Alchembed: A computational method for incorporating multiple proteins into complex lipid geometries – Supporting Information.

Elizabeth Jefferys[1], *Zara A. Sands*[2], *Jiye Shi*[2], *Mark S. P. Sansom*[1] *and Philip W. Fowler*[1]*

1. Department of Biochemistry, University of Oxford, UK

2. UCB NewMedicines, Chemin du Foriest, 1420 Braine-l'Alleud, Belgium

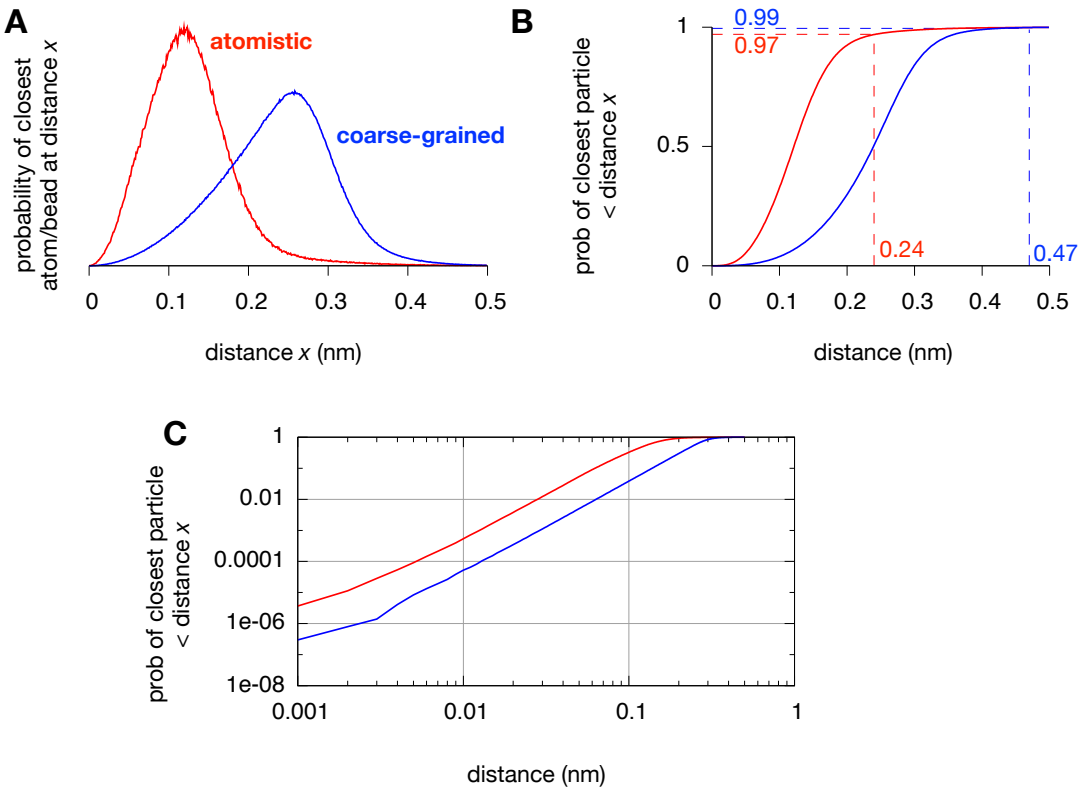

Figure S1: If you randomly place an atom (or coarse-grained bead) inside a typical simulation box there is a high probability it will be within σ of another atom (or bead). A point within the box was randomly selected in the final frame of the simulations of 128 POPC lipids and the distance to the nearest atom calculated and recorded. This was repeated 10 000 times and converted into a (A) probability of finding another atom at a distance x, (B) the probability of finding another atom less than a distance x and (C) the same probability, but plotted on a log-log scale. The analysis was repeated for the coarse-grained simulation, with the main difference being the beads were less densely packed. If we take as typical values of σ 0.24 nm and 0.47 nm for the atomistic and coarse-grained simulations, the probability of find another atom or bead within σ is 0.97 and 0.99, respectively. The results for the atomistic simulation is plotted in red, whilst that of the coarse-grained simulation is plotted in blue.

