Published in final edited form as:
J Chem Theory Comput. 2014 May 13; 10(5): 2165–2175. doi:10.1021/ct500003g.

A Sidekick for Membrane Simulations: Automated Ensemble Molecular Dynamics Simulations of Transmembrane Helices

Benjamin A Hall[1,2], **Khairul Abd Halim**[1], **Amanda Buyan**[1], **Beatrice Emmanouil**[1], and **Mark S P Sansom**[1,*]

[1]Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU

[2]current address: Microsoft Research Cambridge, 21 Station Road, Cambridge, CB1 2FB

Abstract

The interactions of transmembrane (TM) α-helices with the phospholipid membrane and with one another are central to understanding the structure and stability of integral membrane proteins. These interactions may be analysed via coarse-grained molecular dynamics (CGMD) simulations. To obtain statistically meaningful analysis of TM helix interactions, large (N ca. 100) ensembles of CGMD simulations are needed. To facilitate the running and analysis of such ensembles of simulations we have developed Sidekick, an automated pipeline software for performing high throughput CGMD simulations of α-helical peptides in lipid bilayer membranes. Through an end-to-end approach, which takes as input a helix sequence and outputs analytical metrics derived from CGMD simulations, we are able to predict the orientation and likelihood of insertion into a lipid bilayer of a given helix of family of helix sequences. We illustrate this software via analysis of insertion into a membrane of short hydrophobic TM helices containing a single cationic arginine residue positioned at different positions along the length of the helix. From analysis of these ensembles of simulations we estimate apparent energy barriers to insertion which are comparable to experimentally determined values. In a second application we use CGMD simulations to examine self-assembly of dimers of TM helices from the ErbB1 receptor tyrosine kinase, and analyse the numbers of simulation repeats necessary to obtain convergence of simple descriptors of the mode of packing of the two helices within a dimer. Our approach offers proof-of-principle platform for the further employment of automation in large ensemble CGMD simulations of membrane proteins.

Introduction

Molecular dynamics (MD) has been used extensively to understand the structure and dynamics of a wide range of different systems, providing dynamic structural insights into behavior on experimentally inaccessible time- and length scales 1–5. With increasing computing power and sophistication the complexity of systems studied has increased, leading to the application of semi-automated workflows 6 and of high throughput (HT) approaches for performing large comparative analyses of multiple systems 7, 8. HT workflows have been applied widely to docking problems 9–11 and are increasingly being

*to whom correspondence should be addressed mark.sansom@bioch.ox.ac.uk.

Europe PMC Funders Author Manuscripts

applied to MD simulations 12. However, such workflows typically have several manual steps (see the workflow in figure 1), which are both laborious and can potentially introduce variation in the simulation protocol. Alongside this growth in HT approaches, the ongoing successes of MD approaches and related techniques to modelling biomolecular systems has lead to the development of multiple different approaches to automating specific workflows through user-friendly interfaces for performing simulations 13. This in turn has driven the development of web interfaces allowing outside users to access and use data from the simulations in an accessible form for both specialists and non-specialists 7, 14, 15.

High throughput approaches also pose novel issues for data management and analysis. When comparisons between multiple systems are considered (alongside simulation repeats for statistical confidence), these can create problems in terms of simulation and data management from the sheer size and number of individual systems, further slowing the analysis. The MD simulations themselves are computationally intensive, frequently requiring supercomputer time to generate useful data for large systems. Simulation software (such as GROMACS 16, NAMD 17, Charmm18, GROMOS 19 and AMBER20) typically manage scaling over large numbers of processers, and whilst this is becoming increasingly optimized, scaling can break down for small systems and extremely large numbers of processors 16. Coarse grained (CG) simulations increase simulation speed and improve scaling through use of a simplified description of the system, at the cost of some loss of detail 21–23, but still suffer from many of the other disadvantages in terms of set up and analysis.

Here we present the Sidekick software, a grid application that addresses the challenges posed by manual setup, simulation and data management, and scaling for CG MD simulations of helices in a bilayer though deep automation. From an input sequence or set of sequences, entered through a command line interface or web page, simulation systems are built, simulations run, and analysis performed on the resulting trajectories, before being stored in an ordered file structure on a central server. Simulations are performed in replicate (typically > 100 times each) on single processors, giving excellent sampling and creating an “embarrassingly parallelizable” problem which scales perfectly across the computing resource. Sidekick itself is a collection of python scripts and modules, which either individually perform distinct steps in the pipeline or execute external applications (e.g. GROMACS commands). Ensembles of pipelines are performed over a mixed grid/cluster set up, consisting of a mixture of machines, including a dedicated cluster as well as underused workstations (totaling in excess of 300 processors). Job management is performed using Xgrid software and data is shared using NFS (both available as part of the operating system). Raw data is available for detailed analysis across the ensemble of simulations that are generated, alongside web interfaces describing peptide properties based on key metrics (described below). The application of grid based technologies here follows a different approach from that taken in a wide range of disciplines, including the Seti@home 24, folding@home 25, “screensaver lifesaver” 10 (each dedicated to a specific research focus), the UK National Grid Service and OpenMacGrid (general use grids). Our approach is necessarily different from these, due to the computational time required to perform the simulations, and the storage and transport requirements created by large, continuous MD trajectories. In contrast to high throughput approaches driven by specific tasks such as

calculating the PMF of drug binding 26, or the calculation of whole protein dynamics (e.g., Copernicus 27), Sidekick is focused on calculating and presenting a heavily sampled description of the behavior of transmembrane (TM) helices at equilibrium.

To date we have used this software to aid our understanding of NMR data on model peptides 28, 29, of high throughput experimental assays 30, of bacterial signalling systems 31, and of fusion peptides 32, 33. Over four years the application of this approach by ~10 users in our group has generated over 5 TB of data and used ~1,000,000 CPU hours, and the size of the grid/cluster has been several times upgraded to take this into account. The highly modular approach taken in the writing the code has allowed for new features to be added rapidly (including changing starting configurations, adding umbrella potentials, inline forcefield modification). This increases the potential for development of the code in future for novel purposes, and as such we regard Sidekick as both a tool in itself but also as a framework for future tool development.

Here we validate the software by calculating the likelihood of arginine residue ‘guests’ in ‘host’ hydrophobic helices to insert into a phospholipid bilayer. We take 2 related helix systems which have been explored experimentally; poly-leucine helices of length 11 and 12 (here referred to as pL11 and pL12). The energy of insertion is found to be ~1 kCal/mol, consistent with energies measured experimentally in the refolding and insertion of pmOmpA 34 and insertion of helices by the translocon apparatus 35, and compatible with equilibrium calculations of peptide insertion in all atom simulations 36. We further use the pipeline technology to explore the association of ErbB1 TM helix dimers in a DPPC membrane. We analyse convergence of the helix packing modes and show that the major packing mode predicted is in agreement with a recent NMR structure 37 of the ErbB1 TM helix dimer.

Methods

Core software

Individual Sidekick pipelines are based on GROMACS (www.gromacs.org), python and numpy. Optionally, matplotlib can be used to generate plots for each individual run. The management of simulations across the cluster is controlled through Xgrid in our implementation, though other grid managers (such as SGE) have been used successfully. Data (in terms of software and outputs) is stored on a fileserver and in our implementation shared using NFS, though this is performed at the level of the operating system and must be set up manually for each “agent” running the pipelines. Analysis is performed using custom analysis code written in python and numpy. To ensure faithful reproduction of different coarse grained forcefields, the unmodified relevant scripts are run within the pipeline. Specifically, GROMACS tools editconf/trjconv are used for structure/trajectory manipulation, grompp/mdrun for simulation (and pre-processing), genbox for solvation and genion for the addition of ions. Sidekick is available from http://sbcb.bioch.ox.ac.uk/Sidekick/

Single helix simulation pipelines

Individual simulations are performed within independent automated pipelines, where from the input sequence and options are used to generate a CG α-helix, a water/lipid system (with counter ions) and perform the simulation (Fig. 1). An individual pipeline is started using the script CG_Helix.py, and can be run outside of the ensemble for testing purposes. The individual pipelines allow for selection of a wide variety of different aspects of the simulation (see Table 1 for all options). Key parameters for all simulations include the choice between different forcefields (either the “Bond”21 or MARTINI23 CG forcefields) and different available versions of forcefields, system size (selected from predefined system sizes), and initial lipid configurations. The default behaviour is to build a system consisting of a DPPC/water/counterion system around the peptide, with a cuboid box 7 x 7 x 15 nm where lipids are located in a narrow section in the centre of the box, and to use semi-isotropic pressure coupling when performing the simulation (coupling XY and Z separately). This helps ensure that the bilayer is formed in the XY plane and within the first 10-20 ns. Alternatively, a preformed bilayer can be used (with semi-isotropic pressure coupling) or the system can be built as a 9 x 9 x 9 nm cube with lipids evenly distributed throughout the box, and the pressure coupled anisotropically (ie coupling X Y and Z separately), when the “--unbias" option is enabled. The three system configurations alter the behaviour and reliability of the simulations to give stable bilayers in any plane; 100 % of preformed bilayers remain stable, ~95 % of “biased” bilayers form stably in the XY plane and ~75 % of unbiased self assembly protocols form bilayers in the XY, YZ or XZ planes.

Simulation run files are generated based on the input options and a set of standard template run files, which are used to generate the desired system configurations for each simulation. Independence of each run is achieved by altering the starting seed for the generation of velocities in each simulation. However, if the user wishes, it is further possible to vary the systems, either by altering the position and angle of the peptide (which by default is located in the centre of the box and aligned to the Z axis) and by using the seed for each run to seed the positioning of lipids performed by genbox. Alteration of the peptide position and orientation are of particular value when using preformed bilayers. Temperature and pressure are coupled at 323 K and 1 Bar using the Berendsen38 weak coupling algorithm (τ_T = 1 ps and τ_P = 10 ps). The compressibility is set at 3×10^5 (1/bar).

