PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Physical Chemistry Chemical Physics
CREATED USING THE RSC LaTeX PCCP ARTICLE TEMPLATE – SEE www.rsc.org/electronicfiles FOR DETAILS
ARTICLE TYPE www. rsc.org/xxxxxx | XXXXXXXX

Thermodynamic and Kinetic Characterization of Transmembrane Helix Association†

Aiswarya B. Pawar[a], Sneha A. Deshpande[a], Srinivasa M. Gopal[a], Tsjerk A. Wassenaar[b], Chaitanya A. Athale[c], Durba Sengupta[a,*]

Received Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
First published on the web Xth XXXXXXXXXX 20XX
DOI: 10.1039/b000000x

The transient dimerization of transmembrane proteins is an important event in several cellular processes and computational methods are being increasingly used to quantify their underlying energetics. Here, we probe the thermodynamics and kinetics of a simple transmembrane dimer to understand membrane protein association. A multi-step framework has been developed in which the dimerization profiles are calculated from coarse-grain molecular dynamics simulations, followed by meso-scale simulations using parameters calculated from the coarse-grain model. The calculated value of ΔG_{assoc} is approx. -20 kJ/mol and is consistent between three methods. Interestingly, the meso-scale stochastic model reveals low dimer percentages at physiologically-relevant concentrations, despite a favorable ΔG_{assoc}. We identify generic driving forces arising from the protein backbone and lipid bilayer and complementary factors, such as protein density, that govern self-interactions in membranes. Our results provide an important contribution in understanding membrane protein organization and linking molecular, nano-scale computational studies to meso-scale experimental data.

1 Introduction

Membrane proteins play key roles in many cellular processes and understanding their structure and function is crucial for insight into the pathological mechanisms underlying a variety of human diseases[1–3]. Interestingly, several helical membrane proteins require a stable or transient association for their function, and recent evidence points to a ligand-independent association kinetics[4,5]. High resolution microscopy methods have revealed new insights into monomer-dimer equilibrium within membranes[6], but the link between the molecular interactions and the population behavior is still missing. Further, conformational flexibility within the transmembrane dimer state has been shown to be functionally-relevant[7,8], suggesting a complex nature of transmembrane helix association beyond the simple sequence motifs characterized earlier[9]. Importantly, recent quantitative estimates of transmembrane helix association[10] have revealed key differences from the previous estimates in detergent micelles [11–16] and *via* indirect *in vivo* measurements[2,17–19]. Related studies have shown that the lipid bilayer modulates association through sequence independent effects and membrane composition[5,20], and fluidity[21,22] have been suggested to play important roles in helix association. To fully understand membrane-protein association and assembly, one needs to reconcile the roles of kinetics and thermodynamic factors together with the molecular interactions within the membrane[23].

Molecular simulations are being increasingly used to understand the molecular basis of several membrane processes, including membrane protein association[24]. A wide range of models, such as atomistic[8,25–28], coarse-grain[7,29–33] and lower resolution models[34–37] have been used to probe transmembrane helix association. Quantitative analysis of transmembrane helix association by the calculation of the potential of mean force (PMF) along the inter-helical distance has shown similar profiles for several single transmembrane helix dimers[7,25,29–31]. The dimerization energies of alanine and leucine-rich transmembrane peptides represented by the MARTINI force-field have been reported to be in good agreement with previous experimental data[38]. However, favorable ΔG_{assoc} values and the absence of a substantial energetic barrier calculated in several of these studies have been sug-

† Electronic Supplementary Information (ESI) available. For review purposes, see attached file. For the final version see DOI: 10.1039/b000000x/

[a] CSIR-National Chemical Laboratory, Dr. Homi Bhaba Road, Pune-411008
[b] Computational Biology, Department of Biology, University of Erlangen-Nürnberg, Staudtstr. 5, 91058 Erlangen, Germany
[c] Div. of Biology, IISER-Pune, Dr. Homi Bhaba Road, Pune-411008, India
[*] *E-mail: d.sengupta@ncl.res.in*

Keyword: MARTINI force field; transmembrane helix association; lipophobic, coarse grain simulations, potential of mean force

gested to over estimate association kinetics compared to experimentally observed monomer-dimer equilibria[1]. Moreover, recent work suggests that a 1D PMF along the inter-helical distance may result in limited sampling and overestimation of interaction energetics[39–41]. The limited sampling arises from slow membrane dynamics, and biased simulations of membrane partitioning have been recently shown to over- or underestimate the underlying energy surface[42,43].