*to whom correspondence should be addressed: philip.fowler@bioch.ox.ac.uk

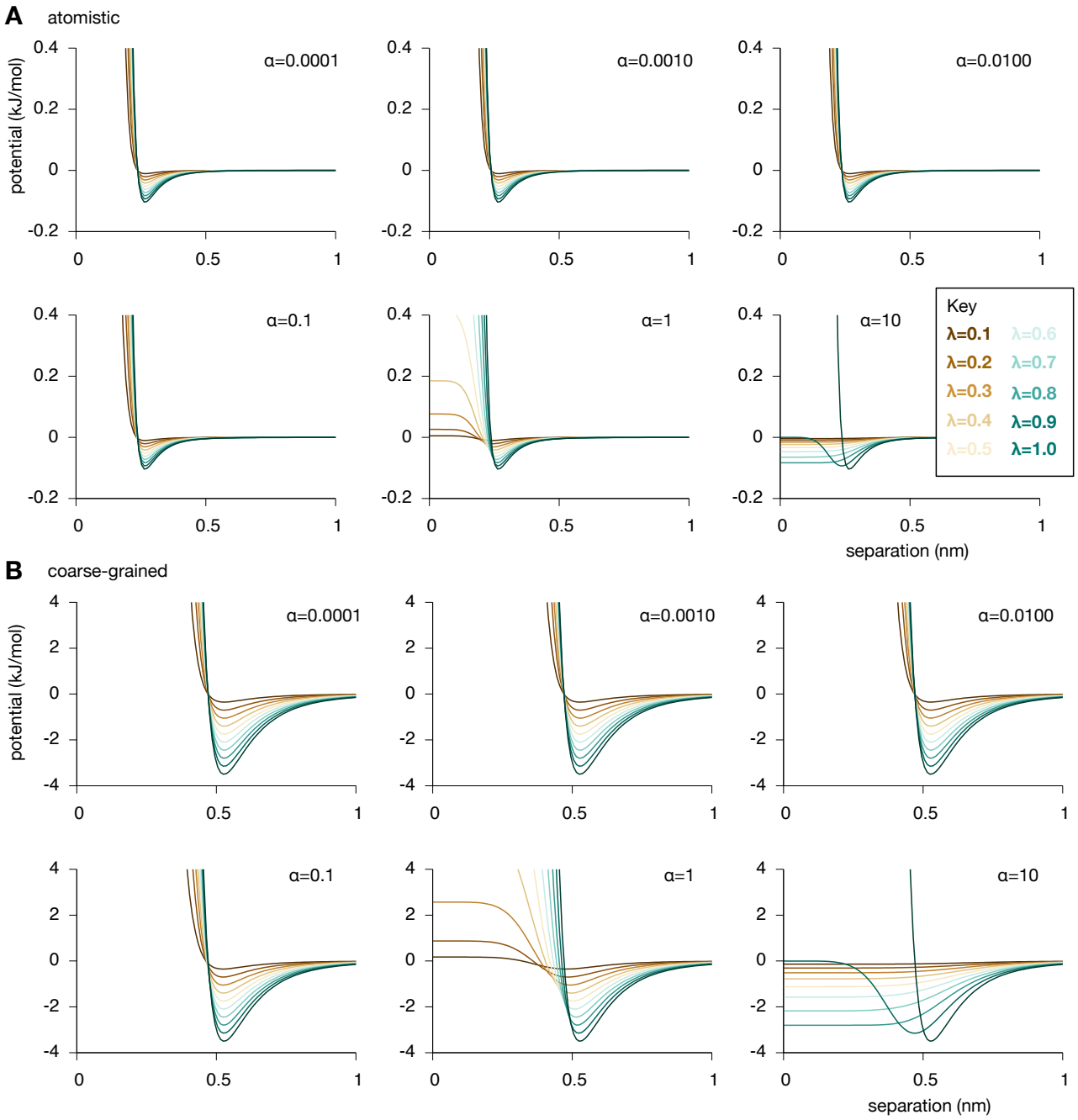

Figure S2: Increasing α slows the rate at which the van der Waals potential is introduced. The form of the soft-core potential in GROMACS and AMBER is used (Equation 4, $a = 1$, $b = 1$ and $c = 6$). Plotted here are the (A) atomistic and (B) coarse-grained potentials for the illustrative cases described in Figure 1. Ten values of λ are plotted ($\lambda = 0.1, 0.2, 0.3 \ldots 1.0$) on each graph using a smoothly varying colour scale.