Individual pipelines are designed to be robust through the use of a) wall clocks and b) automatic restarts with a shorter timestep (Fig. 2). The purpose of these is to mimic the typical CG MD workflow whilst ensuring that simulations stuck in unproductive states are cleared from the cluster to allow new simulations to run. The wall clock ensures that jobs that stop generating data but remain running indefinitely do not block compute resources, which can lead to a net reduction in grid performance over time. The reduction of the timestep reflects the lack of knowledge *a priori* regarding the selection of an appropriate timestep in coarse grain simulations (for extensive discussion, see 39, 40). On completion of mdrun we test output trajectories and logs for successful termination. If this has not occurred, the timestep is reduced (from 20 fs to 10 fs) and the simulation is restarted from the last frame for which position and velocity have been collected. The resulting trajectories

are concatenated at the end of the pipeline prior to analysis. Such crashes are relatively rare and at least partially system dependent.

Multiple helix pipelines

Simulation pipelines also exist to predict the behaviour of systems with multiple helices41. These are always performed using a preformed bilayer, in order to specify helix orientation, and the user is limited to using DPPC, DPPE and DPPG lipid mixes. Simulations are performed as described for single helices, and individual helices in the simulations are analysed using the same metrics (position, tilt and rotation) in addition to new metrics specific to pairwise helix interactions. These new metrics include relative positions of helices, helix contact analyses and termini-to-termini distance (useful when considering helices which are normally connected by a flexible linker region).

The extension of the approach to multi helix systems is supported by the extensive use of object oriented programming, and specifically the reuse of large sections of the code by inheritance, allowing common functions to be shared between the different use cases. Furthermore the code is designed to be highly modular, with discreet functions in the code (e.g. system generation, MD simulation etc.) designed to be easily reused in new code.

Grid architecture and Sidekick ensembles

Simulations are performed across a hybrid grid/cluster architecture (see Supporting Information Fig. S1). Individual pipelines run independently and on single cores, and as such do not require fast networking, and by default single helix simulations write data locally, only transferring output data to the central fileserver at the finish of the pipeline. A mixture of dedicated cluster nodes, and idle workstations are used to perform calculations, connected to the controller and central fileserver by wired or wireless networking. However, access to the core software and the centralised data store is achieved through NFS, reflecting the practical issues with sharing software and output data across a typical grid setup. As such, we refer to our architecture as hybrid grid/cluster as it has features of both HPC paradigms.

Machines can be classed as agents, headnodes and clients depending on their specific functionality. Agents perform simulations as instructed by the grid manager. Headnode tasks (serving files, managing the cluster and presenting a web server interface to the data) can be centralised on a single machine or split between multiple machines; the cluster manager and fileserver need direct access to the agents, and the webserver needs direct access to the fileserver and cluster manager for submitting and retrieving job information. Finally, client machines submit jobs and retrieve data, either through a command line interface or through the website (complete data retrieval is only possible from the command line). Agents can write output data locally, and either copy to the fileserver on completion, or write directly to the fileserver. When data are written locally by the agents, network speeds do not pose a bottleneck for the pipelines, allowing agents to perform simulations over slower networks (e.g. wireless). The concurrent transfer of data from multiple agents however can lead to network congestion, causing problems relating to Xgrid communication. To prevent this, we

Europe PMC Funders Author Manuscripts

can switch the pipeline to writing centrally. Whilst this presents the possibility of reduced performance, no notable slow-downs were found when using gigabit ethernet.

In our implementation, an Xgrid controller initially accepts a batch file (in XML) describing all individual pipelines to be run (Fig. 3). This is generated programmatically, but for SGE job array scripts can be written by hand to generate the ensemble. In principle, this approach could be applied to cloud resources where a MapReduce model 42 could be employed to perform and aggregate the pipeline and outputs. The controller instructs agents (slave nodes in SGE) to perform individual pipelined simulations, which read software from the centralised fileserver, and at the end of a pipeline copy data to the central datastore. Finally, users can view graphics rendered in javascript/HTML5 through the webserver describing the metrics of individual sequences over their ensemble, or the complete output can be accessed through the command line. The ensemble size is limited by available resources (storage and compute). Ensembles up to ~12,000 simulations have been performed.

Simulations presented here

CG MD simulations of the helices in ~128 lipid DPPC bilayer were performed, using a membrane self-assembly protocol successfully employed in the past for single TM helices 30and for more complex membrane proteins 21 . Sequences of each peptide used here are given in Table 2. Simulations are performed using Gromacs v343. The “unbiased” simulation protocol was followed (where the box is cubic and pressure is coupled anisotropically), and the temperature was coupled at 323 K. About 400 simulations were performed for each helix sequence.

In single helix simulations presented here we have used a local modification of the MARTINI forcefield 23, 44 (the “Bond” forcefield), following the protocol used previously for predicting translocon mediated insertion 30. The MARTINI approach reduces system size from ~4 non-hydrogen atoms to a single coarse grained particle. Lennard-Jones interactions between particles are calculated based on 5 interaction levels between 4 classes of particle. These classes reflect polar, charged, mixed polar/apolar and apolar characters, where mixed and charged particle types are further divided into 5 and 4 subclasses respectively to reflect hydrogen bonding capabilities. Lennard-Jones interactions were shifted to zero between 0.9 and 1.2 nm. Electrostatic interactions are treated Coloumbically and were shifted to zero between 0 and 1.2 nm. The peptide backbone in the helix is originally generated from an ideal all atom α-helix and converted to its coarse grained representation. Helicity is maintained in the peptide backbone through harmonic restraints between hydrogen bonded particles.

For simulations of ErbB1 TM helix dimerization, we use the same forcefield and simulation parameters as for single helix simulations. Two parallel helices are inserted into a preformed DPPC membrane, separated from one another by ~5 nm. 50 x 500 ns simulations were performed of helix pairs in total.

Analysis of helix dimer packing interactions

TM helix dimers were characterised in terms of their helix crossing angles. The Cα coordinates of each helix and their respective centroids were calculated in each frame of a

Europe PMC Funders Author Manuscripts

given trajectory. If the distance between helix centroids exceeded 10 Å, the frame was skipped as the helices were considered not to be packed against one another. Otherwise, the first eigenvector of each helix was calculated, and the magnitude of the crossing angle was determined using the cross product of the helices' eigenvectors. To determine its handedness using the standard conventions 45, the helices were rotated so the first helix was aligned to the +x-axis. New eigenvectors were calculated, their cross product was determined and aligned with the +x-axis. The orientation of the eigenvector through the first helix along the z-axis and the x-component of its centroid determined the handedness of the angle.

An iterative jackknife approach 46, 47 was used for assessing the convergence of the crossing angle distributions and of the helix contact residues. This technique allowed the sampling of the whole range of possible subset sizes. It has also been demonstrated that randomly choosing a small number (e.g. 1000) of the total possible combinations is sufficient for sampling purposes 46. This is necessary because, as the number of simulations increases, the number of possible combinations per subset sizes increases such that it becomes unachievable to sample all possible combinations.

To assess the convergence of the list of interhelix contact residues, for each chosen subset, all distance matrices from all the frames of the simulations were averaged together to give an overall distance matrix. Then, for each of the chosen subsets, the six residues which have the lowest Cα distances are tallied and put together into an overall matrix. This overall matrix holds the frequency of the closest contacts that appear in the subsets of the simulations, generating an overall heat map for contacting residues. This is done for each subset size.

Results & Discussion

Performance of the Pipeline

We have evaluated the Sidekick approach via two examples of TM systems which benefit from automated running and analysis of ensembles of simulations: (i) insertion of single TM helices into a bilayer, running and analysing ca. 8000 simulations with a total CG simulation time of ca. 0.8 ms; and (ii) dimerization of TM helices in a preformed lipid bilayer, analysing the convergence of the predicted helix packing as a function of ensemble size.

Single TM Helix Insertion

One major use of Sidekick is to investigate the probabilities of insertion of TM helices into a lipid bilayer. This is of interest in contributing to an understanding of the factors which influence the stability and structure of membrane proteins 48, 49. There have been a number of biochemical 35, 50 and biophysical 51, 52 studies of this problem, and it has also been the subject of a number of computational studies, including both CG 30, 53 and atomistic 36, 54 MD simulations. Of particular interest is the thermodynamics of insertion of otherwise hydrophobic TM helices containing a cationic arginine sidechain 55–58. It has been shown that although complete burial of a cationic sidechain within the hydrophobic core of a bilayer is unfavourable, local distortion of the bilayer and ‘snorkelling’ of the arginine sidechain may occur so the cationic moiety of the sidechain is able to form favourable electrostatic interactions with anionic phosphates of the lipid headgroups. We

have previously shown that the insertion energies of peptides into the membrane may be accurately estimated using ~400 simulations 30.

To evaluate the positional dependence of the insertion of Arg-containing TM helices in a bilayer there have been experimental studies whereby the length of a model hydrophobic TM helix is ‘scanned’ with a single Arg residue and the effect of the cationic sidechain on the insertion propensity of the TM helix evaluated 49. We have therefore used Sidekick to generate a large (ca. 8000) ensemble of simulations, scanning an Arg along two hydrophobic helices of differing length, and performing TM helix/lipid bilayer self-assembly simulations to evaluate profiles for the insertion thermodynamics of the helices.

Simulations of Arg ‘guest’ residues at different positions along two leucine-containing ‘host’ helices were performed. The two systems containing either 11 or 12 leucines forming the hydrophobic cores of 19- or 20-mer helices (Table 2). The two ‘host’ helices were selected based on their relative insertion energies into a DPPC bilayer such that the L11 helix marginally favours an interfacial orientation, whilst the L12 helix favours a transmembrane orientation. As such these two families of helices may be expected to be sensitive to introduction of a single arginine.

For each ensemble of simulations, an individual simulation was run from an initially random arrangement of lipids in a box also containing the helix and waters. The lipids self assemble to form a bilayer, with the helix adopting a TM and/or a surface orientation (Fig. 4; also see Supporting Information Movie 1). For each helix sequence an ensemble of ca. 400 simulations each of duration 100 ns was performed.

Sidekick was used to analyse the position and orientation of the helix relative to the bilayer over the length of the trajectory. Thus each output trajectory of an ensemble of runs was analysed in terms of helix position relative to the centre of the membrane, helix tilt angle in the membrane, and the azimuthal rotation of the helix in the membrane (that is, the rotation angle around the helix axis, also referred to as the direction of tilt) (Fig. 5A). The percentage of helix insertion in the bilayer is quantified based on helix tilt and the depth of the helix in the bilayer. Finally the end state of the bilayer is assessed to determine its location and whether the bilayer has formed a continuous sheet. In addition to writing the specific metric values over time to file, a summary file with information on the end state and percentage insertion is also generated to aid subsequent filtering of data and ensembles.