Several model peptides have been used to study transmembrane helix association, including synthetic peptides with alanine and leucine repeats[44]. In particular, polyalanine peptides allow us to measure sequence-independent effects arising mainly from the protein backbone, beyond the simple sequence motifs characterized earlier[9]. Polyalanine peptides have been used to characterize the thermodynamic parameters related to membrane partitioning[45], helix orientations[46,47], helix-helix interactions [48]. The free energy of association of transmembrane peptides containing polyalanine stretches have been determined to be in the range of -12 to -25 kJ/mol *in vitro*[49,50]. In contrast, in cellular membranes polyalanine peptides are suggested to associate less than naturally-occurring transmembrane dimers such as glycophorin A[51]. In addition, polyalanine is the third-most prevalent homopeptide repeat in eukaryotes[52] and have been linked to human diseases such as congenital malformations and muscular dystrophy[53]. Although, the disease mechanisms are uncertain, it is suggested that the polyalanine tracts lead to aberrant protein-protein interactions[54]. Using polyalanine peptides as a model system, thus provides a method to understand sequence-independent membrane effects as well as their possible role in healthy and diseased conditions.

In this paper, we have developed a framework that can accurately describe the energetics of α-helical dimerization in lipid bilayer environment. We focus on polyalanine peptides to identify those properties that arise from the peptide backbone and are therefore common to all transmembrane proteins. Backbone contributions are important for transmembrane helix association and it has been shown to contribute substantially to the dimerization free energy[31]. However, the larger goal is to provide rules for describing the general behavior of transmembrane proteins in lipid bilayers that would be applicable peptides and proteins of arbitrary amino acid sequence and shape. We have calculated the dimerization free energy from three different methods and show them to be comparable. Further, we have developed a spatial kinetic Monte Carlo model to link the nano-scale coarse-grain results to meso-scale behavior.

2 Methods

2.1 Dimerization Profile Calculations

To calculate the dimerization profile of polyalanine peptides in bilayers, biased and unbiased coarse-grain molecular dynamics simulations were performed. The protein, lipid and water were described by the MARTINI force-field[55,56]. Polyalanine peptides (Ala23) containing 23 successive alanine residues were modeled using standard MARTINI parameters. The conformational space was sampled using three different methodologies: (i) an unbiased simulation that describes *ms* time-scale dynamics (ii) umbrella sampling (iii) multiple μs time-scale simulations.

2.1.1 System Setup

Long Unbiased Simulation: A single long simulation was performed for 320 μs, corresponding to 1.2 *ms* of effective atomistic simulations. Two copies of polyalanine peptides, in the coarse-grained representation, were embedded into a pre-equilibrated DPPC bilayer (protein-lipid ratio = 1:128). The distance between the two peptides was 3 nm.

Multiple short unbiased simulations: 100 independent simulations were carried out for 1 μs each starting with two copies of polyalanine peptides embedded in random orientation and separated by 3 nm in DPPC bilayers. The systems comprised on an average of 94 lipids solvated by 863 coarse-grained water beads, corresponding to a protein-lipid ratio of 1:47. This approach consists of many association simulations performed in parallel, starting from an isotropic distribution of relative orientations. We achieve an optimal starting arrangement and ensure that the mean time to encounter is minimized and is approximately equal for all directions. A triclinic box with a hexagonal tiling has been used. The proteins are placed in such a way that there is maximum distance to their periodic image. This ensures that the time taken by the protein to encounter its periodic image is maximised, with the smallest system size possible. For each simulation, the components are positioned on a grid using a specified distance between the units and giving each unit a random rotation in the plane. Subsequently, each dimer is embedded in a membrane and solvated. The systems thus obtained are subsequently processed using the automated coarse-grained simulation workflow martinate[57].

Biased simulations: A self-assembled dimer conformation extracted from the long unbiased simulation was used as starting structure for calculating PMF using umbrella sampling technique. The system comprises of two polyalanine peptide self as-

sembled in DPPC bilayer of 256 lipids solvated by 4000 coarse-grained water beads, corresponding to a peptide-lipid ratio of 1:128. A harmonic umbrella potential with a force constant of 1000 kJmol^{-1} nm^{-2} applied along the center of mass of peptide backbone. The value of the force constant was optimized to achieve overlap between histograms. For each system, windows were simulated corresponding to a 0.1 nm shift of the monomer per simulation starting from 0.5 nm to 4 nm. Each window was simulated up to 6 μs. The WHAM method[58] was used to unbias the umbrella potentials.

2.1.2 Simulation Protocol

Coarse grain molecular dynamics simulations were performed using the GROMACS software package, version 4.5.5[59], with the scheme developed for the MARTINI model[55,56]. The temperature was coupled (coupling time 0.1 ps) to an external heat bath at temperature T=310 K or 325 K, using a Berendsen thermostat[60]. The time step of the simulation was 20 fs. The pressure was coupled (coupling time 1.0 ps) using a semi-isotropic coupling scheme, in which the lateral and perpendicular pressures were coupled independently to an external pressure of 1 bar using a Berendsen barostat[60]. The non-bonded interactions were treated with a switch function from 0.0 to 1.2 nm for the Coulomb interactions and 0.9 to 1.2 nm for the Lennard Jones interactions (pair-list update frequency of once per 10 steps). The simulation times reported in the manuscript are actual simulation times (which multiplied by a factor of 4 gives the effective times[55]).