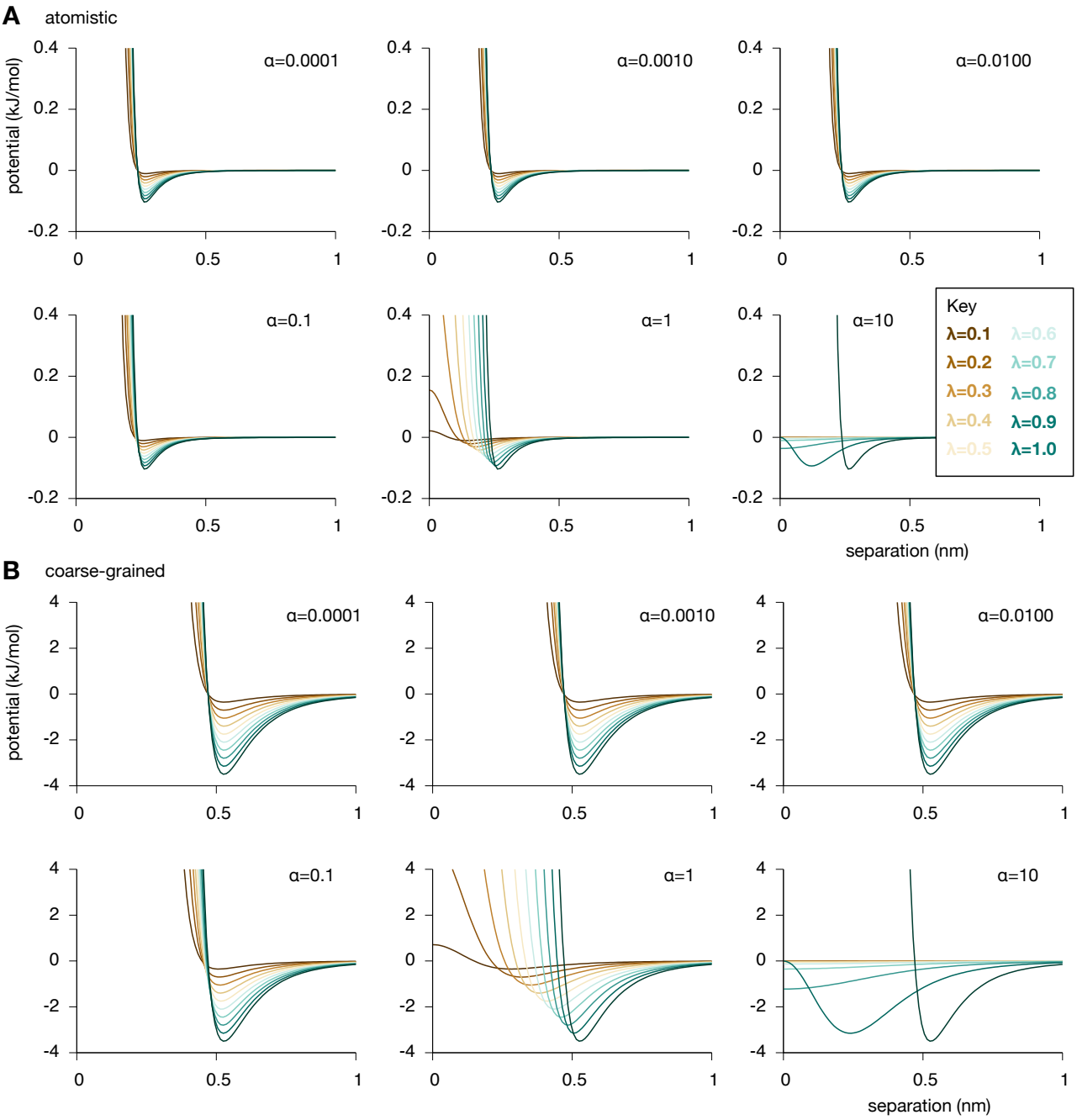

Figure S3: Increasing α slows the rate at which the van der Waals potential is introduced. The form of the soft-core potential in NAMD and CHARMM is used (Equation 5, $a = 1$, $b = 1$ and $c = 2$). Plotted here are the (A) atomistic and (B) coarse-grained potentials for the illustrative cases described in Figure 1. Ten values of λ are plotted ($\lambda = 0.1, 0.2, 0.3 \ldots 1.0$) on each graph using a smoothly varying colour scale.

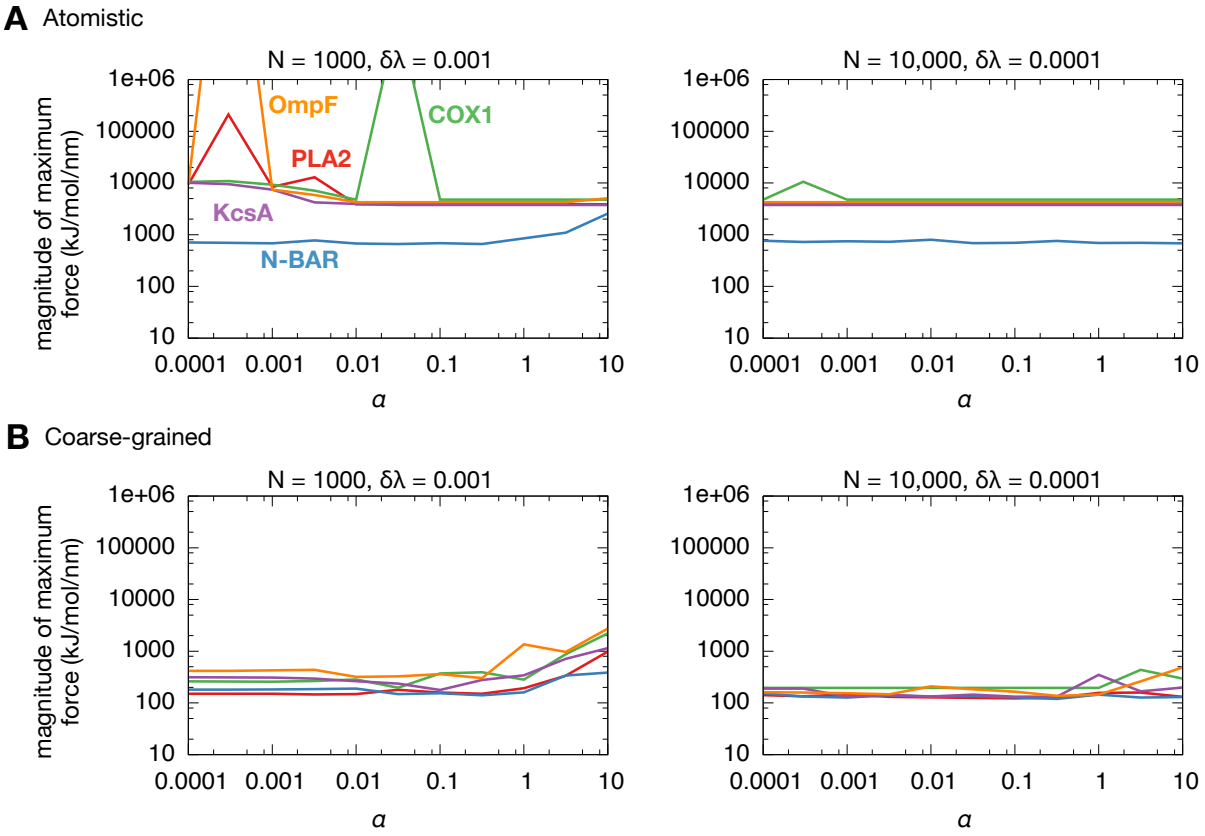

Figure S4: The maximum force experienced by a lipid or water atom (or bead) varies the protein, α, N and whether an (A) atomistic or (B) coarse-grained representation is used. Some of the simulations with very large forces failed and therefore did not complete; if this was the case the largest force up to that point is recorded in the graph. There are three repeats of each of the $N = 1\,000$ simulations. Here $b = 2$; the same analysis for $b = 1$ can be found in Figure 4 in the main body of the paper. Note that both axes have a logarithmic scale.

