP RNA and the NG domain.

Biochemical evidence suggests that SP binding influences the kinetics of downstream targeting events, such as dimerization of the SRP with its receptor, GTPase activity of the SRP, and final release of the RNC from the SRP to the translocon channel [30]. Truncation or modification of the SP has additionally been shown to modulate the rate of these targeting events [31, 32], suggesting that the targeting events serve as checkpoints for different SPs that enforce fidelity of the co-translational targeting pathway [30]. Furthermore, the Anti conformation of the SRP places the universally conserved tetraloop end of its RNA in close proximity

with its catalytic NG domain [13], and it has been shown that this RNA tetraloop is essential for the protein targeting step that involves complex formation between the SRP and its receptor [33–38].

The proposed model shown in Figure 2.1C relates initial SP binding to changes in the SRP conformation, which then impact the downstream targeting events [13, 17, 20, 39–41]. In this model, SP-induced shifts of the SRP to the Anti conformation helps to organize the SRP for necessary catalytic events and interactions with its receptor.

In this study, we investigate the possibility that SP binding drives conformational changes of the SRP associated with the efficiency of subsequent targeting events. A thermodynamic FE cycle is designed to calculate the shift in SRP conformational stability upon substrate binding. Thermodynamic FE cycles have long been applied to the study of biophysical systems, addressing issues that include the solvation FE for small organic molecules [42–44], binding energy for protein-ligand interactions [45–47], the relative stability of protein conformations [48], and the relative binding preference of ligands for different protein conformations [49]. The thermodynamic cycle employed in this study allows us to equate the difference in binding FE of the SP to different conformations of the SRP to SRP conformational changes upon SP binding and reduces the demands of protein conformational sampling. We calculate the allosteric effect induced by SP binding for a range of experimentally studied SPs to investigate the effect of changing the SP amino acid sequence. Furthermore, we reveal details of the large-scale SRP conformational rearrangements by performing several microsecond-timescale (MD) trajectories. Taken together, these calculations provide insight into the sensitivity of the SRP conformational changes to the identity of the SP. In particular, the results of this study reveal that the conformational preference of the SRP for the Anti over the Syn conformation upon SP binding agrees with trends found in experimentally measured targeting efficiencies, suggesting that SRP conformational dynamics are substrate specific and provide a selection mechanism of nascent proteins for the co-translational targeting pathway.

## 2.2 Methods

### Modeling for FE calculations

Initial coordinates for the Syn conformation of the SRP are available from the high-resolution crystal structure of the archeael *S. solfactaricus* species (1QZW) [12].

| SP | Sequence | V(Å$^3$)* | Features |
|---|---|---|---|
| DPAP-WT | GIILVLLIWGTVLLLKSIPH | 3291 | Efficient co-translational targeting sequence |
| DPAP-K464L | GIILVLLIWGTVLLLLSIPH | 3178 | Single-point mutation from WT |
| phoA | IALALLPLLFTPVTKA | 2338 | Inefficient co-translational targeting sequence |

*Excluded volume of the SP, as determined by VMD volmap [50]

Table 2.1: Summary of the sequences and features of studied SPs.

Coordinates for the Anti conformation of the SRP from *S. solfactaricus* have not been reported in a crystallographic study, so they are instead modeled from the available *M. jannaschii* structure (3NDB) [13] as follows. A pairwise sequence alignment of the *S. solfactaricus* and *M. jannaschi* sequences is performed using ClustalX [51] and input into MODELLER [52]. The MODELLER protocol builds coordinates for the *S. solfactaricus* SRP in the Anti conformation that minimizes a set of the spatial restraints generated from the sequence alignment with the existing *M. jannaschii* structure, and a set of stereochemical restraints determined from the CHARMM22 force field [53]. The model coordinates are refined with conjugate gradient minimization and molecular dynamics with simulated annealing. The FE associated with SP binding to either the Syn and Anti conformations of the SRP is calculated for three SPs, listed in Table 2.1. The first is dipeptidyl aminopeptidase B (DPAP-WT), a prototypical example of a peptide sequence that is targeted to the co-translational targeting pathway [54]. The second SP, DPAP-K464L, is obtained from the DPAP-WT sequence by a single lysine-to-leucine mutation at residue 464. This mutation extends the hydrophobic region of the SP relative to the DPAP-WT. The third SP is phoA, which is found to bind SRP *in vitro* [10], but which only weakly targets proteins via the co-translational targeting pathway [30]; *in vivo* systems alternatively select phoA for the post-translational SecB pathway [55]. A crystal structure of the DPAP-WT SP bound to the Syn conformation of the SRP is available (3KL4) [19] and used to initialize coordinates for simulations of the bound DPAP-WT. The coordinates for bound DPAP-K464L are constructed via direct substitution of the leucine side-chain coordinates in DPAP-WT at residue 464 with lysine coordinates, followed by steepest descent minimization to remove steric clashes [56]. The coordinates for bound phoA are constructed by sequence

alignment with the DPAP-WT SP and modeled by the MODELLER protocol described previously. We note that the N-terminal end of the SP sequences are capped using a neutral acetyl group, rather than a positively-charged bare ammonium terminus, to account for the fact that that in the experimental crystal structure [19] the N-terminus of the SP sequence is bonded to additional residues that are not resolved; for the simulations to include a bare ammonium terminus at the N-terminal end of the SP would introduce an unphysical charge at that point in the sequence.

**Free energy sampling**

We perform nine sets of calculations to determine the FE of binding for each of the three SPs to the SRP in the Syn or Anti conformation. For the *i*th SP, these calculations yield $\Delta G^{(i)}_{\text{(aq)}}$, $\Delta G^{(i)}_{\text{S, (SRP)}}$ and $\Delta G^{(i)}_{\text{A, (SRP)}}$, which are respectively the FE of aqueous solvation for the SP, the FE associated with transferring the SP from the vacuum state to binding the solvated SRP in the Syn conformation, and the FE associated with transferring the SP from the vacuum state to binding the solvated SRP in the Anti conformation. These quantities are then used to compute the FE of SP binding, $\Delta G^{(i)}_{\text{S/A}} = \Delta G^{(i)}_{\text{S/A, (SRP)}} - \Delta G^{(i)}_{\text{(aq)}}$. In this study, each calculation of a FE difference is performed using the FE perturbation (FEP) method in two stages.

The first stage involves "turning on" the Lennard-Jones (LJ) interactions between the SP and the remaining atoms in the system. The second stage involves turning on the electrostatic interactions between the SP and the remaining atoms. For each stage, independent sampling trajectories are performed with potential energy functions corresponding to a different value of a parameter $\lambda$, such that

$$V(\lambda) = (1 - \lambda)V_0 + \lambda V_1. \tag{2.1}$$

For the first stage, in which the LJ interactions are turned on, $\text{V}_0$ refers to the non-interacting state, and $\text{V}_1$ refers to the state with fully interacting LJ potentials. To ensure the use of smooth potential energy functions during this process [57, 58], the LJ potential energy function is turned on using a $\lambda$-dependent soft-core potential for $V_1$, as implemented in Gromacs 4.5.3 [59] with $\alpha = 0.5$, $\sigma = 0.3$, and $p = 1$; the full LJ potential energy function is recovered for $\lambda = 1$. For the second stage, $\text{V}_0$ refers to the state with only LJ interactions (i.e. the electrostatically non-interacting state), and $\text{V}_1$ refers to the state with both LJ and electrostatic interactions. The first stage utilizes 10 trajectories corresponding to evenly spaced values of $\lambda$ between 0 and 0.09, 30 trajectories corresponding to evenly spaced values of $\lambda$ from 0.1 to 0.245, and 76 trajectories corresponding to evenly spaced values of $\lambda$ from 0.25 to

1. The second stage utilizes 91 trajectories corresponding to evenly spaced values of $\lambda$ from 0 to 0.9 and 20 trajectories corresponding to evenly spaced values of $\lambda$ between 0.91 and 1.

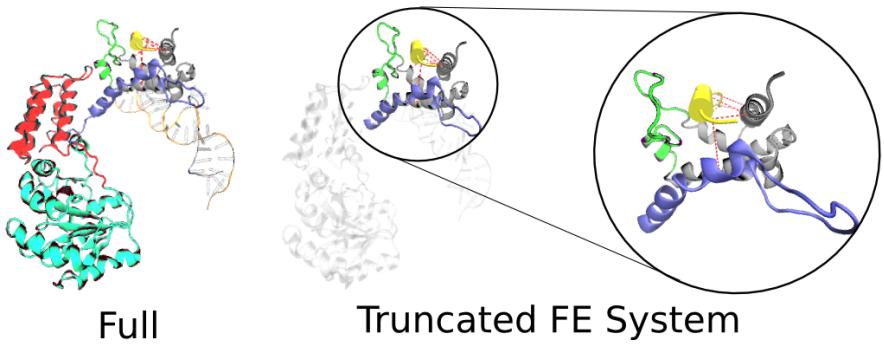



Figure 2.2: Truncation of SRP for FE calculations. The full SRP system includes the N domain (red), G domain (green), linker region (blue), and M domain (gray) with fingerloop (light green) and bound SP (yellow). The truncated system consists of the circled M domain and linker region.

All FE calculations of SP binding to the SRP are performed on a truncated version of the SRP system that consists of the linker region and M-domain (residues 296-432); the NG domain and the RNA are removed to reduce computational cost (Figure 2.2). Consideration of this truncated system is warranted by the fact that all available crystal structures indicate that the SP-binding pocket is fully contained within the combined M-domain and linker region [13, 19]; microsecond simulations of the untruncated system are discussed later. We note that previous experimental studies have focused on the binding of SRP to the full RNC complex whereas the current simulation only considers the binding of the SRP to the SP portion of the RNC complex; the current study thus addresses only the relative impact of the SP on the conformational changes of the SRP.

A truncated octahedral simulation cell is used with periodic boundary conditions. $Na^+$ and $Cl^-$ counterions are included to achieve electroneutrality of the simulation cell at an ion concentration of 100 mM. The total system size varies for the SRP in the Syn and Anti conformation due to different box sizes and different numbers of solvent molecules. For FE simulations of SP binding to the SRP in the Syn conformation, the total system size is 29287 atoms. FE simulations of the SP binding to the SRP in the Anti conformation have a total system size of 39482 atoms. For simulations of the SP in an aqueous environment, the total system size is 9893 atoms.

In all FE calculations, we employ the AMBER03 forcefield [60] and the TIP3P model for water [61].

Two sets of harmonic restraints are applied to limit the sampled configuration space for the SP and SRP molecules in the FE calculations. The first set restrains the SP to the SRP, to avoid dissociation of the SP from the SRP during the course of the long sampling trajectories. Specifically, using the PLUMED plugin [62], each SP is restrained to the SRP by six harmonic restraints between the C$\alpha$ atoms of residues 4, 10, 14, 16, 4, 16 of each SP to the C$\alpha$ atoms of residues L428, M424, E416, E412, A334, and K373 of the SRP M domain, respectively; the force constant of the harmonic restraints is 55.6 kJ mol$^{-1}$ nm$^{-2}$, and the distance of minimum restraint for each pairwise interaction corresponds to the crystal structure distance of the atom pair in the 3KL4 structure [19]. The second set of harmonic restraints limits the ensemble of sampled configurations for the SRP to those associated with either the Syn or Anti structures that are observed in the experimental crystal structures [12, 13]. Specifically, for both the Syn and Anti conformations of the SRP, each C$\alpha$ atom associated with residues in the linker region of the SRP (residues 296-330) is restrained in absolute space to its corresponding position in the experimental crystal structure, using a force constant of 1000 kJ mol$^{-1}$ nm$^{-2}$. For all sampling trajectories associated with a given FE calculation, the applied restraint potentials retain the same minimum and force constant, even as the other molecular interactions are alchemically modified as part of the FEP method. For every sampling trajectory (each of which is associated with a different value of the $\lambda$ parameter in one of the two stages of the FE calculation) the initial configuration of the system is subjected to both relaxation and equilibration before data collection. First, steepest descent minimization [56] is performed to reduce steric clashes to a force tolerance of 100 kJ mol$^{-1}$ nm$^{-1}$. Second, equilibration in the NVT ensemble is performed for 100 ps. Third, for FE calculations with the SRP, we perform 3 ns of equilibration in the NPT ensemble; for FE calculations without the SRP (i.e. aqueous environment), we perform 1 ns of equilibration in the NPT ensemble before data collection. Relaxation, equilibration, and sampling are done within the FEP framework implemented in the Gromacs 4.5.3 package [59]. The system is evolved using Langevin dynamics with a damping constant of 1 ps$^{-1}$ and a temperature of 300 K. The Parrinello-Rahman barostat [63] is applied for pressure coupling at 1 bar. Particle Mesh Ewald (PME) [64, 65] is used to evaluate the electrostatic interactions with a real-space cutoff of 10 Å. The LJ interactions are switched off between 8 Å and 9 Å using a cubic spline and a long-range dispersion correction is applied to the energy and pressure during

the course of the simulation to correct for the cutoff [66]. The simulations are performed with a 2 fs time step with bond distances constrained using LINCS [59]. For FE trajectories associated with turning on the SP interactions in the presence of the SRP, the sampling time for each trajectory ranges from 7-10 ns. For trajectories associated with turning on the SP interactions in the absence of SRP (i.e. aqueous environment) the sampling time is 4 ns. Over 12 $\mu$s of combined simulation time was performed for the FE sampling.

The Bennett acceptance ratio (BAR) method [67] is used to bidirectionally weight the collected data from trajectories sampled in the $i$ and $i+1$ potential, and calculate $\Delta G_{i,i+1}$, the FE difference between the $i$ and $i+1$ states. $\Delta G_{i,i+1}$, is calculated by iteratively solving the following equations

$$\begin{aligned} e^{\beta \Delta G_{i,i+1}} &= \frac{\langle f[-\beta(\Delta U_{i+1,i} - C)]\rangle_{i+1}}{\langle f[\beta(\Delta U_{i,i+1} - C)]\rangle_{i}} e^{(\beta C)}, \\ f(x) &= \frac{1}{1+e^{x}}, \\ C &= \Delta G + \frac{1}{\beta} \ln \frac{N_{i+1}}{N_i}, \end{aligned} \tag{2.2}$$

where $\Delta U_{i,i+1} = U_{i+1}(x_i) - U_i(x_i)$. $U_i(x_i)$ is the potential energy of the configuration $x_i$ sampled with the $i$ potential and evaluated in the $i$ potential. $U_{i+1}(x_i)$ is the potential energy of the configuration $x_i$ sampled with the $i$ potential, but evaluated in the $i+1$ potential. $N_i$ is the number of independently sampled points in trajectory $i$.

For each FE calculation, the convergence of each sampling trajectory and the overlap between neighboring trajectories are examined by plotting the the autocorrelation function (ACF) of $\Delta U_{i,i+1}(t)$ and the relative entropy,

$$D[p_F][p_B] = \int p_F \ln\left(\frac{p_F}{p_B}\right), \tag{2.3}$$

where $p_F = P(\Delta U_{i,i+1})$ and $p_B = P(\Delta U_{i+1,i})$ and where $D = 0$ for identical distributions. In Figure 2.3, we show these convergence and overlap measures for the case of DPAP-WT SP bound to SRP in the Anti conformation, and similar results are seen for all other FE calculations in this study. In Figure 2.3A, we find decorrelation times in the range of 1-900 ns which are used to determine the frequency of sampling for the FE calculations. In all cases the decorrelation time for the individual trajectory is shorter than the total sampling time. The overlap between neighboring trajectories is shown in Figure 2.3B and sufficient windows are added such that $D[p_F][p_B] < 1.5$ to ensure good overlap [68].

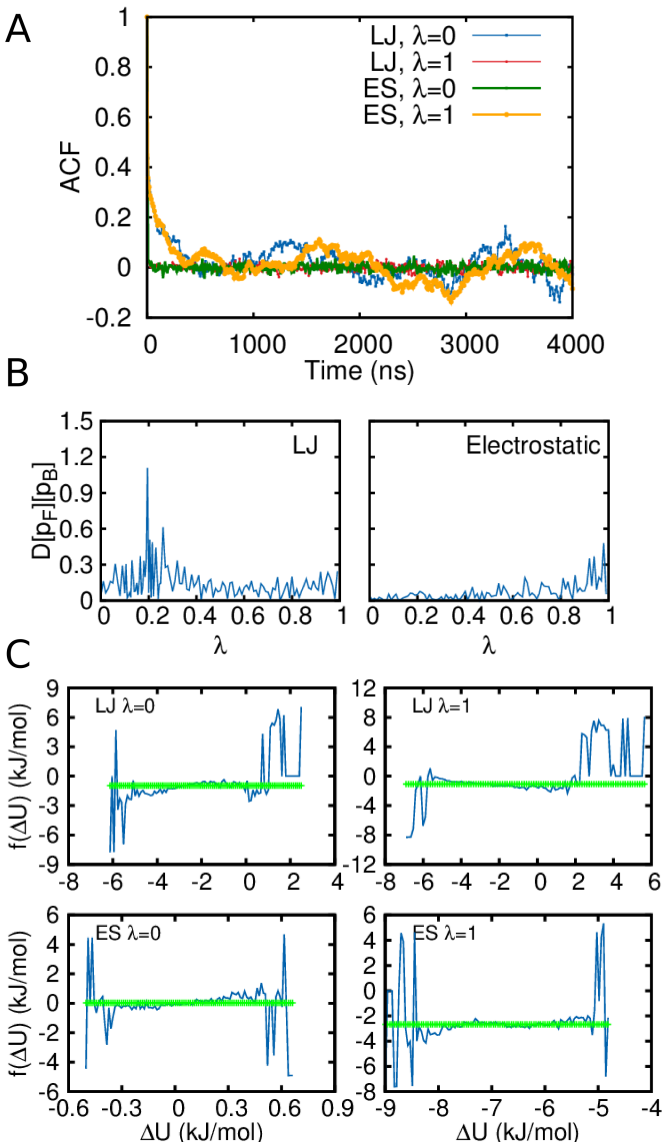



Figure 2.3: A. Autocorrelation of representative sampling trajectories for FE simulations which illustrate decorrelation times in the range of 1-900 ns. These decorrelation estimates are used to determine the frequency of sampling. B. Relative entropy measure, D[p$_F$][p$_B$] (Equation 2.3) between neighboring trajectories for FE simulations is used to determine the spacing of $\lambda$ values for the LJ and electrostatic perturbations such that all neighboring trajectories have a relative entropy less than 1.5 [68]. C. Bennett's overlapping histograms for representative $\Delta G_{i,i+1}$ in FE simulations where $f(\Delta U) = g_b(\Delta U) - g_f(\Delta U)$ is plotted in blue and the BAR result, $\beta\Delta G_{i,i+1}$ is plotted in green. For a region of overlap between the forward and backward $\Delta U$ distribution, $f(\Delta U)$ should be constant and equal to $\beta\Delta G_{i,i+1}$.

To provide an alternate test of the overlap, we use the overlapping histogram method

of Bennett [67] which plots $g_f(\Delta U) - g_b(\Delta U)$, where

$$\begin{aligned} g_f(\Delta U) &= \ln p_F(\Delta U) - \frac{1}{2}\beta\Delta U \\ g_b(\Delta U) &= \ln p_B(\Delta U) + \frac{1}{2}\beta\Delta U. \end{aligned} \tag{2.4}$$

In Figure 2.3C, $g_b(\Delta U) - g_f(\Delta U)$ is compared to the value of $\beta\Delta G$ calculated via the BAR method. For trajectories with sufficient overlap and correct sampling, $g_b(\Delta U) - g_f(\Delta U) = \beta\Delta G$ over a range of $\Delta U$ where $p_F(\Delta U)$ and $p_B(\Delta U)$ have significant overlap. The DPAP-WT sequence is the only SP considered in this study that involves a net charge; the SP sequences phoA and DPAP-K464L are both neutral. Care must be taken when using the PME description of long-range electrostatics to calculate relative free energies that involve the manifestation of a net charge (such as $\Delta G_{\mathrm{S}}^{(\mathrm{DPAP\text{-}WT})}$ and $\Delta G_{\mathrm{A}}^{(\mathrm{DPAP\text{-}WT})}$), since calculations of this kind can be sensitive to the size of the total simulation cell [69]. Although one might expect this effect to be small in simulations as large as those performed here, which range from 29287-39482 atoms in size, we have nonetheless performed two robustness tests to explicitly confirm that such system-size artifacts do not impact the results presented here. In the first robustness test, we utilize the method of Lin *et al.* to compute the leading-order correction to the FE differences due to long-ranged electrostatics (Equation 22 of Ref. [69]), which accounts for the shift in the electrostatic potential due to the finite system size of the simulation cell. The calculated corrections are 1.89 kJ/mol and 1.71 kJ/mol for $\Delta G_{\mathrm{S}}^{(\mathrm{DPAP\text{-}WT})}$ and $\Delta G_{\mathrm{A}}^{(\mathrm{DPAP\text{-}WT})}$, respectively, such that the relative shift in these quantities due to the system-size effect is only 0.18 kJ/mol; since the relative FE of SP binding to the SRP in the Syn versus the Anti conformation is sensitive only on this small relative shift (Equation 2.5), the system-size effect is concluded to be negligible. In the second robustness test, we estimate the error from the constraint imposed on the polarization of the cell as a result of periodic boundary conditions [69]. Using the Delphi 4.0 [70] program, Poisson-Boltzmann (PB) continuum dielectric calculations were performed to estimate the electrostatic FE both with and without periodic boundary conditions. The calculated corrections are 0.26 kJ/mol and 0.22 kJ/mol for $\Delta G_{\mathrm{S}}^{(\mathrm{DPAP\text{-}WT})}$ and $\Delta G_{\mathrm{A}}^{(\mathrm{DPAP\text{-}WT})}$, respectively, such that the relative shift in these quantities due to the system-size effect is only 0.04 kJ/mol. Again, since the relative FE of SP binding to the SRP in the Syn versus the Anti conformation is sensitive only on this small relative shift, the system-size effect associated with the polarization constraint is concluded to be negligible. In summary, the results from both tests suggest that

the conclusions drawn in this paper are robust with respect to possible system-size effects associated with long-range electrostatics.

**Microsecond-timescale trajectories of the untruncated SRP**

Microsecond-timescale trajectories of the SRP system are performed on the Anton supercomputer, a special-purpose machine for long-timescale MD simulations [71, 72]. Unlike the FE calculations, the Anton trajectories are performed without any truncation of the SRP system; the simulations include the full RNA, M, and NG domains of the SRP. Four trajectories are performed as summarized in Table 2.3. Trajectory 1, in which the SRP is initialized from the Syn conformation without a bound SP, is prepared with coordinates from the 1QZW structure [12]. Trajectory 2, in which the SRP is initialized from the Syn conformation with a bound DPAP-WT SP, is prepared with coordinates from the 3KL4 structure [19]. Trajectory 3, in which the SRP is initialized from the Anti conformation without a bound SP, is prepared with coordinates from the 3NDB structure [13] after removal of the model SP. Trajectory 4, in which the SRP is initialized from the Anti conformation with a bound model SP composed of 14 leucine and alanine residues, is prepared from the coordinates of the 3NDB structure [13]. All structures are solvated with explicit water molecules in an orthorhombic simulation cell. $Na^+$ and $Cl^-$ counterions are added to achieve electroneutrality at a salt concentration of 100 mM for a total size of 168500 atoms for the SRP in the Syn conformation and 164294 atoms for the SRP in the Anti conformation. Differences in system size are due to different box sizes and different numbers of solvent molecules. Interactions are described by the CHARMM27 forcefield [73] with TIP3P water.

For each microsecond timescale trajectory, the initial coordinates of the system are equilibrated using the NAMD simulation package [74]. Equilibration of the initial configuration consists of conjugate gradient minimization to reduce steric clashes, thermalization of the system to 300 K, and 10 ns of simulation in the NPT ensemble. The system is evolved using Langevin dynamics with a damping constant of 1 $ps^{-1}$, which also provides temperature control at 300 K. Pressure is maintained at 1 bar using the Nosè-Hoover Langevin piston [75]. Long-range electrostatics are treated using PME with a real space cutoff at 12 Å. LJ interactions are switched off between 10 Å and 12 Å using a cubic spline [74]. A 2 fs time step is used throughout, and all bond lengths are constrained using the SHAKE algorithm [76].

Following equilibration, production trajectories are performed on the Anton system.

Harmonic restraints are introduced to the phosphate backbone of the SRP RNA with a force constant of 418.4 kJ mol$^{-1}$Å$^{-2}$ to prevent conformational rearrangements of the SRP RNA domain and to avoid overall rotations of the complex with respect to the simulation cell that might cause interactions with its periodic images. Molecular dynamics are evolved with the RESPA numerical integration scheme with a 2 fs timestep [77]. The Berendsen coupling scheme maintains the temperature at 300 K and pressure at 1 bar [78]. Long-range electrostatics are treated using the k-space Gaussian Split Ewald method [79] with a real space cutoff of 9.45 Å. LJ interactions are cut off at 9.48 Å. All bond lengths are constrained using the M-SHAKE algorithm [80].

## 2.3 Results and Conclusions

### Coupling between SP binding and SRP conformation

FEP calculations are used to investigate the potential role of SP binding in driving conformational changes in the SRP. For each SP, we consider $\Delta\Delta G^{(i)}$, the relative FE of SP binding to the SRP in the Syn versus the Anti conformation.

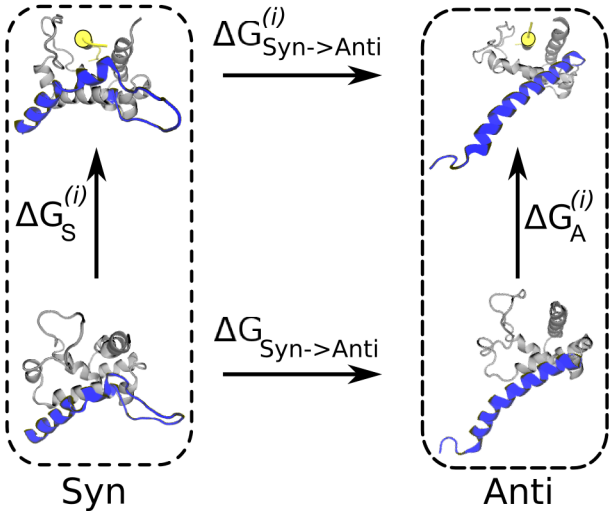



Figure 2.4: A thermodynamic cycle that enables the quantification of coupling between SP binding and SRP linker conformation. Four states are defined by linker conformation (blue) and presence or absence of SP (yellow). $\Delta\Delta G^{(i)}$ is calculated by taking the difference of $\Delta G^{(i)}_{\mathrm{S}}$ and $\Delta G^{(i)}_{\mathrm{A}}$ in the dotted boxes. This is equivalent to the difference of $\Delta G^{(i)}_{\mathrm{Syn}\rightarrow\mathrm{Anti}}$ and $\Delta \mathrm{G}_{\mathrm{Syn}\rightarrow\mathrm{Anti}}$ shown along the horizontal arrows.

As is schematically illustrated using the thermodynamic cycle in Figure 2.4, we

| $i$ | $\Delta G^{(i)}_{(\mathrm{aq})}$ | $\Delta G^{(i)}_{\mathrm{S,\,(SRP)}}$ | $\Delta G^{(i)}_{\mathrm{A,\,(SRP)}}$ | $\Delta G^{(i)}_{\mathrm{S}}$ | $\Delta G^{(i)}_{\mathrm{A}}$ | $\Delta\Delta G^{(i)}$ |
|---|---|---|---|---|---|---|
| DPAP-WT | -310 (2) | -389 (2) | -400 (3) | -79 (3) | -90 (3) | -11 (4) |
| DPAP-K464L | -275 (2) | -340 (2) | -355 (2) | -65 (3) | -80 (3) | -16 (7) |
| phoA | -287 (2) | -326 (2) | -314 (2) | -39 (3) | -27 (3) | 13 (4) |

Table 2.2: Summary of FE calculation results in units of kJ/mol.

calculate $\Delta\Delta G^{(i)}$ for a given SP $i$ using

$$\Delta\Delta G^{(i)} = \Delta G^{(i)}_{\mathrm{A}} - \Delta G^{(i)}_{\mathrm{S}}, \tag{2.5}$$

where $\Delta G^{(i)}_{\mathrm{A}}$ and $\Delta G^{(i)}_{\mathrm{S}}$ are the respective binding FEs of the SP to the Syn and Anti conformations of the SRP. This construction of the thermodynamic cycle avoids the direct calculation of the FE difference between the SRP in the Syn and Anti conformation, illustrated by the horizontal arrows in Figure 2.4.

As is explained in the Methods section, harmonic restraints are employed in the FE calculations to limit the ensemble of sampled configurations for the SRP to those associated with either the Syn or Anti conformations that are observed in the experimental crystal structures [12, 13]. We note that a more common implementation of the thermodynamic cycle approach is to first assume that the different conformations of the biomolecule correspond to basins of stability that are separated by a FE barrier, and then to fully sample the configuration space associated with those basins of stability. However, well-defined FE differences can be computed between any regions of configuration space, and in the current study, we focus on calculating the FE differences between configurations in the vicinity of the experimental crystal structure for SRP in the Syn conformation and configurations in the vicinity of the experimental crystal structure for SRP in the Anti conformation. The advantages of this approach are two-fold: *(i)* it enables us to specifically address the effect of SP binding on the relative stability of SRP in its experimentally observed conformations, and *(ii)* by using the restraints to avoid sampling the slow degrees of freedom associated with the flexible linker domain of the SRP, it enables us to rigorously converge all of the reported FE calculations that are reported in this study.

The primary quantity of interest in this analysis is $\Delta\Delta G^{(i)}$, which reports on the degree to which binding of the SP impacts the conformation of the SRP linker region. The results for the three considered SPs are shown in Table 2.2. For DPAP-WT, a known targeting sequence, we calculate a value of -11 kJ/mol, which indicates substantial coupling of SP binding to the SRP linker region. Furthermore, the negative

sign of $\Delta\Delta G^{(i)}$ for the case of DPAP-WT indicates that SP binding leads to a significant shift in the SRP conformational distribution towards the Anti conformation, which is consistent with the model hypothesis (Figure 2.1C).

The computed value of $\Delta\Delta G^{(i)}$ for DPAP-K464L suggests that a single lysine-to-leucine mutation from the DPAP-WT sequence leads to a modest increase in the induced conformational bias in the SRP distribution. The calculated $\Delta\Delta G^{(i)}$ is -16 kJ/mol, which is greater in magnitude than the conformational preference of the DPAP-WT SP, although this difference is within a standard deviation of the statistical error. The DPAP-K464L sequence was chosen because the lysine-to-leucine mutation at residue 464 removes a charged residue and extends the hydrophobic core of the SP. Previous experimental studies [54, 81] have shown that increased hydrophobicity leads to more efficient targeting, and our results are thus consistent with the interpretation that this increased hydrophobicity also leads to a greater conformational preference for the Anti conformation of the SRP.

Finally, we consider the phoA SP, which is known to successfully bind to the SRP but which provides inefficient targeting along the co-translational pathway [10, 30, 55]. We find (Table 2.2) that binding of the phoA SP leads to a reversed shift in the SRP conformational distribution in comparison with DPAP-WT and DPAP-K464L. The calculated $\Delta\Delta G^{(i)}$ is 13 kJ/mol, indicating that binding of phoA SP stabilizes the Syn conformation of the SRP relative to the Anti conformation. This finding is again in agreement with the model hypothesis in which binding of phoA fails to induce the SRP conformational change that would lead to efficient targeting.

The calculated $\Delta\Delta G^{(i)}$ values demonstrate that SP binding to the SRP is coupled to the linker region and its effects on the SRP conformational distribution are dependent on the sequence of the SP. For all three considered SPs, the degree to which we find that SP binding induces conformational changes in the SRP is fully consistent with the hypothesis that the SRP targeting pathway is triggered by SP binding that induces conformational changes in the SRP from the Syn to the Anti conformation. Furthermore, the results explain the counter-intuitive experimental observation that SPs with favorable binding to the SRP need not lead to efficient downstream protein targeting [41]; indeed, it is seen here that favorable phoA binding induces a shift towards the Syn conformation of SRP, which hinders the kinetics of complex formation between SRP and its receptor. We additionally note that this connection between SRP conformational shifts upon substrate binding provides a mechanism for SP identity to influence the kinetics of processes at large distances from the SP

binding site.

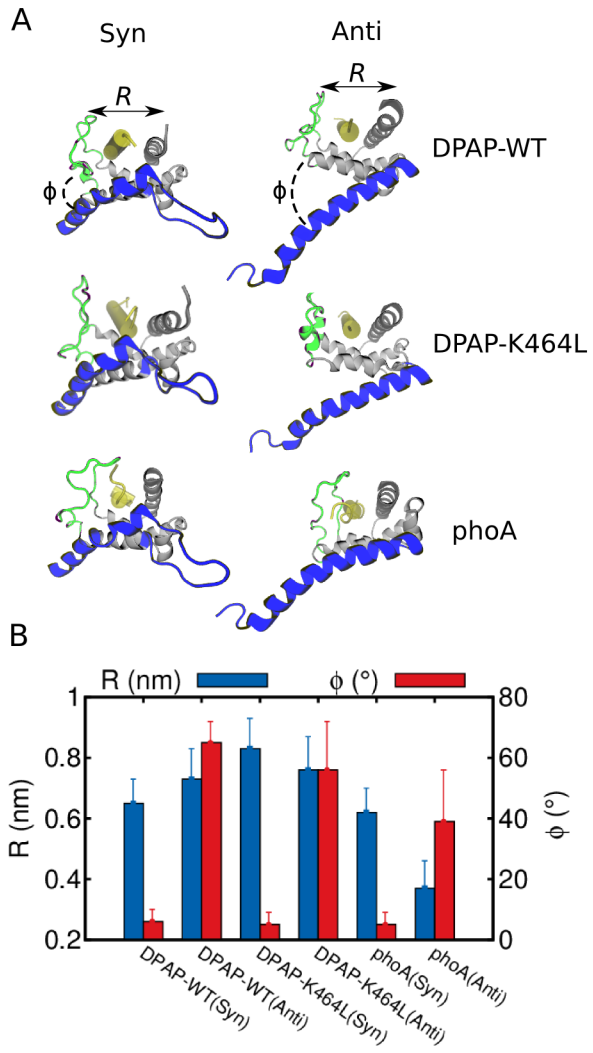



Figure 2.5: A. Ensemble-averaged structures from SP-bound sampling trajectories with the linker region shown in blue, the fingerloop shown in green and the SP shown in yellow. B. Fingerloop displacement $R$ and the relative orientation of the linker and $\alpha$M1 helix $\phi$ are shown for each SP and SRP pair, revealing the structural differences associated with binding efficient and inefficient SPs.