2.1.3 Analysis

PMF Calculation From Unbiased Simulations: The potential of mean force (PMF) was used to describe the interaction strength between two structures. The PMF is a free-energy profile generated by integrating the mean force between two distinct states described by a reaction coordinate along the "path" connecting them. It is expressed as the negative logarithm of the population density between two states

$$P_f^i = -kT log H^i / H^{max}$$

where, P_f^i is the dimerization PMF, H^i and H^{max} the relative populations in the given bin and the reference population, k is the Boltzmann constant and T the temperature. The inter-helical distance, calculated from the centre of mass (COM) separation, r, was chosen as the reaction coordinate for the PMF. The population densities were sampled at a bin size of 0.1 nm.

Interaction Energies: The interaction energies for protein-protein, lipid-lipid and water-water were calculated from the sum of the Lennard-Jones and electrostatic contributions, as a function of the inter-helical distance.

2.2 Monte Carlo Simulations

A rule-based kinetic Monte Carlo model of particles[61] was developed to probe the longer time and length-scale dynamics of polyalanine dimerization in membranes. The scheme comprised of three microscopic events: peptide diffusion, dimerization and dissociation. The membrane was represented by a two-dimensional lattice and the peptides as point particles. The square lattice was initially randomly populated with a given density of peptide. Volume exclusion was implemented to ensure a position on the lattice is occupied by a single peptide or dimer. Each peptide was propagated with displacement steps (δr), derived from the expression of the mean squared displacement ($< r^2 >$) of a two-dimensional random walk, by the relation: $\delta r = \sqrt{4 \cdot D \cdot \delta t}$ where, D is the diffusion coefficient of the monomer or dimer and δt the time interval[61]. In the model, the monomers were allowed to associate if the new site was occupied by another monomer, with the probability of association (P_{on}) given by $P_{on} \approx k_{on} \cdot \delta t$, where k_{on} is the on-rate (1/s). Dimer dissociation was based on the probability $P_{off} \approx k_{off} \cdot \delta t$, where k_{off} is the off rate. We ensured δt is small enough to avoid numerical errors

Simulation parameters and configuration: A square lattice of 100 by 100 lattice points spaced 1 nm apart was modeled with periodic boundary conditions. This lattice corresponds to a two-dimensional cell membrane of area of 10,000 nm^2. Simulations were performed at four peptide:lipid ratios (a)1:20 (1000 receptors), (b) 1:50 (400 receptors), (c) 1:100 (200 receptors) and (d) 1:200 (100 receptors). The peptide:lipid ratios were calculated considering an average area per lipid of 1 nm^2, corresponding to 10,000 lipids in each leaflet of the membrane. The kinetic parameters were determined from the coarse-grain models and are listed in Table 1. Simulation time steps of 10 μs were determined by comparing the diffusion of simulated particles with the analytical mean square displacement to minimize deviation from diffusion theory. *In vivo* estimates of transmembrane helix diffusion were used to determine monomer and dimer diffusion coefficients (Table 1).

3 Results

To study the molecular details of the association of transmembrane proteins, self-assembly simulations of polyalanine peptides embedded in DPPC membranes were performed. The two transmembrane peptides were initially placed in a parallel orientation at a distance of 3 nm from each other. During the simulation, the peptides diffused freely and associated within 700 ns. The evolution of the inter-helical distance of the two peptides is shown in (Figure 1a). No dissociation events were observed for 5 μs after association, and the dimer appeared stable within the time-scales. Interestingly, the dimer state did not appear to exhibit large dynamics and a tight association with an inter-helical distance of 0.75 nm was observed. The contact maps for helix-helix association (shown in Supp. Fig. 1) confirm a tight dimer interface, although flexibility was seen in the the inter-helical angles (Supp. Fig. 2).

To analyze the *ms* time-scale behavior of the dimer species, we extended the simulations to 320 μs, corresponding to 1.2 ms of effective atomistic simulations (Figure 1b). After about 10 μs simulation time, the dimer dissociated and the monomers diffused freely in the membrane. The two monomers re-associated within 500 ns. Subsequently, several dissociation events were observed that appeared to be stochastic. The lifetime of the dimer species was variable and dissociation occurred at a time-scale of approx. 50 μs. All dissociated states re-associated within 1 μs. A few conformations with increased inter-helical distances, that corresponded to large fluctuations in the inter-helical angle, and contacts at either the N- or C-terminal were also observed. Such conformations have been suggested to be intermediates in the association pathway, but they did not lead to any dissociation event in our simulation. In total, ten association/dissociation events were sampled within 320 μs simulation time.