For each timepoint in the simulation, a helix is defined as transmembrane or not transmembrane based on the tilt angle and depth of the helix in the membrane (Fig. 5B). Specifically, if the centre of mass of the helix backbone is within 1 nm of the centre of mass of the bilayer core (along the axis of the bilayer normal) and the helix tilt angle is less than 65 ° relative to the bilayer normal, the helix is considered to be inserted in the membrane. The (apparent) free energy of insertion of a helix across the ensemble is calculated based on the percentages of time in inserted and non-inserted states. The free energy of insertion is defined as:

$$\Delta G_{APP} = -RT \ln K$$

where

$K = \%(inserted)/\%(not\text{-}inserted)$.

This is intended to match the definition used in experimental studies 50 where the *%(inserted)* and *%(not-inserted)* are calculated from the percentage of single glycosylation events relative to double glycosylation events following translocon-mediated insertion

To calculate an energy profile for insertion of an arginine residue at different positions in the host hydrophobic helix we place arginine residues at different positions in the sequence (Table 2) and compare the change in energy relative to the purely hydrophobic 'host' helix (Fig. 6). The resultant profiles are similar for both the L11 and L12 hosts with the important difference that the baseline profile is overall favourable for TM insertion for the L12 host. In each case the peak in the profile is for an Arg residue in the centre of the helix in which location it is unable to effectively snorkel to headgroup phosphates on either side of the bilayer in order to stabilize TM orientation of the helix.

This result is consistent with our previous results based on CFTR-derived helices which predicted an energy barrier within an order of magnitude of the experimentally determined values from translocon mediated helix insertion experiments35, 50 and with refolding experiments of the membrane protein pmOmpA34. Furthermore, these values are consistent with results from all atom simulations at equilibrium36.

ErbB1 TM helix dimerization

As a second example of the use of Sidekick to enable the running and analysis of an ensemble of TM helix simulations we have investigated the dimerization of a CG model of the TM helix domain from the ErbB1 receptor. ErbB1 belongs to the ErbB family of receptor tyrosine kinases, which is important in regulation of cell growth, differentiation and migration. Aberrant signalling by ErbB1 is associated with various cancers 59, 60. The TM region of ErbB1 contains two small-x_3-small motifs 61, one towards its N-terminus and one towards its C-terminus. A recent study by Endres et al 37 has shown the TM region of the ErbB1 is crucial for the activation of the receptor as the intracellular domain alone is monomeric and incapable of independent dimerization. Using an NMR based approach, they showed that the ErB1 TM-juxtamembrane (TM-JM) region formed a contact at the N-terminus with a crossing angle of ~−44°, corresponding to a right-handed packing of the TM helices 37. This packing is mediated via the N-terminal TGxxGA sequence motif and is required for the correct orientation of the JM region, and thus the activation of the receptor. They further showed that mutation of the motif to poly-isoleucine results in significant inhibition of the receptor. We therefore wished to explore the dimerization of the TM helices of ErbB1 in CG simulations, in particular to analyse the size of the simulation ensemble required for reasonable convergence to a structure compatible with the existing experimental 37, 62 and computational (e.g. 63, 64) data.

To this end two ErbB1 TM helices were pre-inserted in a DPPC bilayer at an initial separation of ~6 nm. This system was used to initiate an ensemble of 50 x 500 ns helix dimerization simulations performed using the Sidekick pipeline. In simulations each case the

helices were observed to spontaneously associate during the simulations from their initial separated configuration (Fig. 7).

The ensemble of 50 dimerization simulations was analysed in terms of the helix crossing angle distributions as these allow for ready identification of the frequency of formation of right-handed (i.e. negative crossing angle) helix dimers. Right-handed packing of TM helices is characteristic of interactions via a small-x_3-small sequence motif 61, as exemplified by glycophorin 65. The full ensemble of 50 simulations showed a strong preference for right handed packing of the ErbB1 TM helices, in agreement with NMR structures (Fig. 8A). We therefore used this ensemble to explore convergence of the simulations to this characteristic structure. Our first pass at such analysis consisted of randomly selecting a single subset of 5, of 10, and of 20 simulations from the full ensemble of 50 simulations and analysing the crossing angle distributions for each subset (Fig. 8A). this suggested that running smaller ensembles could result in a biased view of the helix packing model. Thus for the single $N = 10$ subset randomly selected the frequency of right handed crossing was slightly less than for left handed packing of the helices.

To approach this more systematically we next performed a Jackknife analysis of the $N = 50$ ensemble. Based on this analysis we could estimate the convergence of e.g. the probability of a right-handed helix packing and of the crossing angle for right handed packing as a function of sub-ensemble size M (see Fig. 8BC for details). Based on this analysis one can judge that reasonable convergence is attained after a sub-ensemble size of ca. 20 or above. This is a valuable indication of what sizes of ensembles to generate if Sidekick or related methodologies are used to explore different families of related TM helix dimers via CG simulations 66.

We extended this analysis to that of convergence of the residues involved in the ErbB1 TM-TM packing interactions. Thus we used jackknife analysis to explore convergence of the helix-helix contact matrix (Fig. 9A). From this we can see convergence to a preferred motif (‘top 10 contacts’) corresponding to symmetrical interactions of residues T624, G625, G628, A629 in the two TM helices. As with the helix crossing angle distributions, convergence seems to be achieved for subensembles of N = 20 or above. If one selects a representative structure from the ensemble the resultant structure agrees well with those from NMR studies (Fig. 9B). We note that the modal crossing angle from our CG simulations (ca. -26°) is a little smaller than that for the recent NMR structure (ca. −44°). This difference is likely to reflect the absence of the juxtamembrane region from the model of the ErbB1 TM domain used in our simulations, and possibly also the difference between a DPPC bilayer (used in our simulations) and a DHPC/DMPC bicelle (used in the NMR experiments). We further note that this difference is comparable to the variation observed between alternative experiments on the influenza M2 proton channel (Carpenter et al).

Conclusions

We have demonstrated the utility of the Sidekick pipeline for the running and analysis of CG-MD simulations of TM helices. This approach enhances our ability to run high throughput simulations of large ensembles of TM helices and so to explore biophysical

properties of model membrane systems. Such high throughput approaches to biophysics are becoming increasingly commonplace, and their increased automation is a key element facilitating more widespread adoption 12, 27. Thus approaches such as Sidekick will enable e.g. integration of simulation studies within experimental workflows.

The application of Sidekick to analysing Arg residue positional effects on TM helix insertion into bilayers exemplifies the need for automated simulations to enable analysis and comparisons of large ensembles of different simulations. The results of a systematic survey of arginine containing single helices reproduce available experimental and simulation data. And extend our understanding of the fundamental interactions underlying membrane protein folding and assembly.

We have further demonstrated that Sidekick-enabled ensemble simulations are capable of predicting the interactions of the TM regions of ErbB1, and we have highlighted the importance of a more formal analysis of convergence of helix packing models which can be achieved using an automated approach. This is of especial importance if CG-MD and multiscale simulation based modelling are to complement experimental e.g. NMR studies in allowing us to understand TM domain conformational dynamics in RTKs and related receptor proteins. The success of the application to the ErbB1 homo-dimer shows that Sidekick offers a platform for performing systematic ‘parameter sweeps’ of TM helix dimerization. Whilst we have exclusively examined CG systems here, the approach could equally apply to all atom approaches with sufficient computing power, or multiscale approaches6. Such extensions are supported by the wide use of object oriented programming and modular design of the pipelines, allowing for new functions and behaviours to be added relatively easily. In this way it should be possible to explore e.g. changes in structure and stability of dimerization across extended families of TM helices from receptors 66.

Future studies will include the development of increasingly complex pipelines which allow the analysis of distinct properties and larger systems. For transmembrane helices, a valuable addition to the pipelines would be the ability to perform energy calculations (e.g. potential of mean force calculations, steered molecular dynamics, or thermodynamic integration). Preliminary studies have shown that it is possible to add an umbrella potential to single helix simulations, demonstrating that potentials of mean force may be calculated through relatively small modifications to the pipeline. Similarly, the ability to convert between coarse-grained and atomistic representations 6 within a pipeline will add to the usefulness of the tool in future.

The applications of automated simulation extend beyond transmembrane helices. Future efforts will involve the expansion of the types of system which can be studied. Coarse grained nucleic acid/lipid interactions 67 represent a natural future target for pipelined simulations. Protein-protein interactions 68 and protein-membrane interactions 69 also present an attractive goal, but will require advances in the development of more generic metrics to describe the motions and interactions within these rather more complex systems, and may require more powerful machine learning approaches for understanding the resulting aggregated data. The future of pipelined simulations therefore promises the ability to make

biomolecular simulations easier and more reproducible across a broad range of different domains.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank the BBSRC and the Wellcome Trust for support. KAH is supported by The Khazanah-OCIS Merdeka Scholarship Program.