FE calculations not only provide a quantitative measure of the extent of allosteric coupling between SP binding and SRP linker conformation, but the sampling trajectories also provide details of the structural differences in the SRP upon binding of the different SPs. For each SP considered, Figure 2.5A shows the ensemble-averaged configuration of the SRP in the Syn and Anti conformations. In each case,

the fingerloop of the SRP reorganizes to accommodate binding of the SP. The position of the fingerloop, measured in terms of its center-of-mass distance $R$ from the fifth helix of the M-domain ($\alpha$M5) helix at the opposite side of the SP binding pocket, is found to be displaced by 4-9 Å upon SP binding (Figure 2.5A). This is in agreement with previous structural and biochemical work [28, 40], which suggests that SP binding displaces the fingerloop and that fingerloop flexibility is crucial for enabling the SRP to bind to SPs of differing sequences.

The three SPs considered here differ substantially in terms of their size, with DPAP-WT and DPAP-K464L excluding more volume than phoA (Table 2.1). These differences are reflected in the conformation of the SRP bound to the SP, in terms of both the degree of fingerloop displacement and the relative orientation between the linker region and M-domain. These conformational differences are shown in Figure 2.5B, where the relative orientation is monitored by the dihedral angle $\phi$ between the helical axis of the linker and the axis of the first helix of the M-domain ($\alpha$M1), as well as by the fingerloop displacement, $R$. For each SP, the observed dihedral angle associated with binding in the Syn conformation is small. For the cases of DPAP-WT and DPAP-K464L bound to the SRP in the Anti conformation, the value of $\phi$ is much larger than in the Syn conformation due to the loss of contact between the linker and the M-domain; the smaller phoA SP likewise exhibits an increased value of $\phi$ in the Anti conformation, but to a lesser extent. A more striking trend is observed in the fingerloop displacement, $R$. For both of the efficiently targeted sequences, DPAP-WT and DPAP-K464L, the fingerloop is substantially displaced in both the Syn and Anti conformations, whereas the inefficiently targeted phoA SP exhibits a reduced displacement of the fingerloop in the Anti conformation. The differing degree to which the efficient and inefficient SPs impact these structural features of the SRP suggests a mechanistic basis for the observed correlations between SP binding and SRP conformational changes (Table 2.2).

Although the single amino-acid mutation that distinguishes the DPAP-WT and DPAP-K464L SPs appears to have a relatively small impact on the corresponding values of $\Delta\Delta G^{(i)}$ (Table 2.2), and although the binding of these two SPs leads to similar structural changes in the SRP fingerloop displacement and helix orientation (Figure 2.5B), we do note that the binding of these two SPs leads to differences in other structural quantities. In DPAP-WT, the positive charge interacts with the negatively charged residues of the fingerloop (D364, E365) in the Syn conformation, but this interaction is not observed in the Anti conformation; the DPAP-K464L mu-

| Trajectory | Initial SRP Configuration | SP | Time ($\mu s$) |
|---|---|---|---|
| T1 | Syn | no SP | 3.6 |
| T2 | Syn | with SP | 4.5 |
| T3 | Anti | no SP | 4.3 |
| T4 | Anti | with SP | 3.4 |

Table 2.3: Summary of microsecond trajectories.

tant lacks the positive charge for this interaction with the Syn conformation. These different interactions may explain the somewhat stronger bias of the DPAP-K464L mutant for the Anti conformation of the SRP upon binding. Regardless, these results illustrate that specific amino-acid interactions can play a role in determining the conformational shift of the SRP upon SP binding.

**Large-scale conformational changes from microsecond trajectories**

Microsecond trajectories are performed with the Anton supercomputer [71] to investigate the conformational dynamics of the SRP (Table 2.3).

Figure 2.6 plots the fingerloop displacement, $R$, and the dihedral angle between the helical axis of the linker and $\alpha$M1 helix, $\phi$, as a function of simulation time for all four trajectories; these geometric quantities were previously introduced in the discussion of the FE sampling trajectories that involved the truncated SRP system (Figure 2.5). Comparison of Figures 2.5 and 2.6 reveals similar structural responses of the SRP to binding of the various SPs, confirming that $R$ and $\phi$ are relatively local geometric features that are unchanged by inclusion of the NG domain. For Trajectories T1 and T3, neither of which include the bound SP, it is seen in Figure 2.6A that the fingerloop occludes the binding pocket, exhibiting a small value for the displacement distance. For Trajectories T2 and T4, both of which include the bound SP, it is seen that the fingerloop is significantly displaced (4-9 Å) to accommodate the SP; these results are consistent with the observations in Figure 2.5 that SP binding displaces the fingerloop from the binding pocket. Similarly, for Trajectories T1 and T2, both of which are initialized in the Syn conformation, it is seen in Figure 2.6B that the SRP undergoes relaxation to adopt configurations with values of $\phi$ that are consistent with those observed in the equilibrium FE sampling trajectories for SRP in the Syn conformation (Figure 2.5). For Trajectories T3 and T4, both of which are initialized in the Anti conformation, it is seen in Figure 2.6B that the SRP adopts configurations with values of $\phi$ that are consistent with the loss of contact

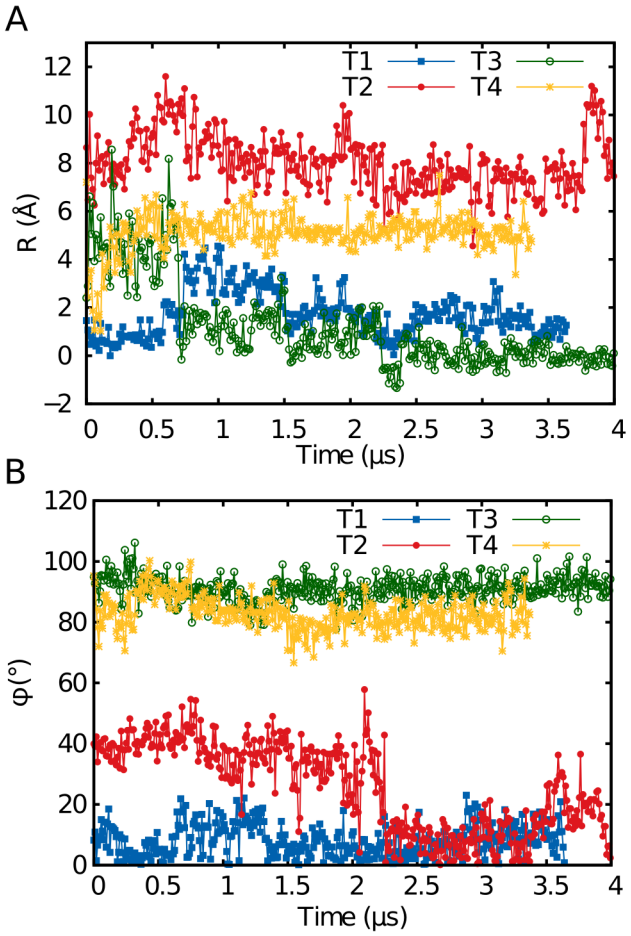



Figure 2.6: A. Fingerloop displacement $R$ as a function of time in the microsecond-timescale MD trajectories. B. The relative orientation of the linker and $\alpha$M1 helix $\phi$ as a function of time in the microsecond-timescale MD trajectories.

between the linker region and the M-domain, which is again consistent with the structural features that are observed for the equilibrium FE sampling trajectories for SRP in the Anti conformation (Figure 2.5).

To further characterize the conformational changes observed in the microsecond-timescale MD trajectories, Figure 2.7 shows snapshots of the system at various times along the trajectories. In addition to illustrating the changes in $R$ and $\phi$ that were discussed previously, Figure 2.7 also illustrates the process by which the fingerloop collapses to occupy the SRP binding pocket in the absence of the bound SP (Trajectory T3).

We now analyze the microsecond-timescale MD trajectories in the context of a re-

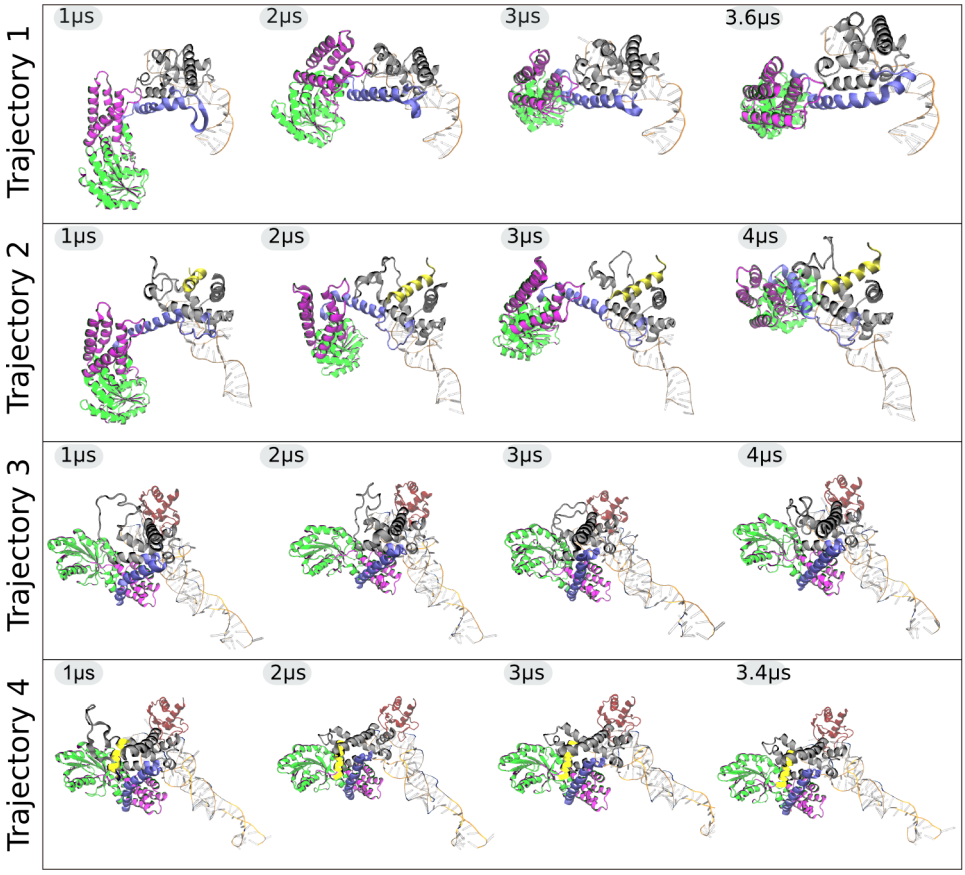



Figure 2.7: Snapshots of full-system SRP trajectories shown at $\mu$s intervals. All structures were aligned by the SRP RNA, shown in orange, with M domain, linker region, N and G domains shown in gray, blue, purple and green respectively. Trajectories T2 and T4 have bound SP shown in yellow.

cent study by Shen *et al.*, which used single-molecule Förster resonance energy transfer (smFRET) experiments to monitor the SRP conformational dynamics [18]. This experimental study labeled the SRP RNA distal end and NG domain and found that the SRP samples both a low-efficiency and a high-efficiency FRET state, with differences in FRET efficiency that correspond to distance changes of approximately 10 nm. Figure 2.8 presents results from the four microsecond-timescale MD simulations of the SRP, plotting the distance between the G domain and RNA corresponding to the donor-acceptor distance measured in the smFRET experiments. Trajectories T1 and T2, which are initialized from the Syn conformation of the SRP, show 6-8 nm changes in donor-acceptor distance during the course of the simulation. Trajectory T2, which has a bound SP, has greater fluctuations in this coordinate

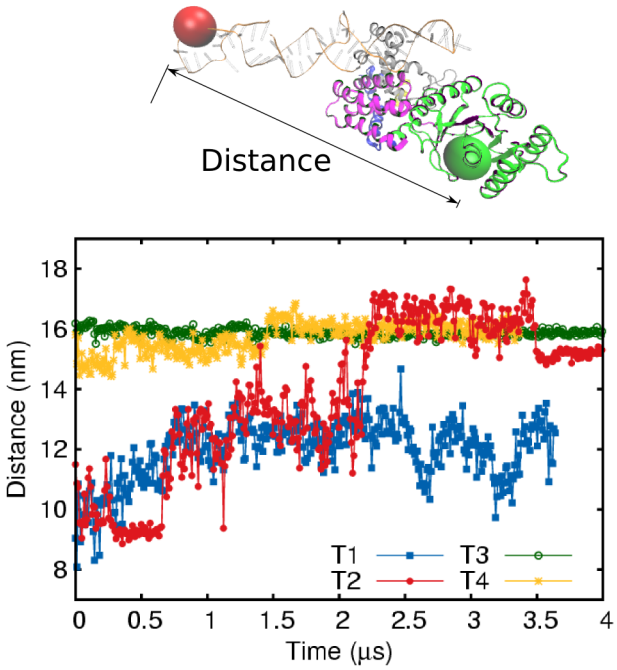



Figure 2.8: Distance between the FRET donor (green) and acceptor (red) positions as a function of time in the microsecond-timescale MD trajectories.

than T1, which does not have a bound SP. Trajectories T3 and T4, which are initialized from the Anti conformation of the SRP, exhibit little deviation from the initial distance. These results support the conclusion that nanometer-lengthscale changes in SRP conformation are intrinsic features of the SRP, although we do not suggest that the conformational states observed in these trajectories correspond directly to the low-FRET and high-FRET states reported by Shen *et al.*

Interestingly, the trajectories reveal that nanometer-lengthscale changes in distance occur via the rigid-body movement of M and NG domains connected by the flexible linker region, while the conformation of each individual domain is unchanged. In Figure 2.9A, for each residue pair $i$ and $j$ in the SRP, we plot $\delta_{ij}(t) = |d_{ij}(0) - d_{ij}(t)|$, the difference in the distance $d_{ij}$ between the two C$\alpha$ atoms of each residue at the initial configuration of the trajectory and at various subsequent times $t$ along Trajectory T2. This metric reveals only small changes in distance between atom pairs that both belong to the M domain or that both belong to the NG domain, which correspond to the indicated diagonal blocks in Figure 2.9A. Large distance changes are only observed in atom pairs for which one atom is in the M domain and the other is in the NG domain, which correspond to the off-diagonal blocks. This

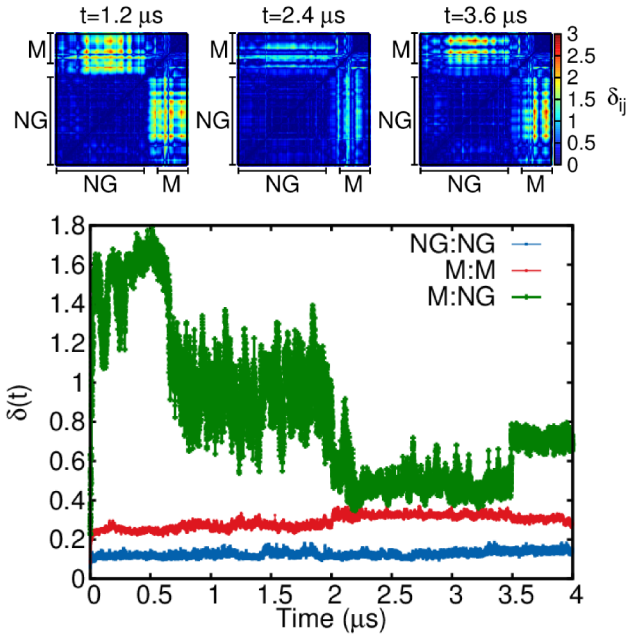



Figure 2.9: A. For each residue pair $i$ and $j$ in the SRP, the difference in the distance $\delta_{ij}$ between the two C$\alpha$ atoms of each residue at the initial configuration of the trajectory and at various subsequent times along trajectory T2. B. The average of $\delta_{ij}$ over NG-NG pairs, M-M pairs and M-NG pairs.

suggests that the observed changes in conformation and donor-acceptor distance are the result of changes in the relative orientation of the M and NG domains, while the M and NG domains individually evolve as rigid bodies. Figure 2.9B further illustrates this point by plotting the average of $\delta_{ij}$ over NG-NG pairs, M-M pairs and M-NG pairs along the trajectory. It is clear that only the M and NG domains undergo large changes in relative orientation in Trajectory T2. The same conclusion holds for the other microsecond trajectories that were performed in this study.

The microsecond-timescale dynamics revealed in these simulations suggests a mechanism for coordinating the spatial and temporal organization between independent functional domains. In SRP, SP binding occurs in the M domain, while downstream targeting events, such as dimerization of the SRP with its receptor and GTPase activity of the SRP involve the NG domain of the SRP [82]. From the earlier FE simulations (Figure 2.4 and Table 2.2), it is evident that SP binding leads to sequence-specific changes in the conformational distribution in the flexible linker domain; Figure 2.9 further reveals that such changes in the linker domain manifest as large lengthscale changes in the relative orientation of the M and NG domains

of the SRP, thus coordinating SP binding at the M domain with downstream steps involving the NG domain.

## 2.4 Conclusions

The fidelity of co-translational protein targeting and recognition is a remarkable feature of cellular signaling pathways that demands understanding from the perspective of molecular interactions. The central component of co-translational targeting, the SRP, is capable of binding a diverse set of SPs while differentiating between sequences with similar biochemical features. In this work, we have demonstrated allosteric coupling between SP binding and conformational changes of the linker region of the SRP. This coupling is sensitive to the sequence identity of the SP, thus providing a mechanism for conferring SP sequence specificity in the signaling pathway. Furthermore, microsecond-timescale simulations reveal that the SRP undergoes large-scale conformational changes that are characterized by rigid-body motion of individual domains connected by a flexible linker, which enables small conformational changes of the linker upon SP binding to be conferred onto larger $\sim$10 nm lengthscales.

## References


(1) Akopian, D.; Shen, K.; Zhang, X.; Shan, S.-o. *Annu. Rev. Biochem.* **2013**, *82*, 693–721.

(2) Pohlschröder, M.; Prinz, W. A.; Hartmann, E.; Beckwith, J. *Cell* **1997**, *91*, 563–566.

(3) Walter, P.; Ibrahimi, I.; Blobel, G. *J.Cell Biol.* **1981**, *91*, 545–550.

(4) Rapoport, T. A.; Jungnickel, B.; Kutay, U. *Annu. Rev. Biochem.* **1996**, *65*, 271–303.

(5) Powers, T.; Walter, P. *EMBO J.* **1997**, *16*, 4880–4886.

(6) Rapoport, T. A. *Nature* **2007**, *450*, 663–669.

(7) Driessen, A. J. M.; Nouwen, N. *Annu. Rev. Biochem.* **2008**, *77*, 643–667.

(8) Zhang, B.; Miller III, T. F. *J. Am. Chem. Soc.* **2012**, *134*, 13700–13707.

(9) Zhang, B.; Miller III, T. F. *Cell Rep.* **2012**, *2*, 927–937.

(10) Gierasch, L. M. *Biochemistry* **1989**, *28*, 923–930.

(11) Von Heijne, G. *J. Mol. Biol.* **1985**, *184*, 99–105.

(12) Rosendal, K. R.; Wild, K.; Montoya, G.; Sinning, L. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 14701–14706.

(13) Hainzl, T.; Huang, S. H.; Merilainen, G.; Brannstrom, K.; Sauer-Eriksson, A. E. *Nat. Struct. Mol. Biol.* **2011**, *18*, 389–391.

(14) Halic, M.; Gartmann, M.; Schlenker, O.; Mielke, T.; Pool, M. R.; Sinning, I.; Beckmann, R. *Science* **2006**, *312*, 745–747.

(15) Schaffitzel, C.; Oswald, M.; Berger, I.; Ishikawa, T.; Abrahams, J. P.; Koerten, H. K.; Koning, R. I.; Ban, N. *Nature* **2006**, *444*, 503–506.

(16) Estrozi, L. F.; Boehringer, D.; Shan, S.-o.; Ban, N.; Schaffitzel, C. *Nat. Struct. Mol. Biol.* **2011**, *18*, 88–90.

(17) Holtkamp, W.; Lee, S.; Bornemann, T.; Senyushkina, T.; Rodnina, M. V.; Wintermeyer, W. *Nat. Struct. Mol. Biol.* **2012**, *19*, 1332–1337.

(18) Shen, K.; Arslan, S.; Akopian, D.; Ha, T.; Shan, S.-o. *Nature* **2012**, *492*, 271–275.

(19) Janda, C. Y.; Li, J; Oubridge, C; Hernandez, H; Robinson, C. V.; Nagai, K *Nature* **2010**, *465*, 507–510.

(20) Hainzl, T.; Huang, S. H.; Sauer-Eriksson, A. E. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 14911–14916.

(21) Ataide, S. F.; Schmitz, N.; Shen, K.; Ke, A.; Shan, S.-o.; Doudna, J. A.; Ban, N. *Science* **2011**, *331*, 881–886.

(22) Wild, K.; Rosendal, K. R.; Sinning, I. *Mol. Microbiol.* **2004**, *53*, 357–363.

(23) Wild, K.; Bange, G.; Bozkurt, G.; Segnitz, B.; Hendricks, A.; Sinning, I. *Acta Crystallogr. D* **2010**, *66*, 295–303.

(24) Noriega, T. R.; Tsai, A.; Elvekrog, M. M.; Petrov, A.; Neher, S. B.; Chen, J.; Bradshaw, N.; Puglisi, J. D.; Walter, P. *J. Biol. Chem.* **2014**, *289*, 19294–19305.

(25) Kurzchalia, T.; Wiedmann, M; Girshovich, A.; Bochkareva, E.; Bielka, H; Rapoport, T. *Nature* **1986**, *320*, 634–636.

(26) Poritz, M. A.; Bernstein, H. D.; Strub, K.; Zopf, D.; Wilhelm, H.; Walter, P. *Science* **1990**, *250*, 1111–1117.

(27) Freymann, D. M.; Keenan, R. J.; Stroud, R. M.; Walter, P. *Nature* **1997**, *385*, 361–364.

(28) Keenan, R. J.; Freymann, D. M.; Walter, P.; Stroud, R. M. *Cell* **1998**, *94*, 181–191.

(29) Batey, R. T.; Rambo, R. P.; Lucast, L.; Rha, B.; Doudna, J. A. *Science* **2000**, *287*, 1232–1239.

(30) Zhang, X.; Rashid, R.; Wang, K.; Shan, S.-o. *Science* **2010**, *328*, 757–760.

(31) Egea, P. F.; Stroud, R. M.; Walter, P. *Curr. Opin. Struct. Biol.* **2005**, *15*, 213–220.

(32) Focia, P. J.; Shepotinovskaya, I. V.; Seidler, J. A.; Freymann, D. M. *Science* **2004**, *303*, 373–377.

(33) Peluso, P.; Shan, S.-o.; Nock, S.; Herschlag, D.; Walter, P. *Biochemistry* **2001**, *40*, 15224–15233.

(34) Siu, F. Y.; Spanggord, R. J.; Doudna, J. A. *RNA* **2007**, *13*, 240–250.

(35) Bradshaw, N.; Neher, S. B.; Booth, D. S.; Walter, P. *Science* **2009**, *323*, 127–130.

(36) Zhang, X.; Schaffitzel, C.; Ban, N.; Shan, S.-o. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 1754–1759.

(37) Zhang, X.; Lam, V. Q.; Mou, Y.; Kimura, T.; Chung, J.; Chandrasekar, S.; Winkler, J. R.; Mayo, S. L.; Shan, S.-o. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 6450–6455.

(38) Shen, K.; Zhang, X.; Shan, S.-o. *RNA* **2011**, *17*, 892–902.

(39) Von Loeffelholz, O.; Knoops, K.; Ariosa, A.; Zhang, X.; Karuppasamy, M.; Huard, K.; Schoehn, G.; Berger, I.; Shan, S.-o.; Schaffitzel, C. *Nat. Struct. Mol. Biol.* **2013**, *20*, 604–610.

(40) Ariosa, A. R.; Duncan, S. S.; Saraogi, I.; Lu, X.; Brown, A.; Phillips, G. J.; Shan, S.-O. *Mol. Biol. Cell* **2013**, *24*, 63–73.

(41) Saraogi, I.; Akopian, D.; Shan, S.-o. *J. Cell Biol.* **2014**, *205*, 693–706.

(42) Jorgensen, W. L.; Ravimohan, C *J. Chem. Phys.* **1985**, *83*, 3050–3054.

(43) Bash, P.; Singh, U.; Langridge, R; Kollman, P. *Science* **1987**, *236*, 564–568.

(44) Rao, S. N.; Singh, U. C.; Bash, P. A.; Kollman, P. A. *Nature* **1987**, *328*, 551–554.

(45) Roux, B.; Nina, M.; Pomès, R.; Smith, J. *Biophys. J.N* **1996**, *71*, 670 –681.

(46) Essex, J. W.; Severance, D. L.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **1997**, *101*, 9663–9669.

(47) Woo, H.-J.; Roux, B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6825–6830.

(48) Jayaram, B.; Sprous, D.; Young, M. A.; Beveridge, D. L. *J. Amer. Chem. Soc.* **1998**, *120*, 10629–10633.

(49) Satpati, P.; Clavaguéra, C.; Ohanessian, G.; Simonson, T. *The Journal of Physical Chemistry B* **2011**, *115*, 6749–6763.

(50) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics Modell.* **1996**, *14*, 33–38.

(51) Larkin, M.; Blackshields, G.; Brown, N.; Chenna, R; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J.; Gibson, T.; Higgins, D. *Bioinformatics* **2007**, *23*, 2947–2948.

(52) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M.; Eramian, D.; Shen, M.-y.; Pieper, U.; Sali, A. In *Curr. Protoc. Bioinf.* 2006.

(53) MacKerell, A. D. et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(54) Ng, D.; Brown, J. D.; Walter, P. *J. Cell Biol.* **1996**, *134*, 269–278.

(55) Huber, D.; Boyd, D.; Xia, Y.; Olma, M. H.; Gerstein, M.; Beckwith, J. *J Bacteriol.* **2005**, *187*, 2983–2991.

(56) Fletcher, R.; Powell, M. J. *Comput. J.* **1963**, *6*, 163–168.

(57) Beutler, T. C.; Mark, A. E.; Schaik, R. C. V.; Gerber, P. R.; Gunsteren, W. F. V. **1994**, *222*, 529–539.

(58) Steinbrecher, T.; Mobley, D. L.; Case, D. A. *J. Chem. Phys.* **2007**, *127*, 214108.

(59) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(60) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(61) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(62) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. *Comput. Phys. Comm.* **2009**, *180*, 1961–1972.

(63) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(64) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(65) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(66) Van Gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 992–1023.

(67) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245–268.

(68) Pohorille, A.; Jarzynski, C.; Chipot, C. *J. Phys. Chem. B* **2010**, *114*, 10235–10253.

(69) Lin, Y.-L.; Aleksandrov, A.; Simonson, T.; Roux, B. *J Chem. Theory Comput.* **2014**, *10*, 2690–2709.

(70) Rocchia, W.; Alexov, E.; Honig, B. *J. Phys. Chem. B* **2001**, *105*, 6507–6514.

(71) Shaw, D. E. et al. *SIGARCH Comput. Archit. News* **2007**, *35*, 1–12.

(72) Shaw, D. et al. In *High Performance Computing Networking, Storage and Analysis, Proceedings of the Conference on*, 2009, pp 1–11.

(73) Foloppe, N.; MacKerell Jr, A. D. *J. Comput. Chem.* **2000**, *21*, 86–104.

(74) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(75) Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613–4621.

(76) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. *J. Comput. Phys.* **1977**, *23*, 327–341.

(77) Tuckerman, M; Berne, B. J.; Martyna, G. J. *J. Chem. Phys.* **1992**, *97*, 1990–2001.

(78) Berendsen, H. J.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(79) Shan, Y.; Klepeis, J. L.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. *J. Chem. Phys.* **2005**, *122*, 054101.

(80) Kräutler, V.; van Gunsteren, W. F.; Hünenberger, P. H. *J. Comput. Chem.* **2001**, *22*, 501–508.

(81) Hatsuzawa, K.; Tagaya, M.; Mizushima, S. *J Biochem.* **1997**, *121*, 270–277.

(82) Egea, P. F.; Shan, S.-o.; Napetschnig, J.; Savage, D. F.; Walter, P.; Stroud, R. M. *Nature* **2004**, *427*, 215–221.

*Chapter 3*

# DEVELOPMENT OF A COARSE-GRAINED SIMULATION MODEL TO STUDY THE ENERGETICS OF MEMBRANE PROTEIN INSERTION AND TOPOLOGY

## 3.1 Introduction

Co-translational protein targeting delivers the ribosome-nascent chain complex (RNC) to the Sec translocon for membrane integration and translocation of the protein [1]. Structural studies have revealed that the RNC proceeds to dock upon the translocon and align its nascent chain exit tunnel with the translocon channel [2–11]. The central part of the Sec translocon is a conserved protein-conducting channel that spans the lipid bilayer and allows both translocation across the membrane and integration of protein domains into the lipid bilayer through a unique lateral opening [12]. The nascent chain is then fed into the channel and oriented to establish the topology of the protein through translocation of either its N-terminus or C-terminus to give rise to type I or type II topology respectively [13] as shown in Figure 3.1.

Despite the ability to assay the final topology of the nascent chain (NC) via biochemical techniques, the mechanism by which the final topologies are reached is still not well understood. In particular, the type II topology requires that the N-terminus be retained in the cytosol, while the C-terminal end is translocated through the channel. This requires a flipping of the nascent chain since the N-terminus is able to contact the channel before the C-terminus [14]. Several groups have hypothesized a step-wise inversion process through a type I intermediate [14–17]; while others hypothesize a looping mechanism [11, 18–21]. Furthermore, many factors have been found to influence the final topology including N-terminal length and folding [16, 22–24], C-terminal length [25, 26], position of charged residues [27–30], translation rate [26] and hydrophobicity [26, 31]. Further complicating matters is the fact that experimental methods which directly interrogate the intermediate stages of the process often use stalled sequences [15, 17], which may not reflect on-pathway intermediates when comparing to the actively translating system *in vivo*.

In this work, we extend a coarse-grained model for protein translocation, previously developed in the Miller group [32], to include the full dimensionality of the system. Furthermore a general mapping procedure that allows for simulation of an arbitrary

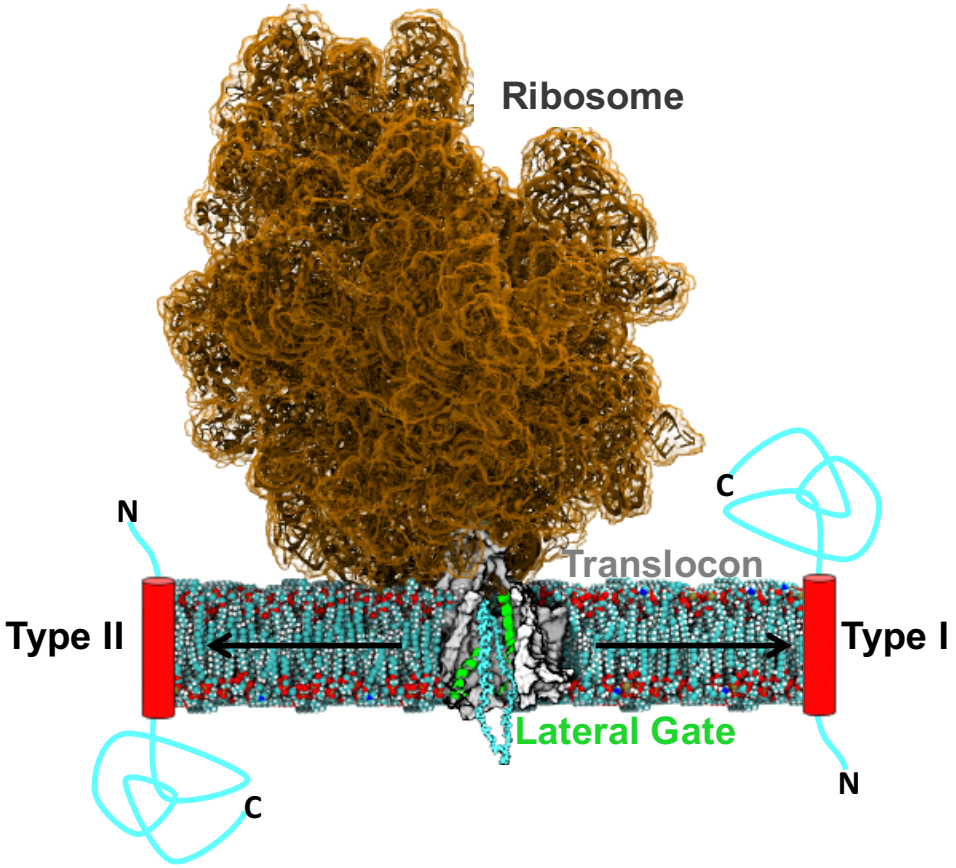



Figure 3.1: Structure of the *E. Coli* ribosome-nascent chain complex docked onto the translocon [7]. The ribosome is shown in brown, the translocon is shown in gray and the nascent chain is shown in light blue. The nascent chain is interacting with the translocon lateral gate helices shown in green. A schematic of the two possible integrated topologies, type I and type II, is shown to the right and left respectively.

amino-acid sequence is introduced that allows for simulations of nonnatural sequences to probe the effects of N-terminal length, flanking charges, and hydrophobicity that have been experimentally observed. The model developed here is shown to reproduce experimentally known trends in membrane integration efficiency [33] and topogenesis [26] validating its accuracy for further applications. We then develop a framework to calculate the binding free energy of the RNC to the translocon as a function of chain length. Previous structural studies have suggested that alternation between tightly bound and loosely bound ribosome-translocon complexes are needed for proper integration [6]. This framework allows for efficient sampling of different topologies and calculation of free energies in a defined topology.

## 3.2 Methods

### 3D model description

In the spirit of a previously developed model of dynamics at the Sec translocon [32], in which all system coordinates are projected onto a two-dimensional (2D) plane, we have develop a CG model of the co-translational Sec machinery that captures the full three dimensionality (3D) of the system. These two models will be referred to as the 2D-CG model and 3D-CG model respectively.

The 2D-CG model has previously been shown to accurately predict the distribution of topologies obtained by signal peptides as a function of C-terminal tail length [26], the probability that a transmembrane segment integrates into the membrane as its hydrophobicity increases [33], and the effect of charge mutations on the topology of the dual-topology protein EmrE [34]. Given these successes, the newly developed 3D-CG model preserves several important features of the 2D-CG model while incorporating new features to address its shortcomings. Key aspects of the 2D-CG model that are retained in the 3D-CG model include: (i) a representation of the nascent polypeptide as a non-overlapping freely-jointed chain, (ii) an approximate 3:1 mapping of amino-acid residues to CG beads, (iii) an implicit representation of the lipid bilayer, (iv) stochastic opening and closing of the translocon lateral gate, and (v) explicit modeling of NC translation during the simulation trajectories.