3.1 Dimerization Profiles

To predict features of the underlying thermodynamics of association, we calculated the PMF of dimerization as a function of the inter-helical distance. The profiles were calculated by three different methods: the self-assembly simulation describing the *ms* time-scale, umbrella sampling calculations and from a series of shorter μs simulations.

3.1.1 Simulations Describing the Millisecond Time-scale Dynamics The dimerization profile was calculated from the trajectory describing the *ms* time-scale dynamics by binning the inter-helical distances sampled during the simulation. The PMF was then expressed as the negative logarithm of the population density between two states (see methods for details). The free-energy of the two well separated monomers is assumed to be zero. Consistent with the self-assembly simulations and previous experimental studies, the dimerized state is observed to be the most favorable. The PMF of dimerization shows a single energy minimum at 0.75 nm (Figure 2a). A negligible small barrier to association (1 kJ/mol) was observed. The free-energy difference between the monomeric state and the dimerized state, related to the ΔG_{asso} is -19 kJ/mol. To probe the convergence of the dimerization profile, we calculated the profiles as a function of the number of dissociations (see Supp. Fig. 3). It is clear that PMFs calculated from a limited number of association/dissociation events differ from the final converged PMF, although the statistical errors calculated are small.

We would like to point out that the PMF calculated here is along a single reaction coordinate, namely the inter-helical distance, but the unbiased simulations performed here do not sample just along this coordinate. Previously, slow membrane dynamics has been implicated to contribute to insufficient sampling in biased simulations along a single reaction coordinate[42,43]. However, this is of less concern in our unbiased simulations since the PMF is built by projection of the data onto the 1D reaction coordinate. In particular, the sampling of multiple association/dissociation events suggests a complete representation of the underlying coarse-grain free energy landscape, and the main features of the PMF of association can be considered converged.

3.1.2 Umbrella Sampling Calculations Umbrella sampling calculations were performed to calculate the PMF using the inter-helical distance as the reaction coordinate. The final PMF calculated is shown in (Figure 2b). The minimum of the association profile is again located at 0.75 nm, with a corresponding free-energy of -19 kJ/mol. To test the convergence, we calculated the profiles at different time intervals (Supp. Fig. 4). The global shape of the profile remains the same for the different sampling times. Increasing the sampling in each window leads to a convergence in the PMF profile. Beyond 2.0 nm, the changes in the free energy levels off.

Although 1D PMFs calculated from biased methods such as umbrella sampling have been extensively used to probe helix-helix association, recent work suggests that 1D PMFs may result in limited sampling and overestimate the energetics[39,40]. The umbrella sampling calculations reported here compare well with the PMF calculated from the unbiased long simulations, suggesting adequate sampling along the reaction coordinate for a simple peptide such as polyalanine. In this case, limited dynamics in the side-chain packing and rotational motion of the peptide contributes to the complete sampling of the dimerization. For

more complex proteins, it is possible, that sampling all degrees of freedom could take longer than current accessible simulation time-scales.

3.1.3 Multiple microsecond time-scale simulations In the third step, multiple unbiased simulations were performed with two polyalanine transmembrane peptides placed at least 3 nm apart in DPPC membranes. The peptides were distributed randomly such that each starting conformation was unique. As the ensemble of simulations progresses over time, the peptides approach each other and influence their relative orientations, according to the underlying energy landscape of interaction. The dimerized state was observed to be stable in the vast majority of the simulations. PMFs were generated from the ensemble of simulations using the population distributions. The PMFs obtained in the current study were generated from 100 simulations, each of 1μs, with the standard error of the mean determined by the bootstrap method. A single minimum is observed in the PMF at 0.75 nm, with corresponding free energy of -20 kJ/mol, and no barrier to association. The free-energy levels off at distances greater than 2.0 nm. To test the convergence, we calculated the profiles taking into account multiple time-frames from ensemble of simulations (Supp. Fig. 5). The global shape of the profile remains the same with different sampling times and convergence is seen at 900 ns .

The PMF calculated from the multiple short simulations closely matches the PMF calculated from the fully sampled long simulations, albeit with larger error estimates in the monomer regime. Despite the absence of any dissociation events in the multiple short simulations, the evolution of the ensemble of simulations allows a close estimate of the landscape sampled by the coarse-grain simulations describing the *ms* time-scale. Unbiased simulations do not bias the underlying energy surface, as represented by the force-field, but how much of that energy landscape we can reproduce is dependent on the sampling issues. In the case of polyalanine association, the presence of relatively flat surfaces with small barriers allows us to estimate the underlying energetics. In addition, the extensive sampling in the multiple simulations allows us to probe the various pathways of association.