References

1. Lindahl E, Sansom MSP. Membrane proteins: molecular dynamics simulations. Curr Opin Struct Biol. 2008; 18:425–431. [PubMed: 18406600]
2. Freddolino PL, Arkhipov AS, Larson SB, McPherson A, Schulten K. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. Structure. 2006; 14:437–449. [PubMed: 16531228]
3. Schäfer LV, de Jong DH, Holt A, Rzepiela AJ, de Vries AH, Poolman B, Killian JA, Marrink SJ. Lipid packing drives the segregation of transmembrane helices into disordered lipid domains in model membranes. Proc Natl Acad Sci USA. 2011; 108:1343–1348. [PubMed: 21205902]
4. Johansson ACV, Lindahl E. Protein contents in biological membranes can explain abnormal solvation of charged and polar residues. Proc Natl Acad Sci USA. 2009; 106:15684–15689. [PubMed: 19805218]
5. Dror RO, Jensen MØ, Borhani DW, Shaw DE. Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations. J Gen Physiol. 2010; 135:555–562. [PubMed: 20513757]
6. Stansfeld PJ, Sansom MSP. From coarse-grained to atomistic: a serial multi-scale approach to membrane protein simulations. J Chem Theor Comp. 2011; 7:1157–1166.
7. Sansom MSP, Scott KA, Bond PJ. Coarse grained simulation: a high throughput computational approach to membrane proteins. Biochem Soc Transac. 2008; 36:27–32.
8. Chetwynd AP, Scott KA, Mokrab Y, Sansom MSP. CGDB: a database of membrane protein/lipid interactions by coarse-grained molecular dynamics simulations. Molec Memb Biol. 2008; 25:662–669.
9. Salwinski L, Eisenberg D. Computational methods of analysis of protein-protein interactions. Curr Opin Struct Biol. 2003; 13:377–382. [PubMed: 12831890]
10. Richards WG. Virtual screening using grid computing: the screensaver project. Nature Rev Drug Discovery. 2002; 1:551–555. [PubMed: 12120261]
11. Schuttelkopf AW, van Aalten DMF. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Cryst D Biol Cryst. 2004; 60:1355–1363.
12. Harvey MJ, De Fabritiis G. High-throughput molecular dynamics: the powerful new tool for drug discovery. Drug Discovery Today. 2012; 17:1059–1062. [PubMed: 22504137]
13. Barrett CP, Noble MEM. Dynamite extended: two new services to simplify protein dynamic analysis. Bioinformatics. 2005; 21:3174–3175. [PubMed: 15855246]
14. Gerstein M, Echols N. Exploring the range of protein flexibility, from a structural proteomics perspective. Curr Opin Chem Biol. 2004; 8:14–19. [PubMed: 15036151]
15. Vohra S, Hall BA, Holdbrook DA, Khalid S, Biggin PC. Bookshelf: a simple curation system for the storage of biomolecular simulation data. Database. 2010:baq033. [PubMed: 21169341]
16. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theor Comp. 2008; 4:435–447.

17. Kalé L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K. NAMD2: Greater scalability for parallel molecular dynamics. J Comp Phys. 1999; 151:283–312.
18. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimisation, and dynamics calculations. J Comp Chem. 1983; 4:187–217.
19. Scott WRP, Hunenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Kruger P, van Gunsteren WF. The GROMOS biomolecular simulation program package. J Phys Chem A. 1999; 103:3596–3607.
20. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, Debolt S, Ferguson D, Seibel G, Kollman P. Amber, a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules. Comp Phys Comm. 1995; 91:1–41.
21. Bond PJ, Holyoake J, Ivetac A, Khalid S, Sansom MSP. Coarse-grained molecular dynamics simulations of membrane proteins and peptides. J Struct Biol. 2007; 157:593–605. [PubMed: 17116404]
22. DeVane R, Shinoda W, Moore PB, Klein ML. Transferable coarse grain nonbonded interaction model for amino acids. J Chem Theory Comp. 2009; 5:2115–2124.
23. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ. The MARTINI coarse grained force field: extension to proteins. J Chem Theor Comp. 2008; 4:819–834.
24. Anderson DP, Cobb J, Korpela E, Lebofsky M, Werthimer D. SETI@home - An experiment in public-resource computing. Comm ACM. 2002; 45:56–61.
25. Beberg, AL.; Ensign, DL.; Jayachandran, G.; Khaliq, S.; Pande, VS. Folding@home: lessons from eight years of volunteer distributed computing. 2009 IEEE International Symposium on Parallel & Distributed Processing; 2009. p. 1624-1631.
26. Buch I, Harvey MJ, Giorgino T, Anderson DP, De Fabritiis G. High-throughput all-atom molecular dynamics simulations using distributed computing. J Chem Inform Model. 2010; 50:397–403.
27. Pronk, S.; Larsson, P.; Pouya, I.; Bowman, GR.; Haque, IS.; Beauchamp, K.; Hess, B.; Pande, VS.; Kasson, PM.; Lindahl, E. Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis (SC '11); New York: ACM; 2011.
28. Vostrikov VV, Hall BA, Greathouse DV, Koeppe RE, Sansom MSP. Changes in transmembrane helix alignment by arginine residues revealed by solid-state NMR experiments and coarse-grained MD simulations. J Amer Chem Soc. 2010; 132:5803–5811. [PubMed: 20373735]
29. Vostrikov VV, Hall BA, Sansom MSP, Koeppe RE. Accommodation of a central arginine in a transmembrane peptide by changing the placement of anchor residues. J Phys Chem B. 2012; 116:12980–12990. [PubMed: 23030363]
30. Hall BA, Chetwynd A, Sansom MSP. Exploring peptide-membrane interactions with coarse grained MD simulations. Biophys J. 2011; 100:1940–1948. [PubMed: 21504730]
31. Hall BA, Armitage JP, Sansom MSP. Mechanism of bacterial signal transduction revealed by molecular dynamics of Tsr dimers and trimers of dimers in lipid vesicles. PLoS Comp Biol. 2012; 8:e1002685.
32. Lindau M, Hall BA, Chetwynd A, Beckstein O, Sansom MSP. Coarse-grain simulations reveal movement of the synaptobrevin C-terminus in response to piconewton forces. Biophys J. 2012; 103:959–969. [PubMed: 23009845]
33. Crowet JM, Parton DL, Hall BA, Steinhauer S, Brasseur R, Lins L, Sansom MSP. Multi-scale simulation of the simian immunodeficiency virus fusion peptide. J Phys Chem B. 2012; 116:13713–13721. [PubMed: 23094791]
34. Moon CP, Fleming KG. Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. Proc Natl Acad Sci USA. 2011; 108:10174–10177. [PubMed: 21606332]
35. Hessa T, Meindl-Beinker NM, Bernsel A, Kim H, Sato Y, Lerch-Bader M, Nilsson I, White SH, von Heijne G. Molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature. 2007; 450:1026–U2. [PubMed: 18075582]

36. Ulmschneider MB, Doux JPF, Killian JA, Smith JC, Ulmschneider JP. Mechanism and kinetics of peptide partitioning into membranes from all-atom simulations of thermostable peptides. J Amer Chem Soc. 2010; 132:3452–3460. [PubMed: 20163187]
37. Endres NF, Das R, Smith AW, Arkhipov A, Kovacs E, Huang YJ, Pelton JG, Shan YB, Shaw DE, Wemmer DE, Groves JT, et al. Conformational coupling across the plasma membrane in activation of the EGF receptor. Cell. 2013; 152:543–556. [PubMed: 23374349]
38. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984; 81:3684–3690.
39. Winger M, Trzesniak D, Baron R, van Gunsteren WF. On using a too large integration time step in molecular dynamics simulations of coarse-grained molecular models. Phys Chem Chem Phys. 2009; 11:1934–1941. [PubMed: 19280004]
40. Marrink SJ, Periole X, Tieleman DP, de Vries AH. Comment on “On using a too large integration time step in molecular dynamics simulations of coarse-grained molecular models” by M. Winger, D. Trzesniak, R. Baron and W. F. van Gunsteren, Phys. Chem. Chem. Phys., 2009, 11, 1934. Phys Chem Chem Phys. 2010; 12:2254–2256. [PubMed: 20165775]
41. Kalli A, BA H, Campbell ID, Sansom MSP. A helix heterodimer in a lipid bilayer: structure prediction of the structure of an integrin transmembrane domain via multiscale simulations. Structure. 2011; 19:1477–1484. [PubMed: 22000516]
42. Lämmel R. Google's MapReduce programming model - revisited. Sci Comp Prog. 2008; 70:1–30.
43. Lindahl E, Hess B, van der Spoel D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. J Molec Model. 2001; 7:306–317.
44. Marrink SJ, Risselada J, Yefimov S, Tieleman DP, de Vries AH. The MARTINI forcefield: coarse grained model for biomolecular simulations. J Phys Chem B. 2007; 111:7812–7824. [PubMed: 17569554]
45. Psachoulia E, Marshall D, Sansom MSP. Molecular dynamics simulations of the dimerization of transmembrane α-helices. Acc Chem Res. 2010; 43:388–396. [PubMed: 20017540]
46. Wilke M. An iterative jackknife approach for assessing reliability and power of fMRI group analyses. PLoS One. 2012; 7:e35578. [PubMed: 22530053]
47. Confalonieri R, Acutis M, Bellocchi G, Genovese G. Resampling-based software for estimating optimal sample size. Environ Model Software. 2007; 22:1796–1800.
48. White SH, von Heijne G. Transmembrane helices before, during, and after insertion. Curr Opin Struct Biol. 2005; 15:378–386. [PubMed: 16043344]
49. White SH, von Heijne G. How translocons select transmembrane helices. Ann Rev Biophys. 2008; 37:23–42. [PubMed: 18573071]
50. Hessa T, Kim H, Bihlmaier K, Lundin C, Boekel J, Andersson H, Nilsson I, White SH, von Heijne G. Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature. 2005; 433:377–381. [PubMed: 15674282]
51. Doherty T, Su Y, Hong M. High-resolution orientation and depth of insertion of the voltage-sensing S4 helix of a potassium channel in lipid bilayers. J Mol Biol. 2010; 401:642–652. [PubMed: 20600109]
52. Tiriveedhi V, Miller M, Butko P, Li M. Autonomous transmembrane segment S4 of the voltage sensor domain partitions into the lipid membrane. Biochim Biophys Acta. 2012; 1818:1698–1705. [PubMed: 22465069]
53. Chetwynd A, Wee CL, Hall BA, Sansom MSP. The energetics of transmembrane helix insertion into a lipid bilayer. Biophys J. 2010; 99:2534–2540. [PubMed: 20959094]
54. Ulmschneider MB, Smith JC, Ulmschneider JP. Peptide partitioning properties from direct insertion studies. Biophys J. 2010; 98:L60–L62. [PubMed: 20550886]
55. Dorairaj S, Allen TW. Free energies of arginine-lipid interactions: potential of mean force of a transmembrane helix through a membrane. Biophys J. 2006; 90:1022–Pos.
56. Bond PJ, Wee CL, Sansom MSP. Coarse-grained molecular dynamics simulations of the energetics of helix insertion into a lipid bilayer. Biochem. 2008; 47:11321–11331. [PubMed: 18831536]
57. Schow EV, Freites JA, Cheng P, Bernsel A, von Heijne G, White SH, Tobias DJ. Arginine in membranes: the connection between molecular dynamics simulations and translocon-mediated insertion experiments. J Memb Biol. 2011; 239:35–48.