Improved features of the 3D-CG model include: (i) a fully three-dimensional representation of the system, (ii) direct mapping of the ribosome-translocon complex geometry from high-resolution cryo-EM and crystallographic data, (iii) a bottom-up parameterization of NC-translocon interactions based on detailed residue-based coarse-grained simulations, and (iv) a well-defined protocol to map an input amino-acid sequence to a CG representation. With the addition of these features, the 3D-CG model provides a more realistic representation of the ribosome/translocon/nascent-chain geometry and accurate residue-specific interactions between the NC and the translocon, while still retaining the ability to reach the long second-minute timescale relevant to studying membrane protein integration.

The improved features of the 3D-CG model are especially important to studying the binding energetics during the establishment of topology. First, the detailed parameterization of NC-translocon interactions provides a more accurate description sequence-specific contributions to the binding energetics. Second, a well-defined mapping protocol removes the need to use model sequences that require *a priori* chemical insight. Third, it has been hypotheiszed that inversion of the signal se-

quence depends on the presence of the ribosome, suggesting that an accurate representation of ribosome geometry and interactions as well as the correct scaling behavior between distance and volume in three-dimensions is necessary [15]. Finally, the preservation of detailed balance during lateral gating makes binding free energy calculations to the translocon with the inclusion of both lateral gate conformations possible.

**Geometry**

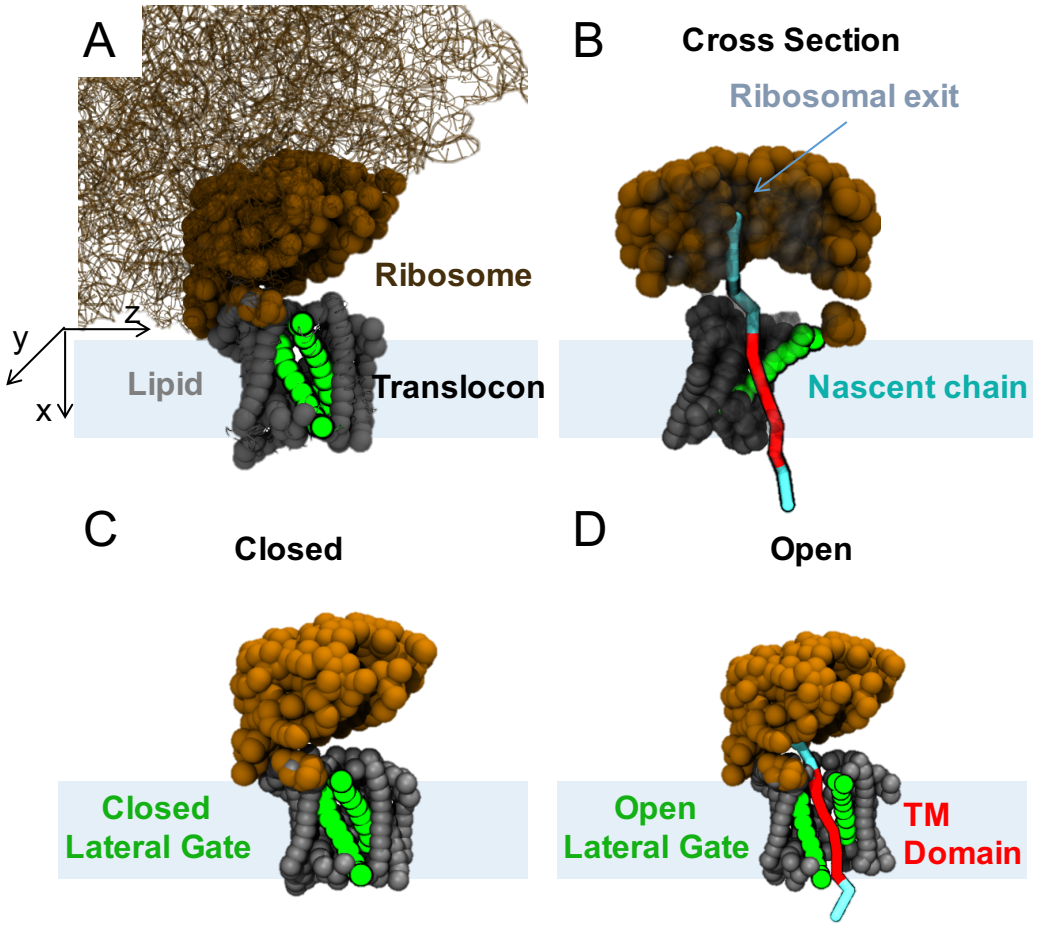



Figure 3.2: Geometry of the 3D-CG model A. Overlaid mapping of the 3D-CG model in CG beads and the 3.4 Åcryo-EM structure of the mammalian ribosome-translocon complex [9] in line representation. B. Cross-sectional side view of the 3D-CG model with the ribosomal exit point of the nascent chain indicated. C. Geometry of the 3D-CG model in the closed lateral gate conformation. D. Geometry of the 3D-CG model in the open lateral gate conformation with nascent chain shown in light blue licorice representation and transmembrane (TM) domain highlighted in red.

Figure 3.2A presents the key molecular components of the 3D-CG model overlaid with an image of the ribosome-translocon complex obtained from a cryo-EM structure [9]. The SecYEG translocon (grey/green), ribosome (brown), and the NC (cyan/red) are represented with explicit CG beads. As in the 2D-CG model [32], each CG bead has a diameter of $\sigma$ = 0.8 nm, the Kuhn length of a polypeptide chain [35, 36] and represents three amino-acid residues; $\sigma$ sets the length scale for the 3D-CG model. The coordinate system is defined such that the origin is placed at the geometric center of the translocon channel C$\alpha$ atoms, the bilayer spans the y-z plane, the translocon channel is aligned with the x-axis (Figure 3.2).

The geometry of the Sec translocon is obtained by mapping all amino-acid residues of the translocon onto CG beads in a 3 residue:1 bead ratio, where the CG bead is positioned at the center of mass of the C$\alpha$ atoms for each sequential triplet of amino-acid residues. The translocon is modeled in two distinct conformations, closed and open (Figure 3.2C,D), whose coordinates are obtained from equilibrated residue-based coarse-grained simulations of the *M. Jannaschii* SecYEG translocon [12] in each state (see Residue-based coarse-grained simulations). Flexible regions of the channel are excluded from the mapping, and an additional bead is included to represent the stable salt bridge between K26 and E421. The closed (Figure 3.2C) and open (Figure 3.2D) conformations of the channel are each described using 156 CG beads (coordinates shown in Appendix B,C). The 3D-CG model of the translocon is oriented such that the y-axis of the simulation coordinate system passes between the helices of the lateral gate when the translocon is in its open conformation (Figure 3.2D).

The geometry of the ribosome is obtained by mapping the ribosome-translocon complex from a recent high-resolution cryo-EM structure [9] (PDB ID: 3J7Q) onto CG beads. Amino-acid residues are mapped onto CG beads in a 3:1 ratio following the same procedure used for the translocon. Every RNA nucleotide in the ribosome is mapped onto two CG beads; one bead represents the sugar-phosphate backbone, while the other bead represents the nucleobase. This mapping is used to properly describe the rigidity of the base pair interactions, and is consistent with previous work on coarse-grained DNA/RNA simulations [37–39].

Because CG bead coordinates for the translocon are obtained from the equilibrated average coordinates of residue-based coarse-grained simulations in the absence of the ribosome, the CG model ribosome must be aligned with the CG model translocon by the procedure shown in Figure 3.3. The ribosome-translocon cryo-EM struc-

ture [9], which contains a translocon in a partially cracked conformation similar to the closed conformation, is aligned with the CG model of the translocon in the closed conformation by minimizing the root mean squared deviation between the LG helices in the CG model translocon and the LG helices mapped from the cryo-EM structure. After alignment, only CG beads of the ribosome that are within 9 $\sigma$ of origin are explicitly retained as CG beads in the final simulation system.

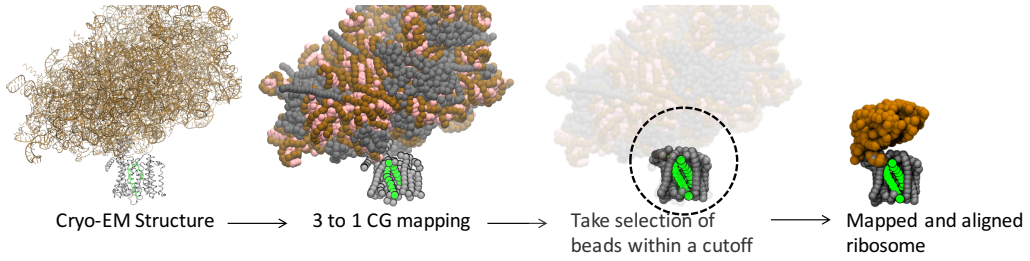



Figure 3.3: Mapping and alignment of ribosome in 3D-CG model. The starting coordinates of the mammalian ribosome-translocon complex from the Voorhees *et al.* cryo-EM structure are shown in the stick representation. These coordinates are mapped in place onto CG beads with CG beads from protein residues shown in gray, CG beads associated with the RNA backbone shown in brown and CG beads associated with RNA nucleobases shown in pink. The lateral gate CG beads, shown in green, are aligned with the equilibrated translocon CG beads from MARTINI simulations and ribosome beads within 9 $\sigma$ of the origin are retained for the final model.

A sphere of diameter 4 $\sigma$ is centered at {-10, -0.5, 1.0} above the NC exit tunnel at the edge of the cut-off region, and represents the remainder of the ribosome to prevent unphysical NC configurations. The same ribosome CG bead coordinates are used for the translocon in the open conformation. In total, the ribosome is represented by 357 CG beads. Complete coordinate files for the CG representation of the ribosome are included in Appendix A.

Two functions are used to define the important geometric regions of the system, $S_{\text{mem}}$ and $S_{\text{chan}}$, shown in Equations 3.1 and 3.2, corresponding to the membrane bilayer and the translocon channel:

$$S_{\text{mem}}(x,r) = S(x)[1 - S(r)], \tag{3.1}$$

$$S_{\text{chan}}(x,r) = S(x)S(r). \tag{3.2}$$

$S_{\text{mem}}$ is 1 in the implicit membrane bilayer and 0 elsewhere, $S_{\text{chan}}$ is 1 inside the translocon channel and 0 elsewhere, and both functions are 0 in the surrounding solvent region.

Smooth switching functions are used to describe the change in value of the geometric functions in boundary regions

$$S(x) = \frac{1}{4}\left[1 + \tanh\left(\frac{x + 2\sigma}{0.25\sigma}\right)\right]\left[1 - \tanh\left(\frac{x - 2\sigma}{0.25\sigma}\right)\right] \tag{3.3}$$

and

$$S(r) = \frac{1}{4}\left[1 + \tanh\left(\frac{r + 1.5\sigma}{0.25\sigma}\right)\right]\left[1 - \tanh\left(\frac{r - 1.5\sigma}{0.25\sigma}\right)\right], \tag{3.4}$$

where $r = \sqrt{(y^2 + z^2)}$ is the radial distance from the origin in the plane of the bilayer. $S(x)$ is approximately 1 for the range -2 $\sigma$ < x < 2 $\sigma$ and 0 elsewhere, while $S(r)$ is approximately 1 for the range -1.5 $\sigma$ < r < 1.5 $\sigma$ and 0 elsewhere.

**Interactions**

All interactions in the 3D-CG model are defined using an energy scale given by $\epsilon$ = $k_B$T, where T is fixed at 310 K to represent physiological temperature. Bonded interactions maintain connectivity between neighboring NC beads and are described using the finite extension nonlinear elastic (FENE) potential (Equation 3.5)

$$U_{\text{FENE}}(r_{ij}) = -\frac{1}{2}\text{KR}_0\ln\left(1 - \frac{r_{ij}^2}{R_0^2}\right), \tag{3.5}$$

where $r_{ij}$ is the distance between NC bead $i$ and $j$, K = 5.33 ($\epsilon/\sigma^2$), and $R_0$ = 2 $\sigma$. Short-ranged non-bonded interactions between pairs of NC beads prevent self-crossing and are modeled using a purely repulsive Lennard-Jones potential (Equation 3.6) which is truncated and shifted to go to zero at $r_{ij} = (2\sigma_{ij})^{1/6}$:

$$U_{\text{LJ}}(r_{ij}) = \left\{ \begin{array}{cc} 4\epsilon\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] & , \; r_{ij} < (2\sigma_{ij})^{1/6} \\ 0 & , \; r_{ij} \geq (2\sigma_{ij})^{1/6} \end{array} \right\}, \tag{3.6}$$

where $\epsilon_{ij} = \epsilon$ and $\sigma_{ij} = \sigma$.

NC bead interactions with the implicit membrane environment are described using a position-dependent potential shown in Equation 3.7,

$$U_{\text{solv}} = gS_{\text{mem}}(x, r), \tag{3.7}$$

where $g$ is the transfer free energy for partitioning the CG bead from water to the membrane and $S_{\text{mem}}$ is the implicit membrane bilayer region geometric function described above.

Non-bonded interactions between NC beads and translocon beads are given by

$$U_{\text{chan}}(r_{ij}) = [1 - S_{\text{mem}}(x_i, r_i)]U_{\text{inner}}(r_{ij}) + [S_{\text{mem}}(x_i, r_i)]U_{\text{outer}}(r_{ij}), \tag{3.8}$$

where $r_{ij}$ is the distance between NC bead $i$ and translocon channel bead $j$. $S_{\text{mem}}$ smoothly interpolates between NC-translocon interactions for which the NC bead is positioned inside the channel ($U_{\text{inner}}$) or positioned in the surrounding membrane, such that it interacts with the channel exterior ($U_{\text{outer}}$). $U_{\text{inner}}$ is defined by four interactions, given by

$$U_{\text{inner}}(r_{ij}) = \left\{ \begin{array}{ccc} \lambda_i^{\text{O}} U_{\text{attr}}^{\text{open}}(r_{ij}) + \left(1 - \lambda_i^{\text{O}} U_{\text{rep}}^{\text{open}}(r_{ij})\right) & , & \text{open channel} \\ \lambda_i^{\text{C}} U_{\text{attr}}^{\text{closed}}(r_{ij}) + \left(1 - \lambda_i^{\text{C}} U_{\text{rep}}^{\text{closed}}(r_{ij})\right) & , & \text{closed channel} \end{array} \right\}, \tag{3.9}$$

where $U_{\text{attr}}^{\text{open}}(r_{ij})$ is the most attractive interaction with a bead in the open channel, $U_{\text{attr}}^{\text{closed}}(r_{ij})$ is the most attractive interaction with a bead in the closed channel, $U_{\text{rep}}^{\text{open}}(r_{ij})$ is the most repulsive interaction with a bead in the open channel, and $U_{\text{rep}}^{\text{closed}}(r_{ij})$ is the most repulsive interaction with a bead in the closed channel. $\lambda_i^{\text{O}}$ and $\lambda_i^{\text{C}}$ scale the interactions of NC bead $i$ between the most attractive and most repulsive interactions in the open and closed channel states respectively.

Interactions between NC beads and CG beads in the ribosome are described by a purely repulsive Lennard-Jones potential (Equation 3.6), with $\epsilon_{ij} = \epsilon$ and $\sigma_{ij} = 1.2\sigma$. To prevent the NC from moving into the part of the ribosome that is not explicitly included in our simulations, a implicit sphere of diameter 4 $\sigma$ is placed at the edge of the cut-off region for explicit CG beads (described in Geometry). A repulsive interaction is applied between NC beads and this implicit sphere,

$$U_{\text{ribo}}(r_{ij}) = \left\{ \begin{array}{ccc} 4\epsilon \left[ \left( \frac{\sigma}{r_{\text{ribo}} - 2\sigma} \right)^{12} - \left( \frac{\sigma}{r_{\text{ribo}} - 2\sigma} \right)^{6} \right] & , & r_{\text{ribo}} - 2\sigma < (2\sigma_{ij})^{1/6} \\ 0 & , & r_{\text{ribo}} - 2\sigma \geq (2\sigma_{ij})^{1/6} \end{array} \right\}, \tag{3.10}$$

where $r_{\text{ribo}}$ is the distance of the NC bead from the implicit sphere centered at {-10, -0.5, 1.0} and 2 $\sigma$ is the radius of the implicit sphere representing the missing part of the ribosome. Electrostatic interactions are described using the Debye-Hückel potential

$$U_{\text{DH}} = \frac{\sigma k_B T q_i q_j}{r_{ij}} \exp\left(-\frac{r_{ij}}{\kappa}\right), \tag{3.11}$$

where the Debye length, $\kappa = \sigma$, is based on the screening of electrostatic interactions at physiological salt concentrations [40, 41], and $q_i$ is the charge of CG bead $i$. The prefactor, $\sigma k_B T$, assumes a Bjerrum length of approximately 0.8 nm, which is appropriate for water.

### Dynamics

The time evolution of NC CG beads is modeled using Brownian dynamics with a first order Euler integrator,

$$x_i(t + \Delta t) = x_i(t) - \beta D \frac{\partial U(x(t))}{\partial x_i} \Delta t + \sqrt{2D\Delta t}\eta_i, \tag{3.12}$$

where $x_i(t)$ is a single Cartesian degree of freedom for NC bead $i$ at time $t$, $U(x(t))$ is the potential energy function for the full system, $\beta = 1/k_B T$, D = 253.0 nm$^2$/s is the diffusion coefficient and $\eta_i$ is a random number drawn from a Gaussian distribution with zero mean and unit variance. The timestep, $\Delta t$ = 300 ns, permits stable integration of the equations of motion with the given value of the diffusion coefficient. All CG beads representing the translocon and ribosome are fixed during the simulations, excluding the switching of the translocon between the open and closed conformations as described below.

NC-dependent conformational gating of the translocon between the closed and open conformation is explicitly included in the 3D-CG model by attempting to stochastically switch the conformation of the translocon at every simulation timestep. The probability of a transition from the closed to open conformation, $p_{\text{open}}$, is given by $p_{\text{open}} = k_{\text{open}}\Delta t$ where

$$k_{\text{open}} = \frac{1}{\tau_{\text{LG}}} \frac{\exp(-\beta \Delta G_{\text{tot}})}{1 + \exp(-\beta \Delta G_{\text{tot}})}. \tag{3.13}$$

Similarly the reverse probability of a transition from the open to the closed conformation $p_{\text{close}}$, is given by $p_{\text{close}} = k_{\text{close}}\Delta t$,

$$k_{\text{close}} = \frac{1}{\tau_{\text{LG}}} \frac{1}{1 + \exp(-\beta \Delta G_{\text{tot}})}, \tag{3.14}$$

where $\tau_{\text{LG}}$ = 500 ns is the timescale for attempting translocon conformational changes and is obtained from prior molecular dynamics simulations [32, 42].

The total free energy change for switching the translocon from the closed to open conformation, $\Delta G_{\text{tot}}$ is given by,

$$\Delta G_{\text{tot}} = \Delta G_{\text{empty}} + \sum_{i=1}^{N_{\text{NC}}} \sum_{j=1}^{N_{\text{chan}}} \Delta U_{\text{chan}}(r_{ij}), \tag{3.15}$$

where $\Delta G_{\text{empty}} = 3\ \epsilon$ is the free energy penalty for opening a closed channel in the absence of a substrate, $N_{\text{NC}}$ is the number of NC CG beads and $N_{\text{chan}}$ is the number of translocon CG beads. $\Delta U_{\text{chan}}(r_{ij})$is the total difference between the energy of NC bead $i$ in the open channel and the energy of NC bead $i$ in the closed channel at its position, $r_{ij}$. Previous simulation studies have revealed that LG opening in the absence of substrate and ribosome is unfavorable, [42], but experimental studies have suggested that inclusion of ribosome or SecA motor binding can prime the channel for opening [8, 43–46]. A direct calculation of $\Delta G_{\text{empty}}$ could be done for future iterations of the model. By explicitly including the energy difference for all NC beads in Equation 3.15, the conformational gating algorithm obeys detailed balance. This is crucial for performing free energy calculations since it ensures the preservation of the proper Boltzmann-weighted ensemble.

Translation is explicitly modeled by adding CG beads to the C-terminus of the NC during a simulation trajectory. At the onset of the simulation, the C-terminal NC bead is fixed at the exit point of the ribosome, {-6.4, 0.0, 1.7}. For each simulation timestep in which translation is performed, the C-terminal bead is moved in the x direction by a distance equal to $\sigma\Delta t/t_{\text{trans}}$, where $t_{\text{trans}}$ is the timescale for translating a single CG bead. $t_{\text{trans}}$ is set to 0.6 seconds to reproduce a translation rate of 5 residues/second unless otherwise specified. The C-terminal NC bead is otherwise not subject to Brownian dynamics, although all interactions between the C-terminal NC bead and other NC beads are included. After a simulation time interval of $t_{\text{trans}}$ , the translation of the C-terminal bead is complete and its dynamics are described using Equation 3.12 for the remainder of the simulation trajectory. The next CG bead in the NC sequence is then positioned at the ribosomal exit point and this process is repeated until the NC beads have been translated.

Based on these dynamics, a series of five steps is iterated to generate a simulation trajectory: (i) forces acting on each NC bead are calculated, (ii) NC bead positions are time-evolved using Brownian dynamics (Equation 3.12), (iii) conformational gating of the translocon is attempted (Equation 3.13 and 3.14), (iv) the simulation is terminated if user-defined conditions are met, and (v) ribosomal translation is performed if necessary.

**Model parameterization**

NC CG bead parameters to describe NC-translocon interaction and solvation interactions must be determined from the underlying amino-acid sequence. In this

model, we use a bottom-up parameterization strategy to utilize more detailed molecular dynamics calculations to inform the parameters of the 3D-CG model.

First, the water-membrane transfer free energy of a NC CG bead is determined from the Wimley-White whole residue hydrophobicity scale [47]. Amino-acid residues that are determined to not fold into secondary structure elements have their transfer free energies further increased by 1.1 kcal/mol. Assuming that NC-channel interactions vary with NC hydrophobicity, potentials of mean force (PMFs) for translocating model hydrophobic and hydrophilic tripeptide substrates across the translocon channel in both the open and closed conformations are then calculated using the MARTINI residue-based coarse-grained force field. The NC-translocon interactions in the 3D-CG model $U_{\text{attr}}^{\text{open}}(r_{ij})$, $U_{\text{rep}}^{\text{open}}(r_{ij})$, $U_{\text{attr}}^{\text{closed}}(r_{ij})$, $U_{\text{rep}}^{\text{closed}}(r_{ij})$, are fit to reproduce these PMFs. Finally, PMFs for translocating a substrate of two intermediate hydrophobicities across the open and closed channel are calculated using MARTINI. The NC-translocon interaction scaling parameters, $\lambda_i^{\text{O}}$ and $\lambda_i^{\text{C}}$, are then parameterized to reproduce the translocation PMFs of all four tripeptides. Using this parameterization strategy, we obtain a complete description of the 3D-CG potential energy function and describe a mapping protocol to assign NC CG bead properties based on the NC amino-acid sequence.

**Water-membrane transfer free energy**

The water-membrane transfer free energy of an NC CG bead is quantified by calculating the total transfer free energy of the associated trio of amino-acid residues using the Wimley-White octanol-water hydrophobicity scale [47]. The Wimley-White octanol-water scale is a whole-residue hydrophobicity scale, and therefore is the sum of the unfavorable free energy cost for transferring the peptide bond into octanol and the cost for transferring each amino-acid side chain. However, the free energy cost for transferring peptide bonds into the membrane environment is reduced significantly (~4.3 kcal/mol [48, 49]) if the peptide bond forms hydrogen bonds via secondary structure formation [48, 50]. The Wimley-White octanol-water scale expressly measures the partitioning of peptide bonds that have formed hydrogen bonds [47]. Therefore, the transfer free energy of any amino-acid residue that does not form secondary structural elements must have its transfer free energy further increased to account for these missing hydrogen bonds, as will be discussed below. This scale is an appropriate choice for this application because it: (i) experimentally measures the partitioning of pentapeptides, which are similar in size

to the tripeptides modeled here, (ii) correlates well with the partitioning of amino acid side chains between water and lipid calculated with computational methods [51], (iii) is widely used to predict transmembrane protein domains [33], and (iv) appears to be approximately additive [48].

**Residue-based coarse-grained simulations**

Detailed residue-based coarse-grained simulations are performed to obtain equilibrated conformations of the open and closed channel, as well as to obtain potentials of mean force as input for the parameterization of the 3D-CG model. All simulations are done with the v2.2P version of the MARTINI force field with the MARTINI polarizable water model [52, 53]. All simulations are performed using version 5.0 of the Gromacs simulations package [54] and version 2.1.0 of the PLUMED library [55]. All MARTINI simulations include the translocon embedded within a lipid bilayer containing 368 palmitoyloleoylphosphatidylcholine (POPC) lipids and solvated by 6,225 coarse-grained water molecules in an electroneutral 50 mM NaCl salt solution. The ribosome is not included due to its large size.

The initial configuration for equilibration trajectories of the translocon is obtained directly from the crystal structure of the SecYEG from emphM. Jannaschii (PDB ID: 1RHZ) [12] using the martinize.py script. The translocon is then restrained during simulations to either the closed or open conformation by applying a biasing potential; the minimum distance between the center-of-mass of each LG helix is constrained to be 0.88 nm in the closed conformation and 1.75 nm in the open conformation [42].

Using the MARTINI model, PMFs for translocating homogeneous tripeptide substrates across the translocon are calculated from umbrella-sampling simulations to inform the parameterization of $U_{\text{attr}}^{\text{open}}(r_{ij})$, $U_{\text{attr}}^{\text{closed}}(r_{ij})$, $U_{\text{rep}}^{\text{open}}(r_{ij})$, and $U_{\text{rep}}^{\text{closed}}(r_{ij})$. The reaction coordinate for umbrella sampling is defined as the distance, $d_x$, between the center-of-mass of the tripeptide and the center-of-mass of the six pore residues that line the center of the translocon, projected along the x-axis of the simulation box. To fully sample the channel axis, 40 umbrella-sampling trajectories are performed in which $d_x$ is restrained to values in the interval from -5 to 4.5, spaced by 0.25. The substrate is further confined within a cylinder with a radius of 1.5 $\sigma$ that is aligned along the x-axis to prevent lateral diffusion into the membrane or into the surrounding solvent. Each translocation PMF is obtained from reweighting the corresponding umbrella-sampling trajectories using the Weighted

Histogram Analysis Method [56].

A key assumption underlying the parameterization of NC-channel interactions is that the hydrophobicity of the NC determines the sign and magnitude of the non-bonding interaction with the channel. Following this assumption, three homogeneous tripeptide substrates of varying hydrophobicities are chosen to span a range of possible channel interactions; tripeptides are selected to match the 3:1 mapping of amino-acid residues to CG beads in the 3D-CG model. Using the Wimley-White scale as a measure of substrate hydrophobicity, residues with a negative transfer free energy for partitioning from water to octanol are hydrophobic, while residues with a positive transfer free energy are hydrophilic; transfer free energies are assumed to be additive when calculating the total transfer free energy of a substrate [33, 48]. An LLL tripeptide is chosen as the model hydrophobic substrate because leucine is the most hydrophobic aliphatic amino-acid residue according to the Wimley-White scale. A DDD tripeptide is chosen as the model hydrophilic substrate because aspartate is the most hydrophilic amino-acid residue according to the Wimley-White scale. Finally, a QQQ tripeptide is chosen as a substrate with a intermediate hydrophobicity lying in between the two extremes.

In MARTINI, each residue is split into a backbone bead and one or more side chain beads. The side chain bead type is determined by the given amino-acid, but the backbone bead type depends on the assumed secondary structure element of the residue [52]. As discussed in Water-membrane transfer free energies, peptide bonds that do not participate in hydrogen bonds face a larger free energy penalty for partitioning into apolar environments, leading to the assignment of a more hydrophilic MARTINI bead type to backbone beads that are in a random coil. To explicitly consider the extremes of hydrophobicity, the more hydrophobic helix backbone type is assigned to the LLL MARTINI substrate, and the more hydrophilic coil backbone type is assigned to the DDD MARTINI substrate. To test the effect of this assignment, the QQQ substrate is simulated twice, once with the helix backbone type and once with the coil backbone type. In total, this yields a set of four tripeptide substrates (Table 3.1) that span a range of effective substrate hydrophobicities, allowing translocation PMFs to be calculated to inform NC-channel interactions.

Figure 3.4 shows PMFs calculated from the MARTINI simulations for the translocation of all four substrates across both the closed and open channels. Each PMF is shifted such that the average value for the range $4.0 < x < 4.5$ is zero. The PMFs are each divided by a factor of four relative to the raw measurements from MARTINI to

| Tripeptide | Secondary Structure | Wimley-White Transfer Free Energy ($\epsilon$) |
|---|---|---|
| LLL | Helical | -6.09 |
| QQQ | Helical | 3.75 |
| QQQ | Coil | 9.11 |
| DDD | Coil | 23.10 |

Table 3.1: Summary of the tripeptides studied by the MARTINI simulations

account for the difference in partitioning free energies between bulk alkane, which MARTINI measures, and biological environments [51]. The PMFs confirm that the hydrophobic LLL substrate and hydrophilic EEE substrate demonstrate qualitatively different behaviors in both channel conformations; LLL is attracted to the center of the channel, which is lined with hydrophobic residues, while EEE is repelled. The PMF for the QQQ substrate with a helix backbone is more attractive than the QQQ substrate with a coil backbone, consistent with expectations, while both profiles lie in between the LLL and EEE PMFs. These results support the assumption that NC-channel interactions depend on substrate hydrophobicity. The profiles qualitatively agree with the atomistic simulations performed by Gumbart et al. [57], which identified a large negative free energy change for transferring a polyleucine helix from water into the channel and a positive free energy change for transferring a single arginine residue into the channel. Comparing the profiles between the closed and open channels shows that the LLL tripeptide is stabilized in the open channel relative to the closed channel, providing a driving force for translocon conformational gating according to Equation 3.13.

### Fitting

NC-channel interactions $U_{\text{attr}}^{\text{open}}(r_{ij})$, $U_{\text{attr}}^{\text{closed}}(r_{ij})$, $U_{\text{rep}}^{\text{open}}(r_{ij})$, and $U_{\text{rep}}^{\text{closed}}(r_{ij})$ are then fit to the PMFs calculated via the residue-based coarse-grained simulations. Each interaction is defined using a soft-core Lennard-Jones potential [58] with a generic form of

$$U(r_{ij}) = \left\{ \begin{array}{cc} 4\epsilon_j^{int}\left[\frac{\sigma_j^{12}}{\left(r_{ij}^6+\alpha_j\right)^2} - \frac{\sigma_j^6}{\left(r_{ij}^6+\alpha_j\right)}\right] + \epsilon_j^{\text{cr}} & , \; r_{ij} < r_{ij}^{\text{cr}} \\ 0 & , \; r_{ij} \geq r_{ij}^{\text{cr}} \end{array} \right\}, \tag{3.16}$$

where $\epsilon_j^{int}$ is the depth of the Lennard-Jones potential when interacting with translocon bead $j$, $\sigma_j$ sets the length scale of the potential, $\epsilon_j^{\text{cr}}$ is the value of the Lennard-Jones potential at the right cut-off radius, $r_{ij}^{\text{cr}}$ is the value of the right cut-off radius,

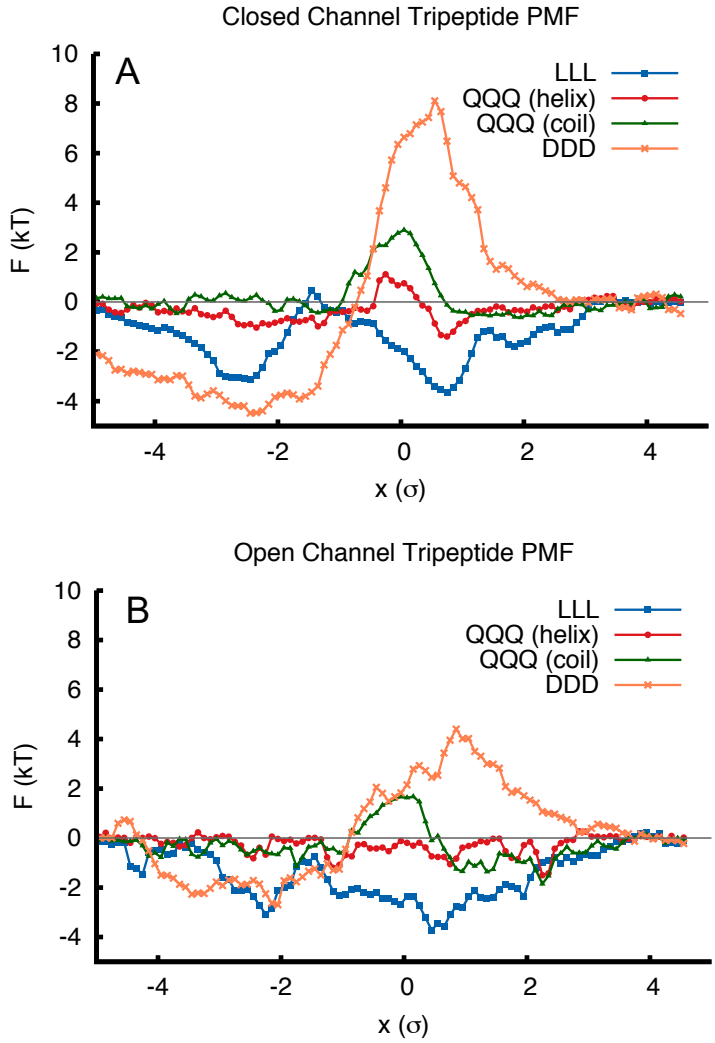



Figure 3.4: Residue-based coarse-grained calculations for tripeptide translocation PMFs. A. Potential of mean force (PMF) for the translocation of model tripeptides through the translocon in the closed conformation. B. Potential of mean force (PMF) for the translocation of model tripeptides through the translocon in the open conformation.

and $\alpha_j$ sets the maximum value of the repulsive potential. The soft-core potential is used to prevent infinite energies during the stochastic gating of the translocon conformation.