3.2 Interaction energies of protein association

To gain further insight into the driving forces of transmembrane helix association, we decomposed the free energy into protein-protein, protein-lipid and lipid-lipid interaction energies as a function of inter-helical distance. The protein-protein interaction energy decreases as the transmembrane peptides approach each other. The minimum in the free energy at the associated state (0.75 nm) can not be discerned for any of the three sets of simulations. The protein-lipid and lipid-lipid interaction energies can not be directly compared between the simulation methodologies, due to the difference in the number of lipid molecules, although the trends can be compared. The protein-lipid interaction energies increase as the peptides are de-lipidated upon association, for all three protocols. Consequently, the lipid-lipid interaction energies decrease in the same regime. Interestingly, a minimum in the lipid-lipid interaction energy is seen at 0.75 nm, corresponding to the minimum in the free-energy of dimerization. In the single long simulation, the sampling of the protein-lipid and lipid-lipid energies in the monomer regime (beyond 1.5 nm) is seen to be limited and longer sampling in the monomer regime is required.

3.3 Meso-scale Stochastic Simulations

To probe the experimentally-relevant, *ms-s* time-scale dynamics of multiple transmembrane peptides, we developed a spatial kinetic Monte Carlo model. The receptors evolve during the simulations based on a rule-based algorithm discussed in the Methods. The kinetic rate constants, k_{on} and k_{off} were estimated from the long time-scale coarse-grain simulations. Due to the high k_{on} and the absence of a barrier to association, the probability of association at each time step was considered to be 0.9. The value of k_{off} was calculated based on an average dimer lifetime of 50 μs, corresponding to 20,000 s^{-1}. The value was estimated from the long unbiased simulation, since it is the only method that samples binding/unbinding events. The diffusion constants were taken from previous *in vivo* measurements. The simulations were performed at four peptide-lipid ratios: 1:20, 1:50, 1:100 and 1:200. Figure 4 (a) shows the evolution of the percentage of dimers over the simulation time at a peptide-lipid ratio 1:100, and (b) shows the average dimer percentages obtained from the simulations at different protein-lipid ratios (1:20, 1:50, 1:100 1:200). The time evolution of the percentage of dimers during the course of the simulations at varying peptide-lipid ratios is shown in Supp. Fig. 6. The dimer percentage is low at the start of the simulation and reaches a steady-state within a few time steps. At physiologically relevant peptide-lipid ratios (1:100, 1:200), extremely low (≈ 5) dimer percentages are observed. The dimer percentage increases as the peptide-lipid ratio is increased. At the highest peptide-lipid ratios studied (1:20), the dimer percentage is below 20%. Although the on rate of dimer formation directly affects dimer percentages, it is interesting that even at high values of k_{on}, the dimer percentages remain low at physiologically relevant concentrations.

4 Discussion

Transmembrane helix organization is of significant interest due to its central role in cellular signaling and membrane protein folding. Specific sequence motifs have been probed extensively but non-specific effects are only now being recognized as important. Here, we have used long time-scale coarse-grain and meso-scale simulations to probe the association between two polyalanine peptides within the lipid bilayer and extracted the thermodynamic and kinetic factors. A favorable dimerization energy was observed with no barrier to association, and the PMF of association is similar in topology to previously calculated profiles of membrane receptor dimerization[7,25,29–31]. The value of ΔG_{asso} is consistent with *in vitro* estimates of related polyalanine peptides[49,50]. At the *ms-s* time-scale, we observed a low population of dimers under physiological conditions, in agreement with previous *in vivo* results. Although the association rate, k_{on} is fast, the relatively high *off* rates results in low dimer percentages. The results highlight the importance of non-sequence based metrics in modulating helix association within membranes and provide a link between molecular-level simulation studies with population-level experimental data.

The first interesting feature of our work is the role of the lipid bilayer in driving transmembrane helix association in the absence of sequence motifs. Early work on transmembrane helix association focussed extensively on characterizing the specific protein motifs, such as the GxxG motif and the leucine zipper[2,11–15,17–19], but recent studies have questioned such a simplistic view[1,9]. Non-specific contributions that could play an important role are only now being recognized as important and further studies are required to quantify their effect[1]. Although, the bilayer has been previously implicated as an important factor in transmembrane helix dimerization[30,31], our current study provides a direct quantitative estimate of the lipid contributions in the absence of sequence motifs. Further, our calculations reveal favorable backbone-backbone interactions that together with the lipid effects act as a generic driving force of transmembrane helix association. Backbone self-interactions in membranes have not been studied earlier due to the inherent focus on side-chain packing. In addition, since the peptide studied here does not contain flanking residues such as lysine and tryptophan, we are able to probe the non-specific effects of the acyl chains in the absence of the complex contributions of the head group. Our work thus provides a comprehensive understanding of the generic driving forces governing transmembrane helix association.