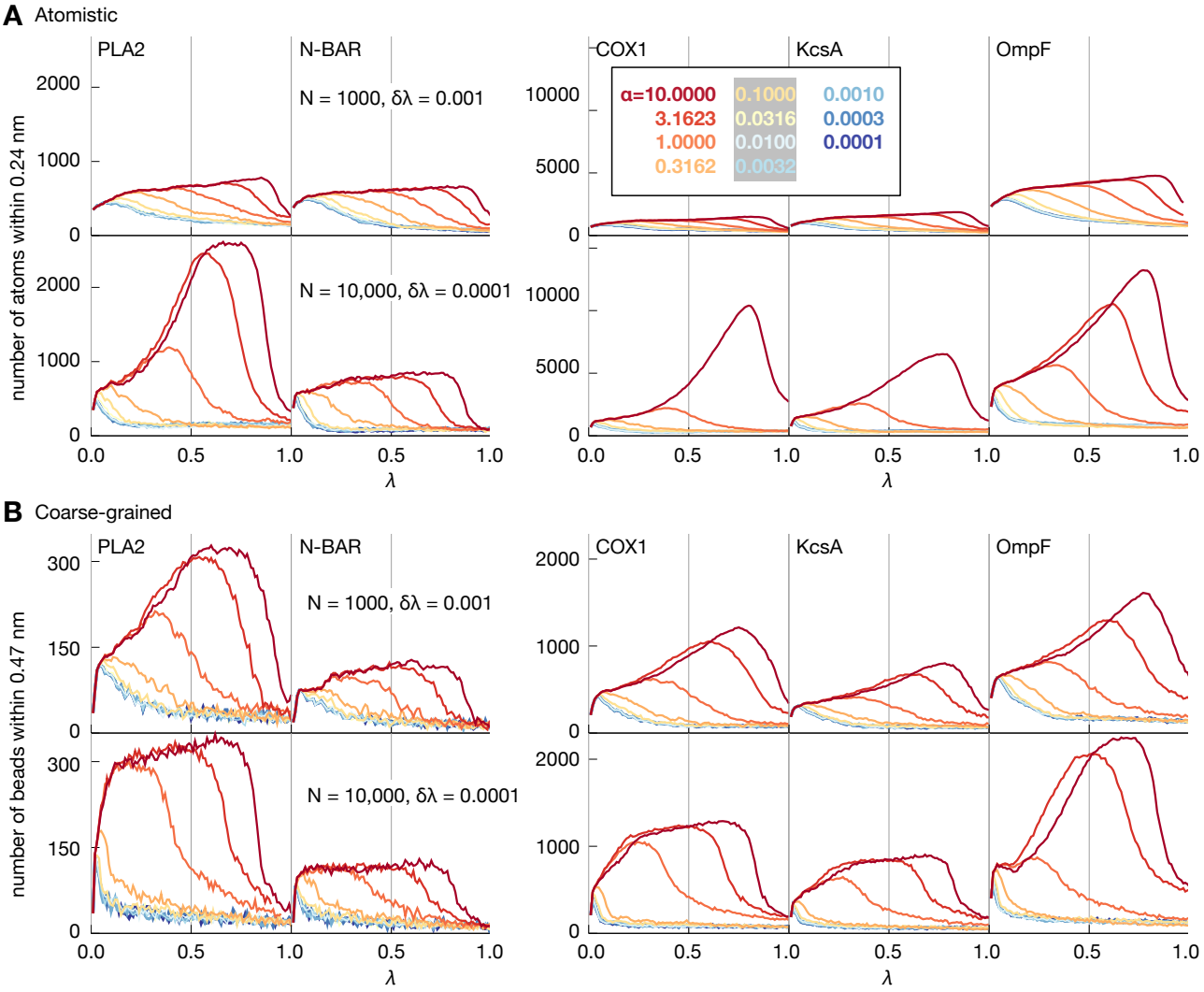

Figure S5: The number of atoms (or beads) within a set distance of the embedding protein varies as the van der Waals interactions between the protein and the remainder of the system are 'switched on' (i.e. as $\lambda \to 1$) . The five test proteins are modelled using both (A) atomistic and (B) coarse-grained representations. In all cases the results of eleven simulations spanning values of α from 0.0001 to 10 are plotted using a colour scale that smoothly progresses from red to blue. Each set is also run for either $N = 1\,000$ or $N = 10\,000$ timesteps. For clarity the results of only one of the three repeats for the shorter simulations are plotted. Here $b = 2$; the same analysis for $b = 1$ can be found in Figure 6 of the main body of the paper.