The NC-translocon interactions in the 3D-CG model are determined by least squares fitting of the PMFs shown in Figure 3.5. Parameters for the NC-translocon attractive

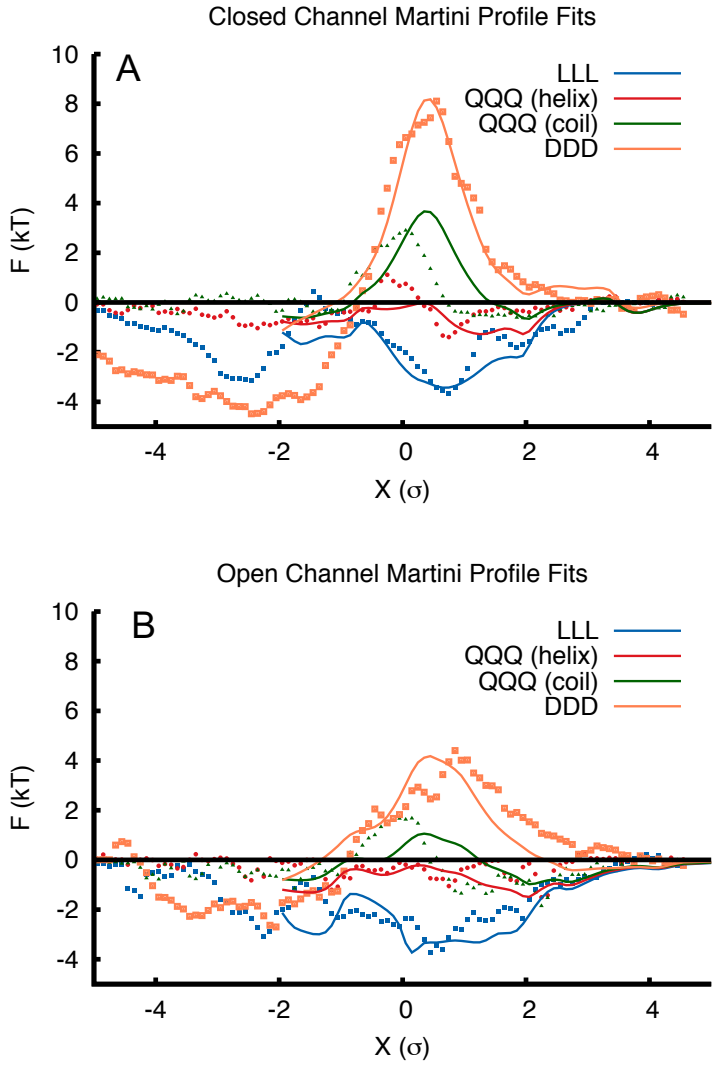



Figure 3.5: Fitting CG parameters to MARTINI potentials of mean force. A. Analytical PMFs for tripeptide translocation across the translocon in the closed channel conformation calculated from the best fit parameters of the CG model (lines) compared to the MARTINI potentials of mean force (points). B. Analytical PMFs for tripeptide translocation across the translocon in the open channel conformation calculated from the best fit parameters of the CG model (lines) compared to the MARTINI potentials of mean force (points).

interactions, $U_{\text{attr}}^{\text{open}}$ and $U_{\text{attr}}^{\text{closed}}$, are determined from the LLL PMFs in the open and closed channel, respectively. Similarly, parameters for the NC-translocon repulsive interactions, $U_{\text{rep}}^{\text{open}}$ and $U_{\text{rep}}^{\text{open}}$, are determined from the EEE PMFs. The MARTINI

PMFs are fit by analytically calculating the corresponding translocation PMFs with the 3D-CG model potential energy function and fitting the interaction energy, $\epsilon_j^{int}$, between the NC and translocon beads using the lmfit function in Python yielding the parameters summarized in Table 3.2. The lmfit module performs a weighted least squares fit between $-2 < x < 4$ where the error is exponentially weighted to prioritize fitting the peaks and valleys. The region $x < -2$ is not used for the fit because this region contains contributions from flexible loop regions of the translocon in the MARTINI PMF simulations; these loop regions have been removed in the 3D-CG model and therefore features in this region are not expected to be captured in the 3D-CG model. In order to match the MARTINI residue-based coarse-grained simulations for which the plug domain has been removed and the ribosome is not present, the CG beads corresponding to the ribosome and the plug domain are removed during the fitting procedure. The 5 CG beads corresponding to the plug domain in the open and closed conformations are indicated in Appendix B and C. The PMFs with these plug CG beads and the ribosome CG beads included are shown in Appendix D.

| Potential | $\epsilon_j^{int}$ | $\epsilon_j^{\mathrm{cr}}$ | $r_j^{\mathrm{cr}}$ | $\alpha_j$ | $\sigma_j$ |
|---|---|---|---|---|---|
| $U_{\mathrm{attr}}^{\mathrm{open}}$ | 0.46 | 0.0075 | 2.5 | 0.1268 | 1.0 |
| $U_{\mathrm{attr}}^{\mathrm{closed}}$ | 0.3 | 0.0049 | 2.5 | 0.1037 | 1.0 |
| $U_{\mathrm{rep}}^{\mathrm{open}}$ | 0.3 | 0.0049 | 2.5 | 0.1037 | 1.0 |
| $U_{\mathrm{rep}}^{\mathrm{closed}}$ | 0.3 | 0.0049 | 2.5 | 0.1037 | 1.0 |
| $U_{\mathrm{attr}}^{\mathrm{open\text{-}confined}}$ | 1.38 | 0.0225 | 2.5 | 0.2089 | 1.0 |
| $U_{\mathrm{attr}}^{\mathrm{closed\text{-}confined}}$ | 1.41 | 0.0229 | 2.5 | 0.2109 | 1.0 |
| $U_{\mathrm{rep}}^{\mathrm{open\text{-}confined}}$ | 9.85 | 0.9784 | 1.0745 | 0.4609 | 1.2 |
| $U_{\mathrm{rep}}^{\mathrm{closed\text{-}confined}}$ | 0.51 | 0.8577 | 1.1097 | 0.1330 | 1.2 |
| $U_{\mathrm{outer}}$ | 0.5 | 0.0082 | 2.5 | 0.1317 | 1.0 |

Table 3.2: Summary of the best-fit channel parameters

The translocation profiles for LLL and EEE can only be reasonably fit if two NC bead types are defined—one default bead type and one "confined" bead type. Confined CG bead types for the repulsive potential are chosen as beads that have $x$ between -0.1 and 1.1 and $r < 2.2$ in both the open and closed conformations. Confined CG bead types for the attractive potential are beads that have $x$ between 0.3 and 1.0 for and $r < 1.5$ in both the open and closed conformations. All remaining channel beads, including plug domain CG beads are assigned to the default bead type. These confined beads allow the CG model to capture the large peaks

in the EEE PMFs and valleys in the LLL PMFs. The functional form for the non-bonding interactions with the purely repulsive channel beads is also identical to the non-bonding interactions with beads in the ribosome. Analytical PMFs, using the best-fit values of $\epsilon_j^{int}$, are presented in Figure 3.5. A full description of the channel bead positions and parameters is given in Appendices B and C.

The values of $\lambda^{\mathrm{O}}$ and $\lambda^{\mathrm{C}}$ are set to 1 for the LLL substrate and 0 for the EEE substrate by construction, since the MARTINI PMFs for these substrates determine the parameters for the limiting attractive and repulsive channel interactions, respectively. The values of $\lambda^{\mathrm{O}}$ and $\lambda^{\mathrm{C}}$ are set for the two QQQ substrates, which differ only in the assignment of backbone bead type, by calculating analytical PMFs using Equation 3.8 with $U_{\mathrm{attr}}^{\mathrm{open}}$, $U_{\mathrm{attr}}^{\mathrm{closed}}$, $U_{\mathrm{rep}}^{\mathrm{open}}$, $U_{\mathrm{rep}}^{\mathrm{closed}}$, $U_{\mathrm{attr}}^{\text{open-confined}}$, $U_{\mathrm{attr}}^{\text{closed-confined}}$, $U_{\mathrm{rep}}^{\text{open-confined}}$, $U_{\mathrm{rep}}^{\text{closed-confined}}$, and $U_{\mathrm{outer}}$ defined using the parameters in Table 3.2. The analytical PMFs are fit to the MARTINI PMFs using the lmfit module in Python with $\lambda^{\mathrm{O}}$ and $\lambda^{\mathrm{C}}$ as fit parameters. As before, the lmfit module performs a weighted least squares fit between $-2 < x < 4$ where the error is exponentially weighted to prioritize fitting the peaks and valleys.

The NC-channel interaction scaling parameters, $\lambda^{\mathrm{O}}$ and $\lambda^{\mathrm{C}}$, are generalized by assuming that NC-channel interactions scale with hydrophobicity, as confirmed by the MARTINI simulations. Since each of the four substrates simulated with MARTINI can be assigned values of the transfer free energy, $g$, and corresponding values of $\lambda^{\mathrm{O}}$ and $\lambda^{\mathrm{C}}$, the scaling parameters are determined from $g$ by linearly interpolating between these known values,

$$\lambda^{\mathrm{O}} = \left\{ \begin{array}{lcc} m_a^{\mathrm{O}} g + b_a^{\mathrm{O}} & , & g < g_{\mathrm{QQQ}}^{helix} \\ m_b^{\mathrm{O}} g + b_b^{\mathrm{O}} & , & g_{\mathrm{QQQ}}^{helix} \leq g < g_{\mathrm{QQQ}}^{coil} \\ m_c^{\mathrm{O}} g + b_c^{\mathrm{O}} & , & g_{\mathrm{QQQ}}^{coil} \leq g \end{array} \right\} \tag{3.17}$$

and

$$\lambda^{\mathrm{C}} = \left\{ \begin{array}{lcc} m_a^{\mathrm{C}} g + b_a^{\mathrm{C}} & , & g < g_{\mathrm{QQQ}}^{helix} \\ m_b^{\mathrm{C}} g + b_b^{\mathrm{C}} & , & g_{\mathrm{QQQ}}^{helix} \leq g < g_{\mathrm{QQQ}}^{coil} \\ m_c^{\mathrm{C}} g + b_c^{\mathrm{C}} & , & g_{\mathrm{QQQ}}^{coil} \leq g \end{array} \right\}, \tag{3.18}$$

where the values for the parameters for Equations 3.18 and 3.17 are summarized in Table 3.1 and 3.3.

The value of $g$ for each substrate is determined by summing the Wimley-White transfer free energies of each residue using the Wimley-White hydrophobicity scale [47]. The EEE and QQQ substrates with coil backbones have an additional 5.6 $\epsilon$

| Parameter | Values |
|---|---|
| $m_a^{\mathrm{O}}$ | -0.022 |
| $m_a^{\mathrm{C}}$ | -0.025 |
| $m_b^{\mathrm{O}}$ | -0.035 |
| $m_b^{\mathrm{C}}$ | -0.062 |
| $m_c^{\mathrm{O}}$ | -0.042 |
| $m_c^{\mathrm{C}}$ | -0.03 |
| $b_a^{\mathrm{O}}$ | 0.864 |
| $b_a^{\mathrm{C}}$ | 0.845 |
| $b_b^{\mathrm{O}}$ | 0.913 |
| $b_b^{\mathrm{C}}$ | 0.981 |
| $b_c^{\mathrm{O}}$ | 0.974 |
| $b_c^{\mathrm{C}}$ | 0.693 |

Table 3.3: Summary of the parameters for $\lambda^{\mathrm{O}}$ and $\lambda^{\mathrm{C}}$ interpolation

added to the transfer free energy; this penalty is equal to approximately one-quarter of the cost for partitioning peptide bonds that lack hydrogen bonds between water and alkane [47].

From the values of $g$ and corresponding values of $\lambda^{\mathrm{O}}$ and $\lambda^{\mathrm{C}}$, all the parameters for $U_{\mathrm{inner}}$ in Equation 3.9 are derived. This procedure defines a mapping between generic substrate hydrophobicity and corresponding channel interactions. The set of four tripeptide substrates used to define this mapping is just one part of a much larger set of possible tripeptides. In principle, it is possible to simulate all tripeptide combinations, including heterogeneous tripeptides, but due to the lack of computational resources we instead use the trilinear interpolation scheme specified here.

The final set of parameters to be defined determine $U_{\mathrm{outer}}(r_{ij})$, which describes interactions between NC beads and the outside of the channel. This interaction is again defined by the generic Lennard-Jones potential given in Equation 3.16. It is assumed that all interactions with the outside of the channel are approximately equivalent, as interactions in the membrane will be largely dominated by hydrophobic interactions. Parameters are thus assigned based on comparison to the MARTINI parameters for hydrophobic beads and summarized in Table 3.2.

### Mapping

Every NC bead in the 3D-CG model is fully specified by four properties that are used to determine the interactions with the rest of the system: $g$, the water-membrane

transfer free energy; $q$, the charge; $\lambda^{\mathrm{O}}$, the scaling parameter for NC interactions with the interior of the translocon in its open conformation; and $\lambda^{\mathrm{C}}$, the scaling parameter for NC interactions with the interior of the translocon in its closed conformation. All of these properties are directly obtained from the amino-acid sequence as follows. $g$ is equal to the sum of the transfer free energies of the three amino-acid residues associated with the NC CG bead according to the Wimley-White octanol-water whole residue hydrophobicity scale. $q$ is equal to the sum of the charges of the three associated amino-acid residues. It is assumed that arginine and lysine residues always bear a +1 charge, glutamate and aspartate residues always bear a -1 charge, and all other residues are neutral. The N- and C-terminal CG beads are assigned an additional +1 and -1 charge, respectively, and have 6 $\epsilon$ added to their transfer free energies to account for the additional charge [47, 49]. $\lambda^{\mathrm{O}}$ and $\lambda^{\mathrm{C}}$ are then determined from Equations 3.17 and 3.18 using the value of $g$ for the NC bead.

Only the charge must be mapped to CG beads in the ribosome and translocon from the underlying amino-acid sequence. A CG bead in the ribosome or translocon is assigned a +1 charge if its corresponding trio of amino acids has a net positive charge, and -1 charge if the triplet has a net negative charge; otherwise the bead is neutral. Beads representing the RNA sugar-phosphate backbone in the ribosome carry a -1 charge and beads representing the nucleobase are neutral.

## 3.3 Results and Discussion

The ability of the 3D-CG model to reproduce the key features of co-translational membrane protein insertion and translocation is validated by comparison to experiments. We focus on two seminal experiments done by von Heijne and coworkers and Spiess and coworkers that examine the stop-transfer efficiency of a transmembrane domain with varying hydrophobicity [33], and the topology of inserted transmembrane domains as a function of chain length as well as translation rate. Finally, we develop a binding free energy framework to calculate the interaction between the RNC and translocon as a function of nascent chain length and interrogate the interactions during the establishment of type II topology.

### Stop-transfer efficiency

The Sec translocon facilitates the integration of hydrophobic transmembrane segments into the lipid bilayer. When a putative transmembrane segment reaches the translocon it can either translocate across the membrane bilayer, or it can laterally

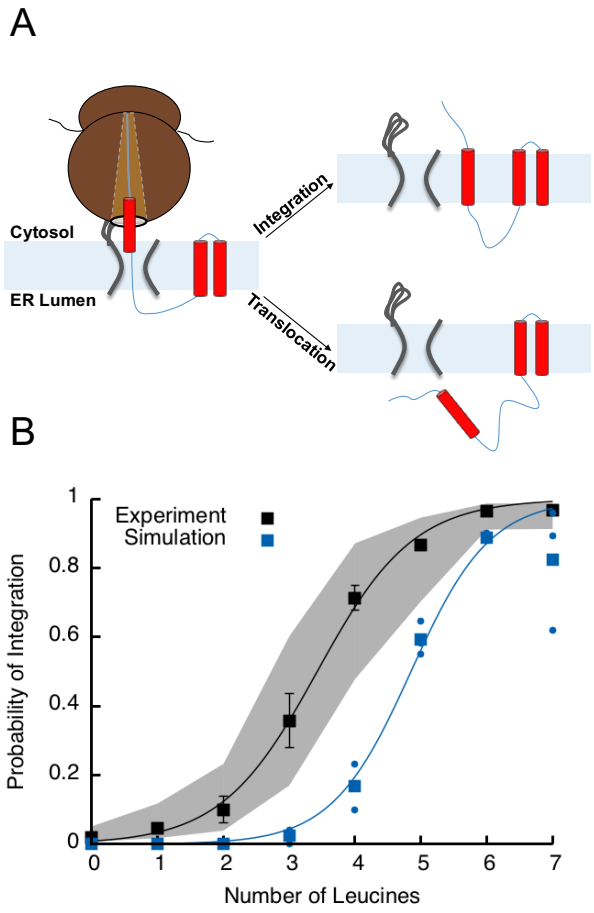



Figure 3.6: Stop-transfer efficiency. A. Schematic of the two possible products of each trajectory where the studied transmembrane domain can either integrate into the membrane or be secreted through translocation into the ER Lumen. B. Comparison of experimental assay [33] results, shown in black, for the probability of integration as a function of the number of leucines in the transmembrane domain to the probability of integration from simulated results, shown in blue. Results for all three frameshifts are shown in blue diamonds and the averaged result is shown in blue squares. A model function corresponding to a system in an apparent two-state equilibrium (Equation 3.19) is fit through both the experimental results (black line) and the averaged simulation results (blue line).

partition into the membrane bilayer (Figure 3.6A). Previous work has shown that the integration probability depends on the hydrophobicity of a putative transmembrane segment via an apparent two state equilibrium [32, 33, 59] (Figure 3.6B, black).

The co-translational integration of transmembrane segments with varying hydrophobicities is simulated using the newly developed 3D-CG model. The protein se-

quence is directly mapped from the Lep protein construct (full sequence and mapping given in Appendix E) which has three known transmembrane domains identified via the UniProt database [60]. All other regions are assigned a coil-like secondary structure and the transfer free energies reflect the additional free energy cost of transferring the peptide bonds into the membrane. The first two transmembrane domains have identical sequences in all constructs while the number of leucines in the third transmembrane domain is modified. The insertion probability of the third transmembrane sequence is assayed as shown in Figure 3.6.

The trajectories are initialized with the first two transmembrane domains translated (initial coordinates given in Appendix F). These domains have the native sequence and are assumed to integrate with high efficiency. The translation of the remaining 32 beads and transmembrane helix of interest is performed at 5 res/s until the whole sequence has been translated. A trajectory is terminated when the end conditions for the integration state or translocation state are reached. The integrated state is defined as a fully translated NC where all transmembrane beads lie in the membrane bilayer region defined as $-2 \leq x \leq 2$ and all transmembrane beads have diffused at least 10 $\sigma$ away from the origin. The translocated state is defined as having all beads on the lumen side of the membrane defined as $x > 2.5$. For each point a total of 50 independent trajectories are run. For each given 3:1 mapping from the amino acid sequence, there are three possible frameshifts that lead to three distinct NC CG bead sequences. For all the studied sequences we plot all three frameshifts and the average stop-transfer value.

In agreement with previous results, we find that integration probability increases with the number of leucines in the h-segment. The integration probability, $p_{\text{int}}$, can be fit to an apparent two state equilibrium (shown in Figure 3.6B) given by Equation 3.19

$$P_{\text{int}}(n_{\text{Leu}}) = \frac{1}{1 + \exp[-\beta\alpha\Delta G(n_{\text{Leu}}) + \gamma]}, \tag{3.19}$$

where $\Delta G(n_{\text{Leu}}) = n_{Leu}(-2.84\epsilon)$ and -2.84 $\epsilon$ is the difference in water-membrane transfer free energy obtained by mutating an alanine residue to a leucine residue. For the experimental results [33], the parameters of the two-state equilibrium are $\alpha = 1.37$ and $\gamma = 4.69$; for the simulated results the parameters are $\alpha = 1.66$ and $\gamma = 8.07$. The difference in $\gamma$ values reflects the ~2 leucine rightward shift in the probability of integration when comparing the simulation results to the experimental results (Figure 3.6B). The origin of this shift is not yet well understood but may be due to unaccounted for effects such as formation of helical structure [48, 61–63].

Nevertheless, the qualitative agreement confirms the suitability of the 3D-CG model in capturing the hydrophobicity dependence of stop-transfer for co-translationally integrated membrane domains.

### Topogenesis

Once a transmembrane segment reaches the translocon it can orient either in an $N_{cyt}/C_{ER}$ topology, or in an $N_{ER}/C_{cyt}$ topology (Figure 3.7 A). Experiments in the Spiess group have shown that the final topology is kinetically determined; slowing down NC translation via the addition of the antibiotic cyclohexamide increases the fraction of the $N_{cyt}/C_{ER}$ topology (Figure 3.7B). Given the additional time to equilibrate, the NC will adapt the thermodynamically favored $N_{cyt}/C_{ER}$ topology with a higher probability. Furthermore, a dependence of topology on the length of the C-terminal loop was observed (Figure 3.7B) [32], where an increase in C-terminal loop length increases $N_{cyt}/C_{ER}$ topology up until a certain point where the percent of $N_{cyt}/C_{ER}$ topology plateaus.

The topogenesis of a modifed ASGP receptor (H1$\Delta$22) sequence studied by Speiss and coworkers [26] is simulated using the 3D-CG model. The secondary structure of the sequence is assumed to be helical for the first 100 residues and coil for the remaining residues. This assignment was determined from the PSIPRED secondary structure prediction server [64, 65]. The complete sequence mapping is given in Appendix E. Each of the trajectories are initialized with four CG beads translated. In the default trajectories, translation proceeds at 5 res/s; in the slow trajectories, translation proceeds at 1.25 res/s to match the effect of cyclohexamide in the experimental results. After translation the trajectories are continued until the end conditions for either $N_{cyt}/C_{ER}$ or $N_{ER}/C_{cyt}$ topologies are met. The end conditions of the $N_{cyt}/C_{ER}$ topology are defined as integration of all the transmembrane domain CG beads into the membrane region and the CG bead N-terminal to the transmembrane domain in the cytosol ($x$ < -2) and the bead C-terminal to the transmembrane domain in the lumen ($x$ > 2). Similarly, the $N_{ER}/C_{cyt}$ topology is defined as integration of all the transmembrane beads into the membrane bilayer region with the CG bead C-terminal to the transmembrane domain in the cytosol ($x$ < -2) and the bead N-terminal to the transmembrane domain in the lumen ($x$ > 2). Simulations are run until one of these end conditions is met. A translocated end condition is also considered, but no trajectories lead to this state. Each of the calculated points in Figure 3.7C are the result of 100 individual trajectories.

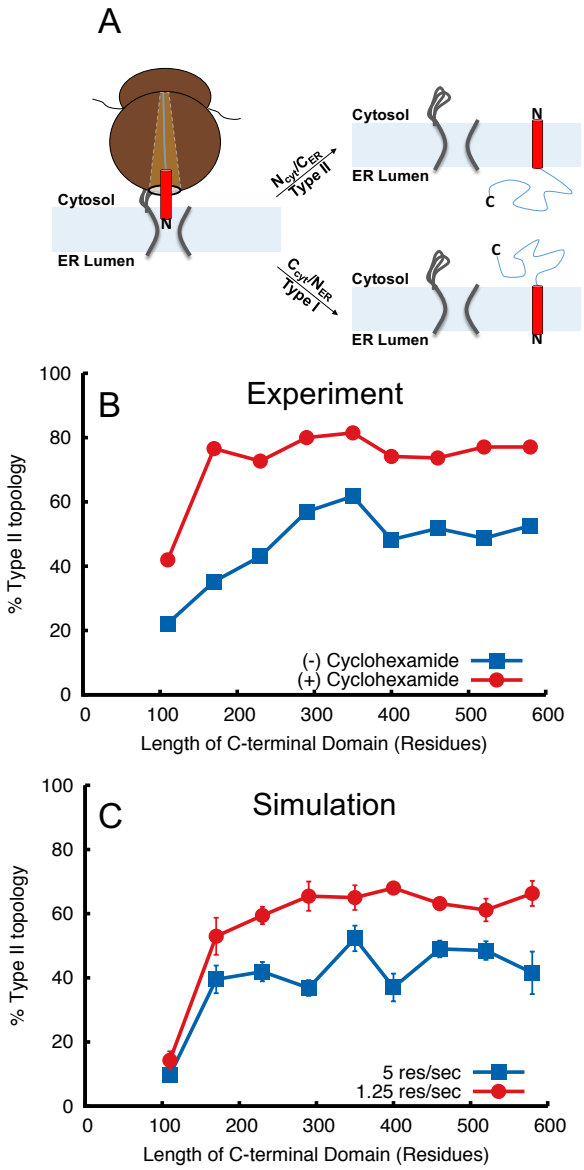



Figure 3.7: Topogenesis simulations. A. Schematic of the two possible products of each trajectory where the studied transmembrane domain can either integrate into the membrane in a $N_{cyt}/C_{ER}$ (Type II) topology or the $N_{ER}/C_{cyt}$ (Type I) topology. B. Experimental results [26] for the probability of Type II topology as a function of C-terminal length. The experiments are done without cyclohexamide (blue) and with cyclohexamide, an antibiotic that slows the translation rate. C. Simulation results for probability of Type II topology as a function of C-terminal length. The simulations are performed with a 5 res/sec translation rate (blue) and a 1.25 res/sec translation rate (red).

Figure 3.7C shows that the simulation results are in good agreement with the experimental results in both the dependence on C-terminal length and the dependence on ribosomal translation rate. Our previous work [32] suggests that the mechanism behind the C-terminal loop length dependence is the limited space available in the ribosome-translocon junction. Once the NC has filled the available space it is forced to move away from the translocon, and thus the topology cannot further equilibrate. However, due to the low dimensionality of the previous model, the scaling between length and volume was incorrect, leading to a premature plateau in the amount of type II topology observed. In this work, we observe this same mechanism, and the improved agreement of simulation results and experimental results reflects the realistic geometry used for the ribosome-translocon junction in the 3D-CG model.

**Free energy calculations**

To enable the direct comparison to experimental measurements of the binding affinity of the ribosome-nascent chain complex to the translocon, we develop a framework to calculate binding free energies of the RNC to the translocon and membrane using the 3D-CG model and alchemical free energy calculations. While the effects of ribosome-translocon interactions in the free energy calculation are not explicitly modeled, this analysis can be applied to relative free energies between NC sequences with differing lengths and assuming that the ribosome-translocon binding interaction is insensitive to the NC length.

Despite the reduced computational cost of the 3D-CG model, long-lived metastable states corresponding to different topologies of the NC are challenging to sample during the timescales of the trajectories. As a result the straightforward application of the alchemical free energy perturbation method is computationally challenging. Instead, we define discrete states corresponding to the different topologies that are separated by kinetic barriers and calculate the free energy of binding in these discrete states. By choosing these states to span all of the configurational space, the free energies can be recombined to give a total free energy of binding. Restrained simulations of this type have previously been applied to rotameric states of sidechains and orientational states of small ligands [66, 67].

For these calculations, we define three states, $\{0, 1, 2\}$ as shown in Figure 3.8. Each of the states are defined by the position of a c-bead shown in blue. The c-bead for a sequence is chosen as the bead with the highest Wimley-White transfer free energy of the five beads following the transmembrane domain. Selection of an alternate

c-bead in principle does not affect the total binding free energy but can affect the sampling of individual states. State 1 corresponds to configurations associated with the $N_{cyt}/C_{ER}$ or type I topology and state 2 corresponds to configurations associated with the $N_{ER}/C_{cyt}$ or type II topology.

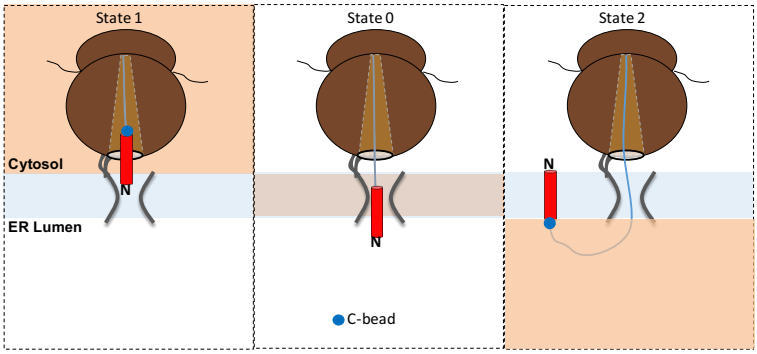



Figure 3.8: Discrete states for free energy of binding calculations shown in orange where the state is defined by the position of the c-bead.

The total binding free energy, $\Delta G_{\text{bind}}(l)$, of a given NC sequence with length $l$ is then obtained from the discrete state free energies via Equation 3.20

$$\Delta G_{\text{bind}}(l) = -kT \ln \left[ \exp(-\beta \Delta G_1(l)) + \exp(-\beta \Delta G_2(l)) + \exp(-\beta \Delta G_0(l)) \right] , \tag{3.20}$$

where $\Delta G_1$ is the free energy between the RNC, in the absence of the translocon and membrane, and the fully interacting system for configurations where $x_c < -2$. $\Delta G_0$ is the free energy of binding between the RNC, in the absence of the translocon and membrane, and the fully interacting system for configurations where $-2 \leq x_c \leq 2$. $\Delta G_2$ is the free energy of binding between the RNC, in the absence of the translocon and membrane, and the fully interacting system for configurations where $x_c > 2$ . This allows for adequate sampling of each state without requiring frequent sampling between states that are separated by large barriers associated with translocation of the hydrophilic c-bead across the membrane.

The discrete state binding free energy, $\Delta G_{s\in\{0,1,2\}}$, is the sum of a purely entropic term $\Delta G_s^{\text{entr}}$ associated with confining the NC configurations to those of state $s$, and an enthalpic term $\Delta G_s^{\text{inter}}$ associated with the interactions of the NC configurations in the state $s$ with the translocon and membrane environment. Each of these terms are calculated separately and combined for each state to calculate $\Delta G_{\text{bind}}(l)$.

### Calculation of entropic term, $\Delta G_s^{\text{entr}}$

The entropic term, $\Delta G_s^{\text{entr}}(l)$ associated with the confinement of the NC with length $l$ to configurations of a given state (Figure 3.8), is calculated via unbiased simulations. The unbiased simulations are used to calculate the probability of being in each state, $p_s(l)$ followed by Boltzmann inversion of the probabilities to obtain the associated free energies, $\Delta G_s^{\text{entr}}(l) = -kT\ln[p_s(l)]$.

For a given sequence of length $l$, trajectories are initialized with four beads translated as shown in Appendix F. Ribosomal translation is modeled by introducing CG beads at the exit point, $r = \{-6.4, 0.01.4\}$, sequentially at the C-terminus until $l$ beads have been translated. The trajectory is then "stalled" at length $l$ by halting further translation and leaving the last translated bead fixed at the exit point. All other beads are allowed to evolve according to Equation 3.12. Statistics for the position of the c-bead are collected for 15 seconds after translation. For the calculation of $\Delta G_s^{\text{entr}}$, all interactions associated with NC-translocon and NC-membrane interactions (Equations 3.7 and 3.8) are removed.

Figure 3.9 shows the calculated $\Delta G_s^{\text{entr}}$ for a model sequence shown in Appendix E.

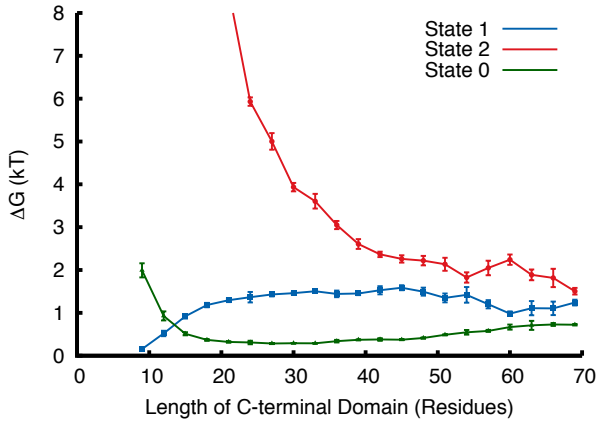



Figure 3.9: Entropic free energy for discrete states.

The calculated $\Delta G_s^{\text{entr}}$ for State 1 has a minimum at short C-terminal lengths and eventually plateaus at a value of ~1.4 kT. For the very short C-terminal domains, the total possible length of the chain from the stalled c-terminal bead to the c-bead

lies within State 1, and therefore there is no free energy cost associated with confinement to State 1. For State 0, this corresponds to the confinement of the c-bead to values of $-2 < x < 2$. This has a minimum when the radius of gyration for the c-bead fits into the State 0 space and incurs a penalty for both shorter and longer lengths. Finally, for State 2, there is a large penalty at short lengths associated with the finite extension of the bond lengths. At longer C-terminal lengths, the chain becomes long enough to reach the $x \geq 2$ state. The entropic terms calculated here are expected to be insensitive to the amino-acid sequence changes since ribosome-nascent chain interactions are identical except for charge interactions.

### Calculation of interaction term, $\Delta G_s^{\text{inter}}$

We calculate the interaction term $\Delta G_s^{\text{inter}}(l)$ associated with interaction of the NC with the translocon and membrane in the bound RNC-translocon complex for NC configurations in the state $s$ using the alchemcal free energy perturbation method and uncoupling the interactions of the translocon and membrane.

The free energy difference $\Delta G_s^{\text{inter}}(l)$ is calculated from independent sampling trajectories performed with potential energy functions corresponding to a different value of a parameter $\zeta$, such that

$$V(\zeta) = (1 - \zeta)V_0 + \zeta V_1. \tag{3.21}$$

$V_0$ refers to the state without channel and membrane bilayer interactions, and $V_1$ refers to the state with fully interacting channel and membrane potentials (Equations 3.7 and 3.8). The Bennett acceptance ratio (BAR) method [68] is used to bidirectionally weight the collected data from trajectories sampled in the $i$ and $i + 1$ potential, and calculate $\Delta G_{i,i+1}$, the FE difference between the $i$ and $i + 1$ states. $\Delta G_m ail.comi, i + 1$, is calculated by iteratively solving the following equations

$$\begin{aligned} e^{\beta \Delta G_{i,i+1}} &= \frac{\langle f[-\beta(\Delta U_{i+1,i} - C)]\rangle_{i+1}}{\langle f[\beta(\Delta U_{i,i+1} - C)]\rangle_i} e^{(\beta C)}, \\ f(x) &= \frac{1}{1 + e^x}, \\ C &= \Delta G + \frac{1}{\beta} \ln \frac{N_{i+1}}{N_i}, \end{aligned} \tag{3.22}$$

where $\Delta U_{i,i+1} = U_{i+1}(x_i) - U_i(x_i)$. $U_i(x_i)$ is the potential energy of the configuration $x_i$ sampled with the $i$ potential and evaluated in the $i$ potential. $U_{i+1}(x_i)$ is the potential energy of the configuration $x_i$ sampled with the $i$ potential, but evaluated in

the $i+1$ potential. $N_i$ is the number of independently sampled points in trajectory $i$. For each value of $\zeta \in \{0.0, 0.01, 0.03, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$ a stalled CG trajectory is performed to sample the configuration space in the assigned state $s$.