The thermodynamic dimerization profile was calculated here by three different methods: a long unbiased simulation, umbrella sampling calculations and multiple short unbiased simulations. Unbiased simulations sample the entire conformational space without an external bias (potential/force)[62]. However, the system only samples the thermodynamically accessible energy landscape and large barriers (larger than kT) are not easily overcome. The major limitation of long unbiased simulations is therefore, the adequate sampling of transitions between states and all possible conformations within a state. The statistical significance of a single long trajectory is arguably low, and an exhaustive search of the conformational space is needed that is computationally very expensive. In most cases, due to the time-scale issues the system's properties can only be obtained with meaningful statistics by enhanced sampling approaches, such as umbrella sampling. Enhanced sampling approaches are usually biased and out of equilibrium simulations due to the external force/potential that is applied to improve the sampling along a reaction coordinate[62]. A careful analysis has to be carried out to correctly estimate the underlying unbiased true energy landscape. Often only a single reaction coordinate is used to describe transmembrane dimerization, and has been suggested to over estimate the energies due to limited sampling of the remaining degrees of freedom, such as the inter-helical angle[39–41]. The third approach used is the multiple short unbiased simulations that have been observed to reproduce the dimer states, but their representation of the underlying energy landscape has not yet been examined. The advantages are the absence of a pre-defined reaction coordinate and the sampling of multiple binding pathways and all thermodynamically-accessible peptide degrees of freedom. In addition, performing multiple shorter simulations is computationally more advantageous than generating longer simulations and allows us to sample hundreds of shorter trajectories in a regular high-performance computing cluster. The main limitation is the time-scale issue and in this particular example, the simulations do not sample unbinding events. Care must be taken to ensure the convergence of the ensemble.

The position and the value of the free energy minimum is similar for the profiles calculated by the three different approaches. Interestingly, no additional minima were observed in any of the PMFs. Additional minima have been observed in glycophorin A only upon increased sampling and suggested to arise due to favorable lipid-lipid interactions[31]. The profiles calculated here are well sampled but the lack of additional minima suggests the absence of non-native dimer states with favorable lipid packing in polyalanine. In the first method reported here, we have performed 320 μs of coarse-grain simulations, corresponding to effective 1.2 *ms* of effective atomistic simulations. On a standard HPC cluster, this long simulation requires several months. Additionally, related transmembrane receptors have been suggested to have dimer lifetimes in the range of *ms-s*[6], that is difficult to access with current computational resources. To overcome the computational load and limited statistical sampling in related scenarios, we have tested the framework of multiple unbiased simulations. In this particular case, the simulation cost of the method is half or

third of the other methods. The total simulation time for the multiple short simulation method is 100 μs, compared to 320 μs of the long simulation and 216 μs in the biased simulations. It is encouraging to see the direct correspondence of the three methods, suggesting that umbrella sampling calculations can be used in the future to probe similar processes. Care should however to be taken to test the convergence of the simulations and not simply report the apparent standard error of the mean, which appears to be low in all cases. In particular, sampling around the rotational axes should be tested when sampling is enhanced only along the inter-helical distance. The short unbiased simulations used in our study appears promising and further work is required to assess whether they would help predict the energetics for different transmembrane dimers.

The Monte Carlo simulations reported here indicate that relatively low dimer percentages exist *in vivo* at physiological peptide concentrations, despite a favorable dimerization energy. These results are consistent with the association constants obtained from *in vitro* measurements[49,50] and low dimer percentages seen in *in vivo* studies[51]. Our results thus appear to resolve a seemingly contradictory aspect of transmembrane helix association. Previous Monte Carlo simulations have suggested that large-scale organization of membrane proteins is low unless driven by long-range "lipophobic" interactions[36]. In addition, corral-induced clustering has been implicated to increase dimerization propensity of transmembrane proteins[63]. Mathematical models have further implicated the role of diffusion and encounter frequency in the organization of membrane receptors[64]. Taken together, our results indicate that favorable driving forces are not adequate to corroborate high dimer populations *in vivo*. At physiological protein densities, complementary factors such as surface coverage and encounter frequency are important. The dynamic equilibrium of membrane protein association is thus governed by both thermodynamic and kinetic factors. This interplay between the thermodynamic and kinetic parameters within the context of the membrane is an important aspect of membrane receptor association that remains to be further unraveled.

5 Conclusions

In conclusion, we have performed coarse-grain and meso-scale simulations, that have allowed us to predict features of the underlying landscape of association of polyalanine peptides in lipid bilayers. Transmembrane polyanaine dimers are observed even in the absence of any sequence motif, highlighting the importance of non-specific effects in membrane protein association. Additionally, features of the dynamics are observed to be robust to the sampling protocols used. Short unbiased simulations, although not directly sampling the helix dissociation events, appears to correctly estimate dimerization profiles and closely matches the estimates from umbrella sampling and unbiased coarse-grain simulations describing the *ms* time-scale regime. Spatial Monte Carlo methods developed here allow us to probe the longer time-scale phenomenon for multiple peptides. We observe that at high peptide-lipid ratios, high dimer percentages are observed in accordance with the favorable dimerization free energies. At low peptide-lipid ratios, the overall dimer percentages are low and emphasizing the role of physiological conditions in modulating membrane protein association. The study constitutes an important step towards relating nano-scale computational studies with the meso-scale experimental studies of membrane protein organization.