58. Li LB, Vorobyov I, Allen TW. The different interactions of lysine and arginine side chains with lipid membranes. J Phys Chem B. 2013; 117:11906–11920. [PubMed: 24007457]
59. Fry WHD, Kotelawala L, Sweeney C, Carraway KL. Mechanisms of ErbB receptor negative regulation and relevance in cancer. Exp Cell Res. 2009; 315:697–706. [PubMed: 18706412]
60. Normanno N, De Luca A, Bianco C, Strizzi L, Mancino M, Maiello MR, Carotenuto A, De Feo G, Caponigro F, Salomon DS. Epidermal growth factor receptor (EGFR) signaling in cancer. Gene. 2006; 366:2–16. [PubMed: 16377102]
61. Lemmon MA, Treutlein HR, Adams PD, Brunger AT, Engelman DMA. dimerisation motif for transmembrane a helices. Nature Struct Biol. 1994; 1:157–163. [PubMed: 7656033]
62. Mineev KS, Bocharov EV, Pustovalova YE, Bocharova OV, Chupin VV, Arseniev AS. Spatial structure of the transmembrane domain heterodimer of ErbB1 and ErbB2 receptor tyrosine kinases. J Mol Biol. 2010; 400:231–243. [PubMed: 20471394]
63. Prakash A, Janosi L, Doxastakis M. Self-association of models of transmembrane domains of ErbB receptors in a lipid bilayer. Biophys J. 2010; 99:3657–3665. [PubMed: 21112290]
64. Prakash A, Janosi L, Doxastakis M. GxxxG motifs, phenylalanine, and cholesterol guide the self-association of transmembrane domains of ErbB2 receptors. Biophys J. 2011; 101:1949–1958. [PubMed: 22004749]
65. MacKenzie KR, Prestegard JH, Engelman DM. A transmembrane helix dimer: structure and implications. Science. 1997; 276:131–133. [PubMed: 9082985]
66. Finger C, Escher C, Schneider D. The single transmembrane domain of human tyrosine kinases encode self interactions. Science Signal. 2009; 2:ra56, 1–8.
67. Khalid S, Bond PJ, Holyoake J, Hawtin RW, Sansom MSP. DNA and lipid bilayers: self assembly and insertion. J Roy Soc Interface. 2008; 5:S241–S250. [PubMed: 18765335]
68. Hall BA, Sansom MSP. Coarse-grained MD simulations and protein-protein interactions: the cohesin-dockerin system. J Chem Theor Comput. 2009; 5:2465–2471.
69. Kalli AC, Devaney I, Sansom MSP. Interactions of phosphatase and tensin homologue (PTEN) proteins with phosphatidylinositol phosphates: insights from molecular dynamics simulations of PTEN and voltage sensitive phosphatase. Biochem. 2014; 53:1724–1732. [PubMed: 24588644]

Europe PMC Funders Author Manuscripts

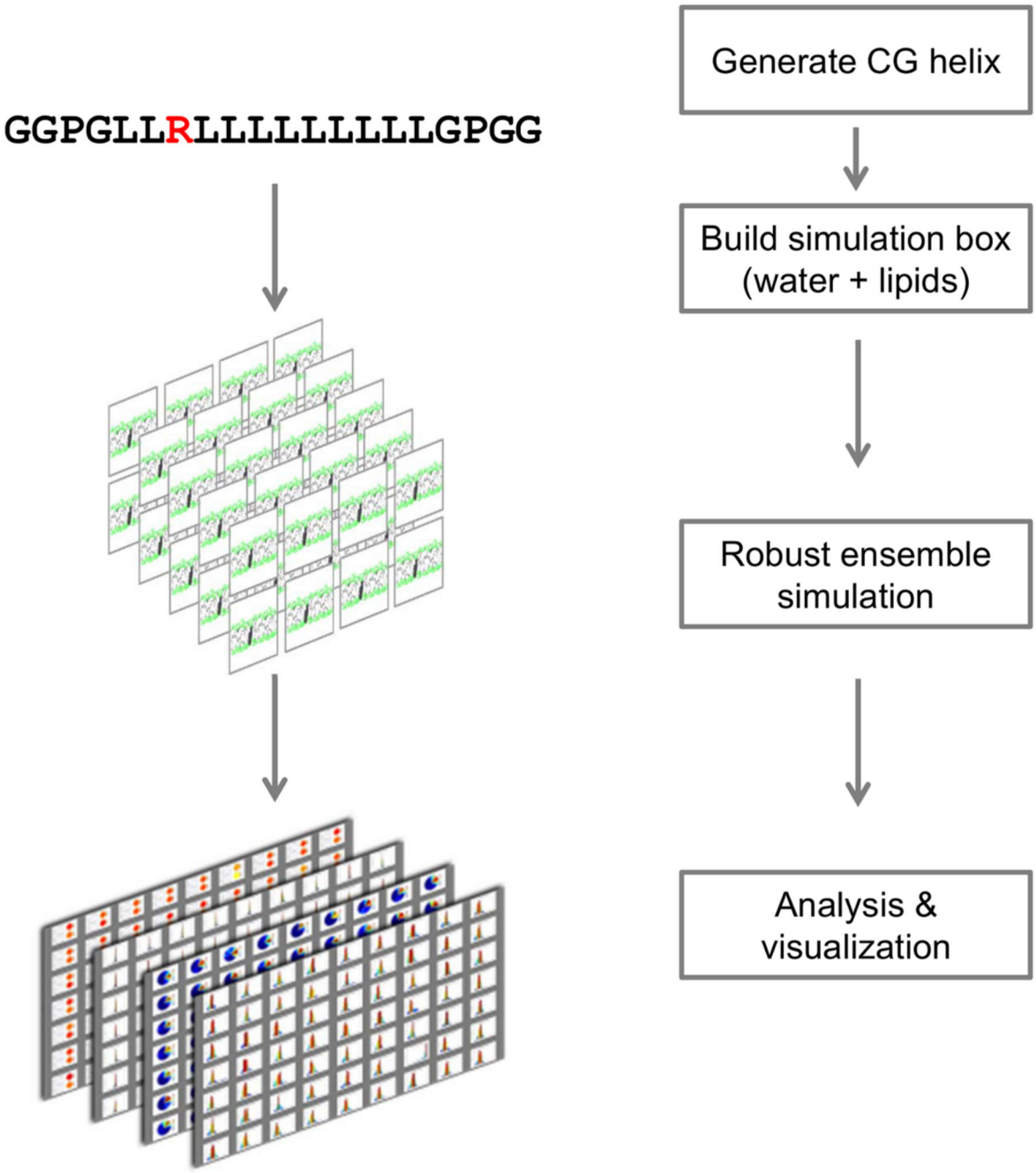

Figure 1.

Schematic of a Sidekick pipeline. A single script runs an ensemble of simulations from a handful of input values, including the amino acid sequence of a helix, the forcefield, and a random seed. From this a CG model of an α-helix is built, it is placed in a box with randomly placed water, lipids and counter ions, and an ensemble of molecular dynamics simulation is performed. The resulting trajectories are visualized and analysed in terms of e.g. the helix depth in the membrane, helix tilt angle, and helix screw rotation angle.

Europe PMC Funders Author Manuscripts

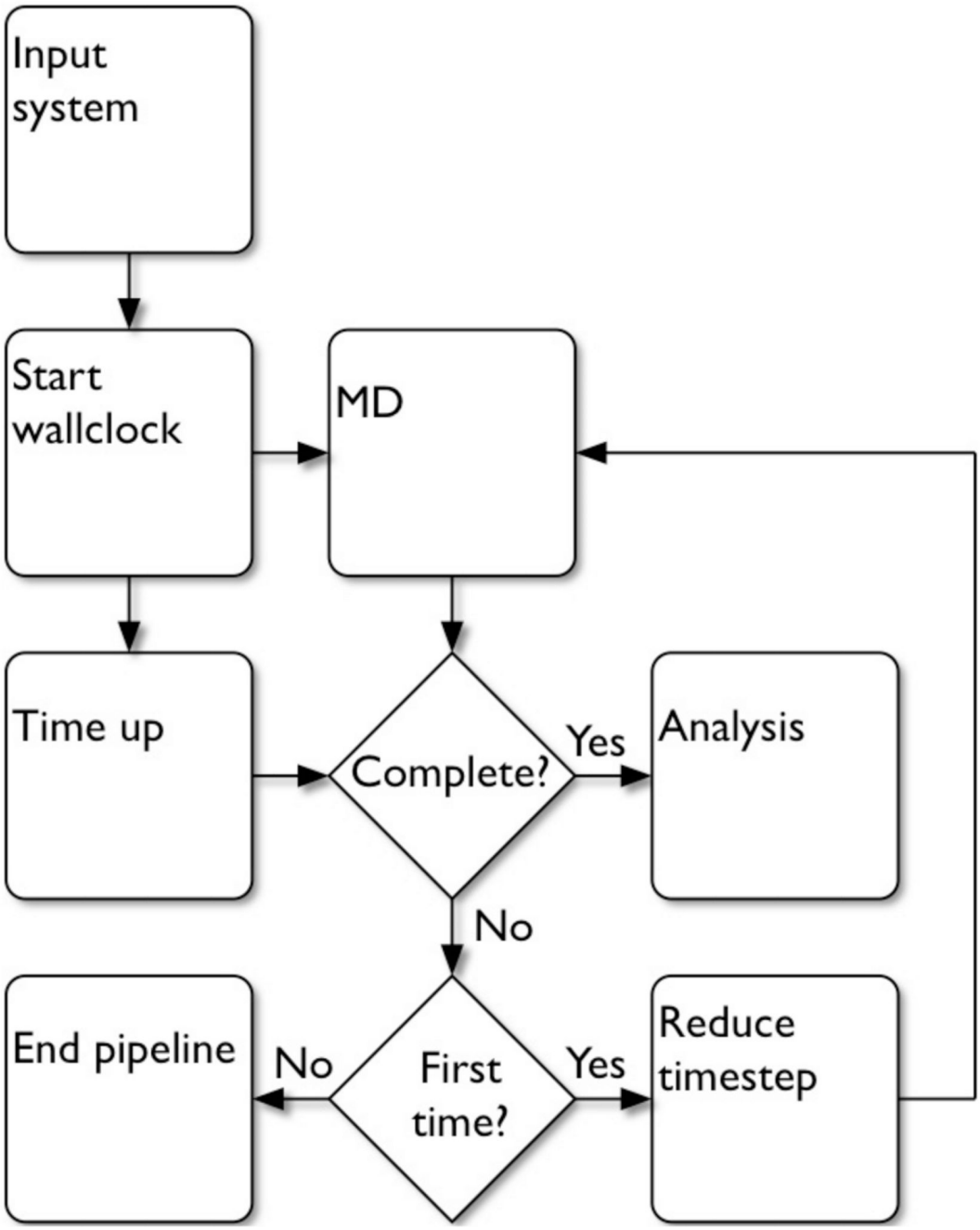

Figure 2.