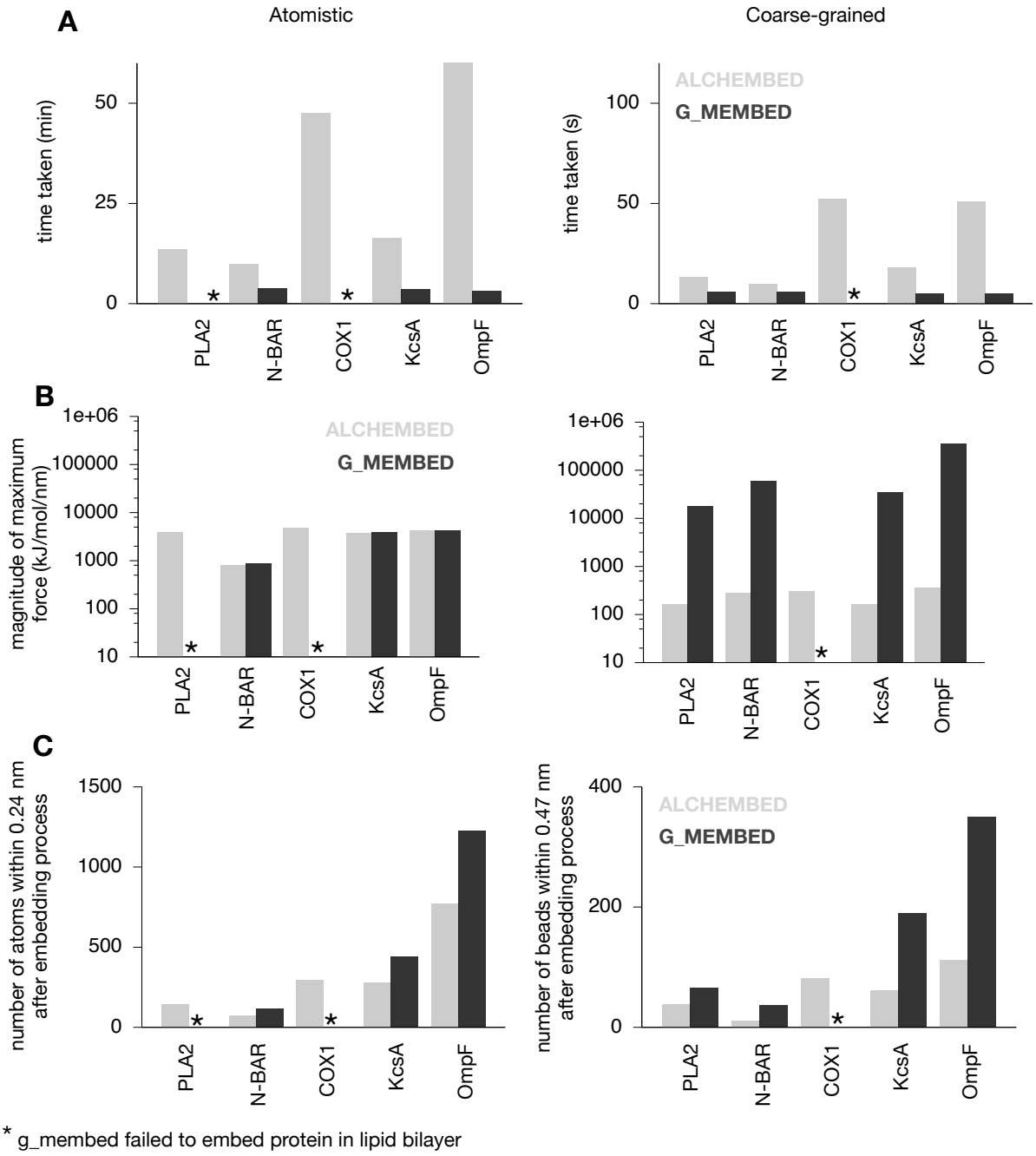

Figure S6: We embedded the five test membrane proteins into both the coarse-grained and atomistic simple POPC bilayer using g_membed. The initial configuration was identical to that used to start the alchembed simulations. We were unable to get g_membed working for GROMACS 4.6.x or 5.0.x so used version 4.5.5. (A) Alchembed is slower than g_membed in GROMACS but works for all the test cases, (B) has smaller forces during embedding and (C) leads to a smaller number of lipid or water atoms (or beads) being within 0.24 nm (or 0.47 nm) of the protein at the end of the embedding process. The variation of the number of close contacts with λ can be found in Figure S7 and the timings are given in Table S1. Both cases used $N = 1000$.