6 Supporting Information Available

Supporting information is available. Additional graphics: Additional information on the characterization of the dimer state (Supplementary Figure 1, 2) and convergence testing of the PMF calculations (Supplementary Figure 3-5). Additional details of the meso-scale simulations are shown in Supplementary Figure 6. Additional Table: Overview of simulations performed.

7 Acknowledgements

D.S. gratefully acknowledges the support of the Ramalingaswami Fellowship from the Dept. of Biotechnology (D.B.T), and the Fast Track grant from the Dept. of Science and Technology (D.S.T) Govt. of India. The authors thank the Multi-Scale Simulation and Modeling project - MSM (CSC0129) and Center of excellence in polymers (SPIRIT) at CSIR-NCL for computational time. We thank Xavier Prasanna for critical reading of the manuscript.

References

1 A. Rath and C. M. Deber, *Ann. Rev. Biophys.*, 2012, **41**, 135–155.
2 C. Finger, C. Escher and D. Schneider, *Sci Signaling*, 2009, **2**, ra56.
3 J. B. Casaletto and A. I. McClatchey, *Nat. Rev. Cancer*, 2012, **12**, 1387–1400.
4 N. Jura, N. F. Endres, K. Engel, R. Das, S. Deindl, M. H. Lamers, D. E. Wemmer, X. Zhang and J. Kuriyan, *Cell*, 2009, **137**, 1293–1307.
5 C. Matsushita, H. Tamagaki, Y. Miyazawa, S. Aimoto, S. O. Smith and T. Sato, *Proc. Natl. Acad. Sci.*, 2013, **110**, 1646–1651.
6 I. Chung, R. Akita, R. Vandlen, D. Toomre, J. Schlessinger and I. Mellman, *Nature*, 2010, **464**, 783–710.
7 A. J. Brooks, W. Dai, M. L. O'Mara, D. Abankwa, Y. Chhabra, R. A. Pelekanos, O. Gardon, K. A. Tunny, K. M. Blucher, C. J. Morton, M. W. Parker, E. Sierecki, Y. Gambin, G. A. Gomez, K. Alexandrov, I. A. Wilson, M. Doxastakis, A. E. Mark and M. J. Waters, *Science*, 2014, **344**, 6185.
8 A. Arkhipov, Y. Shan, R. Das, N. Endres, M. P. Eastwood, D. E. Wemmer, J. Kuriyan and D. E. Shaw, *Cell*, 2013, **152**, 557–569.
9 E. Li, W. C. Wimley and K. Hristova, *Biophys. Biochim. Acta*, 2012, **1818**, 183–193.
10 H. Hong, T. M. Blois, Z. Cao and J. U. Bowie, *Proc. Natl. Acad. Sci. USA*, 2010, **107**, 19802–19807.
11 L. E. Fisher, D. M. Engelman and J. N. Sturgis, *Biophys. J.*, 2003, **85**, 3097–3105.
12 K. G. Fleming, C.-C. Ren, A. K. Doura, M. E. Eisley, F. J. Kobus and A. M. Stanley, *Biophys. Chem.*, 2004, **108**, 43–49.
13 L. Cristian, J. D. Lear and W. F. DeGrado, *Protein Science*, 2003, **12**, 1732–1740.
14 A. K. Doura and K. G. Fleming, *J. Mol Biol.*, 2004, **343**, 1487–1497.
15 A. Z. Ebie and K. G. Fleming, *J. Mol Biol.*, 2007, **366**, 517–524.
16 A. J. Beevers, A. Nash, M. Salazar-Cancino, D. J. Scott, R. Notman and A. M. Dixon, *Biochem.*, 2012, **51**, 2558–2568.
17 W. Ruan, E. Lindner and D. Langosch, *Prot. Science*, 2009, **13**, 555–559.
18 W. P. Russ and D. M. Engelman, *Proc. Natl. Acad. Sci. U. S.Á.*, 1999, **96**, 863–868.
19 C. Finger, T. Volkmer, A. Prodohl, D. E. Otzen, D. M. Engelman and D. Schneider, *J. Mol. Biol.*, 2006, **358**, 1221–1228.
20 H. Hong and J. U. Bowie, *J. Am. Chem. Soc.*, 2011, **133**, 11389–11398.
21 V. Anbazhagan, C. Munz, L. Tome and D. Schneider, *J. Mol. Biol.*, 2010, **404**, 773–777.
22 V. Anbazhagan and D. Schneider, *Biochim. Biophys. Acta*, 2010, **1798**, 1899–1907.