"Robust" simulation flowchart. Initially a wallclock is started to ensure that individual pipelines do not becomes halted where calculations become trapped in an unproductive infinite loop. Subsequently, CG simulations are started with a timestep of 20 fs. Once the simulation finishes (either due to correct completion or crash), or the wallclock runs out, the resulting files are testing for simulation completion. If no crashes have occurred, the resulting trajectories are analysed. If the simulation has crashed, and if this is the first crash,

the timestep is reduced to 10 fs and the simulation is restarted from the last stored frame with positions and velocities. If the simulation has crashed once before, the pipeline ends.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

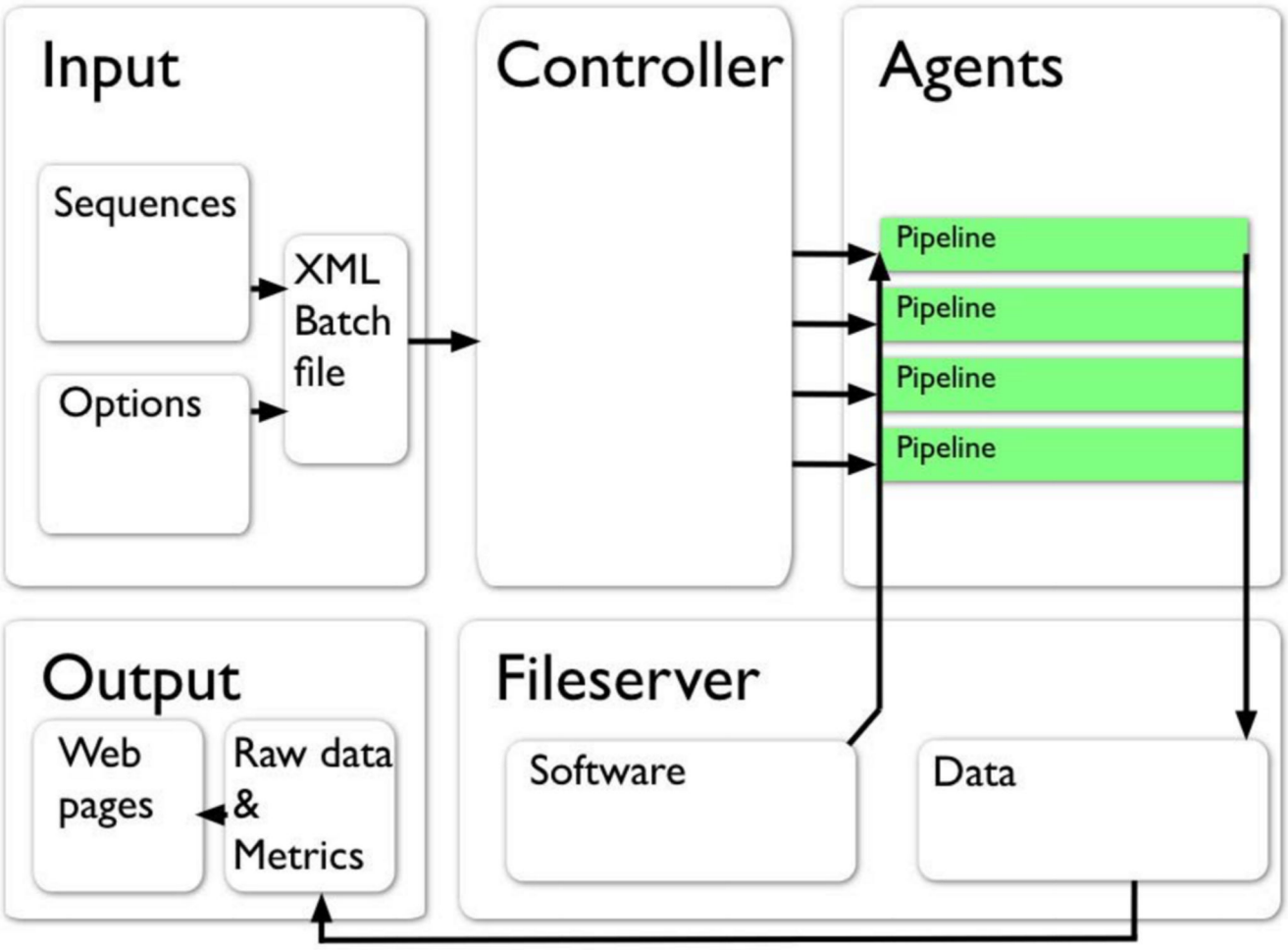

Figure 3.

Schematic of the ensemble simulation workflow (arrows represent the flow and transformation of information at distinct stages of the ensemble). To generate an ensemble of Sidekick jobs a collection of helix amino acid sequences and simulation options (e.g. number of jobs, choice of forcefield, etc.) are used to generate an XML input file for submission to the controller, a machine which coordinates jobs across the grid. From this file the controller starts jobs on all available “agents”, which run individual pipelines, reading required software from a centralised fileserver and depositing output files onto the same fileserver. Files may be written locally during the simulation and copied at the end of the job, or written as the job is running. Individual simulations are analysed within each pipeline to generate metrics to describe the orientation of the helix over the course of the simulation. Both raw data (trajectories and other outputs) and descriptive metrics (e.g. the position and orientation of the helix in the membrane) can be accessed on the fileserver or presented through a web interface.

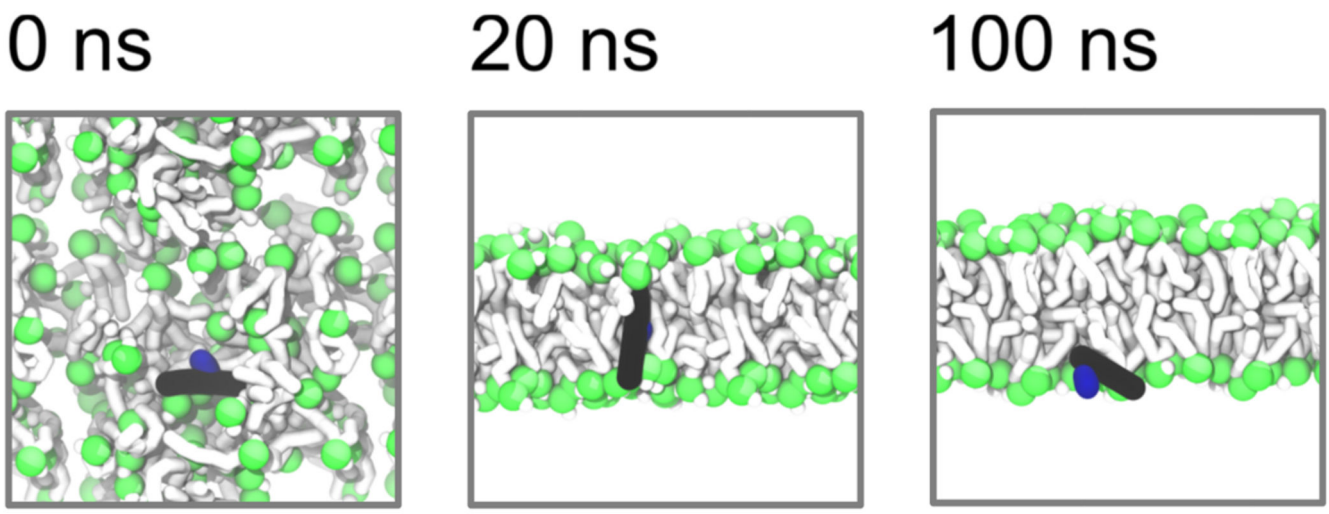

Figure 4.

Dynamics of a single helix insertion simulation illustrated for an L11 helix (see main text and Table 1). Snapshots are shown from a single simulation pipeline at 0, 20 and 100 ns. Lipids are rendered as grey tubes with green spheres for phosphate particles. The helix backbone is rendered as a black tube, and the arginine sidechain rendered in blue. At 20 ns a bilayer has formed, with the peptide in a transmembrane orientation, but between 20 and 100 ns the helix exits the bilayer and remains interfacial for the remainder of the simulation. The entire trajectory is shown in the Supporting Information Movie 1.