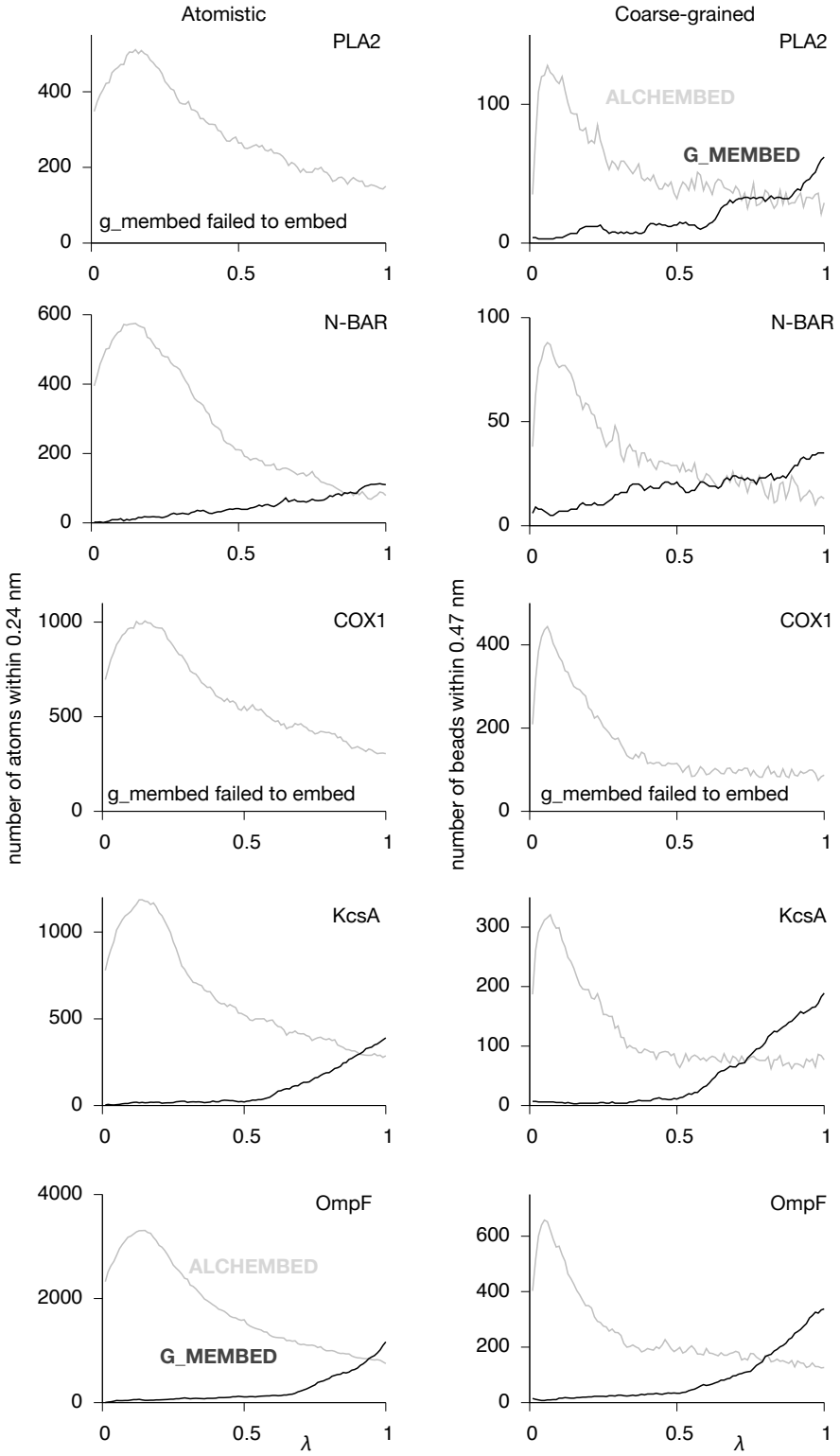

Figure S7: The number of atoms (or beads) within 0.24 nm (or 0.47 nm) of the protein is plotted as a function of λ for both alchembed (grey line, $\alpha = 0.1$) and g_membed (black line). The results are summarised in Figure S6.

	Atomistic		Coarse-grained	
Proteins	alchembed	g_membed	alchembed	g_membed
PLA2	13.4 min	–	13 s	5.8 s
N-BAR	9.8 min	3.8 min	9.8 s	5.8 s
COX1	47.6 min	–	52 s	–
KcsA	16.7 min	3.5 min	18 s	4.8 s
OmpF	107 min	3.1 min	55 s	4.7 s

Table S1: The g_membed approach is faster than alchembed, however, only alchembed successfully embeds all the test proteins into the simple POPC lipid bilayer. The timings are those reported by GROMACS and a dash (–) indicates that the method failed to embed the protein. GROMACS version 4.5.5 has used for g_membed (we could not get more recent versions to work) whereas version 5.0.2 was used for alchembed. All simulations used a single core of an Intel Xeon E5 workstation.

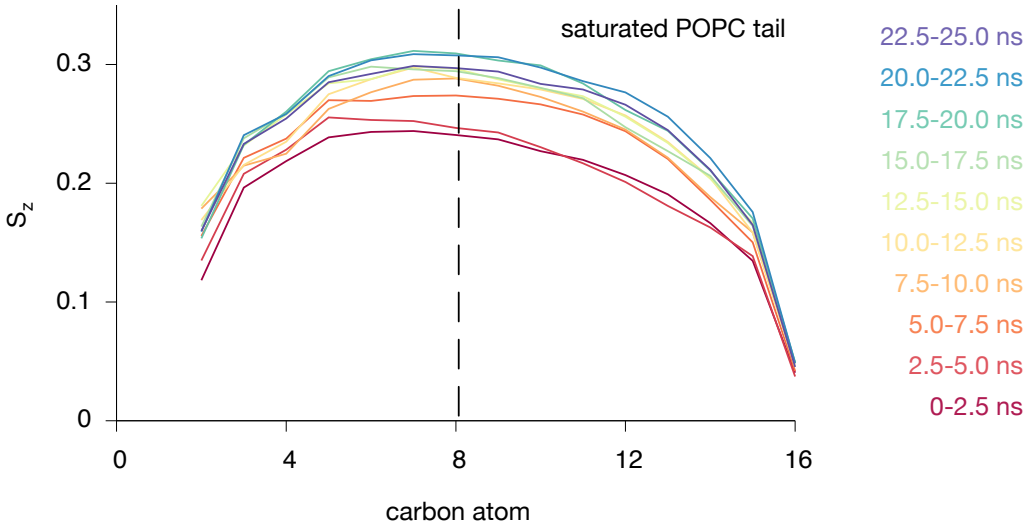

Figure S8: The deuterium lipid order parameter (S_z) for the sn-1 chain of POPC changes during the 25 ns simulation of the small patch of 128 POPC lipids. Broadly, the magnitude of the order parameter increases with simulation time, suggesting the lipids are becoming more ordered. The 25 ns trajectory was divided into ten equal blocks and S_z calculated for all frames within each block. The resulting ten traces are plotted and coloured using a scheme that progresses from red at the start of the simulation to blue at the end. To facilitate plotting some measure of lipid order as a time series, the value of S_z is plotted at carbon 8 in Figure 7 of the main body of the paper.