Trajectories are initialized with four beads translated as shown in Appendix F. Ribosomal translation is modeled at 5 res/s by introducing CG beads at the exit point, $r = \{-6.4, 0.01.4\}$, sequentially at the C-terminus until $L$ beads have been translated and the trajectory is then "stalled" at length $L$ for sampling. Statistics for $\Delta U_{i,i+1}$ are collected for 1 minute in each potential $U_i$. To ensure the proper sampling for each state space, we apply restraint potentials $U^1_{\text{res}}, U^0_{\text{res}}, U^2_{\text{res}}$ associated with each state given by

$$U^1_{\text{res}}(x_c) = \left\{ \begin{array}{cc} \frac{1}{2}k_{\text{res}}(x_c+2)^2 & , \quad x_c > -2 \\ 0 & , \quad x_c \leq -2 \end{array} \right\}, \tag{3.23}$$

$$U^2_{\text{res}}(x_c) = \left\{ \begin{array}{cc} \frac{1}{2}k_{\text{res}}(x_c-2)^2 & , \quad x_c < 2 \\ 0 & , \quad x_c \geq 2 \end{array} \right\}, \tag{3.24}$$

$$U^0_{\text{res}}(x_c) = \left\{ \begin{array}{cc} \frac{1}{2}k_{\text{res}}(x_c+2)^2 & , \quad x_c < -2 \\ \frac{1}{2}k_{\text{res}}(x_c-2)^2 & , \quad x_c > 2 \end{array} \right\}, \tag{3.25}$$

where the restraint potential is applied on the c-bead. The potential is constructed to be zero when the c-bead is in the correct state and applies a harmonic restoring potential when it is not. As described in Chapter 2, checks for the convergence of each sampling trajectory and the overlap between neighboring trajectories will be done. We check the autocorrelation function (ACF) of $\Delta U_{i,i+1}(t)$ to find decorrelation times which determine the frequency of sampling for the FE calculations. The overlap between neighboring trajectories is checked via Bennett's overlapping histograms and sufficient windows are added to ensure good overlap [69].

## 3.4 Conclusions

The proper insertion and topology of membrane proteins is established during co-translational integration at the Sec translocon. We have developed a new 3D-CG model to directly simulate the long-timescales of this process and to interrogate the binding energetics of the RNC and translocon. The 3D-CG model is shown to reproduce key features of membrane protein integration, including the stop-transfer efficiency as a function of transmembrane domain hydrophobicity and effects of translation rate and C-terminal length on topogenesis. The 3D-CG model also rigorously preserves detailed balance to allow for the calculation of free energies through

the discrete state binding free energy framework described. Furthermore, the direct structural mapping approach for the ribosome and translocon allows for the future investigation of species-specific effects as additional structural information is obtained from experiments. The bottom-up parameterization strategy provides a way to probe translocon mutations to identify key residues of the translocon. Finally the inclusion of the full-dimensionality of the system offers the possibility to incorporate secondary and tertiary structure effects into future simulation models of membrane protein integration.

## References


(1) Keenan, R. J.; Freymann, D. M.; Stroud, R. M.; Walter, P. *Ann. Rev. of Biochem.* **2001**, *70*, 755–775.

(2) Mitra, K.; Schaffitzel, C.; Shaikh, T.; Tama, F.; Jenni, S.; Brooks, C. L.; Ban, N.; Frank, J. *Nature* **2005**, *438*, 318–324.

(3) Menetret, J.-F.; Schaletzky, J.; Clemons, W. M. J.; Osborne, A. R.; Skanland, S. S.; Denison, C.; Gygi, S. P.; Kirkpatrick, D. S.; Park, E.; Ludtke, S. J.; Rapoport, T. A.; Akey, C. W. *Mol. Cell* **2007**, *28*, 1083–1092.

(4) Becker, T.; Bhushan, S.; Jarasch, A.; Armache, J.-P.; Funes, S.; Jossinet, F.; Gumbart, J.; Mielke, T.; Berninghausen, O.; Schulten, K.; Westhof, E.; Gilmore, R.; Mandon, E. C.; Beckmann, R. *Science* **2009**, *326*, 1369–1373.

(5) Park, E.; Ménétret, J.-F.; Gumbart, J. C.; Ludtke, S. J.; Li, W.; Whynot, A.; Rapoport, T. A.; Akey, C. W. *Nature* **2013**, *506*, 102–106.

(6) Bischoff, L; Wickles, S; Berninghausen, O; van der Sluis, E. O.; Beckmann, R *Nat Commun* **2014**, *5*, 4103.

(7) Frauenfeld, J.; Gumbart, J.; van der Sluis, E. O.; Funes, S.; Gartmann, M.; Beatrix, B.; Mielke, T.; Berninghausen, O.; Becker, T.; Schulten, K.; Beckmann, R. *Nat. Struct. Mol. Biol.* **2011**, *18*, 614–621.

(8) Gogala, M.; Becker, T.; Beatrix, B.; Armache, J.-P.; Barrio-Garcia, C.; Berninghausen, O.; Beckmann, R. *Nature* **2014**, *506*, 107–10.

(9) Voorhees, R. M.; Fernández, I. S.; Scheres, S. H. W.; Hegde, R. S. *Cell* **2014**, *157*, 1632–1643.

(10) Pfeffer, S.; Burbaum, L.; Unverdorben, P.; Pech, M.; Chen, Y.; Zimmermann, R.; Beckmann, R.; Förster, F. *Nat. Commun.* **2015**, *6*, 8403.

(11) Li, L.; Park, E.; Ling, J.; Ingram, J.; Ploegh, H.; Rapoport, T. A. *Nature* **2016**, *531*, 395–399.

(12) Berg, B. V. D.; Clemons Jr, W. M.; Collinson, I.; Modis, Y.; Hartmann, E.; Harrison, S. C.; Rapoport, T. A.; Van den Berg, B.; Clemons, W. M. *Nature* **2004**, *427*, 36–44.

(13) Von Heijne, G. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 909–918.

(14) Higy, M.; Junne, T.; Spiess, M. *Biochemistry* **2004**, *43*, 12716–12722.

(15) Devaraneni, P. K.; Conti, B.; Matsumura, Y.; Yang, Z.; Johnson, A. E.; Skach, W. R. *Cell* **2011**, *146*, 134–147.

(16) Kocik, L.; Junne, T.; Spiess, M. *J. Mol. Biol.* **2012**, *424*, 368–378.

(17) Vermeire, K.; Bell, T. W.; Van Puyenbroeck, V.; Giraut, A.; Noppen, S.; Liekens, S.; Schols, D.; Hartmann, E.; Kalies, K. U.; Marsh, M. *PLoS Biol.* **2014**, *12*, e1002011.

(18) Mothes, W; Heinrich, S. U.; Graf, R; Nilsson, I; von Heijne, G; Brunner, J; Rapoport, T. a. *Cell* **1997**, *89*, 523–533.

(19) Jungnickel, B.; Rapoport, T. A. *Cell* **1995**, *82*, 261–270.

(20) Shaw, A. S.; Rottier, P. J.; Rose, J. K. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 7592–6.

(21) Rapoport, T. A. *Nature* **2007**, *450*, 663–669.

(22) Denzer, A. J.; Nabholz, C. E.; Spiess, M. *EMBO J.* **1995**, *14*, 6311–6317.

(23) Wahlberg, J. M.; Spiess, M. *J. Cell Biol.* **1997**, *137*, 555–562.

(24) Lundin, C.; Kim, H.; Nilsson, I.; White, S. H.; von Heijne, G. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 15702–15707.

(25) Goder, V.; Spiess, M. *FEBS Lett.* **2001**, *504*, 87–93.

(26) Goder, V.; Spiess, M. *EMBO J.* **2003**, *22*, 3645–3653.

(27) Von Heijne, G. *Nature* **1989**, *341*, 456–458.

(28) Beltzer, J. P.; Fiedler, K.; Fuhrer, C.; Geffen, I.; Handschin, C.; Wessels, H. P.; Spiess, M. *J. Biol. Chem.* **1991**, *266*, 973–978.

(29) Parks, G. D.; Lamb, R. A. *Cell* **1991**, *64*, 777–787.

(30) Kida, Y.; Morimoto, F.; Mihara, K.; Sakaguchi, M. *J. Biol. Chem.* **2006**, *281*, 1152–1158.

(31) Rothman, R. E.; Andrews, D. W.; Calayag, M. C.; Lingappa, V. R. *J. Biol. Chem.* **1988**, *263*, 10470–10480.

(32) Zhang, B.; Miller III, T. F. *J. Am. Chem. Soc.* **2012**, *134*, 13700–13707.

(33) Hessa, T.; Kim, H.; Bihlmaier, K.; Lundin, C.; Boekel, J.; Andersson, H.; Nilsson, I.; White, S. H.; von Heijne, G. *Nature* **2005**, *433*, 377–381.

(34) Van Lehn, R. C.; Zhang, B.; Miller, T. F. *Elife* **2015**, *4*, 1–23.

(35) Staple, D. B.; Payne, S. H.; Reddin, A. L. C.; Kreuzer, H. J. *Phys. Rev. Lett.* **2008**, *101*, 1–4.

(36) Hanke, F.; Serr, a.; Kreuzer, H. J.; Netz, R. R. *EPL (Europhysics Lett.* **2010**, *92*, 53001.

(37) Maffeo, C.; Ngo, T. T. M.; Ha, T.; Aksimentiev, A. *J. Chem. Theory Comput.* **2014**, *10*, 2891–2896.

(38) Drukker, K.; Schatz, G. C. *J. Phys. Chem. B* **2000**, *104*, 6108–6111.

(39) Sales-Pardo, M.; Guimerà, R.; Moreira, A. A.; Widom, J.; Amaral, L. A. N. *Phys. Rev. E* **2005**, *71*, 1–13.

(40) Spitzer, J.; Poolman, B. *Microbiol. Mol. Biol. Rev.* **2009**, *73*, 371–88.

(41) Ando, T.; Skolnick, J. *Proc. Natl. Acad. Sci.* **2010**, *107*, 18457–18462.

(42) Zhang, B.; Miller, T. F. *Proc. Natl. Acad. Sci.* **2010**, *107*, 5399–5404.

(43) Zimmer, J.; Nam, Y.; Rapoport, T. A. *Nature* **2008**, *455*, 936–43.

(44) Park, M. S.; Ma, S. B.; Lee, D. J.; Im, D.; Doo, S.-G.; Yamamoto, O. *Sci. Rep.* **2014**, *4*, 3815.

(45) Ge, Y.; Draycheva, A.; Bornemann, T.; Rodnina, M. V.; Wintermeyer, W. *Nat. Commun.* **2014**, *5*, 5263.

(46) Egea, P. F.; Stroud, R. M. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 17182–17187.

(47) Wimley, W. C.; White, S. H. *Nature* **1996**, *3*, 842–848.

(48) White, S. H.; Wimley, W. C. *Ann. Rev. Biophys. Biomol. Str.* **1999**, *28*, 319–365.

(49) Ben-Tal, N.; Ben-Shaul, A.; Nicholls, A.; Honig, B. *Biophys. J.* **1996**, *70*, 1803–1812.

(50) Cymer, F.; Von Heijne, G.; White, S. H. *J. Mol. Biol.* **2015**, *427*, 999–1022.

(51) MacCallum, J. L.; Tieleman, D. P. *Trends Biochem. Sci.* **2011**, *36*, 653–662.

(52) De Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schäfer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. *J. Chem. Theory Comput.* **2013**, *9*, 687–697.

(53) Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. *PLoS Comput. Biol.* **2010**, *6*, 1–17.

(54) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(55) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. *Comput. Phys. Comm.* **2009**, *180*, 1961–1972.

(56) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comp. Chem.* **1992**, *13*, 1011–1021.

(57) Gumbart, J.; Chipot, C.; Schulten, K. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 3596–3601.

(58) Beutler, T. C.; Mark, A. E.; Schaik, R. C. V.; Gerber, P. R.; Gunsteren, W. F. V. **1994**, *222*, 529–539.

(59) Hessa, T.; Meindl-Beinker, N. M.; Bernsel, A.; Kim, H.; Sato, Y.; Lerch-Bader, M.; Nilsson, I.; White, S. H.; von Heijne, G. *Nature* **2007**, *450*, 1026–1030.

(60) The UniProt Consortium *Nucleic Acids Res.* **2014**, *43*, D204–12.

(61) White, S. H.; von Heijne, G. *Annu. Rev. Biophys.* **2008**, *37*, 23–42.

(62) Bowie, J. U. *Curr. Opin. Struct. Biol.* **2011**, *21*, 42–49.

(63) Bhushan, S.; Gartmann, M.; Halic, M.; Armache, J.-P.; Jarasch, A.; Mielke, T.; Berninghausen, O.; Wilson, D. N.; Beckmann, R. *Nat. Struct. Mol. Biol.* **2010**, *17*, 313–317.

(64) Jones, D. T. *J. Mol. Biol.* **1999**, *292*, 195–202.

(65) Buchan, D. W. A.; Minneci, F.; Nugent, T. C. O.; Bryson, K.; Jones, D. T. *Nucleic Acids Res.* **2013**, *41*, 349–357.

(66) Boresch, S; Tettinger, F; Leitgeb, M; Karplus, M *J. Phys. Chem. B* **2003**, *107*, 9535–9551.

(67) Mobley, D. L.; Chodera, J. D.; Dill, K. A. *J. Chem. Phys.* **2006**, *125*, 0–16.

(68) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245–268.

(69) Pohorille, A.; Jarzynski, C.; Chipot, C. *J. Phys. Chem. B* **2010**, *114*, 10235–10253.

*Chapter 4*

# DEVELOPMENT OF A SIMULATION MODEL TO STUDY LITHIUM DENDRITE GROWTH IN SECONDARY BATTERIES

## 4.1 Introduction

Lithium metal is an attractive battery anode material due to its potential for high energy density storage applications [1–3]. The direct reduction of the highly electropositive lithium ions onto the metal surface offers a theoretical energy density of 3860 mAh $g^{-}1$ [4, 5], significantly higher than intercalation materials used in lithium-ion batteries. However the development of lithium metal batteries in secondary (rechargeable) batteries has been hindered by the propensity of lithium metal to form dendritic depositions during the battery recharging process [6–10]. The growth of these lithium dendrites leads to battery degradation and loss of cycling efficiency via side reactions with the electrolyte and fragmentation of the metal surface; as well as more extreme failure mechanisms through battery shorting and thermal runaway [8, 11, 12].

The molecular mechanisms associated with dendrite initiation and growth have yet to be fully elucidated [2], and improved understanding is needed to guide development of dendrite mitigation strategies. Experimental studies have shown that formation of dendritic growth during battery recharging is dependent on a wide range of factors including current density [7, 8, 13, 14], electrolyte composition [15–19], additives [20, 21], mechanical properties of electrode-electrolyte interface [22, 23], temperature [24], and cycling protocol [4, 25–27]. In particular, these studies revealed the importance of a surface film formed during the initial cycling steps. This film, first described by Peled and coworkers [28], often called the solid electrolyte interphase (SEI) layer is an ionically conducting but electrically insulating layer that forms from the reaction of the commonly used organic electrolytes with the lithium metal surface [29]. The SEI layer has a protective effect by preventing further reactions with the electrolyte and has been shown to be critical for stable cycling [29]. The dependence on these electrochemical, mechanical and kinetic properties highlight the need to incorporate the effects of nonequilibrium dynamics and heterogenous electrolyte environment into models of dendrite formation in the lithium metal battery.

Computational approaches for studying dendrite growth have commonly included (i) continuum methods to model the evolution of the electrode-electrolyte surface [30–36], (ii) calculation of concentration gradients of the ions [7, 37], (iii) and detailed density functional theory (DFT) calculations to examine the stability of lithium metal atoms [38, 39]. However, there have been few atomistic simulation studies of the lithium deposition and dissolution process. Studying the growth of lithium dendrites poses a unique challenge to atomistic simulation methods due to the complexity of the interfacial electrode-electrolyte system, long timescales of battery cycling processes, and long length-scales inherent in studying the micron-sized dendrite growth.

Early atomistic simulation studies of metal electrodeposition used reaction-diffusion type models to understand the effects of current density on the deposited structure [40–43]. Previous work in the Miller group developed a particle-based reaction-diffusion model of a lithium metal anode that studied the effects of different pulse-charging waveforms on deposition morphology. Other work has included the effects of electric field concentration at the tips and diffusion of the plated Li atoms [24, 44, 45]. Studies have also established the connection between the continuum and particle based approaches [46, 47].

In this work, we extend the previously developed reaction-diffusion CG model of lithium deposition [48] to model the lithium electrodissolution process that results in unplating of lithium metal atoms. Experimental studies have shown that the configuration of the metal surface at the end of the dissolution step impacts the structure of the metal deposits during the subsequent plating step [49–53]. Furthermore, addition of lithium electrodissolution allows for modeling of both the plating and unplating processes that contribute to loss of cycling efficiency. Experimental studies have observed the ionic mass transfer away from the surface during electrodissolution is found to follow simple diffusion models [54] suggesting that the physics underlying dissolution can similarly be modeled within the reaction-diffusion framework. We also extend the model to explicitly capture the heterogeneity of the SEI layer to observe the effect of mechanical barriers, which has been shown experimentally to inhibit dendrite growth [22, 23, 55]. Simulations performed here provide insight to the effect of current density, pulse charging, reverse pulse charging, and nanoparticle additives in the inhibition of dendritic growth and improvement of cycling efficiency. We leverage these insights to study optimal charging protocols for the prevention of dendrite growth while balancing

the practical need for fast charging.

## 4.2 Methods

We extend a previously developed CG model of lithium dendrite formation which enabled the simulations of the long length-scales and timescales relevant to dendrite formation [48]. We retain the key features of the previous developed CG model for lithium deposition including i) off-lattice dynamics of the lithium ion at the electrode interface where the interaction with counterions and electrolyte is treated implicitly, ii) stochastic reaction events associated with electron-transfer at the electrode surface and plating of the lithium atoms, iii) a particle bath that reproduces the bulk cation concentration far from the electrode surface. This model has shown to predict the effect of current density and pulse-charging waveforms on the dendrite formation propensity [48, 56]. However, the previous model only studied the reduction of lithium ions during the recharging process in a homogeneous electrolyte environment. In the following section, we describe the extension of the CG model to incorporate dissolution of the lithium metal surface with the applications of a reverse potential. This allows for the modeling of a cycling protocol and for the examination of diverse pulse sequences which include short reverse pulses. Furthermore, we introduce barrier particles to model the inhomogeneity in the electrolyte environment from nanostructured artificial SEI layers which have shown great promise in inhibiting dendritic growth [22, 23].

### System

We model lithium deposition and dissolution at the planar electrode surface of a lithium metal battery (Figure 4.1). All simulations are performed using a simulation box of size $d_s$ = 24 nm and with periodic boundary conditions in the x and y coordinates. The box has a height $d_h$ = 12 nm to a particle bath region of an additional 1 nm. Lithium cations (Figure 4.1, red) in the SEI and metal atoms (Figure 4.1, black) that are formed via reductive deposition of the cations are explicitly represented.

The SEI is an electrically insulating but ionically conducting surface film formed from the reaction of the organic electrolytes with the lithium metal surface [28], and has been reported to have thickness between 5 to 50 nm [57–59] via atomic force microscopy, X-ray photon spectroscopy, and scanning probe microscopy. The lithium ions are believed to lose their electrolyte solvation shell in bulk solution and migrate through the SEI layer through Shottkey vacancies to reach the electrode

surface [29]. In all simulations, the SEI is assumed to extend uniformly through the simulation cell. The particle bath region (Figure 4.1, green) keeps the bulk concentration of ions in the region $d_{\mathrm{h}}$ from the surface at a constant concentration $C_{\mathrm{Li}}$. For the indicated simulations, we introduce barrier particles to model the effect of artificial SEI layers that provide mechanical resistance to dendrite formation.

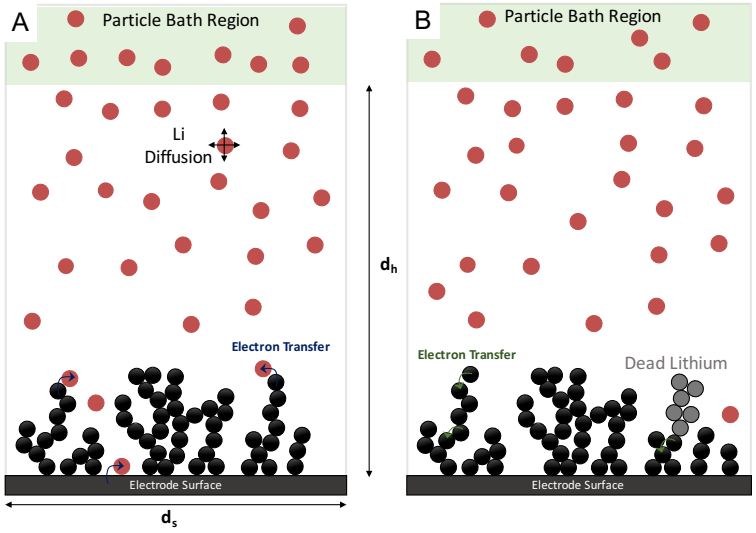



Figure 4.1: CG modeling of dendrite formation in a lithium metal battery. A. Lithium ion plating events after electron transfer shown in blue. B. Lithium dissolution events after electron transfer shown in green. Formation of detached lithium after dissolution shown in gray.

### Ion and blocker interactions and dynamics

Lithium metal ions evolve according to an overdamped Brownian dynamics which is simulated using the forward Euler integration scheme shown in Equation 4.1.

$$\mathbf{r_i}(t + \Delta t) = \mathbf{r_i}(t) + (2D\Delta t)^{\frac{1}{2}}\mathbf{g_i}, \tag{4.1}$$

where $\mathbf{r}_i(t)$ is the 3D position vector for lithium cation $i$ at time $t$, $\mathbf{g}_i$ is a 3D vector of normally distributed random variables, $D$ is the cation diffusion coefficient, and $\Delta t$ is the time step. The ion-ion interactions are assumed to be screened at the bulk concentration $C_{\mathrm{Li}}$ and are treated as effectively non-interacting.

Barrier particles evolve according to an overdamped Brownian dynamics as shown in Equation 4.2 where

$$\mathbf{r_b}(t + \Delta t) = \mathbf{r_b}(t) - \nabla U(\mathbf{r}_b)D/k_{\mathrm{B}}T + (2D_b\Delta t)^{\frac{1}{2}}\mathbf{g_b}, \tag{4.2}$$

| parameter | value ranges |
|---|---|
| particle radius, $a$ | 0.119 nm |
| diffusion coefficient, $D_i$ | 1.40 x $10^{-14}$ $m^2$ $s^{-1}$ |
| concentration, $C_i$ | 0.174 M |
| time step, $\Delta t$ | 2.53 x $10^{-7}$ s |
| reaction probability, $p$ | 0.001-1 |
| dissolution probability, $p_d$ | 0.0001-1 |
| bond energy, $k_n$ | 10 $k_B$T |
| particle bath timescale, $\tau$ | 2.53 x $10^{-3}$ s |
| anode on time, $t_{on}$ | 0.253 $\mu$s - 25.3 s |
| anode off time, $t_{off}$ | 0.253 $\mu$s - 25.3 s |
| anode reverse time, $t_{rev}$ | 0.253 $\mu$s- 0.253 s |
| blocker diffusion coefficient, $D_b$ | 0.001 $D_i$ - 0.1 $D_i$ |
| blocker length, $r_b$ | 11.9 nm |

Table 4.1: Summary of parameters for the lithium metal battery cycling model and their studied ranges.

where $\mathbf{r}_b(t)$ is the three-dimensional position vector for barrier particle $b$ at time $t$, $\mathbf{g}_b$ is a 3D vector of normally distributed random variables, $D_b$ is the barrier particle diffusion coefficient, and $\Delta t$ is the time step and $U(\mathbf{r}_b)$ is defined in Equation 4.3.

$$U(\mathbf{r}_b) = \sum_{a\neq b}^{a=N_b} 4\epsilon \left[ \frac{1}{(\mathbf{r}_b - \mathbf{r}_a)^{12}} - \frac{1}{(\mathbf{r}_b - \mathbf{r}_a)^{6}} \right] + \sum_{m=0}^{N_m} 4\epsilon \left[ \frac{1}{(\mathbf{r}_b - \mathbf{r}_m)^{12}} - \frac{1}{(\mathbf{r}_b - \mathbf{r}_m)^{6}} \right], \tag{4.3}$$

Here, a barrier particle, $b$, interacts with all other barrier particles, $a$, and plated lithium metal atoms, $m$, according to a WCA potential [60]. Ions are allowed to freely diffuse through the blockers as is expected for an ionically conductive SEI layer [23].

The parameters and system geometry described here are designed to model a proposed battery coin cell architecture such as that described by Yang and coworkers [4] in a lithium hexafluorophosphate in ethylene carbonate/dimethyl carbonate electrolyte medium.

**Lithium reduction and oxidation at the electrode**

Electrodeposition of lithium cations at the metal electrode occurs after an electron-transfer event to the electrode interface that causes reduction of the cation. We model this event by a stochastic process that is attempted when an MD move gen-

erated by Equation 4.1 results in an ion at a distance, $r$, from the electrode surface or a previously reduced lithium atom, which is less than its radius, $a$ (Figure 4.1A). The reduction event is accepted with a probability $p$ given by Equation 4.4:

$$p = 1 - e^{k_{\mathrm{ET}}(r,\eta)\Delta t}. \tag{4.4}$$

This reaction probability $p$ takes into account the material-dependent properties of the battery via the rate of electron transfer $k_{\mathrm{ET}}$ defined in Equation 4.5 which also depends on the applied overpotential, $\eta$:

$$k_{\mathrm{ET}}(r,\eta) = \tau_{\mathrm{sol}}^{-1} k_{\mathrm{el}}(r,\eta) e^{-\Delta G^{\ddagger}(r,\eta)/k_{\mathrm{B}}T}. \tag{4.5}$$

Here, the prefactor terms $\tau_{\mathrm{sol}}$ and $\kappa_{\mathrm{el}}(r,)$ are related to the timescale for solvent fluctuations and the probability of reactive electron tunneling, respectively. $\Delta G^{\ddagger}(r,\eta)$ is the free energy for the solvent reorganization to the electron-transfer transition state. Previous studies have also proposed alternate interpretations for the stochastic reduction probabilities [41–43, 61].

If the reduction attempt fails, then the position update from Equation 4.1 is rejected, and the cation is left in its previous position. If the reduction attempt is successful, the lithium cation is then switched to a reduced lithium metal atom and deposited on the electrode or other metal atoms. Upon deposition, the position of the new metal atom is adjusted to avoid overlaps and account for the finite volume of the metal atoms [42, 48] via the following protocol. If the new metal atom overlaps with the electrode, then the position of the atom is shifted outward until the overlap is eliminated. If the volume of the new metal atom overlaps with that of another metal atom, the position of the new metal atom is shifted along the vector $r_{new} - r_{old}$ until the overlap is eliminated; where $r_{new}$ and $r_{old}$ indicate the respective positions of the new and previously plated metal atoms. This cycle is iterated until either the newly deposited atom has no overlaps with the electrode or previously plated atoms, or if 100 cycles have been performed without eliminating all overlaps. In the second case, the reduction event is then rejected and the lithium cation is returned to its position from the previous step.

After plating, the positions of reduced metal atoms remain fixed, implicitly assuming that the timescale for diffusion of the metals on the surface is slow in comparison to the other timescales in the system. Experimental evidence confirms that lithium diffusion in metallic lithium has a diffusion coefficient of 0.006-0.31 $\mathrm{cm}^2/\mathrm{s}^{-1}$ seconds so this assumption seems reasonable [62]. Further extensions of the model

could incorporate diffusion on the surface if necessary. Position updates and reduction attempts are applied sequentially to ion particles in these simulations, where both steps are performed for a given particle before the next ion particle position is updated.

Electrochemical dissolution of lithium atoms from the metal surface is similarly the result of oxidation of lithium metal via an electron-transfer event. Dissolution at the metal surface has previously been studied via Monte Carlo simulation of 2D lattice models [63, 64] in the context of crystal dissolution into a solvent. We similarly model this event by a stochastic process that is attempted by all plated lithium atoms with a probability defined in Equation 4.6

$$p_{\text{rev}} = (1 - e^{k_{\text{ET}}(\eta)\Delta t})e^{-E_{\text{bond}}n_{\text{bond}}}. \tag{4.6}$$

This reaction probability $p_{\text{rev}}$ has the same form as the plating probability $p$ where the material-dependent properties of the battery are accounted for by the rate of electron transfer $k_{\text{ET}}$. This rate also accounts for the effect of the applied overpotential, $\eta$. The rate is also dependent on the number of bonds, $n_{\text{bond}}$, defined as all atoms in contact with the given atom and a bond energy $E_{\text{bond}}$ that describes the favorable interactions in the lattice between neighboring atoms [38, 39, 65].

If the dissolution step is rejected, no changes to the atom position or state are made. If the dissolution attempt is successful, the lithium metal atom is then switched to a lithium ion at the same position and resumes dynamics according to Equation 4.1 in the following step. Following a successful dissolution step, the connectivity of its remaining neighbors is checked for attachment to the surface (Figure 4.1B). If connectivity is not maintained, the number of particles in the disconnected cluster of lithium metal atoms $n_{\text{cluster}}$ is counted. We define a minimum nucleus size $N_{\text{dead}} = 5$. For clusters, which $n_{\text{cluster}} \geq N_{\text{dead}}$, the lithium metal particles are switched to an electrically disconnected state where they no longer participate in either reduction or oxidation steps. For clusters where $n_{\text{cluster}} < N_{\text{dead}}$, the lithium metal is assumed to be dissolved as well and returned to the lithium ion state at the same position. This formation of disconnected lithium metal clusters that no longer participate has been observed experimentally and causes a loss of cycling efficiency in the battery due to the loss of active material [13, 51, 66–68]. The choice of $N_{\text{dead}}$ has not been found to affect the qualitative results of shown here, but future calculations can be done to obtain an accurate estimation of this parameter.

For simulations that include barrier particles, the rate of electron transfer to cations

in the barrier particle is greatly reduced. This is reflected in the reaction probability $p$, which is set to 0 for ions diffusing in a barrier particle. This accounts for the fact that these barriers are typically electrically insulating, preventing electron transfer reactions at the metal surface [21, 23, 28].

**Battery charging and cycling trajectories**

Battery charging trajectories are run for a series of different time-dependent voltages. In all trajectories, no plated particles are in the initial simulation cell and ions are randomly placed with uniform probability at the given ion concentration. For battery charging simulations, the trajectories are terminated when the highest plated particle reaches the bottom edge of the particle bath region. For battery cycling simulations, the trajectories are terminated after a fixed amount of time has passed. Although these trajectories only examine the relatively short-timescale dendrite initiation processes, results from the previous model [48] have shown that an analysis of the plated structures at these lengths correlated with experimental results on much longer timescales [44].

We characterize the dendrite propensity $\alpha$ of the plated structure defined as

$$\alpha = \left[ V_{\text{cell}}^{-1} \int_0^{d_a} \rho(z) \delta z \right]^{-1}, \tag{4.7}$$

where $\rho(z)$ is the number density of plated particles at height $z$, $V_{\text{cell}}$ is the volume of the simulation cell and $d_{\text{bath}}$ is the distance from the electrode surface to defined distance from the surface $d_\alpha = 1.19$ nm. We quantify the current density $i$ defined as

$$i(\Delta t) = \frac{n_{\text{plated}}(\Delta t) F}{A_{\text{cell}} \Delta t}, \tag{4.8}$$

where $n_{\text{plated}}(t)$ is the number of particles plated during the time $\Delta t$, $A_{\text{cell}}$ is the area of the electrode surface and $F$ is Faraday's constant.

## 4.3 Results and Discussion

In this work, we reexamine continuous charging and pulse charging protocols for both the plated structure, $\alpha$ and current density, $i$. We then apply the extended model to study reverse pulse charging protocols and the effect of mechanical barriers on dendrite formation. Finally, we study the loss of efficiency during battery cycling from the formation of electrically disconnected "dead" lithium.

**Lithium plating during continuous charging**

The CG model is applied to study continuous charging as a function of probability (i.e. applied voltage) as was done in the previous model [48] for $p \in \{1, 0.91, 0.5, 0.1, 0.091, 0.05, 0.01, 0.001\}$. Observations of the resulting plated structures (Figure 4.2) reveal that trajectories with high plating probabilities result in uneven, loosely-plated structures; and trajectories with low plating probabilities lead to uniform, densely-plated structures.

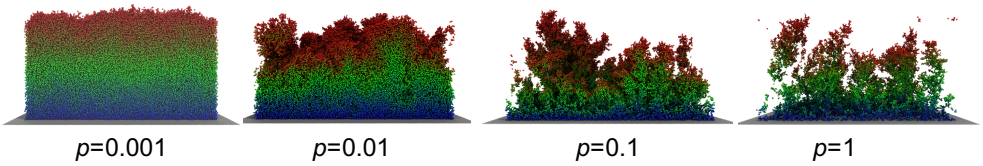



Figure 4.2: Snapshots of the plated structure from representative trajectories at the given reaction probability.

Figure 4.3 plots the dendrite propensity and the average current density as a function of the reaction probability $p$. Figure 4.3A shows that the density of the plated structure, quantified by $\alpha$, decreases with increasing probability. However, Figure 4.3B shows that the average current density, which directly indicates the total time needed to charge the cell, decreases with decreasing probability. This is in agreement with experimental studies that find that dendrite growth is correlated to the current density [7, 8]. This correlation between the dendrite propensity and the current density limits the ability to leverage low current density charging when time limitations are placed on the charging process. Figure 4.3B also reveals that for $p > 0.1$, the plating process is essentially diffusion-limited, where increasing the reaction probability has little effect on the average number of plated particles per unit time. In this regime, the plated structures are uneven and loosely plated. Conversely for low voltage charging, the plating is reaction-limited and the plated structure is uniform and dense.