Europe PMC Funders Author Manuscripts

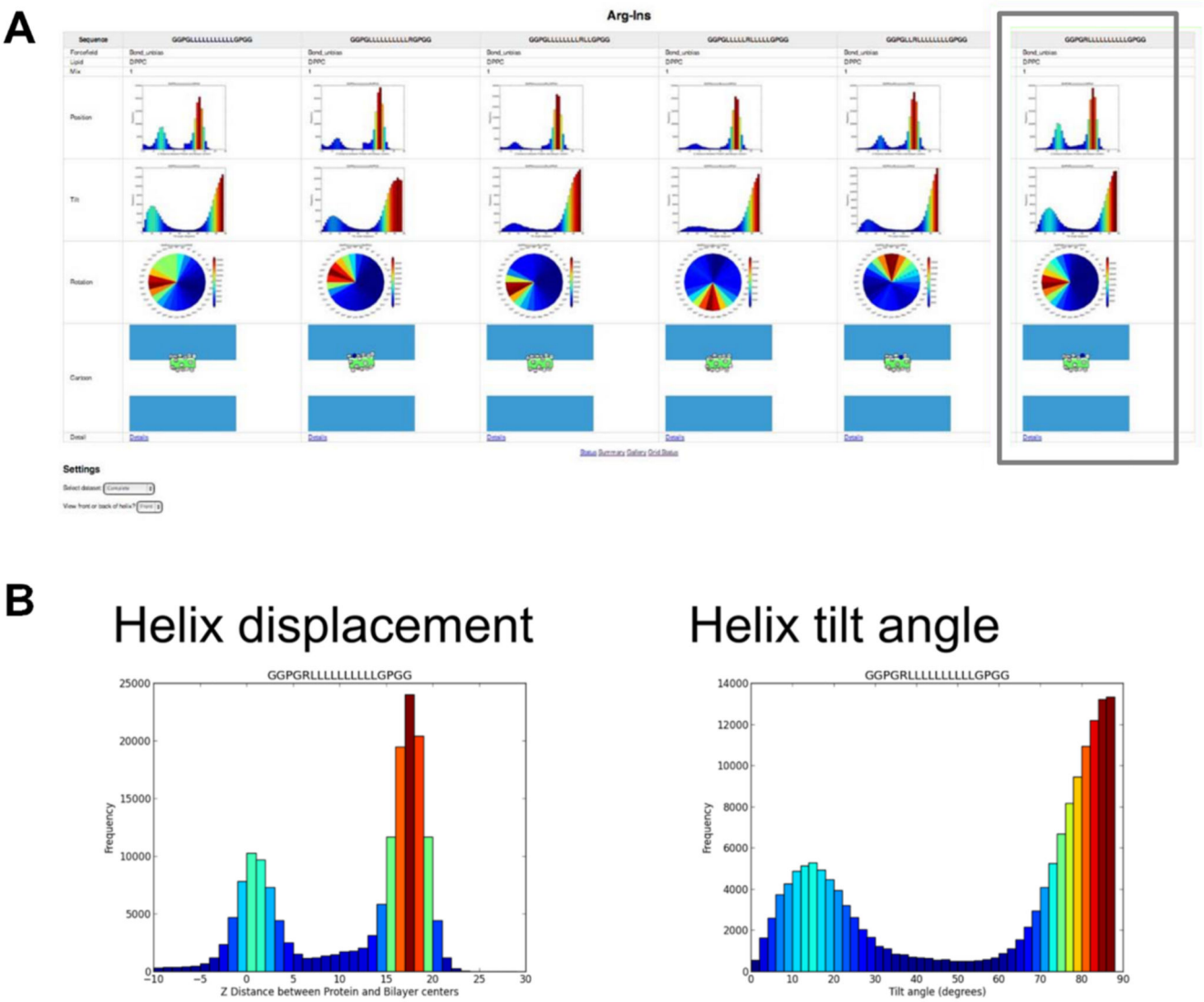

Figure 5.

Analysis of a single helix insertion simulation ensemble. **A** shows part of a screen snapshot from analysis of an ensemble of single helix insertion simulations. **B** shows the helix displacement and helix tilt angle distributions for a single simulation from that ensemble.

Europe PMC Funders Author Manuscripts

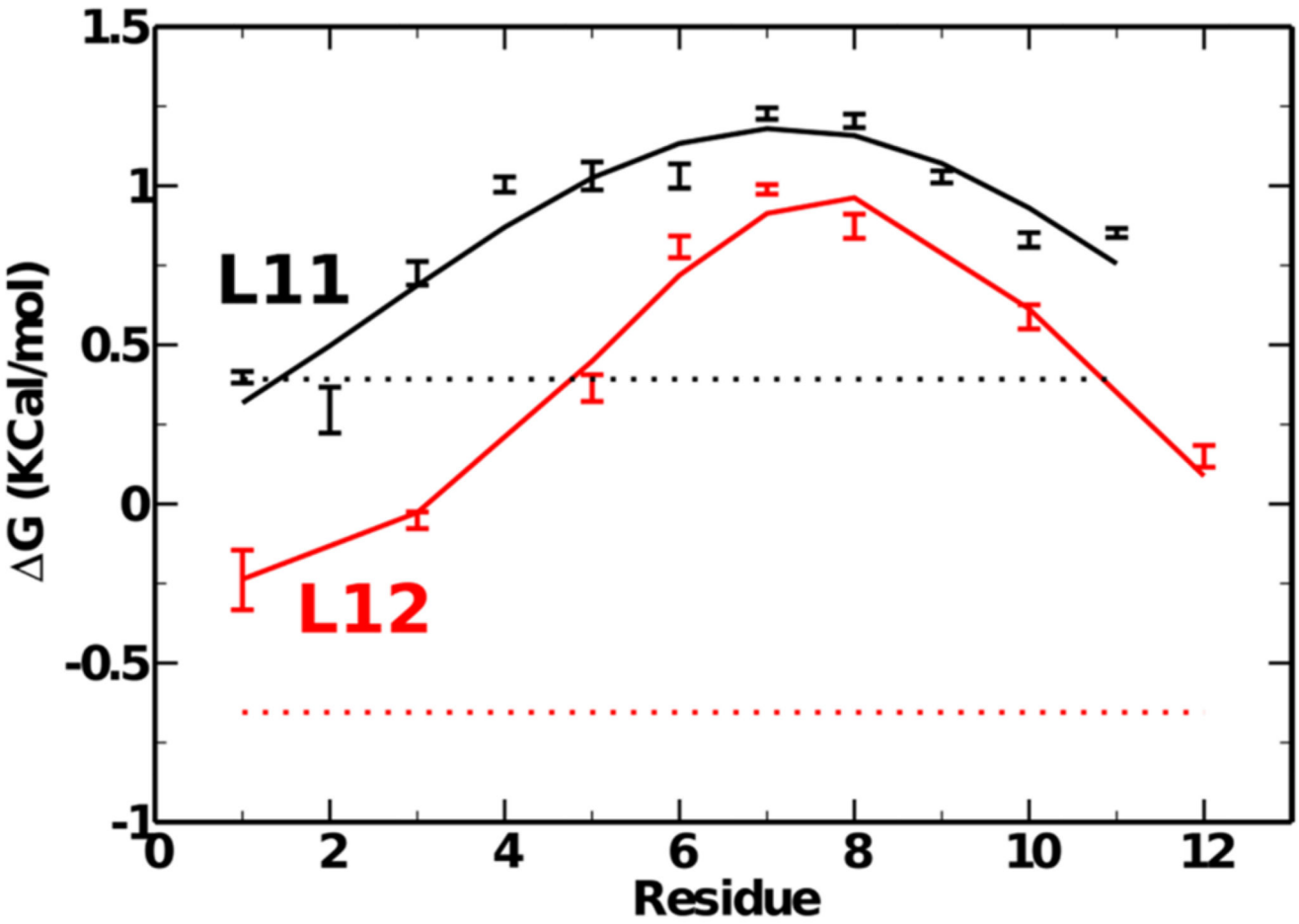

Figure 6.

Insertion free energies of the model hydrophobic helices of the L11 and L12 series (see Table 1) with arginine at different positions along the peptide length (shown on the x-axis). The energies of the purely hydrophobic L11 and L12 ‘parent’ sequences are plotted as broken horizontal black and red lines for comparison. Free energies of insertion were calculated as $\Delta G = -RT \ln K$ where $K = \%(inserted)/\%(not\text{-}inserted)$. Ensembles of ca. 400 simulations were run for each sequence.

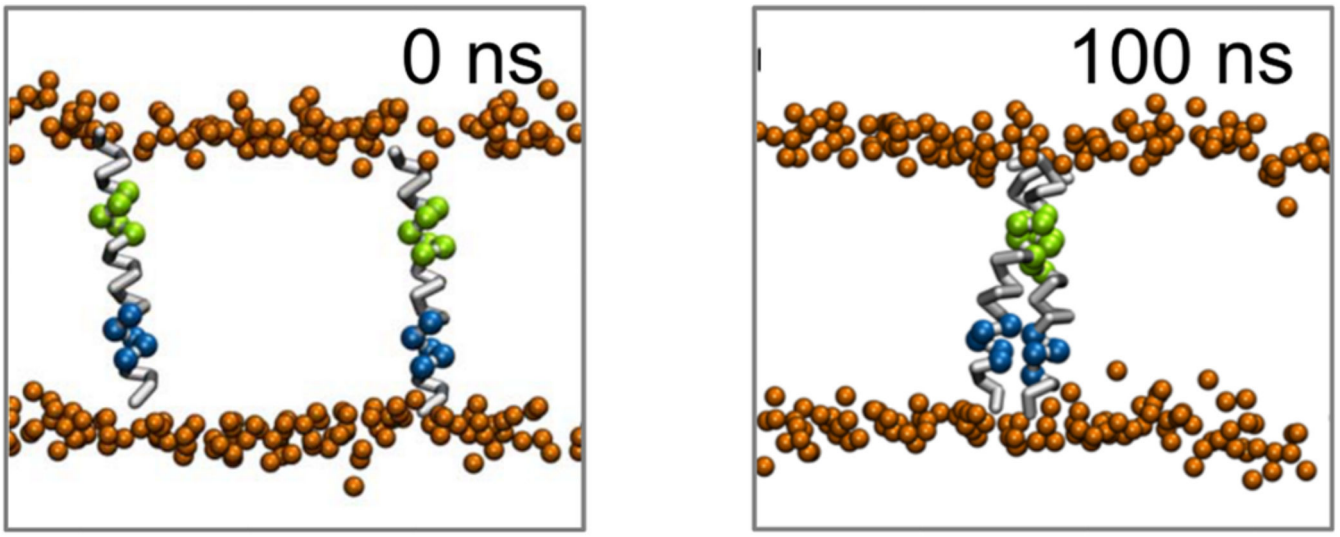

Figure 7.

Simulations of ErbB1 TM helix dimerization. **A** Snapshots of a single simulation at 0 and 500 ns, showing the two helices inserted in a lipid bilayer. Phosphate groups of the phospholipids are shown as orange spheres, TM helix backbones as white chains, with the N- and C- terminal regions rendered in green and blue respectively. The two helices associate spontaneously to form a specific interaction around the N-terminal motif.