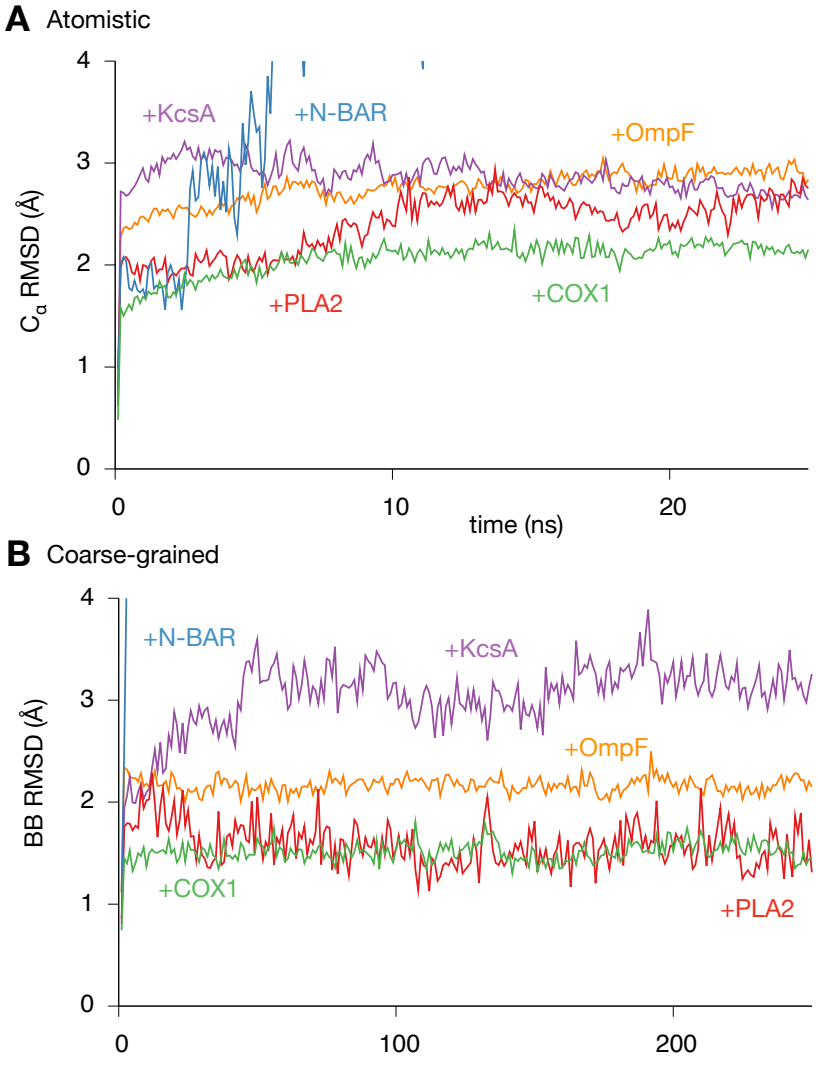

Figure S9: The test proteins are stable in molecular dynamics simulations following insertion by the alchembed method. This is measured by the (A) C_α RMSD for the atomistic simulations or (B) the RMSD of the backbone beads (BB RMSD) for the coarse-grained simulations. Shown are the values for the KcsA tetramer and OmpF trimer. The RMSD values for the monomers are, as expected, lower. Since N-BAR is very dynamic, the RMSD values are very large.