**Lithium plating during pulse charging**

Next, the CG model is applied to study time-dependent voltages which can alter the balance between diffusive and reactive timescales without needing to alter the applied voltage. Here, the reaction probability $p = 1$ for all times $t_{\text{on}}$ and $p = 0$ for all times $t_{\text{off}}$. The ratio of these pulse times is given by $\gamma = t_{\text{off}}/t_{\text{on}}$. Figure 4.4 plots the dendrite propensity $\alpha$ and average current $i$ for a series of pulse sequences

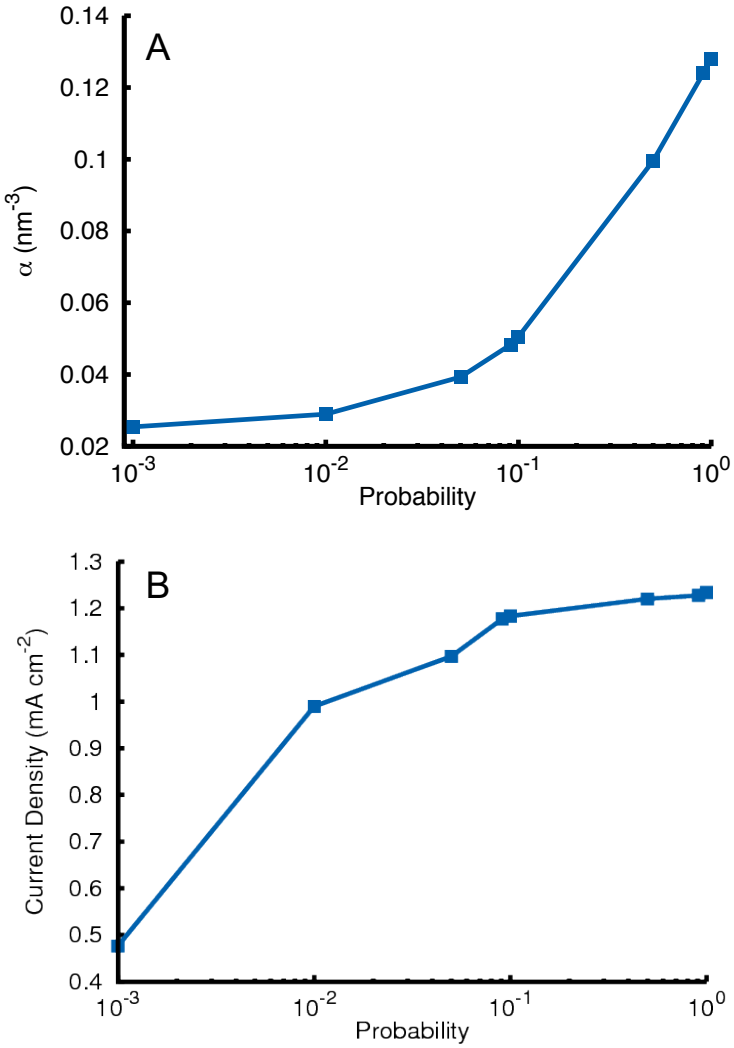



Figure 4.3: Dendrite formation and current density during continuous plating. A. Dendrite propensity $\alpha$ during continuous charging as a function of reaction probability $p$. B. Current density during continuous charging as a function of reaction probability $p$.

where $t_{\text{on}} \in \{0.253\ \mu\text{s}, 2.53\ \mu\text{s}, 25.3\ \mu\text{s}, 253\ \mu\text{s}, 2.53\ \text{ms}, 25.3\ \text{ms}, 253\ \text{ms}\}$ and $\gamma \in \{0.1, 1, 10, 100\}$.

Figure 4.4A shows that, at long pulse lengths, the addition of off pulses by increaseing $\gamma$ makes little difference on the plated structure. However, as the pulse length decreases and $\gamma$ increases, the dendrite propensity is reduced leading to more uniform growth. In the infinitely short pulse-length limit, the effect of pulsing becomes identical to continuous charging with a lowered reaction probability equal to $1/\gamma$.

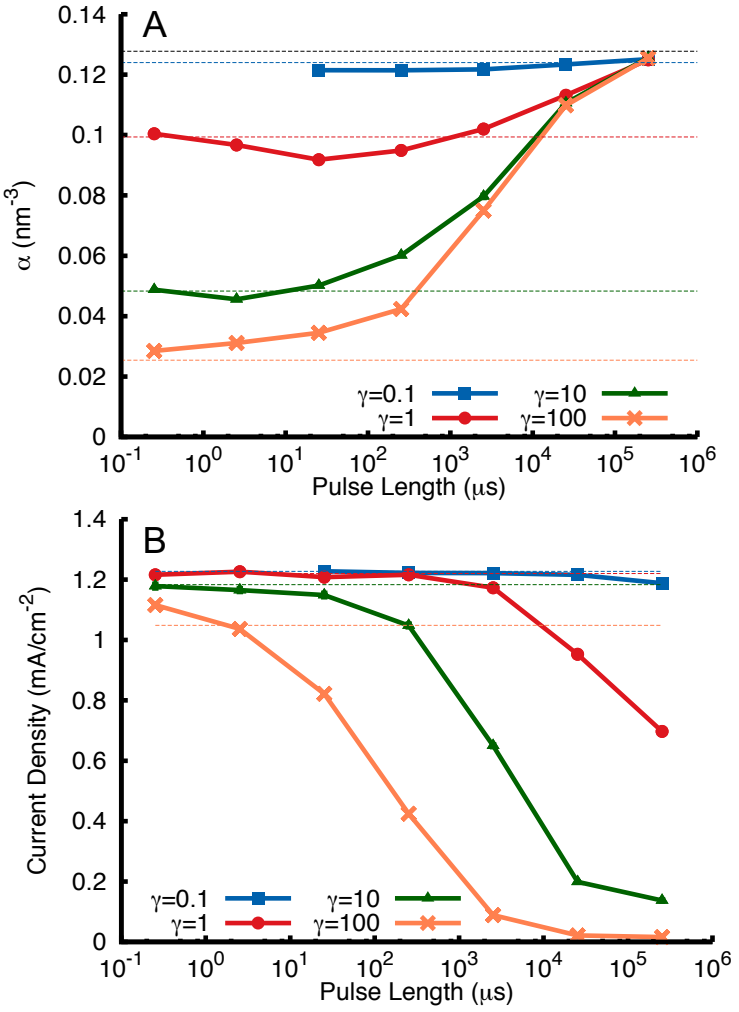



Figure 4.4: Dendrite formation and current density during pulse plating. A. Dendrite propensity $\alpha$ during pulse charging as a function of $t_{\text{on}}$ pulse length. The reference result for infinitely fast pulsing corresponds to a reaction probability with the given $\gamma$ value and is shown for each pulse ratio in dotted lines. B. Current density $i$ during pulse charging as a function of $t_{\text{on}}$ pulse length.

This limit is shown in dotted lines in Figure 4.4A. This analysis allows us to identify certain pulse sequences (e.g. $\gamma = 1$, $t_{\text{on}} = 25$ $\mu$s) where pulsing improves upon the plating in the infinitely short pulse-length limit. These results are in agreement with the previous models and experiments [44, 48].

Figure 4.4B plots the average current density during pulse charging and shows that, for the given parameters, decreasing the pulse length leads to increased current den-

sity and faster charging. By going to extremely short pulses, uniform growth can be obtained without sacrificing a relatively high current density. However these pulse lengths are quite short ($\mu s$) and pose a challenge to implementation in a realistic battery system [69].

**Lithium plating during reverse pulse charging**

The extended CG model is then applied to study reverse pulse charging. Reverse pulse charging replaces the time $t_{\text{off}}$, from pulse charging where no voltage is applied, with a reverse voltage for time $t_{\text{rev}}$, leading to lithium metal electrodissolution. For $t_{\text{rev}} < t_{\text{on}}$, this can lead to net charging. This has been proposed by Yang et al. [4] as a potential strategy to improve cycling efficiency and reduce dendrite growth without needing to wait for long times $t_{\text{off}}$.

Figure 4.5 shows the results for trajectories using these reverse pulse sequences with $t_{\text{on}} \in \{0.0253 \text{ s}, 0.253 \text{ s}, 2.53 \text{ s}, 25.3 \text{ s}\}$ and with varying ratios $\gamma \in \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}\}$, where $\gamma = t_{\text{rev}}/t_{\text{on}}$. These ratios are much smaller than unity in order to lead to a net plating of the lithium ions. The reaction probability $p = 1$ during $t_{\text{on}}$ pulses, and the reverse reaction probability $p_{\text{rev}} = 1$ for the $t_{\text{rev}}$ pulses.

Figure 4.5A reveals that the dendrite propensity decreases with both decreasing pulse length and increasing ratio $\gamma$. This is analogous to the trends seen for pulse charging. However, the mechanism for decreasing dendrite propensity is not the result of increasing time for diffusion since $t_{\text{rev}}$ is very short. Instead, the improvement is due to the removal of plated particles with few neighbors. The reverse pulse allows for the removal of these particles and allows them to diffuse to a new location with additional neighbors. By biasing the position of the plated atoms to positions with additional neighbors, the final plated structure is more uniform and densely packed. By increasing the ratio, $\gamma$, more particles are removed during the $t_{\text{rev}}$ step relative to the $t_{\text{on}}$ step. Shorter $t_{\text{on}}$ time leads to fewer particles plated per pulse and less dead lithium formation, which also leads to a lower dendrite propensity.

Promisingly, the effective current density shown in Figure 4.5B is not as sensitive to either the pulse length or $\gamma$, as was seen for pulse charging. Reverse pulse charging appears to be a promising way to increase the density of the plating while maintaining a high average current density. Further studies will perform tests for the effects of the bond strength parameter, $k_n$, and the crystal size parameter, $N_{\text{dead}}$. These parameters may also be experimentally modified with the addition of additives or through the use of microstructured supports [20, 55].

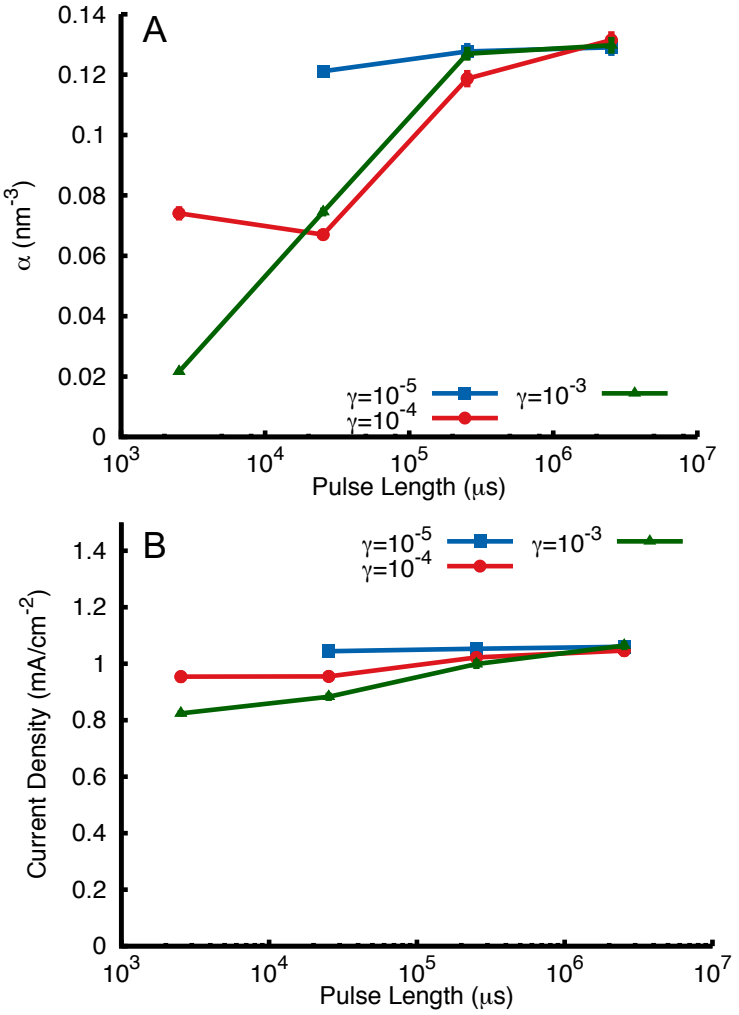



Figure 4.5: Dendrite formation and current density during reverse pulse plating. A. Dendrite propensity $\alpha$ during reverse pulse charging as a function of $t_{on}$ pulse length. B. Current density $i$ during reverse pulse charging as a function of $t_{on}$ pulse length.

**Lithium plating in the presence of a mechanical barrier**

The model is applied to simulate a mechanical barrier that acts as an electrically insulating but ionically conducting layer as shown in Figure 4.6A. Here the barrier particle spans the area of the simulation cell and allows for ion diffusion through the particle but prevents electron transfer processes [23]. Figure 4.6B plots the effect of barrier diffusion coefficient on the dendrite propensity for $D_B \in \{0.1D_0,$

$0.01D_0$, $0.001D_0$}. Slower diffusion coefficients, corresponding to more mechanically stiff (higher modulus) barriers, lead to less dendrite growth in agreement with experimental observations [22, 32].

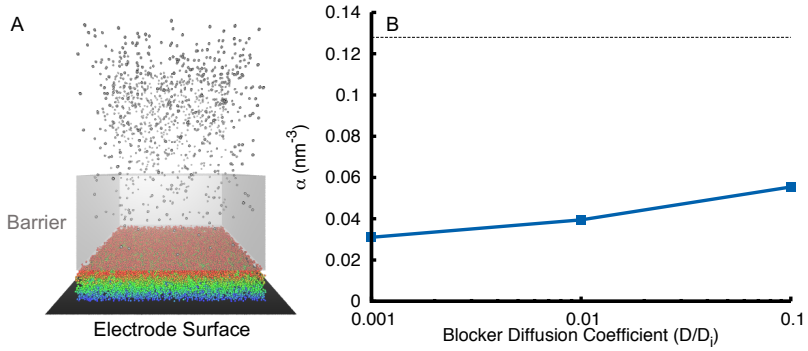



Figure 4.6: Plating with mechanical barrier. A. Snapshot of a plating trajectory with a barrier particle shown in transparent gray volume. The barrier particle is ionically conductive to lithium ion particles shown in gray spheres but electrically insulating. Plated particles are colored from blue to red as a function of their plating time. B. Dendrite propensity $\alpha$ as a function of the blocker diffusion coefficient, with the continuous plating reference without any blocker particles shown in dashed line.

### Quantifying efficiency losses from dead lithium

One contributing factor to the loss of cycling efficiency in batteries is the electrical disconnection of lithium metal particles [13, 52, 56]. This effect can be quantified by studying cycling trajectories where the ratio of $t_{\text{on}}$ to $t_{\text{rev}}$ time is equal, corresponding to repeated charge-discharge cycles of battery operation. This has been studied experimentally [27, 56], and it has been observed that the fraction of dead lithium increases as the cycle length increases. Figure 4.7 plots the fraction of dead lithium after cycling. For these simulations, the reaction probabiltiy $p = 1$ for $t_{\text{on}}$ and $p_{\text{rev}} = 1$ for $t_{\text{rev}}$. In the simulations, it is found that increasing cycle length leads to an increase in the fraction of dead lithium.

However the increase in fraction of dead lithium is much larger than expected, going up to as high is 60 % for the longest studied cycle lengths. This suggests that factors that prevent the formation of dead lithium, including diffusion effects of lithium metal atoms and possible reconnection of previously disconnected particles, are missing from the current model. These features will be implemented and explored in future models.

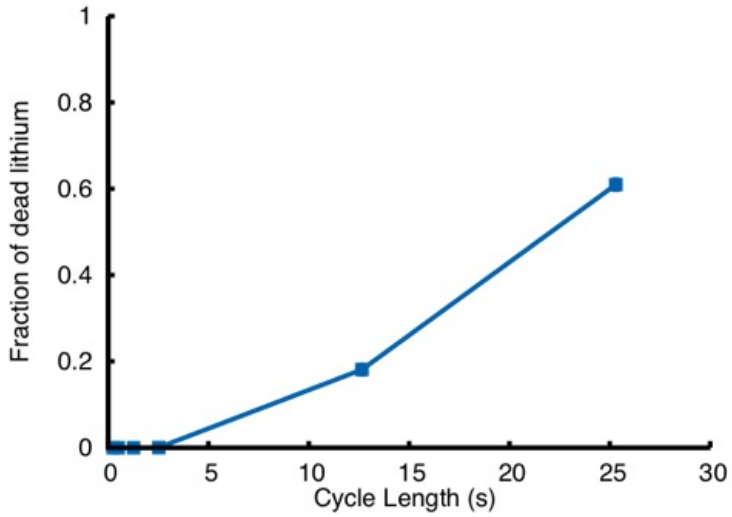



Figure 4.7: (A) Fraction of dead lithium as a function of cycle length

## 4.4 Conclusions

Understanding the factors that contribute to dendrite growth in lithium metal batteries is critical to development of this battery chemistry for commercial rechargeable applications. We've extended a previously developed particle-based simulation model to incorporate both deposition and dissolution of lithium ions onto the surface. This allows for charging protocols that use reverse pulse charging methods which lead to more uniform, dense growth even at relatively long $t_{on}$ pulses. Furthermore, the model is applied to study the effect of mechanical barriers on dendrite growth and found that slower diffusing particles have a stronger impact on the deposited structure. Finally, the model allows for the simulation of the loss of cycling efficiency due to formation of electrically disconnected lithium from the surface. This model provides a computationally efficient and flexible model that can be systematically improved though more detailed DFT or molecular dynamics calculations to have a more accurate treatment of surface energetics, dissolution and deposition rates, and effects of novel additives and SEI heterogeneity.

## References

(1) Tarascon, J. M.; Armand, M *Nature* **2001**, *414*, 359–367.

(2) Li, Z.; Huang, J.; Yann Liaw, B.; Metzler, V.; Zhang, J. *J. Power Sources* **2014**, *254*, 168–182.

(3) Choi, J. W.; Aurbach, D. *Nat. Rev. Mater.* **2016**, *1*, 1–16.

(4) Yang, H.; Fey, E. O.; Trimm, B. D.; Dimitrov, N.; Whittingham, M. S. *J. Power Sources* **2014**, *272*, 900–908.

(5) Vaughey, J. T.; Liu, G.; Zhang, J.-G. *MRS Bull.* **2014**, *39*, 429–435.

(6) Brissot, C.; Rosso, M.; Chazalviel, J.; Baudry, P.; Lascaud, S. *Electrochim. Acta* **1998**, *43*, 1569–1574.

(7) Brissot, C.; Rosso, M.; Chazalviel, J. N.; Lascaud, S. *J. Power Sources* **1999**, *81*, 925–929.

(8) Aurbach, D.; Zinigrad, E.; Teller, H.; Dan, P. *J. Electrochem. Soc.* **2000**, *147*, 1274–1279.

(9) Rosso, M; Gobron, T; Brissot, C; Chazalviel, J; Lascaud, S *J. Power Sources* **2001**, *98*, 804–806.

(10) Rosso, M.; Brissot, C.; Teyssot, A.; Dollé, M.; Sannier, L.; Tarascon, J. M.; Bouchet, R.; Lascaud, S. *Electrochim. Acta* **2006**, *51*, 5334–5340.

(11) Aurbach, D.; Zinigrad, E.; Cohen, Y.; Teller, H. *Solid State Ionics* **2002**, *148*, 405–416.

(12) Amine, K.; Kanno, R.; Tzeng, Y. *MRS Bull.* **2014**, *39*, 395–401.

(13) Arakawa, M.; Tobishima, S.-i.; Nemoto, Y.; Ichimura, M.; Yamaki, J.-i. *J. Power Sources* **1993**, *43*, 27–35.

(14) Sano, H.; Sakaebe, H.; Senoh, H.; Matsumoto, H. *J. Electrochem. Soc.* **2014**, *161*, A1236–A1240.

(15) Crowther, O.; West, A. C. *J. Electrochem. Soc.* **2008**, *155*, A806–A811.

(16) Stark, J. K.; Ding, Y; Kohl, P. A. *J. Electrochem. Soc.* **2013**, *160*, D337–D342.

(17) Lu, Y.; Korf, K.; Kambe, Y.; Tu, Z.; Archer, L. A. **2014**, 488–492.

(18) Miao, R.; Yang, J.; Feng, X.; Jia, H.; Wang, J.; Nuli, Y. *J. Power Sources* **2014**, *271*, 291–297.

(19) Qian, J.; Henderson, W. A.; Xu, W.; Bhattacharya, P.; Engelhard, M.; Borodin, O.; Zhang, J.-G. *Nat. Commun.* **2015**, *6*, 1–9.

(20) Ding, F.; Xu, W.; Chen, X.; Zhang, J. G.; Shao, Y.; Engelhard, M. H.; Zhang, Y.; Blake, T. A.; Graff, G. L.; Liu, X.; Zhang, J. G. *J. Phys. Chem. C* **2014**, *118*, 4043–4049.

(21) Yan, K.; Lu, Z.; Lee, H.-W.; Xiong, F.; Hsu, P.-C.; Li, Y.; Zhao, J.; Chu, S.; Cui, Y. *Nat. Energy* **2016**, *1*, 16010.

(22) Stone, G. M.; Mullin, S. A.; Teran, A. A.; Hallinan, D. T.; Minor, A. M.; Hexemer, A.; Balsara, N. P. *J. Electrochem. Soc.* **2012**, *159*, A222–A227.

(23) Zheng, G.; Lee, S. W.; Liang, Z.; Lee, H.-W.; Yan, K.; Yao, H.; Wang, H.; Li, W.; Chu, S.; Cui, Y. *Nat. Nanotechnol.* **2014**, *9*, 618–623.

(24) Aryanfar, A.; Brooks, D. J.; Colussi, A. J.; Merinov, B. V.; Goddard III, W. A.; Hoffmann, M. R. *Phys. Chem. Chem. Phys.* **2015**, *17*, 8000–8005.

(25) Chandrasekar, M. S.; Pushpavanam, M. *Electrochim. Acta* **2008**, *53*, 3313–3322.

(26) Li, J.; Murphy, E.; Winnick, J.; Kohl, P. A. *J. Power Sources* **2001**, *102*, 302–309.

(27) Harry, K. J.; Hallinan, D. T.; Parkinson, D. Y.; Macdowell, A. A.; Balsara, N. P. *Nat. Mater.* **2013**, *13*, 69–73.

(28) Peled, E *J. Electrochem. Soc.* **1979**, *126*, 2047–2051.

(29) Cheng, X.-B.; Zhang, R.; Zhao, C.-Z.; Wei, F.; Zhang, J.-G.; Zhang, Q. *Adv. Sci.* **2015**, *3*, 1500213.

(30) Fleury, V.; Chazalviel, J. N.; Rosso, M. *Phys. Rev. E* **1993**, *48*, 1279–1295.

(31) Monroe, C.; Newman, J. *J. Electrochem. Soc.* **2003**, *150*, A1377.

(32) Monroe, C.; Newman, J. *J. Electrochem. Soc.* **2005**, *152*, A396.

(33) Liang, L.; Qi, Y.; Xue, F.; Bhattacharya, S.; Harris, S. J.; Chen, L. Q. *Phys. Rev. E* **2012**, *86*, 1–5.

(34) Ely, D. R.; Jana, A.; García, R. E. *J. Power Sources* **2014**, *272*, 581–594.

(35) Tikekar, M.; Archer, L.; Koch, D. *J. Electrochem. Soc.* **2014**, *161*, A847–A855.

(36) Tan, J.; Tartakovsky, A. M.; Ferris, K.; Ryan, E. M. *J. Electrochem. Soc.* **2016**, *163*, A318–A327.

(37) Chazalviel, J. N. *Phys. Rev. A* **1990**, *42*, 7355–7367.

(38) Jäckle, M.; Groß, A. *J. Chem. Phys.* **2014**, *141*, 0–7.

(39) Ozhabes, Y.; Gunceler, D.; Arias, T. A. *arXiv* **2015**, *1504.05799*, 1–7.

(40) Diggle, J. W.; Despic, A. R.; Bockris, J. O. *J. Electrochem. Soc.* **1969**, *116*, 1503.

(41) Voss, R. F.; Tomkiewicz, M. *J. Electrochem. Soc.* **1985**, *132*, 371–375.

(42) Magan, R. V.; Sureshkumar, R.; Lin, B. *J. Phys. Chem. B* **2003**, *107*, 10513–10520.

(43) Magan, R. V.; Sureshkumar, R. *Nanotechnology* **2005**, *16*, S545–S553.

(44) Aryanfar, A.; Brooks, D.; Merinov, B. V.; Goddard, W. A.; Colussi, A. J.; Hoffmann, M. R. *J. Phys. Chem. Lett.* **2014**, *5*, 1721–1726.

(45) Aryanfar, A.; Cheng, T.; Colussi, A. J.; Merinov, B. V.; Goddard, W. A.; Hoffmann, M. R. *J. Chem. Phys.* **2015**, *143*, 0–5.

(46) Nagy, G.; Sugimoto, Y.; Denuault, G. *J. Electroanal. Chem.* **1997**, *433*, 167–173.

(47) Welland, M. J.; Lau, K. C.; Redfern, P. C.; Liang, L.; Zhai, D.; Wolf, D.; Curtiss, L. A. *J. Chem. Phys.* **2015**, *143*, DOI: 10.1063/1.4936410.

(48) Mayers, M. Z.; Kaminski, J. W.; Miller, T. F. *J. Phys. Chem. C* **2012**, *116*, 26214–26221.

(49) Shiraishi, S; Kanamura, K; Takehara, Z *J. Electrochem. Soc.* **1999**, *146*, 1633–1639.

(50) Gireaud, L.; Grugeon, S.; Laruelle, S.; Yrieix, B.; Tarascon, J. M. *Electrochem. Commun.* **2006**, *8*, 1639–1649.

(51) Steiger, J.; Kramer, D.; Mönig, R. *Electrochim. Acta* **2014**, *136*, 529–536.

(52) Steiger, J.; Kramer, D.; Mönig, R. *J. Power Sources* **2014**, *261*, 112–119.

(53) Sacci, R. L.; Dudney, N. J.; More, K. L.; Parent, L. R.; Arslan, I.; Browning, N. D.; Unocic, R. R. *Chem. Commun.* **2014**, *50*, 2104.

(54) Nishikawa, K.; Fukunaka, Y.; Sakka, T.; Ogata, Y. H.; Selman, J. R. *J. Electroanal. Chem.* **2005**, *584*, 63–69.

(55) Liu, Y.; Lin, D.; Liang, Z.; Zhao, J.; Yan, K.; Cui, Y. *Nat. Commun.* **2016**, *7*, 10992.

(56) Aryanfar, A.; Brooks, D. J.; Colussi, A. J.; Hoffmann, M. R. *Phys. Chem. Chem. Phys.* **2014**, *16*, 24965–70.

(57) Xu, K. *Chem. Rev.* **2004**, *104*, 4303–4417.

(58) Lee, J. T.; Nitta, N.; Benson, J.; Magasinski, A.; Fuller, T. F.; Yushin, G. *Carbon N. Y.* **2013**, *52*, 388–397.

(59) Xiong, S.; Xie, K.; Blomberg, E.; Jacobsson, P.; Matic, A. *J. Power Sources* **2014**, *252*, 150–155.

(60) John D. Weeks, David Chandler, H. C. A. *J. Chem. Phys.* **1971**, *54*, 5237–5247.

(61) Zeng, Y.; Smith, R. B.; Bai, P.; Bazant, M. Z. *J. Electroanal. Chem.* **2014**, *735*, 77–83.

(62) Heitjans, P; Korblein, A; Ackermann, H; Dubbers, D; Fujara, F; Stockmann, H. J. *J. Phys. F Met. Phys.* **1985**, *15*, 41.

(63) Lasaga, A. C.; Luttge, A *Science (80-. ).* **2001**, *291*, 2400–2404.

(64) Alekseev, G. Y.; Alekseev, Y. V.; Bityurin, V. A. *Prot. Met.* **2005**, *41*, 221–233.

(65) Alekseev, Y. V.; Alekseev, G. Y.; Bityurin, V. A. *Prot. Met.* **2002**, *38*, 517–529.

(66) Yamaki, J.-i.; Tobishima, S.-i.; Hayashi, K.; Keiichi Saito; Nemoto, Y.; Arakawa, M. *J. Power Sources* **1998**, *74*, 219–227.

(67) Ota, M.; Izuo, S.; Nishikawa, K.; Fukunaka, Y.; Kusaka, E.; Ishii, R.; Selman, J. R. *J. Electroanal. Chem.* **2003**, *559*, 175–183.

(68) Bhattacharyya, R.; Key, B.; Chen, H.; Best, A. S.; Hollenkamp, A. F.; Grey, C. P. *Nat. Mater.* **2010**, *9*, 504–510.

(69) Bard, A. J.; Faulkner, L. R. *Electrochem. Methods, 2nd ed.; Wiley New York* **2001**.

*Appendix A*

# 3D-CG MODEL RIBOSOME COORDINATES

| *x* | *y* | *z* | *q* |
|---|---|---|---|
| -7.514784 | -1.683994 | -0.886238 | 1 |
| -8.104101 | -1.63385 | -1.649206 | -1 |
| -7.586482 | -1.685579 | -2.602041 | -1 |
| -6.807804 | -1.552141 | -2.66721 | 1 |
| -6.934504 | -0.932227 | -2.063273 | 1 |
| -7.18765 | -0.476263 | -0.88001 | 1 |
| -7.55626 | -0.161701 | 0.103331 | 0 |
| -8.111252 | 0.693923 | 0.096178 | 1 |
| -8.848511 | 0.62285 | -0.422665 | 1 |
| -8.726295 | 0.53167 | -0.967654 | 0 |
| -8.581862 | 0.502603 | -1.522917 | 0 |
| -8.457825 | 0.468627 | -2.079291 | 1 |
| -8.009384 | 0.26054 | -2.760692 | 0 |
| -7.346953 | 0.800207 | -2.699984 | 1 |
| -7.480787 | 1.183157 | -2.275319 | 1 |
| -7.708429 | 1.55205 | -1.88268 | 1 |
| -7.925132 | 1.970529 | -1.599212 | -1 |
| -8.096656 | 2.451092 | -1.720752 | 1 |
| -8.35369 | 1.756134 | -0.918598 | 1 |
| -7.90209 | 1.209179 | -0.784119 | 1 |
| -7.450047 | 0.889351 | -0.92273 | -1 |
| -7.062152 | 0.575055 | -1.221159 | 0 |
| -6.755085 | 0.266189 | -1.568044 | 1 |
| -6.435247 | -0.068135 | -1.914705 | -1 |
| -6.123912 | -0.450625 | -2.175959 | -1 |
| -5.779541 | -0.832027 | -2.385568 | 1 |
| -6.296489 | -0.319211 | -3.007501 | -1 |
| -6.896547 | 0.036104 | -3.14476 | -1 |
| -7.776091 | -0.284098 | -2.537796 | -1 |

| *x* | *y* | *z* | *q* |
|---|---|---|---|
| -8.223771 | -0.55361 | -1.583764 | 0 |
| -8.557389 | -0.388812 | -0.385669 | 1 |
| -8.371603 | -0.342078 | 0.739254 | 1 |
| -8.359682 | 0.447649 | 1.657952 | 1 |
| -8.65766 | 0.520874 | 0.941092 | 0 |
| -8.106941 | -0.114718 | -0.031991 | 0 |
| -7.753336 | -0.446516 | -1.108682 | -1 |
| -7.407497 | -0.677052 | -2.310091 | 0 |
| -7.356667 | -0.55923 | -3.468588 | -1 |
| -5.203617 | 3.362772 | 2.155831 | 1 |
| -5.057833 | 3.780104 | 2.505895 | 1 |
| -4.659746 | 4.36749 | 0.899629 | -1 |
| -4.277495 | 3.735466 | 1.011242 | -1 |
| -4.174492 | 3.505845 | 1.547386 | -1 |
| -4.094404 | 3.119773 | 2.145225 | 0 |
| -3.881707 | 3.519771 | 2.967959 | 1 |
| -3.744357 | 3.958434 | 2.637524 | 1 |
| -3.68001 | 4.408961 | 2.298991 | 1 |
| -3.473564 | 0.749758 | 4.749194 | 0 |
| -3.552078 | 1.447487 | 3.85495 | -1 |
| -3.425063 | 1.921056 | 4.114153 | 0 |
| -3.059727 | 2.233526 | 4.353235 | 1 |
| -1.975106 | 1.344479 | 4.755291 | -1 |
| -2.01388 | 0.757324 | 4.649225 | 1 |
| -2.509607 | 0.81366 | 4.173068 | 0 |
| -4.964982 | -4.756104 | 0.572706 | 1 |
| -3.784516 | -3.992066 | 2.205 | 0 |
| -4.687457 | -3.752044 | 2.793225 | 1 |
| -5.303025 | -3.610978 | 2.785104 | 1 |
| -4.640579 | -4.00463 | 2.149108 | 0 |
| -3.816852 | -4.102335 | 1.272858 | 0 |
| -4.344304 | -3.714395 | 0.804874 | 1 |
| -4.619082 | -2.628932 | 0.266636 | 0 |
| -4.161883 | -2.298399 | 0.313689 | 0 |

| *x* | *y* | *z* | *q* |
|---|---|---|---|
| -4.24987 | -3.118298 | 1.02813 | 0 |
| -4.897228 | -3.952685 | 1.569452 | 0 |
| -6.722128 | 4.293695 | -2.432271 | 0 |
| -6.608298 | 3.961377 | -1.620074 | -1 |
| -6.244474 | 4.755228 | -0.8346 | 0 |
| -5.506506 | 4.457987 | -2.205013 | 1 |
| -4.922037 | 3.671612 | -1.74164 | -1 |
| -5.353688 | 2.894681 | -1.646003 | -1 |
| -5.88365 | 3.144233 | -2.135169 | 1 |
| -5.917623 | 4.198153 | -1.966941 | 0 |
| -2.95147 | -0.376184 | 4.647504 | 1 |
| -3.088004 | -0.100077 | 4.172251 | 1 |
| -3.195552 | 0.198445 | 3.678947 | 0 |
| -2.933055 | -0.706494 | 3.376565 | 0 |
| -2.975758 | -1.145552 | 3.757329 | 1 |
| -3.22458 | -1.393489 | 4.283186 | 1 |
| -6.066441 | 0.424702 | 4.810067 | 1 |
| -5.98403 | 0.247252 | 4.246869 | 1 |
| -6.006935 | 0.025124 | 3.439772 | 1 |
| -5.526089 | -0.773988 | 2.839593 | 0 |
| -5.450147 | -1.42372 | 2.645922 | 1 |
| -5.481937 | -2.187127 | 2.581109 | 1 |
| -6.280627 | -2.211244 | 2.772688 | 1 |
| -6.485852 | -1.135558 | 2.682083 | 1 |
| -6.836171 | -0.525602 | 3.355149 | 1 |
| -6.958267 | 0.446914 | 3.618493 | 1 |
| -7.416985 | 1.232612 | 3.702267 | 1 |
| -7.536821 | 0.804196 | 4.323363 | 1 |
| -6.536097 | -4.932045 | -0.155972 | -1 |
| -5.976399 | -4.554326 | 0.018835 | -1 |
| -5.828768 | -4.928628 | 0.608428 | -1 |
| -6.271484 | -4.422851 | 1.377006 | -1 |
| -7.71665 | -2.72091 | 3.574988 | -1 |
| -7.15486 | -2.198583 | 3.45203 | -1 |