Europe PMC Funders Author Manuscripts

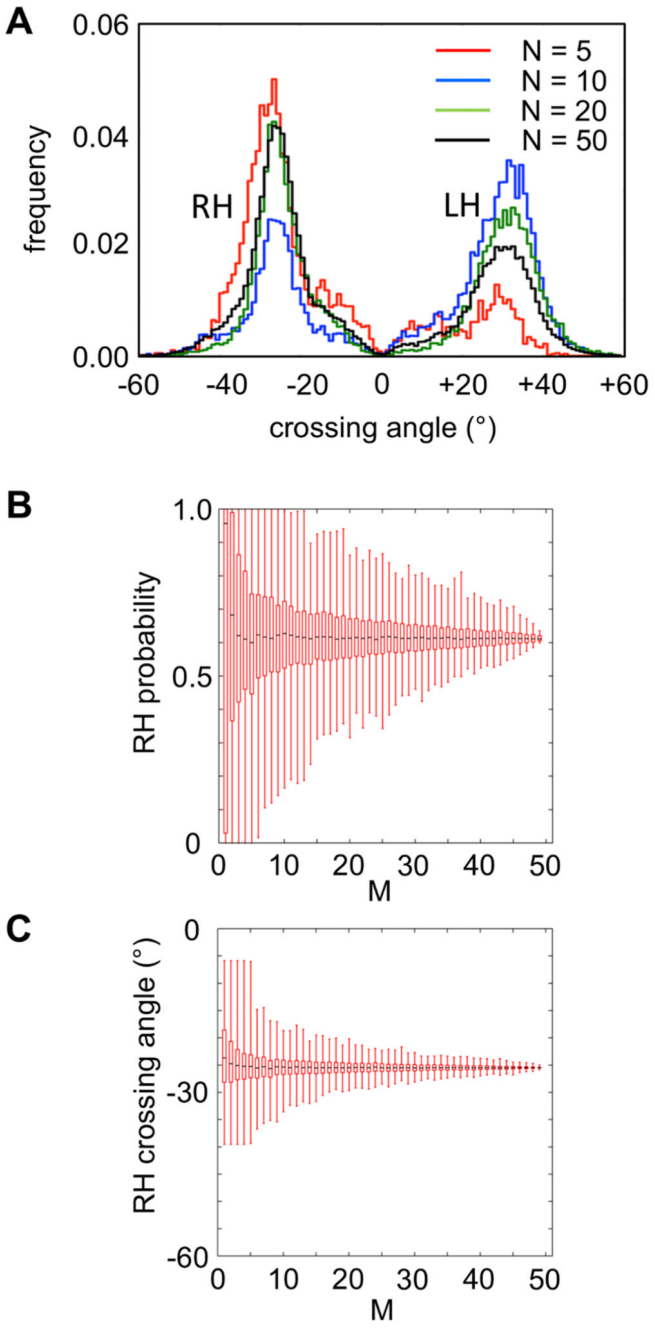

Figure 8.

Convergence analysis of the ErbB1 TM helix dimerization ensemble.

A Helix crossing angle distributions for different ensembles sizes (N = 5, 10, 20, and 50). Both the N = 20 and N = 50 ensembles demonstrate clearly the preference for the right handed (RH) over the left handed (LH) dimer packing.

B and **C** show the results of jackknife analysis (see text for details) for the estimated probability (**B**) and mean crossing angle (**C**) for the RH packing mode as a function of the

subset ensemble size M (where a sample of 1000 ensembles of size M are taken as subsets from the $N = 50$ ensemble).

Europe PMC Funders Author Manuscripts

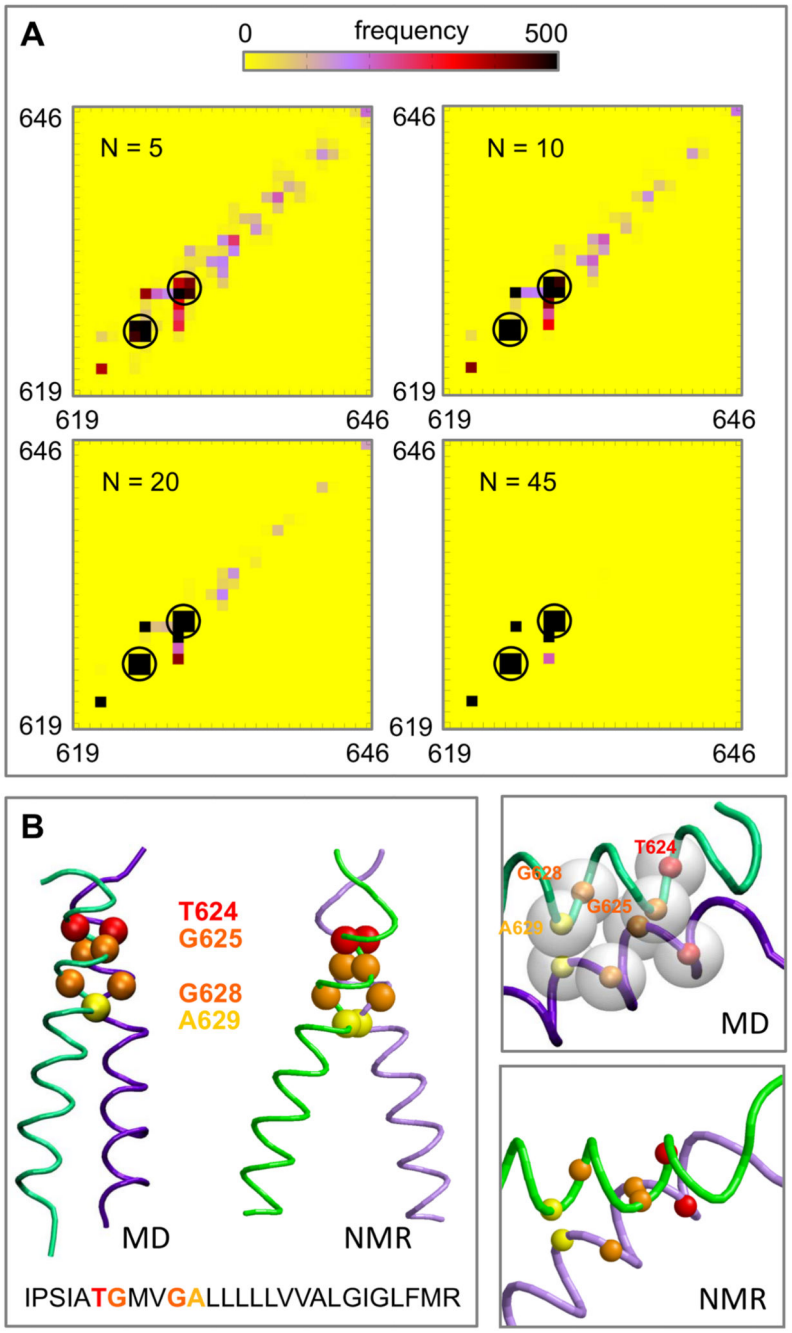

Figure 9.

Analysis of ErbB1 TM helix dimer interactions.

A Convergence analysis of the ErbB1 TM helix-helix contact interactions for the RH packing mode, via jackknife analysis for different ensemble sizes (N = 5, 10, 20, and 45) drawn from the 50 simulations. The helix interactions are shown on a contact matrix where the heatscale indicates the relative frequency of interactions. Circles indicate the interactions observed for the small-small-x-x-small-small motif (T624-G625-x-x-G628-A629) in the NMR structure of the ErbB1 TM-JM helix dimer 37.

B Representative structures of the RH and LH dimer populations with the T624 and G628 contact residues (in red) shown, along with the sequence of the ErbB1 TM region.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 1

Options for Running an Individual Simulation Pipeline

Option	Default	Effect
-t, --type	MARTINI	Specify which forcefield or version of the forcefield to use
-j, --job_name		Specifies the location of the results
-b, --batch-mode	False	Instructs the pipeline to store the resulting files on the fileserver
-s, --sequence		Sequence used to build the helix
-l, --length	100 ns	Simulation length
-u, --unbias	False	Use a cubic box with lipids distributed evenly across the box. Simulations to be run with anisotropic pressure coupling
-r, --random-seed	5	Seed used to generate initial velocities
--randomize_lipids	False	Use the seed to generate different lipid configurations across the ensemble
-a, --angle	0	Initial angle of the peptide relative to z axis
--change_temperature	323	Temperature for the simulation
-p, --position	0	Initial z position of the peptide in the box, relative to the centre of the box
--system_size	“XS”	Preselected box system sizes and numbers of lipids. “XS” contains 128 lipids, “S” 256 lipids
--preformed_bilayer	False	Use a preformed bilayer rather than self assembly
--special		Additional options to pass to the CG scripts
--lipid_headgroup_mix	DPPC	Mixture of lipids to be used in the simulation. Type and ratio are specified
--wallclock	48	Maximum length MD simulations should run

Table 2

Model Hydrophobic Helix Sequences

L11 Series
GGPGLLLLLLLLLLGPGG
GGPGLLLLLLLLL**R**GPGG
GGPGLLLLLLLLL**R**LGPGG
GGPGLLLLLLLL**R**LLGPGG
GGPGLLLLLLL**R**LLLGPGG
GGPGLLLLLL**R**LLLLGPGG
GGPGLLLLL**R**LLLLLGPGG
GGPGLLLL**R**LLLLLLGPGG
GGPGLLL**R**LLLLLLLGPGG
GGPGLL**R**LLLLLLLLGPGG
GGPGL**R**LLLLLLLLLGPGG
GGPG**R**LLLLLLLLLLGPGG
L12 Series
GGPGLLLLLLLLLLLLGPGG
GGPGLLLLLLLLLLL**R**GPGG
GGPGLLLLLLLLLL**R**LLGPGG
GGPGLLLLLLLL**R**LLLLGPGG
GGPGLLLLLLL**R**LLLLLGPGG
GGPGLLLLLL**R**LLLLLLGPGG
GGPGLLLLL**R**LLLLLLLGPGG
GGPGLLLL**R**LLLLLLLLGPGG
GGPGLL**R**LLLLLLLLLGPGG
GGPG**R**LLLLLLLLLLLGPGG

For each series, the ‘parent’ hydrophobic sequence is shown followed by a set of sequences in which the arginine residues is progressively moved along the hydrophobic core sequence.