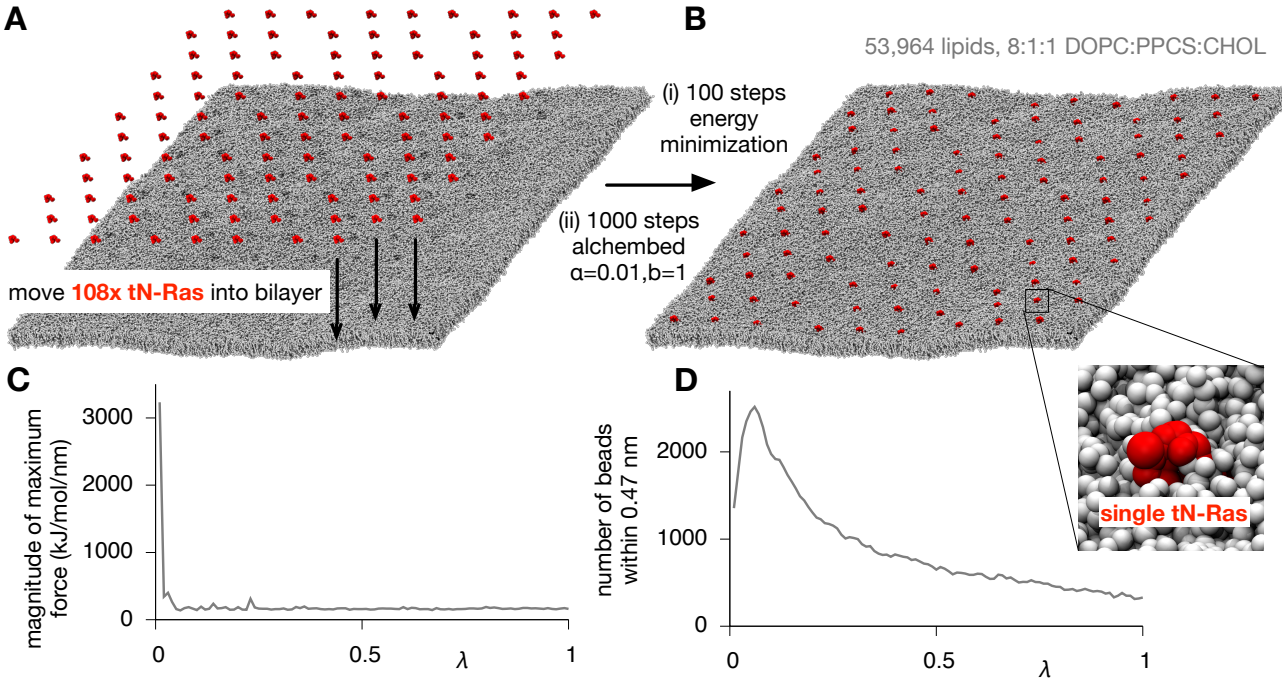

Figure S10: *Alchembed* easily and rapidly embeds 108 copies of tN-Ras into a very large undulating coarse-grained ternary lipid bilayer. This is a challenging test due to its size: the lipid bilayer contains 53,964 coarse-grained DOPC, sphingomyelin and cholesterol lipids in the ratio 8:1:1 and is 120 nm square. The truncated form of N-Ras (tN-Ras) is missing the soluble G-domain of the full length protein and has two lipid anchors covalently attached to two cysteines. (A) The proteins were arranged in a trigonal array above the bilayer. Each protein was moved towards the bilayer until the number of waters close to its lipid anchors was < 3, thereby embedding the lipid anchors in the bilayer. (B) The energy of the system was first minimised for 100 steps before a 1000 timestep *alchembed* simulation using the standard parameter set. A zoomed image of a single tN-Ras protein is shown in the inset. (C) Large forces are recorded early on in the embedding process, but this is still within an order of magnitude of the average maximum magnitude and so does not result in failure of the simulation. (D) As seen for the five test membrane proteins, the number of beads within 0.47 nm of any of the 108 tN-Ras proteins initially increases, but then rapidly falls as the van der Waals interactions between the proteins and the remainder of the system are "switched-on". The entire process took 48 minutes on a 6-core multi-threaded Intel Xeon E5 workstation CPU, which is fast given there are 2.1 million coarse-grained beads in the simulation. To demonstrate that the proteins are stably inserted into the undulating lipid bilayer a 250 ns molecular dynamics simulation was subsequently run starting from the final alchembed configuration. As expected the proteins diffuse on the surface of the bilayer whilst the bilayer continues to undulate (Figure S9).

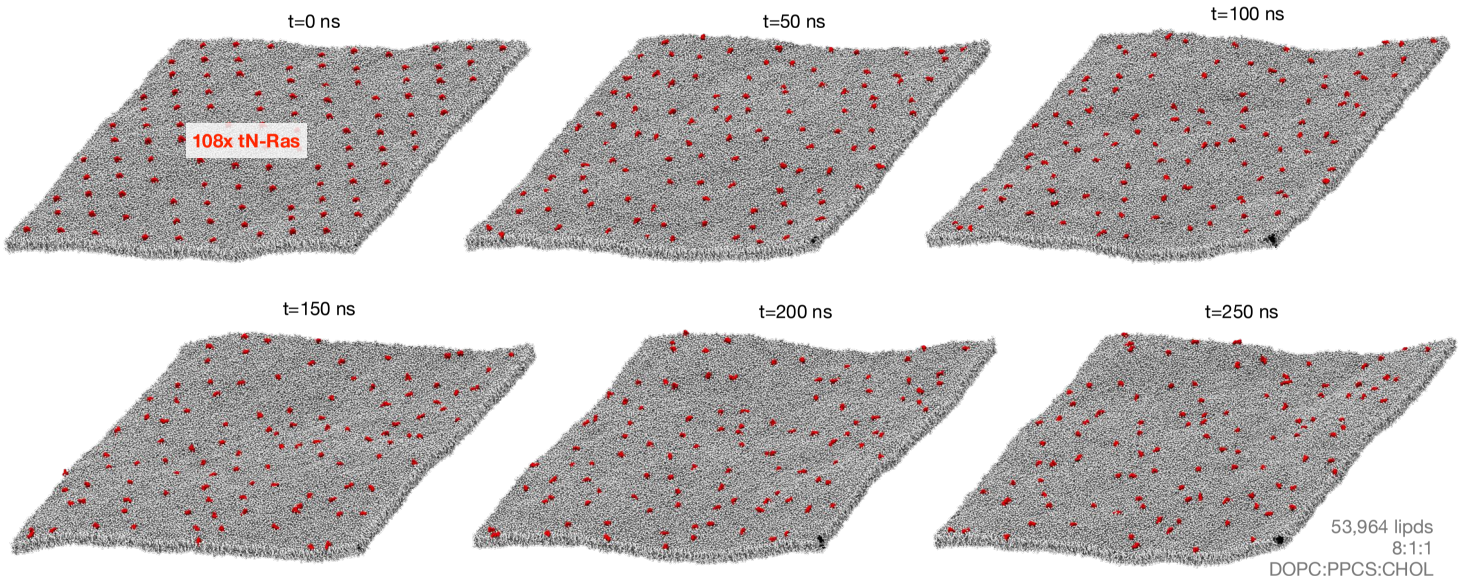

Figure S11: The number of atoms (or beads) within a set distance of the embedding protein varies as the van der Waals interactions between the protein and the remainder of the system are 'switched on' (i.e. as $\lambda \rightarrow 1$) . The five test cases shown in both (A) atomistic and (B) coarse-grained representations. In all cases the results of six simulations with different values of α (0.25, 0.50, 0.75, 1.00, 1.50, 2.00 and 2.50) are plotted using a colour scale that progresses from red to blue. Each set is also run with either $N = 1000$ or $N = 10000$ timesteps, with $\delta\lambda$ changing accordingly. Here $b = 2$; the same analysis for $b = 1$ can be found in Figure 6 of the main body of the paper.