| *x* | *y* | *z* | *q* |
|---|---|---|---|
| -7.141372 | -1.659361 | 3.855717 | -1 |
| -7.527959 | -0.960305 | 4.401132 | -1 |
| -7.637008 | -0.3876 | 3.964318 | -1 |
| -8.156937 | -1.806358 | 2.66701 | -1 |
| -7.668999 | -2.361648 | 2.468448 | -1 |
| -7.306376 | -2.747416 | 1.938752 | -1 |
| -7.028788 | -2.8606 | 1.280898 | -1 |
| -6.673718 | -3.394406 | 1.052037 | -1 |
| -6.258715 | -3.70643 | 1.429786 | -1 |
| -5.621276 | -3.33408 | 1.501403 | -1 |
| -5.333146 | -2.630675 | 1.342324 | -1 |
| -5.237078 | -2.692106 | 0.686341 | -1 |
| -5.541985 | -1.994575 | 0.823444 | -1 |
| -5.41071 | -1.601363 | 0.129616 | -1 |
| -4.703253 | -1.421517 | 0.061419 | -1 |
| -4.522466 | -1.132499 | -0.595294 | -1 |
| -4.71096 | -0.736705 | -1.200532 | -1 |
| -5.194466 | -0.250149 | -1.382146 | -1 |
| -5.693535 | 0.248671 | -1.326739 | -1 |
| -5.926606 | 0.657955 | -0.84206 | -1 |
| -5.594098 | 0.886976 | -0.205503 | -1 |
| -5.877913 | 0.355341 | 0.225119 | -1 |
| -6.488565 | -0.056376 | 0.418027 | -1 |
| -6.822438 | -0.45539 | -0.107536 | -1 |
| -6.703855 | -0.813423 | -0.750199 | -1 |
| -6.451704 | -1.136321 | -1.366479 | -1 |
| -5.946532 | -1.661798 | -1.496548 | -1 |
| -5.450998 | -2.089452 | -1.194633 | -1 |
| -5.295523 | -2.515441 | -0.638843 | -1 |
| -5.421827 | -2.952168 | -0.100591 | -1 |
| -5.906164 | -3.447366 | 0.10389 | -1 |
| -6.360958 | -3.772298 | -0.310527 | -1 |
| -6.800495 | -3.24539 | -0.30469 | -1 |
| -6.924729 | -2.468665 | -0.238479 | -1 |

| *x* | *y* | *z* | *q* |
|---|---|---|---|
| -7.504069 | -2.646235 | -0.207135 | -1 |
| -7.1954 | -2.572556 | 0.420018 | -1 |
| -6.852505 | -1.860004 | 0.631903 | -1 |
| -6.749088 | -1.611204 | 1.330265 | -1 |
| -7.483751 | -1.658048 | 1.487298 | -1 |
| -7.420054 | -2.253952 | 1.097353 | -1 |
| -7.643354 | -2.897773 | 0.896772 | -1 |
| -8.23551 | -3.082632 | 0.56464 | -1 |
| -8.007391 | -3.391332 | -0.083527 | -1 |
| -7.792698 | -3.583428 | 0.492075 | -1 |
| -7.653847 | -4.04409 | 1.0295 | -1 |
| -7.488973 | 4.386803 | 0.476207 | -1 |
| -6.974173 | 4.497345 | -0.028547 | -1 |
| -6.314054 | 4.466783 | -0.293445 | -1 |
| -5.64645 | 4.183054 | -0.257591 | -1 |
| -5.159357 | 3.710169 | 0.033073 | -1 |
| -4.97578 | 3.127054 | 0.362632 | -1 |
| -5.078357 | 2.503043 | 0.408509 | -1 |
| -5.403494 | 1.966671 | 0.022428 | -1 |
| -5.891536 | 2.014917 | -0.446438 | -1 |
| -6.442219 | 2.398866 | -0.199806 | -1 |
| -6.858199 | 2.973667 | 0.033725 | -1 |
| -6.951132 | 3.280901 | 0.566211 | -1 |
| -6.733089 | 3.079576 | 1.19988 | -1 |
| -6.107416 | 3.218011 | 1.505388 | -1 |
| -5.590692 | 3.727521 | 1.509426 | -1 |
| -5.373679 | 4.379465 | 1.331005 | -1 |
| -7.140101 | 1.52343 | 4.520563 | -1 |
| -6.484765 | 1.343568 | 4.243617 | -1 |
| -6.154464 | 1.549369 | 3.706455 | -1 |
| -6.214878 | 2.192671 | 3.534136 | -1 |
| -5.628982 | 2.565035 | 3.474425 | -1 |
| -5.08478 | 2.518856 | 2.997688 | -1 |
| -4.755755 | 2.21971 | 2.436349 | -1 |

| *x* | *y* | *z* | *q* |
|---|---|---|---|
| -4.794829 | 1.747837 | 1.903509 | -1 |
| -5.232614 | 1.317697 | 1.476462 | -1 |
| -5.785913 | 1.039741 | 1.120354 | -1 |
| -6.443512 | 0.867644 | 1.006615 | -1 |
| -6.946198 | 0.46557 | 1.278333 | -1 |
| -7.524759 | 0.761048 | 1.708184 | -1 |
| -7.749891 | 1.388167 | 1.99614 | -1 |
| -7.89888 | 1.920661 | 1.866636 | -1 |
| -8.259165 | 1.88862 | 1.252839 | -1 |
| -7.765631 | 1.966711 | 0.738204 | -1 |
| -7.007113 | 1.979911 | 0.687916 | -1 |
| -6.32377 | 1.953812 | 0.982415 | -1 |
| -5.909273 | 2.350881 | 1.337321 | -1 |
| -6.203032 | 2.413928 | 2.033609 | -1 |
| -6.398174 | 2.167531 | 2.718719 | -1 |
| -6.718493 | 1.686687 | 2.891984 | -1 |
| -6.639184 | 1.013736 | 2.628623 | -1 |
| -6.07764 | 0.567834 | 2.546147 | -1 |
| -5.362495 | 0.484193 | 2.684855 | -1 |
| -4.767292 | 0.642137 | 3.036855 | -1 |
| -4.488918 | 1.132235 | 3.512292 | -1 |
| -4.47151 | 1.720222 | 4.015646 | -1 |
| -4.225412 | 3.013111 | 3.462607 | -1 |
| -4.821633 | 2.972402 | 3.539732 | -1 |
| -1.792457 | 4.192908 | 1.802374 | -1 |
| -2.273721 | 3.969026 | 1.657157 | -1 |
| -2.443933 | 3.553965 | 2.265809 | -1 |
| -2.517962 | 3.512583 | 2.892537 | -1 |
| -6.317763 | 2.917716 | 4.055382 | -1 |
| -7.536799 | 2.594871 | 3.456135 | -1 |
| -7.414093 | 2.178888 | 4.078855 | -1 |
| -8.078485 | 1.90426 | 3.050885 | -1 |
| -7.137582 | 3.950227 | 2.241132 | -1 |
| -7.580428 | 3.828594 | 1.745175 | -1 |

| *x* | *y* | *z* | *q* |
|---|---|---|---|
| -8.196905 | 2.83377 | 1.579647 | -1 |
| -8.1212 | 3.191681 | 0.905403 | -1 |
| -8.140512 | 3.489059 | 0.176471 | -1 |
| -8.533443 | 2.205634 | -0.494449 | -1 |
| -8.224976 | 1.865681 | -0.002543 | -1 |
| -8.339338 | 1.536986 | 0.570457 | -1 |
| -8.831997 | 1.250419 | 0.943654 | -1 |
| -7.6311 | 2.827169 | -3.604202 | -1 |
| -7.898903 | 2.17726 | -3.279931 | -1 |
| -8.468894 | -2.043725 | -0.994409 | -1 |
| -8.199324 | -1.506185 | -0.4852 | -1 |
| -8.367395 | -1.109616 | 0.191242 | -1 |
| -8.688595 | -1.032649 | 0.848077 | -1 |
| -7.545294 | -3.79611 | 3.027882 | -1 |
| -4.301856 | -1.271655 | 4.336357 | -1 |
| -4.132359 | -0.873028 | 3.879213 | -1 |
| -4.60682 | -0.735845 | 3.374181 | -1 |
| -5.282943 | -0.685583 | 3.911227 | -1 |
| -6.011865 | -0.618472 | 4.039129 | -1 |
| -6.524541 | -1.106429 | 4.23966 | -1 |
| -6.642923 | -1.740789 | 4.52109 | -1 |
| -4.538108 | -2.93975 | 3.535973 | -1 |
| -4.749557 | -2.526708 | 2.941739 | -1 |
| -4.648716 | -2.125967 | 2.451489 | -1 |
| -3.970624 | -1.82837 | 2.426444 | -1 |
| -3.359661 | -1.670779 | 2.791632 | -1 |
| -2.927843 | -1.830642 | 3.29976 | -1 |
| -2.638407 | -2.225136 | 3.725828 | -1 |
| -2.329792 | -2.668124 | 4.203341 | -1 |
| -6.496789 | -4.716638 | 0.640529 | 0 |
| -6.779921 | -1.259887 | 3.46215 | 0 |
| -8.119832 | -1.326066 | 3.312723 | 0 |
| -7.974395 | -1.781079 | 3.267931 | 0 |
| -5.998244 | -3.92093 | 0.835609 | 0 |

| *x* | *y* | *z* | *q* |
|---|---|---|---|
| -5.878362 | -3.473498 | 0.854982 | 0 |
| -5.9235 | -2.606388 | 1.01707 | 0 |
| -5.790868 | -3.022346 | 0.693857 | 0 |
| -6.068253 | -1.713491 | 0.123272 | 0 |
| -4.852971 | -1.800694 | -0.428626 | 0 |
| -6.193727 | 1.081392 | -0.074265 | 0 |
| -6.321375 | 0.651557 | -0.104082 | 0 |
| -6.246907 | -0.180137 | -0.340065 | 0 |
| -5.904243 | -0.931086 | -1.018585 | 0 |
| -5.700676 | -1.811987 | -0.66911 | 0 |
| -6.701517 | -3.32009 | 0.324054 | 0 |
| -6.412655 | -2.457697 | 0.279719 | 0 |
| -6.656309 | -2.831745 | 0.512132 | 0 |
| -5.993111 | 4.202652 | 0.336639 | 0 |
| -5.727914 | 3.802648 | 0.357435 | 0 |
| -5.64139 | 3.311362 | 0.480418 | 0 |
| -6.025177 | 1.529169 | -0.016707 | 0 |
| -6.188142 | 2.76961 | 0.517731 | 0 |
| -7.591218 | 3.149637 | 0.511594 | 0 |
| -6.235822 | 4.862846 | 0.825341 | 0 |
| -5.061862 | 1.788373 | 2.883296 | 0 |
| -6.484519 | 0.523559 | 1.714571 | 0 |
| -6.887608 | 0.853843 | 1.894332 | 0 |
| -7.145812 | 1.750259 | 1.751892 | 0 |
| -7.345952 | 2.376116 | 1.140742 | 0 |
| -6.860781 | 2.498829 | 3.128004 | 0 |
| -6.349154 | 1.514196 | 2.250842 | 0 |
| -5.535549 | 1.113728 | 2.536548 | 0 |
| -4.604117 | 2.350896 | 3.770006 | 0 |
| -6.024326 | 2.128851 | 4.501024 | 0 |
| -7.237659 | 2.622502 | 2.747479 | 0 |
| -6.948192 | 4.388572 | 1.785636 | 0 |
| -8.545564 | 2.433865 | 0.08014 | 0 |
| -5.069972 | -1.428156 | 4.245971 | 0 |

| *x* | *y* | *z* | *q* |
|---|---|---|---|
| -5.944548 | -0.981078 | 3.575226 | 0 |
| -5.286698 | -1.540122 | 4.643025 | 0 |
| -3.774873 | -3.225493 | 3.412559 | 0 |
| -3.834031 | -1.826369 | 3.241078 | 0 |
| -2.943088 | -2.809045 | 3.789674 | 0 |
| -7.925856 | -2.940932 | 1.977474 | 0 |
| -6.479814 | -2.561975 | 1.40539 | 0 |
| -5.190837 | -0.985244 | -0.94705 | 0 |
| -5.381341 | -0.54683 | -0.914128 | 0 |
| -5.777449 | -0.074602 | -0.876403 | 0 |
| -5.934462 | -2.734176 | -0.182619 | 0 |
| -6.117429 | -3.182134 | -0.290751 | 0 |
| -8.287777 | -2.834607 | 1.152856 | 0 |
| -7.360594 | -4.821617 | 0.675504 | 0 |
| -7.272323 | 4.594426 | 1.050451 | 0 |
| -6.88502 | 4.624709 | 0.610055 | 0 |
| -6.217396 | 3.079555 | 0.828905 | 0 |
| -5.962256 | 3.50087 | 0.959108 | 0 |
| -5.76453 | 3.991985 | 0.985249 | 0 |
| -5.951146 | 1.529605 | 4.469975 | 0 |
| -6.364599 | 1.886429 | 4.094899 | 0 |
| -5.125878 | 1.690732 | 2.447476 | 0 |
| -5.825441 | 1.399328 | 1.766945 | 0 |
| -6.000735 | 2.326462 | 0.554317 | 0 |
| -6.539887 | 2.07111 | 1.539151 | 0 |
| -5.173102 | 1.07294 | 2.839422 | 0 |
| -1.872933 | 3.530912 | 1.478161 | 0 |
| -1.72592 | 3.563965 | 2.033372 | 0 |
| -6.729544 | 2.449076 | 3.714398 | 0 |
| -8.499943 | 2.699635 | 0.726953 | 0 |
| -5.094084 | -1.35017 | 3.793779 | 0 |
| -5.945683 | -1.723154 | 4.617084 | 0 |
| -4.127012 | -2.941854 | 3.184443 | 0 |
| -4.504827 | -2.955887 | 2.841288 | 0 |

| *x* | *y* | *z* | *q* |
|---|---|---|---|
| -7.769296 | -2.17839 | 3.197433 | 0 |
| -5.061087 | -1.401259 | -0.54776 | 0 |
| -6.602327 | -3.850346 | 0.245682 | 0 |
| -6.838672 | -2.141513 | 1.341771 | 0 |
| -6.475547 | 4.467861 | 0.380579 | 0 |
| -5.575267 | 2.898046 | 0.289547 | 0 |
| -7.358808 | 2.820414 | 0.725244 | 0 |
| -5.705316 | 1.976958 | 3.772397 | 0 |
| -5.326215 | 2.036001 | 3.365598 | 0 |
| -7.11304 | 1.256801 | 1.917716 | 0 |
| -7.847661 | 2.494391 | 1.035457 | 0 |
| -6.085258 | 1.179069 | 2.417604 | 0 |
| -4.358676 | 2.429805 | 3.285313 | 0 |
| -1.937104 | 3.724993 | 2.562046 | 0 |
| -6.796621 | -3.510665 | 3.244972 | 0 |
| -4.490643 | -1.283536 | 3.447536 | 0 |
| -6.121812 | -1.645802 | 4.221483 | 0 |
| -4.326107 | -1.78784 | 2.945969 | 0 |
| -2.732378 | -3.235149 | 3.723224 | 0 |
| -6.489749 | -4.290117 | 0.370716 | 0 |
| -5.763154 | -4.330254 | 0.916114 | 0 |
| -8.055938 | -2.884471 | 2.486406 | 0 |
| -6.244216 | -3.016944 | 1.424362 | 0 |
| -6.16785 | -2.15623 | 1.174618 | 0 |
| -5.773971 | 0.257979 | -0.455472 | 0 |
| -6.664624 | 0.247884 | -0.125235 | 0 |
| -6.09332 | -0.523011 | -0.81761 | 0 |
| -5.598016 | -1.37788 | -1.009892 | 0 |
| -5.815771 | -2.262044 | -0.349743 | 0 |
| -7.89573 | -2.816642 | -0.712389 | 0 |
| -6.272456 | -2.042078 | 0.304944 | 0 |
| -6.078686 | -1.694709 | 1.444735 | 0 |
| -7.455103 | -1.335061 | 2.114964 | 0 |
| -8.00478 | -3.09959 | 1.356185 | 0 |

| *x* | *y* | *z* | *q* |
|---|---|---|---|
| -8.315866 | -2.377597 | 0.998085 | 0 |
| -7.531027 | -4.557671 | 0.082742 | 0 |
| -7.812402 | -4.157337 | 0.210793 | 0 |
| -5.671765 | 2.547065 | -0.064088 | 0 |
| -6.090719 | 1.931305 | 0.252579 | 0 |
| -5.897148 | 4.517216 | 0.778676 | 0 |
| -5.496696 | 1.405125 | 2.105996 | 0 |
| -6.087466 | 1.541493 | 1.393252 | 0 |
| -7.016854 | 2.205034 | 1.381815 | 0 |
| -6.69127 | 1.806978 | 2.072568 | 0 |
| -5.029071 | 1.411171 | 3.289052 | 0 |
| -5.1066 | 1.652303 | 3.764121 | 0 |
| -5.216187 | 1.735289 | 4.195725 | 0 |
| -7.600863 | 2.387145 | 2.541475 | 0 |
| -8.642822 | 2.090551 | 0.436745 | 0 |
| -8.820483 | -1.3874 | -0.554431 | 0 |
| -8.59129 | -1.964193 | 1.022906 | 0 |
| -4.129321 | -1.438781 | 3.531117 | 0 |
| -6.081944 | -1.441138 | 3.882743 | 0 |
| -4.752401 | -1.773834 | 3.009018 | 0 |
| -3.527188 | -2.17389 | 3.402824 | 0 |
| -3.25212 | -2.51944 | 3.466838 | 0 |

*Appendix B*

# 3D-CG MODEL CHANNEL CLOSED CONFORMATION COORDINATES AND PARAMETERS

| $x$ | $y$ | $z$ | $q$ | $\epsilon_{\text{in}}^{\text{attr}}$ | $\epsilon_{\text{in}}^{\text{rep}}$ | $\epsilon_{\text{out}}$ | $\text{rcut}_{\text{in}}^{\text{attr}}$ | $\text{rcut}_{\text{in}}^{\text{rep}}$ | $\text{rcut}_{\text{out}}$ |
|---|---|---|---|---|---|---|---|---|---|
| -2.67 | -1.339 | -0.315 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.5113 | 2.1073 | -2.7643 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.8779 | 1.5424 | -2.9151 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.1628 | 1.1049 | -2.6302 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.2303 | 0.4143 | -2.1891 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.0198 | -0.1829 | -1.89 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.4412 | -0.5858 | -1.2127 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.6112 | -1.4984 | -1.1072 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.3073 | -2.1133 | -1.2513 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.7324 | -2.0118 | -1.2396 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.1594 | -1.8757 | -1.2657 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.6108 | -1.7448 | -1.3727 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.0384 | -1.6547 | -1.4439 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.5436 | -1.5561 | -1.4453 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.1679 | -1.426 | -1.3661 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.7797 | -0.9288 | -1.5097 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.0635* | -0.8695 | 0.3158 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.2747* | -0.9453 | 0.5613 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.014* | -0.6495 | -0.1318 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.7796* | -0.9698 | -0.6901 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.2241* | -0.7939 | -1.0031 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.9182 | -0.0398 | -1.099 | 0 | 1.41 | 0.51 | 0.5 | 2.5 | 0 | 2.5 |
| 0.3794 | 0.5076 | -0.8477 | 0 | 1.41 | 0.51 | 0.5 | 2.5 | 0 | 2.5 |
| -0.0447 | 0.8438 | -0.645 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.4481 | 1.2061 | -0.4454 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.8166 | 1.5927 | -0.1784 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.1102 | 1.9527 | 0.1016 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.6085 | 1.964 | 0.3505 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |

| $x$ | $y$ | $z$ | $q$ | $\epsilon_{\text{in}}^{\text{attr}}$ | $\epsilon_{\text{in}}^{\text{rep}}$ | $\epsilon_{\text{out}}$ | $\text{rcut}_{\text{in}}^{\text{attr}}$ | $\text{rcut}_{\text{in}}^{\text{rep}}$ | $\text{rcut}_{\text{out}}$ |
|---|---|---|---|---|---|---|---|---|---|
| -1.5992 | 2.5459 | 0.2429 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.1593 | 2.12 | -0.4052 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.6236 | 1.8063 | -1.042 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.1241 | 1.9 | -1.2477 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.5912 | 1.7352 | -1.4127 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.0743 | 1.5087 | -1.5517 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.529 | 1.2917 | -1.6006 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.0734 | 1.1674 | -1.6181 | 0 | 0.3 | 0.51 | 0.5 | 2.5 | 0 | 2.5 |
| 0.635 | 1.1748 | -1.573 | 0 | 0.3 | 0.51 | 0.5 | 2.5 | 0 | 2.5 |
| 1.2105 | 1.1687 | -1.5598 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.7879 | 1.1262 | -1.5354 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.361 | 1.0658 | -1.4313 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.6926 | 1.0217 | -0.997 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.8374 | 0.4926 | -1.4668 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.9797 | 0.4632 | -2.3865 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.5441 | 0.1814 | -2.3998 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.9752 | 0.1715 | -2.2849 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.4118 | 0.1978 | -2.1507 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.8575 | 0.1836 | -1.9667 | 0 | 0.3 | 0.51 | 0.5 | 2.5 | 0 | 2.5 |
| 0.2981 | 0.0355 | -1.7937 | 0 | 0.3 | 0.51 | 0.5 | 2.5 | 0 | 2.5 |
| -0.2025 | -0.2178 | -1.6584 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.69 | -0.4969 | -1.5285 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.2078 | -0.717 | -1.3485 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.7047 | -0.8848 | -1.1127 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.5533 | -1.2428 | -0.5885 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.7789 | -0.783 | -0.5138 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.3592 | -0.9393 | -0.2203 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.1042 | -1.1368 | 0.0496 | 0 | 0.3 | 0.51 | 0.5 | 2.5 | 0 | 2.5 |
| 0.6142 | -1.3548 | 0.2304 | 0 | 1.41 | 0.51 | 0.5 | 2.5 | 0 | 2.5 |
| 1.0887 | -1.7016 | 0.3276 | 0 | 0.3 | 0.51 | 0.5 | 2.5 | 0 | 2.5 |
| 1.4638 | -2.1206 | 0.3842 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.7496 | -2.5607 | 0.4296 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.2684 | -2.3537 | 0.1591 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.3376 | -2.5923 | 0.93 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |

| $x$ | $y$ | $z$ | $q$ | $\epsilon_{in}^{attr}$ | $\epsilon_{in}^{rep}$ | $\epsilon_{out}$ | $rcut_{in}^{attr}$ | $rcut_{in}^{rep}$ | $rcut_{out}$ |
|---|---|---|---|---|---|---|---|---|---|
| 2.613 | -2.6848 | 1.4181 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.9141 | -2.8247 | 1.8809 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 3.2152 | -2.9607 | 2.3172 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 3.008 | -2.134 | 2.356 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.6547 | -1.4996 | 2.0528 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.064 | -1.4158 | 2.0857 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.4534 | -1.3508 | 2.105 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.8766 | -1.2345 | 2.096 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.327 | -1.031 | 2.1048 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.2131 | -0.8162 | 2.1637 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.7639 | -0.6276 | 2.2352 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.465 | -0.5806 | 2.4032 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.5009 | -0.3402 | 2.3661 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.6031 | -0.5199 | 1.7764 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.6964 | 0.0064 | 1.9286 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.1756 | 0.1981 | 1.5459 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.9438 | 0.8197 | 1.098 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.4018 | 0.9706 | 0.9641 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.1403 | 1.1134 | 0.7673 | 0 | 0.3 | 0.51 | 0.5 | 2.5 | 0 | 2.5 |
| 0.6362 | 1.193 | 0.47 | 0 | 1.41 | 0.51 | 0.5 | 2.5 | 0 | 2.5 |
| 1.1325 | 1.2347 | 0.1659 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.6561 | 1.2676 | -0.0939 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.1874 | 1.3556 | -0.349 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.6199 | 1.6992 | -0.527 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.7457 | 2.1653 | -0.6731 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.23 | 2.1426 | 0.0994 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.6781 | 1.4764 | 0.4488 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.946 | 0.6844 | 0.4018 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.6367 | 0.938 | 0.9292 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.3489 | 1.7477 | 1.0495 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.9943 | 1.374 | 1.2292 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.5961 | 0.9388 | 1.2256 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.9164 | -0.0377 | 1.4298 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.1933 | -0.2184 | 2.0499 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |

| $x$ | $y$ | $z$ | $q$ | $\epsilon_{\text{in}}^{\text{attr}}$ | $\epsilon_{\text{in}}^{\text{rep}}$ | $\epsilon_{\text{out}}$ | $\text{rcut}_{\text{in}}^{\text{attr}}$ | $\text{rcut}_{\text{in}}^{\text{rep}}$ | $\text{rcut}_{\text{out}}$ |
|---|---|---|---|---|---|---|---|---|---|
| 2.4738 | 0.5908 | 2.5804 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.2922 | 1.3443 | 2.5796 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.7537 | 1.3228 | 2.3034 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.2105 | 1.387 | 2.0566 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.6545 | 1.5246 | 1.9041 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.0964 | 1.6626 | 1.8448 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.4824 | 1.7328 | 1.7676 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.0699 | 1.8097 | 1.6598 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.6016 | 1.9909 | 1.4074 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.0441 | 2.2547 | 1.2053 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.316 | 1.5793 | 0.9333 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.8166 | 1.2633 | 1.724 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.3704 | 1.1509 | 2.2571 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.7537 | 1.1709 | 2.3919 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.1555 | 1.0987 | 2.3144 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.6376 | 0.8946 | 2.1694 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.1194 | 0.6316 | 2.0663 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.3959 | 0.3763 | 2.0508 | 0 | 0.3 | 0.51 | 0.5 | 2.5 | 0 | 2.5 |
| 0.9328 | 0.2065 | 2.0202 | 0 | 0.3 | 0.51 | 0.5 | 2.5 | 0 | 2.5 |
| 1.4795 | -0.0322 | 1.9718 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.5132 | -0.7385 | 1.5985 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.206 | -0.4954 | 1.2001 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.576 | -0.4147 | 1.1393 | 0 | 1.41 | 0.51 | 0.5 | 2.5 | 0 | 2.5 |
| 0.0188 | -0.5251 | 1.1491 | 0 | 0.3 | 0.51 | 0.5 | 2.5 | 0 | 2.5 |
| -0.5415 | -0.7228 | 1.0808 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.0554 | -0.9535 | 0.9441 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.5949 | -1.1055 | 0.7481 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.1556 | -1.1548 | 0.5755 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.7359 | -1.1803 | 0.4701 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.1953 | 3.7845 | 2.4824 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.3828 | 3.2466 | 2.697 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.3878 | 2.6898 | 2.8515 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.3317 | 2.1188 | 2.972 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.2745 | 1.5449 | 3.034 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |

| $x$ | $y$ | $z$ | $q$ | $\epsilon_{in}^{attr}$ | $\epsilon_{in}^{rep}$ | $\epsilon_{out}$ | $rcut_{in}^{attr}$ | $rcut_{in}^{rep}$ | $rcut_{out}$ |
|---|---|---|---|---|---|---|---|---|---|
| -2.2587 | 0.9621 | 3.1273 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.2532 | 0.3922 | 3.2742 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.7542 | -0.3486 | 3.2091 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.5963 | -1.0978 | 3.5624 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.3811 | -1.7686 | 3.133 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.134 | -1.7847 | 2.6235 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.9202 | -1.8133 | 2.0852 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.7139 | -1.9153 | 1.5601 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.41 | -2.0104 | 1.0696 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.0721 | -2.0921 | 0.6193 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.244 | -2.1773 | 0.1397 | 0 | 0.3 | 0.51 | 0.5 | 2.5 | 0 | 2.5 |
| 0.5527 | -2.2622 | -0.3308 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.9667 | -2.3754 | -0.7483 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.4455 | -2.5395 | -1.057 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.9644 | -2.6205 | -1.3323 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.451 | -2.6728 | -1.65 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.824 | 1.1425 | -2.159 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.7277 | 0.3809 | -2.5512 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.6161 | -0.6326 | -2.4527 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.9765 | -0.9035 | -2.5684 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.4197 | -1.0733 | -2.629 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.156 | -1.2103 | -2.596 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.736 | -1.2705 | -2.5301 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.3183 | -1.2605 | -2.5268 | -1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.8634 | -1.2436 | -2.5641 | 0 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.4241 | -1.4873 | -2.2277 | 1 | 0.3 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |

*Plug Domain Beads

*Appendix C*

# 3D-CG MODEL CHANNEL OPEN CONFORMATION COORDINATES AND PARAMETERS

| $x$ | $y$ | $z$ | $q$ | $\epsilon_{\text{in}}^{\text{attr}}$ | $\epsilon_{\text{in}}^{\text{rep}}$ | $\epsilon_{\text{out}}$ | $\text{rcut}_{\text{in}}^{\text{attr}}$ | $\text{rcut}_{\text{in}}^{\text{rep}}$ | $\text{rcut}_{\text{out}}$ |
|---|---|---|---|---|---|---|---|---|---|
| -2.67 | -1.339 | -0.315 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.5275 | 1.4302 | -3.45 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.9584 | 0.8157 | -3.4543 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.2838 | 0.455 | -3.1128 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.4232 | -0.1182 | -2.5284 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.231 | -0.7009 | -2.0665 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.7465 | -1.1579 | -1.4649 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.7803 | -1.6665 | -0.6742 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.3478 | -2.1423 | -0.5986 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.7897 | -1.9991 | -0.6427 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.2257 | -1.8639 | -0.7423 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.6609 | -1.8604 | -0.9167 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.0799 | -1.9195 | -1.015 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.5055 | -1.9429 | -1.0513 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.106 | -1.6693 | -1.0444 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.8412 | -1.2782 | -1.3441 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.9864* | -0.2225 | -0.1783 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.4477* | -0.6955 | 0.3493 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.1439* | -0.4618 | -0.1886 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.9672* | -0.9471 | -0.6052 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.2062* | -1.0057 | -1.1293 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.8727 | -0.2919 | -1.1306 | 0 | 1.38 | 9.85 | 0.5 | 2.5 | 0 | 2.5 |
| 0.3182 | 0.3049 | -1.1274 | 0 | 1.38 | 9.85 | 0.5 | 2.5 | 0 | 2.5 |
| -0.1245 | 0.6757 | -1.0043 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.5362 | 1.0737 | -0.9667 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.8819 | 1.5448 | -0.9078 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.1539 | 2.0045 | -0.8545 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.6159 | 2.181 | -0.6243 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |

| $x$ | $y$ | $z$ | $q$ | $\epsilon_{in}^{attr}$ | $\epsilon_{in}^{rep}$ | $\epsilon_{out}$ | $rcut_{in}^{attr}$ | $rcut_{in}^{rep}$ | $rcut_{out}$ |
|---|---|---|---|---|---|---|---|---|---|
| -1.4883 | 2.6473 | -1.0144 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.1479 | 2.1998 | -1.3457 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.6588 | 1.7466 | -1.7977 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.158 | 1.674 | -1.9896 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.6583 | 1.399 | -2.0735 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.1772 | 1.0795 | -2.1139 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.6559 | 0.8009 | -2.0898 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.0901 | 0.6116 | -2.0551 | 0 | 0.46 | 9.85 | 0.5 | 2.5 | 0 | 2.5 |
| 0.4856 | 0.514 | -2.0616 | 0 | 0.46 | 9.85 | 0.5 | 2.5 | 0 | 2.5 |
| 1.0502 | 0.4241 | -2.1326 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.6087 | 0.3031 | -2.2001 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.1788 | 0.1897 | -2.1812 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.6302 | 0.252 | -1.9122 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.769 | -0.2442 | -2.4217 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.5743 | -0.6367 | -3.1954 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.1635 | -0.9092 | -2.9477 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.6533 | -0.8097 | -2.6838 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.1571 | -0.6591 | -2.4012 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.6758 | -0.5228 | -2.108 | 0 | 0.46 | 9.85 | 0.5 | 2.5 | 0 | 2.5 |
| 0.1421 | -0.4675 | -1.8352 | 0 | 0.46 | 9.85 | 0.5 | 2.5 | 0 | 2.5 |
| -0.4082 | -0.5417 | -1.6144 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.9345 | -0.6858 | -1.4524 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.4688 | -0.8144 | -1.2875 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.9852 | -0.9145 | -1.0374 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.7775 | -0.9913 | -0.4182 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.0045 | -0.513 | -0.4213 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.5825 | -0.7429 | -0.0985 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.0963 | -0.9745 | 0.1348 | 0 | 0.46 | 9.85 | 0.5 | 2.5 | 0 | 2.5 |
| 0.4174 | -1.1851 | 0.2916 | 0 | 1.38 | 9.85 | 0.5 | 2.5 | 0 | 2.5 |
| 0.9224 | -1.4418 | 0.4531 | 0 | 1.38 | 9.85 | 0.5 | 2.5 | 0 | 2.5 |
| 1.342 | -1.775 | 0.6836 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.6456 | -2.1714 | 0.8532 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.1747 | -2.0302 | 0.5505 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.2235 | -2.2089 | 1.2988 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |

| $x$ | $y$ | $z$ | $q$ | $\epsilon_{\text{in}}^{\text{attr}}$ | $\epsilon_{\text{in}}^{\text{rep}}$ | $\epsilon_{\text{out}}$ | $\text{rcut}_{\text{in}}^{\text{attr}}$ | $\text{rcut}_{\text{in}}^{\text{rep}}$ | $\text{rcut}_{\text{out}}$ |
|---|---|---|---|---|---|---|---|---|---|
| 2.5691 | -2.3497 | 1.7352 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.9166 | -2.524 | 2.1545 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 3.2615 | -2.6834 | 2.5413 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 3.0455 | -1.8519 | 2.6987 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.637 | -1.235 | 2.4435 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.0352 | -1.2265 | 2.4177 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.4206 | -1.2407 | 2.3736 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.8377 | -1.2017 | 2.321 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.2692 | -1.0498 | 2.2992 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.2726 | -0.8582 | 2.3235 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.8263 | -0.6873 | 2.3747 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.5282 | -0.5844 | 2.4639 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.5349 | -0.175 | 2.3849 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.6028 | -0.5582 | 1.8935 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.6287 | -0.0476 | 1.9973 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.084 | 0.193 | 1.6469 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.6679 | 0.8129 | 1.2675 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.0692 | 0.9531 | 1.21 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.507 | 1.0973 | 1.1163 | 0 | 0.46 | 9.85 | 0.5 | 2.5 | 0 | 2.5 |
| 1.0759 | 1.2073 | 0.9393 | 0 | 0.46 | 9.85 | 0.5 | 2.5 | 0 | 2.5 |
| 1.6226 | 1.25 | 0.7192 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.174 | 1.2514 | 0.5384 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.7338 | 1.3053 | 0.3927 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 3.1922 | 1.6073 | 0.2288 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 3.4374 | 1.9706 | 0.157 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.7791 | 2.1212 | 0.7589 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 3.1687 | 1.6465 | 1.2901 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 3.5793 | 0.9669 | 1.5288 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.9844 | 1.1539 | 1.6603 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.55 | 1.9114 | 1.7365 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.2347 | 1.4694 | 1.8032 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.9177 | 0.9289 | 1.5382 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.8044 | 0.007 | 1.4959 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.2275 | -0.3175 | 1.817 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |

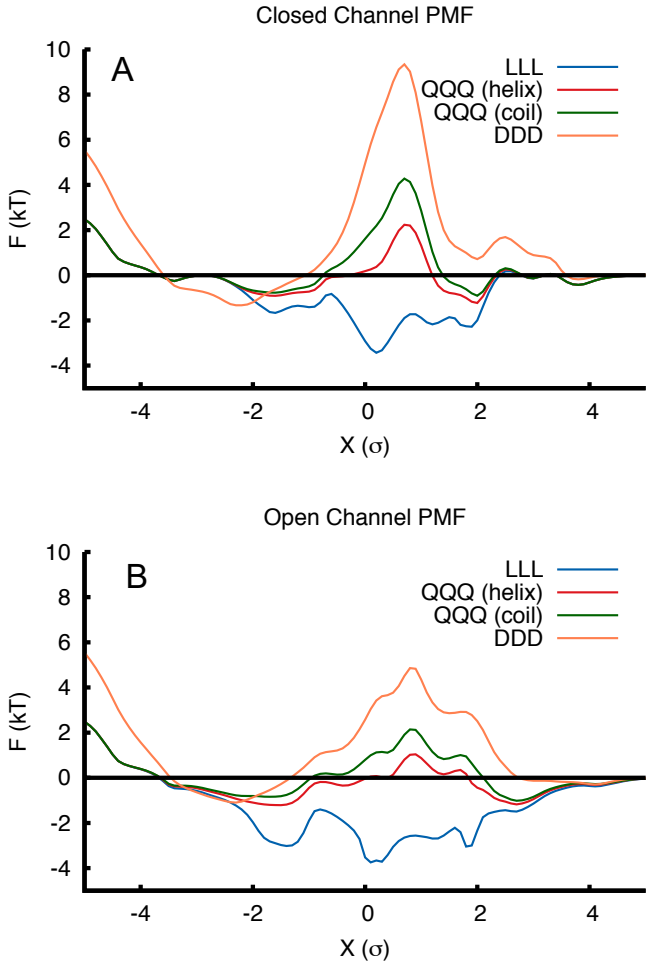
| $x$ | $y$ | $z$ | $q$ | $\epsilon_{\text{in}}^{\text{attr}}$ | $\epsilon_{\text{in}}^{\text{rep}}$ | $\epsilon_{\text{out}}$ | $\text{rcut}_{\text{in}}^{\text{attr}}$ | $\text{rcut}_{\text{in}}^{\text{rep}}$ | $\text{rcut}_{\text{out}}$ |
|---|---|---|---|---|---|---|---|---|---|
| 2.5297 | 0.1485 | 2.53 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.4371 | 0.8098 | 2.8737 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.8879 | 0.9879 | 2.6109 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.3693 | 1.2062 | 2.4406 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.8365 | 1.4179 | 2.3381 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.2788 | 1.5662 | 2.2272 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.2763 | 1.6495 | 2.0729 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.835 | 1.7654 | 1.8853 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.3063 | 1.9726 | 1.5809 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.6766 | 2.2263 | 1.3135 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.0425 | 1.6005 | 1.0322 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.6226 | 1.3145 | 1.7235 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.2172 | 1.1815 | 2.3016 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.6192 | 1.1655 | 2.4848 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.0214 | 1.0563 | 2.4548 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.507 | 0.8143 | 2.354 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.0066 | 0.5145 | 2.29 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.4886 | 0.2181 | 2.2869 | 0 | 0.46 | 9.85 | 0.5 | 2.5 | 0 | 2.5 |
| 1.0217 | 0.0036 | 2.2486 | 0 | 0.46 | 9.85 | 0.5 | 2.5 | 0 | 2.5 |
| 1.5563 | -0.2037 | 2.0984 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.4496 | -0.8875 | 1.6475 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.1387 | -0.5423 | 1.3345 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.5425 | -0.5057 | 1.2655 | 0 | 1.38 | 9.85 | 0.5 | 2.5 | 0 | 2.5 |
| -0.0242 | -0.6124 | 1.2579 | 0 | 0.46 | 9.85 | 0.5 | 2.5 | 0 | 2.5 |
| -0.5806 | -0.8247 | 1.2275 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.1031 | -1.0925 | 1.096 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.6 | -1.3379 | 0.8912 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.1214 | -1.5456 | 0.7032 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.6779 | -1.7278 | 0.6109 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.79 | 3.7306 | 2.5519 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.0442 | 3.2522 | 2.7887 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.1342 | 2.7157 | 2.9712 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.1698 | 2.151 | 3.1138 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.205 | 1.5795 | 3.1793 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |

| $x$ | $y$ | $z$ | $q$ | $\epsilon_{in}^{attr}$ | $\epsilon_{in}^{rep}$ | $\epsilon_{out}$ | $rcut_{in}^{attr}$ | $rcut_{in}^{rep}$ | $rcut_{out}$ |
|---|---|---|---|---|---|---|---|---|---|
| -2.2788 | 1.0185 | 3.2419 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.3511 | 0.4685 | 3.3487 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.0677 | -0.327 | 3.3462 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.9095 | -1.0035 | 3.3317 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.5742 | -1.706 | 3.3083 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.2621 | -1.7614 | 2.8304 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.9755 | -1.8444 | 2.3203 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.7231 | -1.98 | 1.8233 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.3827 | -2.0682 | 1.3406 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.0129 | -2.1107 | 0.9011 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.3352 | -2.1537 | 0.455 | 0 | 0.46 | 9.85 | 0.5 | 2.5 | 0 | 2.5 |
| 0.6975 | -2.2855 | 0.0131 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.0978 | -2.5045 | -0.3528 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.5658 | -2.7656 | -0.5852 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.0823 | -2.9144 | -0.8224 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.5685 | -3.0879 | -1.1014 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -2.0105 | 0.5737 | -2.7206 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.8558 | -0.2969 | -2.8982 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.7118 | -1.1873 | -2.4057 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -1.0659 | -1.4792 | -2.2478 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| -0.5205 | -1.6845 | -2.286 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.0448 | -1.8463 | -2.2289 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 0.6163 | -1.958 | -2.1525 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.1972 | -2.0227 | -2.153 | -1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 1.7505 | -2.0873 | -2.1963 | 0 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |
| 2.3684 | -2.3206 | -2.0905 | 1 | 0.46 | 0.3 | 0.5 | 2.5 | 2.5 | 2.5 |

*Plug Domain Beads

*Appendix D*

# ANALYTICAL PMFS FOR 3D-CG MODEL INCLUDING THE RIBOSOME AND PLUG DOMAINS




Fitting CG parameters to MARTINI potentials of mean force. A. Analytical potentials of mean force for tripeptide translocation across the translocon in the closed channel conformation with CG beads corresponding to the plug domain and ribosome. B. Analytical potentials of mean force for tripeptide translocation across the translocon in the open channel conformation with the CG beads corresponding to the plug domain and ribosome.

*Appendix E*

# SEQUENCE INFORMATION FOR SELECTED PROTEINS

| Construct | Sequence* |
|---|---|
| Lep[0L] | *MANMFALILVIATLVTGILWCVDKFFFAPKRRERQAAAQAAAGDSLDKA* *TLKKVAPKPG*WLETGASVFPVLAIVLIVRSFIYEPFQIPSGSMMPTLLI GDFILVEKFAYGIKDPIYQKTLIETGHPKRGDIVVFKYPEDPKLDYIKR AVGLPGDKVTYDPVSKELTIQPGCSSGQACENALPVTYSNVEPSDFVQT FSRRNGGEATSGFFEVPKNETKENGIRLSETSGGPG[AAAAAAAAAAAA AAAAAAA]GGPGGTQQPGQQLATWIVPPGQYFMMGDNRDNSADSRYWGF VPEANLVGRATAIWMSFDKQEGEWPTGLRLSRIGGIH |
| [1L] | [AAAAAAAAALAAAAAAAAA] |
| [2L] | [AAAALAAAAAAAAALAAAA] |
| [3L] | [AAAAAAALALALAAAAAAA] |
| [4L] | [AAAALALAAAAALALAAAA] |
| [5L] | [AAAALALAALAALALAAAA] |
| [6L] | [AAAALALALALALALAAAA] |
| [7L] | [ALAALALAALAALALAALA] |

*Italicized residues not explicit modeled

Amino-acid sequence of the modified leader peptidase (Lep) construct used in stop-transfer simulations. The third transmembrane domain shown between brackets is systematically mutated to introduce increasing numbers of leucine residues.

| Construct | Sequence |
|---|---|
| H1Δ22(110) | MGPRLLLLLLLLLLLLLLLLLLLLLGSQNSQLQEELRGLRETFSNFTA<br>STEAQVKGLSTQGGNVGRKMKSLESQLEKQQKDLSEDHSSLLLHVKQFV<br>SDLRSLSCQMAALQGNGSERTCCPVNETELDKASQEPPLL |
| H1Δ22(170) | MGPRLLLLLLLLLLLLLLLLLLLLLGSQNSQLQEELRGLRETFSNFTA<br>STEAQVKGLSTQGGNVGRKMKSLESQLEKQQKDLSEDHSSLLLHVKQFV<br>SDLRSLSCQMAALQGNGSERTCCPVNWVEHERSCYWFSRSGKAWADADN<br>YCRLEDAHLVVVTSWEEQKFVQHHIGPVNTWMGLHDQETELDKASQEPP<br>LL |
| H1Δ22(230) | MGPRLLLLLLLLLLLLLLLLLLLLLGSQNSQLQEELRGLRETFSNFTAS<br>TEAQVKGLSTQGGNVGRKMKSLESQLEKQQKDLSEDHSSLLLHVKQFVSD<br>LRSLSCQMAALQGNGSERTCCPVNWVEHERSCYWFSRSGKAWADADNYCR<br>LEDAHLVVVTSWEEQKFVQHHIGPVNTWMGLHDQNGPWKWVDGTDYETGF<br>KNWRPEQPDDWYGHGLGGGEDCAHFTDDGRWNDDVCQRPYRWVCETELDK<br>ASQEPPLL |
| H1Δ22(290) | MGPRLLLLLLLLLLLLLLLLLLLLLGSQNSQLQEELRGLRETFSNFTAS<br>TEAQVKGLSTQGGNVGRKMKSLESQLEKQQKDLSEDHSSLLLHVKQFVSD<br>LRSLSCQMAALQGNGSERTCCPVNWVEHERSCYWFSRSGKAWADADNYCR<br>LEDAHLVVVTSWEEQKFVQHHIGPVNTWMGLHDQERSCYWFSRSGKAWAD<br>ADNYCRLEDAHLVVVTSWEEQKFVQHHIGPVNTWMGLHDQNGPWKWVDGT<br>DYETGFKNWRPEQPDDWYGHGLGGGEDCAHFTDDGRWNDDVCQRPYRWVC<br>ETELDKASQEPPLL |
| H1Δ22(350) | GPRLLLLLLLLLLLLLLLLLLLLLGSQNSQLQEELRGLRETFSNFTAST<br>EAQVKGLSTQGGNVGRKMKSLESQLEKQQKDLSEDHSSLLLHVKQFVSDL<br>RSLSCQMAALQGNGSERTCCPVNWVEHERSCYWFSRSGKAWADADNYCRL<br>EDAHLVVVTSWEEQKFVQHHIGPVNTWMGLHDQNGPWKWVDGTDYETGFK<br>NWRPEQPDDWYGHGLGGGEDCAHFTDDGRWNDDVCQRPYRWVCETELDKA<br>SQEPPLLERSCYWFSRSGKAWADADNYCRLEDAHLVVVTSWEEQKFVQHH<br>IGPVNTWMGLHDQNGPWKWVDGTDYETGFKNWRPEQPDDWYGHGLGGGED<br>CAHFTDDGRWNDDVCQRPYRWVCETELDKASQEPPLL |
| H1Δ22(400) | MGPRLLLLLLLLLLLLLLLLLLLLLGSQNSQLQEELRGLRETFSNFTAS<br>TEAQVKGLSTQGGNVGRKMKSLESQLEKQQKDLSEDHSSLLLHVKQFVSD<br>LRSLSCQMAALQGNGSERTCCPVNWVEHERSCYWFSRSGKAWADADNYCR<br>LEDAHLVVVTSWEEQKFVQHHIGPVNTWMGLHDQETELDKASQEPPLLQN<br>SQLQEELRGLRETFSNFTASTEAQVKGLSTQGGNVGRKMKSLESQLEKQQ<br>KDLSEDHSSLLLHVKQFVSDLRSLSCQMAALQGNGSERTCCPVNWVEHER<br>SCYWFSRSGKAWADADNYCRLEDAHLVVVTSWEEQKFVQHHIGPVNTWMG<br>LHDQNGPWKWVDGTDYETGFKNWRPEQPDDWYGHGLGGGEDCAHFTDDGR<br>WNDDVCQRPYRWVCETELDKASQEPPLL |

| Construct | Sequence |
|---|---|
| H1Δ22(460) | GPRLLLLLLLLLLLLLLLLLLLLLGSQNSQLQEELRGLRETFSNFTA<br>STEAQVKGLSTQGGNVGRKMKSLESQLEKQQKDLSEDHSSLLLHVKQF<br>VSDLRSLSCQMAALQGNGSERTCCPVNWVEHERSCYWFSRSGKAWADA<br>DNYCRLEDAHLVVVTSWEEQKFVQHHIGPVNTWMGLHDQNGPWKWVDG<br>TDYETGFKNWRPEQPDDWYGHGLGGGEDCAHFTDDGRWNDDVCQRPYR<br>WVCETELDKASQEPPLLQNSQLQEELRGLRETFSNFTASTEAQVKGLS<br>TQGGNVGRKMKSLESQLEKQQKDLSEDHSSLLLHVKQFVSDLRSLSCQ<br>MAALQGNGSERTCCPVNWVEHERSCYWFSRSGKAWADADNYCRLEDAH<br>LVVVTSWEEQKFVQHHIGPVNTWMGLHDQNGPWKWVDGTDYETGFKNW<br>RPEQPDDWYGHGLGGGEDCAHFTDDGRWNDDVCQRPYRWVCETELDKA<br>SQEPPLL |
| H1Δ22(520) | MGPRLLLLLLLLLLLLLLLLLLLLLGSQNSQLQEELRGLRETFSNFTA<br>STEAQVKGLSTQGGNVGRKMKSLESQLEKQQKDLSEDHSSLLLHVKQFV<br>SDLRSLSCQMAALQGNGSERTCCPVNWVEHERSCYWFSRSGKAWADADN<br>YCRLEDAHLVVVTSWEEQKFVQHHIGPVNTWMGLHDQERSCYWFSRSGK<br>AWADADNYCRLEDAHLVVVTSWEEQKFVQHHIGPVNTWMGLHDQNGPWK<br>WVDGTDYETGFKNWRPEQPDDWYGHGLGGGEDCAHFTDDGRWNDDVCQR<br>PYRWVCETELDKASQEPPLLQNSQLQEELRGLRETFSNFTASTEAQVKG<br>LSTQGGNVGRKMKSLESQLEKQQKDLSEDHSSLLLHVKQFVSDLRSLSC<br>QMAALQGNGSERTCCPVNWVEHERSCYWFSRSGKAWADADNYCRLEDAH<br>LVVVTSWEEQKFVQHHIGPVNTWMGLHDQNGPWKWVDGTDYETGFKNWR<br>PEQPDDWYGHGLGGGEDCAHFTDDGRWNDDVCQRPYRWVCETELDKASQ<br>EPPLL |
| H1Δ22(580) | MGPRLLLLLLLLLLLLLLLLLLLLLGSQNSQLQEELRGLRETFSNFTA<br>STEAQVKGLSTQGGNVGRKMKSLESQLEKQQKDLSEDHSSLLLHVKQFV<br>SDLRSLSCQMAALQGNGSERTCCPVNWVEHERSCYWFSRSGKAWADADN<br>YCRLEDAHLVVVTSWEEQKFVQHHIGPVNTWMGLHDQNGPWKWVDGTDY<br>ETGFKNWRPEQPDDWYGHGLGGGEDCAHFTDDGRWNDDVCQRPYRWVCE<br>TELDKASQEPPLLERSCYWFSRSGKAWADADNYCRLEDAHLVVVTSWEE<br>QKFVQHHIGPVNTWMGLHDQNGPWKWVDGTDYETGFKNWRPEQPDDWYG<br>HGLGGGEDCAHFTDDGRWNDDVCQRPYRWVCETELDKASQEPPLLQNSQ<br>LQEELRGLRETFSNFTASTEAQVKGLSTQGGNVGRKMKSLESQLEKQQK<br>DLSEDHSSLLLHVKQFVSDLRSLSCQMAALQGNGSERTCCPVNWVEHER<br>SCYWFSRSGKAWADADNYCRLEDAHLVVVTSWEEQKFVQHHIGPVNTWM<br>GLHDQNGPWKWVDGTDYETGFKNWRPEQPDDWYGHGLGGGEDCAHFTDD<br>GRWNDDVCQRPYRWVCETELDKASQEPPLL |

Amino-acid sequence of the modified ASGP receptor constructs used in topogenesis simulations

| Tripeptide | TFE | $\lambda_c$ | $\lambda_o$ | $q$ | helix |
|---|---|---|---|---|---|
| W | -3.56 | 1.00 | 1.00 | 0 | 0 |
| LET | 4.269 | 0.718 | 0.762 | -1 | 0 |
| GAS | 3.425 | 0.758 | 0.787 | 0 | 0 |
| VFP | -3.295 | 0.929 | 0.938 | 0 | 0 |
| VLA | -1.964 | 0.895 | 0.908 | 0 | 0 |
| IVL | -4.594 | 0.962 | 0.967 | 0 | 0 |
| IVR | 5.731 | 0.628 | 0.710 | 1 | 0 |
| SFI | 1.510 | 0.807 | 0.830 | 0 | 0 |
| YEP | 10.325 | 0.383 | 0.539 | -1 | 0 |
| FQI | 2.013 | 0.794 | 0.819 | 0 | 0 |
| PSG | 8.198 | 0.476 | 0.622 | 0 | 0 |
| SMM | 3.929 | 0.739 | 0.774 | 0 | 0 |
| PTL | 3.961 | 0.737 | 0.773 | 0 | 0 |
| LIG | 3.377 | 0.759 | 0.788 | 0 | 0 |
| DFI | 6.672 | 0.570 | 0.676 | -1 | 0 |
| LVE | 8.474 | 0.459 | 0.612 | -1 | 0 |
| KFA | 7.938 | 0.492 | 0.631 | 1 | 0 |
| YGI | 4.253 | 0.719 | 0.762 | 0 | 0 |
| KDP | 16.039 | 0.212 | 0.297 | 0 | 0 |
| IYQ | 3.636 | 0.753 | 0.783 | 0 | 0 |
| KTL | 8.279 | 0.471 | 0.619 | 1 | 0 |
| IET | 9.838 | 0.398 | 0.559 | -1 | 0 |
| GHP | 11.234 | 0.356 | 0.500 | 1 | 0 |
| KRG | 14.708 | 0.252 | 0.354 | 2 | 0 |
| DIV | 8.701 | 0.445 | 0.604 | -1 | 0 |
| VFK | 6.380 | 0.588 | 0.687 | 1 | 0 |
| YPE | 10.325 | 0.383 | 0.539 | -1 | 0 |
| DPK | 16.039 | 0.212 | 0.297 | 0 | 0 |
| LDY | 8.084 | 0.483 | 0.626 | -1 | 0 |
| IKR | 11.023 | 0.362 | 0.509 | 2 | 0 |
| AVG | 7.289 | 0.532 | 0.654 | 0 | 0 |
| LPG | 5.422 | 0.647 | 0.721 | 0 | 0 |
| DKV | 15.065 | 0.241 | 0.338 | 0 | 0 |
| TYD | 10.519 | 0.378 | 0.530 | -1 | 0 |

| Tripeptide | TFE | $\lambda_c$ | $\lambda_o$ | $q$ | helix |
|---|---|---|---|---|---|
| PVS | 5.584 | 0.637 | 0.715 | 0 | 0 |
| KEL | 13.766 | 0.280 | 0.393 | 0 | 0 |
| TIQ | 5.195 | 0.661 | 0.729 | 0 | 0 |
| PGC | 7.419 | 0.524 | 0.650 | 0 | 0 |
| SSG | 8.718 | 0.444 | 0.604 | 0 | 0 |
| QAC | 7.386 | 0.526 | 0.651 | 0 | 0 |
| ENA | 13.442 | 0.290 | 0.407 | -1 | 0 |
| LPV | 2.808 | 0.774 | 0.801 | 0 | 0 |
| TYS | 5.357 | 0.651 | 0.723 | 0 | 0 |
| NVE | 11.883 | 0.337 | 0.473 | -1 | 0 |
| PSD | 12.240 | 0.326 | 0.458 | -1 | 0 |
| FVQ | 3.084 | 0.767 | 0.795 | 0 | 0 |
| TFS | 3.734 | 0.750 | 0.780 | 0 | 0 |
| RRN | 12.614 | 0.315 | 0.442 | 2 | 0 |
| GGE | 14.984 | 0.243 | 0.342 | -1 | 0 |
| ATS | 7.321 | 0.530 | 0.653 | 0 | 0 |
| GFF | 1.672 | 0.803 | 0.826 | 0 | 0 |
| EVP | 10.731 | 0.371 | 0.521 | -1 | 0 |
| KNE | 17.175 | 0.177 | 0.249 | 0 | 0 |
| TKE | 16.201 | 0.207 | 0.290 | 0 | 0 |
| NGI | 6.786 | 0.563 | 0.672 | 0 | 0 |
| RLS | 7.013 | 0.549 | 0.664 | 1 | 0 |
| ETS | 12.403 | 0.321 | 0.451 | -1 | 0 |
| GGP | 9.318 | 0.414 | 0.581 | 0 | 0 |
| GAA | 3.490 | 0.757 | 0.786 | 0 | 1 |
| AAA | 2.435 | 0.783 | 0.809 | 0 | 1 |
| AAA | 2.435 | 0.783 | 0.809 | 0 | 1 |
| AAA | 2.435 | 0.783 | 0.809 | 0 | 1 |
| AAA | 2.435 | 0.783 | 0.809 | 0 | 1 |
| AAA | 2.435 | 0.783 | 0.809 | 0 | 1 |
| AAG | 3.490 | 0.757 | 0.786 | 0 | 1 |
| GPG | 9.318 | 0.414 | 0.581 | 0 | 0 |
| GTQ | 8.880 | 0.434 | 0.598 | 0 | 0 |
| QPG | 8.701 | 0.445 | 0.604 | 0 | 0 |

| Tripeptide | TFE | $\lambda_c$ | $\lambda_o$ | $q$ | helix |
|---|---|---|---|---|---|
| QQL | 5.828 | 0.622 | 0.706 | 0 | 0 |
| ATW | 3.182 | 0.764 | 0.793 | 0 | 0 |
| IVP | 3.019 | 0.769 | 0.796 | 0 | 0 |
| PGQ | 8.701 | 0.445 | 0.604 | 0 | 0 |
| YFM | 0.341 | 0.837 | 0.856 | 0 | 0 |
| MGD | 12.045 | 0.332 | 0.466 | -1 | 0 |
| NRD | 15.584 | 0.225 | 0.316 | 0 | 0 |
| NSA | 8.295 | 0.470 | 0.619 | 0 | 0 |
| DSR | 14.951 | 0.244 | 0.343 | 0 | 0 |
| YWG | 2.679 | 0.777 | 0.804 | 0 | 0 |
| FVP | 2.062 | 0.793 | 0.818 | 0 | 0 |
| EAN | 13.442 | 0.290 | 0.407 | -1 | 0 |
| LVG | 4.448 | 0.707 | 0.755 | 0 | 0 |
| RAT | 9.513 | 0.408 | 0.573 | 1 | 0 |
| AIW | 0.958 | 0.821 | 0.842 | 0 | 0 |
| MSF | 2.240 | 0.788 | 0.814 | 0 | 0 |
| DKQ | 17.062 | 0.181 | 0.254 | 0 | 0 |
| EGE | 19.010 | 0.122 | 0.172 | -2 | 0 |
| WPT | 2.597 | 0.779 | 0.806 | 0 | 0 |
| GLR | 8.133 | 0.480 | 0.625 | 1 | 0 |
| LSR | 7.013 | 0.549 | 0.664 | 1 | 0 |
| IGG | 7.273 | 0.533 | 0.655 | 0 | 0 |
| IH | 7.321 | 0.530 | 0.653 | 1 | 0 |

Example CG bead mapping of the Lep[0Leu] sequence for stop transfer simulations in a given tripeptide frame

| Tripeptide | TFE | $\lambda_c$ | $\lambda_o$ | $q$ | helix |
|---|---|---|---|---|---|
| MGP | 6.364 | 0.589 | 0.687 | 1 | 0 |
| RLL | -1.120 | 0.874 | 0.889 | 1 | 1 |
| LLL | -6.088 | 1.000 | 1.000 | 0 | 1 |
| LLL | -6.088 | 1.000 | 1.000 | 0 | 1 |
| LLL | -6.088 | 1.000 | 1.000 | 0 | 1 |
| LLL | -6.088 | 1.000 | 1.000 | 0 | 1 |
| LLL | -6.088 | 1.000 | 1.000 | 0 | 1 |
| LLL | -6.088 | 1.000 | 1.000 | 0 | 1 |
| LLG | -2.192 | 0.901 | 0.913 | 0 | 1 |
| SQN | 3.377 | 0.759 | 0.788 | 0 | 0 |
| SQL | -0.032 | 0.846 | 0.865 | 0 | 0 |
| QEE | 13.036 | 0.302 | 0.424 | -2 | 0 |
| LRG | 2.776 | 0.775 | 0.802 | 1 | 0 |
| LRE | 6.802 | 0.562 | 0.672 | 0 | 0 |
| TFS | -1.623 | 0.887 | 0.900 | 0 | 0 |
| NFT | -0.990 | 0.870 | 0.886 | 0 | 0 |
| AST | 1.964 | 0.795 | 0.820 | 0 | 0 |
| EAQ | 7.955 | 0.491 | 0.631 | -1 | 0 |
| VKG | 5.666 | 0.632 | 0.712 | 1 | 0 |
| LST | -0.877 | 0.868 | 0.883 | 0 | 0 |
| QGG | 4.984 | 0.674 | 0.736 | 0 | 0 |
| NVG | 2.500 | 0.782 | 0.808 | 0 | 0 |
| RKM | 6.396 | 0.587 | 0.686 | 2 | 0 |
| KSL | 3.263 | 0.762 | 0.791 | 1 | 0 |
| ESQ | 7.890 | 0.495 | 0.633 | -1 | 0 |
| LEK | 8.409 | 0.463 | 0.615 | 0 | 0 |
| QQK | 7.045 | 0.547 | 0.663 | 1 | 0 |
| DLS | 4.627 | 0.696 | 0.749 | -1 | 0 |
| EDH | 15.584 | 0.225 | 0.316 | -1 | 0 |
| SSL | -0.536 | 0.859 | 0.876 | 0 | 0 |
| LLH | -0.276 | 0.852 | 0.870 | 1 | 0 |
| VKQ | 5.049 | 0.670 | 0.734 | 1 | 0 |
| FVS | -2.776 | 0.916 | 0.926 | 0 | 0 |
| DLR | 12.175 | 0.328 | 0.460 | 0 | 0 |

| Tripeptide | TFE | $\lambda_c$ | $\lambda_o$ | $q$ | helix |
|---|---|---|---|---|---|
| SLS | 4.821 | 0.684 | 0.742 | 0 | 0 |
| CQM | 5.487 | 0.643 | 0.718 | 0 | 0 |
| AAL | 4.951 | 0.676 | 0.737 | 0 | 0 |
| QGN | 9.854 | 0.398 | 0.558 | 0 | 0 |
| GSE | 13.864 | 0.277 | 0.389 | -1 | 0 |
| RTC | 8.669 | 0.447 | 0.606 | 1 | 0 |
| CPV | 4.805 | 0.685 | 0.743 | 0 | 0 |
| NET | 13.036 | 0.302 | 0.424 | -1 | 0 |
| ELD | 15.130 | 0.239 | 0.336 | -2 | 0 |
| KAS | 11.461 | 0.349 | 0.491 | 1 | 0 |
| QEP | 12.727 | 0.311 | 0.437 | -1 | 0 |
| PLL | 1.526 | 0.807 | 0.830 | 0 | 0 |

Example CG bead mapping of the H1Δ22(110) sequence for topogenesis simulations

| Tripeptide | TFE | $\lambda_c$ | $\lambda_o$ | $q$ | helix |
|---|---|---|---|---|---|
| MGP | 6.364 | 0.589 | 0.687 | 0 | 0 |
| RLL | -1.120 | 0.874 | 0.889 | 0 | 1 |
| LLG | -2.192 | 0.901 | 0.913 | 0 | 1 |
| LLG | -2.192 | 0.901 | 0.913 | 0 | 1 |
| LLG | -2.192 | 0.901 | 0.913 | 0 | 1 |
| LLG | -2.192 | 0.901 | 0.913 | 0 | 1 |
| LLG | -2.192 | 0.901 | 0.913 | 0 | 1 |
| LLG | -2.192 | 0.901 | 0.913 | 0 | 1 |
| LLG | -2.192 | 0.901 | 0.913 | 0 | 1 |
| SQN | 3.377 | 0.759 | 0.788 | 0 | 0 |
| SQL | -0.032 | 0.846 | 0.865 | 0 | 0 |
| QEE | 13.036 | 0.302 | 0.424 | 0 | 0 |
| LRG | 2.776 | 0.775 | 0.802 | 0 | 0 |
| LRE | 6.802 | 0.562 | 0.672 | 0 | 0 |
| TFS | -1.623 | 0.887 | 0.900 | 0 | 0 |
| LRE | 6.802 | 0.562 | 0.672 | 0 | 0 |
| LRE | 6.802 | 0.562 | 0.672 | 0 | 0 |

| Tripeptide | TFE | $\lambda_c$ | $\lambda_o$ | $q$ | helix |
|---|---|---|---|---|---|
| LRE | 6.802 | 0.562 | 0.672 | 0 | 0 |
| LRE | 6.802 | 0.562 | 0.672 | 0 | 0 |
| LRE | 6.802 | 0.562 | 0.672 | 0 | 0 |
| LRE | 6.802 | 0.562 | 0.672 | 0 | 0 |
| LRE | 6.802 | 0.562 | 0.672 | 0 | 0 |
| LRE | 6.802 | 0.562 | 0.672 | 0 | 0 |
| LRE | 6.802 | 0.562 | 0.672 | 0 | 0 |
| LRE | 6.802 | 0.562 | 0.672 | 0 | 0 |
| LRE | 6.802 | 0.562 | 0.672 | 0 | 0 |
| LRE | 6.802 | 0.562 | 0.672 | 0 | 0 |
| LRE | 6.802 | 0.562 | 0.672 | 0 | 0 |
| LRE | 6.802 | 0.562 | 0.672 | 0 | 0 |
| LRE | 6.802 | 0.562 | 0.672 | 0 | 0 |
| LRE | 6.802 | 0.562 | 0.672 | 0 | 0 |
| LRE | 6.802 | 0.562 | 0.672 | 0 | 0 |
| LRE | 6.802 | 0.562 | 0.672 | 0 | 0 |
| LRE | 6.802 | 0.562 | 0.672 | 0 | 0 |
| LRE | 6.802 | 0.562 | 0.672 | 0 | 0 |
| LRE | 6.802 | 0.562 | 0.672 | 0 | 0 |

CG sequence for binding free energy simulations

*Appendix F*

# INITIAL NC COORDINATES FOR STOP TRANSFER SIMULATIONS

| *x* | *y* | *z* |
|---|---|---|
| -1.4331 | 10.7241 | -1.6470 |
| -0.4331 | 10.7241 | -1.6470 |
| 0.5731 | 10.7241 | -1.6470 |
| 1.5731 | 10.7241 | -1.6470 |
| 2.5731 | 10.7241 | -1.6470 |
| 3.5731 | 10.7241 | -1.6470 |
| 4.5731 | 10.7241 | -1.6470 |
| 5.5731 | 10.7241 | -1.6470 |
| 6.5731 | 10.7241 | -1.6470 |
| 7.5731 | 10.7241 | -1.6470 |
| 8.5731 | 10.7241 | -1.6470 |
| 9.5731 | 10.7241 | -1.6470 |
| 9.5731 | 10.7241 | -2.6470 |
| 9.5731 | 10.7241 | -3.6470 |
| 9.5731 | 10.7241 | -4.6470 |
| 9.5731 | 10.7241 | -5.6470 |
| 9.5731 | 10.7241 | -6.6470 |
| 9.5731 | 10.7241 | -7.6470 |
| 9.5731 | 10.7241 | -8.6470 |
| 9.5731 | 9.7241 | -8.6470 |
| 9.5731 | 9.7041 | -7.6470 |
| 9.5731 | 9.7041 | -6.6470 |
| 9.5731 | 9.7041 | -5.6470 |
| 9.5731 | 9.7041 | -4.6470 |
| 9.5731 | 9.7041 | -3.6470 |
| 9.5731 | 9.7041 | -2.6470 |
| 9.5731 | 9.7041 | -1.6470 |
| 9.5731 | 8.7241 | -1.6470 |

| *x* | *y* | *z* |
|---|---|---|
| 9.5731 | 7.7241 | -1.6470 |
| 9.5731 | 6.7241 | -1.6470 |
| 9.5731 | 5.7241 | -1.6470 |
| 9.5731 | 4.7241 | -1.6470 |
| 9.5731 | 3.7241 | -1.6470 |
| 9.5731 | 2.7241 | -1.6470 |
| 9.5731 | 1.7241 | -1.6470 |
| 9.5731 | 0.7241 | -1.6470 |
| 8.5731 | 0.7241 | -1.6470 |
| 7.5731 | 0.7241 | -1.6470 |
| 6.5731 | 0.7241 | -1.6470 |
| 5.5731 | 0.7241 | -1.6470 |
| 4.5731 | 0.7241 | -1.6470 |
| 3.5731 | 0.7241 | -1.6470 |
| 2.6464 | 0.6605 | -1.1212 |
| 2.7886 | -0.0316 | -0.3667 |
| 1.8128 | 0.3522 | -0.4123 |
| 1.4733 | 0.1749 | 0.5260 |
| 0.3980 | 0.1401 | 0.0957 |
| -0.6623 | 0.2125 | -0.1465 |
| -1.5822 | -0.3358 | 0.0148 |
| -1.6248 | 0.1265 | -0.9069 |
| -1.5064 | 0.9153 | -0.3291 |
| -2.0159 | 0.7458 | 0.6708 |
| -2.7542 | 0.3496 | 1.3444 |
| -3.6719 | 1.0077 | 0.9427 |
| -4.3456 | 0.6682 | 1.6756 |
| -5.1240 | -0.1785 | 1.8581 |
| -5.9268 | -0.6679 | 1.3577 |

Initial coordinates for stop transfer simulations

| *x* | *y* | *z* |
|---|---|---|
| -2.8829 | 0.2046 | 1.1142 |
| -3.4874 | 0.1911 | 2.0138 |
| -4.5403 | 0.1595 | 1.3695 |
| -5.4242 | -0.3420 | 1.3956 |

Initial coordinates for topogenesis and free energy simulations