23 A. Cambi and D. S. Lidke, *ACS Chem. Biol.*, 2012, **7**, 39–149.
24 M. Baaden and S. J. Marrink, *Curr. Opin. Struct. Biol.*, 2013, **23**, 878–886.
25 J. Henin, A. Pohorille and C. Chipot, *J. Am. Chem. Soc.*, 2005, **127**, 8478–8484.
26 J. Zhang and T. Lazaridis, *Biophys. J.*, 2006, **91**, 1710–1723.
27 V. Becker, D. Sengupta, R. Ketteler, G. M. Ullmann, J. C. Smith and U. Klingmueller, *Biochem.*, 2008, **47**, 11771–11782.
28 S. Sharma and A. H. Juffer, *J. Am. Chem. Soc.*, 2013, **135**, 2188–2197.
29 N. Castillo, L. Monticelli, J. Barnoud and D. P. Tieleman, *Chem. Phys. Lip.*, 2013, **169**, 95–105.
30 X. Prasanna, P. J. Praveen and D. Sengupta, *Phys. Chem. Chem. Phys.*, 2013, **15**, 19031–19041.
31 D. Sengupta and S. J. Marrink, *Phys. Chem. Chem. Phys.*, 2010, **12**, 12987–12996.
32 D. Sengupta, A. Rampioni and S. J. Marrink, *Mol. Memb. Biol.*, 2009, **26**, 422–434.
33 A. C. Kalli, B. A. Hall, I. D. Campbell and M. S. P. Sansom, *Structure*, 2011, **19**, 1477–1484.
34 K. Mayawala, D. G. Vlachos and J. S. Edwards, *Biophys. Chem.*, 2006, **121**, 194–208.
35 F. J. M. de Meyer, M. Venturoli and B. Smit, *Biophys. J.*, 2008, **95**, 1851–1865.
36 J.-P. Duneau and J. N. Sturgis, *Eur. Biophys. J.*, 2012, **42**, 843–50.
37 K. Gowrishankar, S. Ghosh, S. Saha, C. Rumamol, S. Mayor and M. Rao, *Cell*, 2012, **149**, 1353–67.
38 M. Nishizawa and K. Nishizawa, *J. Chem. Phys.*, 2014, **141**, 075101.
39 W. Im and S. Park, *J. Chem. Theor. Comp.*, 2013, **9**, 13–17.
40 P. C. Li, N. Miyashita, W. Im, S. Ishido and Y. Sugita, *J. Comp. Chem.*, 2014, **35**, 300–308.
41 S. Park and T. K. W. Im, *Phys. Chem. Lett.*, 2013, **108**, 108102–.
42 C. Neale, C. Madill, S. Rauscher and R. Pomes, *J. Comp. Theor. Chem.*, 2013, **9**, 3686–3703.
43 J. P. M. Jaembeck and A. P. Lyubartsev, *J. Phys. Chem. Let.*, 2013, **4**, 1781–1787.
44 L. V. Schaefer, D. H. de Jong, A. Holt, A. J. Rzepiela, A. H. de Vries, B. Poolman, J. A. Killian and S. J. Marrink, *Proc. Natl. Acad. Sci.*, 2011, **108**, 1343–1348.
45 Y. Yano, N. Shimai and K. Matsuzaki, *J. Phy. Chem.*, 2010, **114**, 1925–1931.
46 B. Bechinger, *Biophys. J.*, 2001, **81**, 2251–2256.
47 D. Sengupta, L. Meinhold, D. Langosch, G. M. Ullmann and J. C. Smith, *Proteins: Struc. Func. Bioinfo.*, 2005, **58**, 913–922.
48 Y. Yano, T. Takemoto, S. Kobayashi, H. Yasui, H. Sakurai, W. Ohashi, M. Niwa, S. Futaki, Y. Sugiura and K. Matsuzaki, *Biochem.*, 2002, **41**, 3073–3080.
49 Y. Yano and K. Matsuzaki, *Biochem.*, 2006, **45**, 3370–3378.
50 Y. Yano, A. Yamamoto, M. Ogura and K. Matsuzaki, *Biochem.*, 2011, **50**, 6806–6814.
51 R. Gurezka, R. Laage, B. Brosig and D. Langosch, *J. Biol. Chem.*, 1999, **274**, 9265–9270.
52 N. G. Faux, S. P. Bottomley, A. M. Lesk, J. A. Irving, J. R. Morrison, M. G. de la Banda and J. C. Whisstock, *Genome Res.*, 2005, **15**, 537–551.
53 L. Y. Brown and S. A. Brown, *Trends Genet.*, 2004, **20**, 51–58.
54 C. Messaed and G. Rouleau, *Neurobiol. Dis.*, 2009, **34**, 397–405.
55 S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman and A. H. de Vries, *J. Phys. Chem. B*, 2007, **11**, 7812–7824.
56 L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman and S. J. Marrink, *J. Comp. Theor. Chem.*, 2008, **4**, 819–834.

57 T. A. Wassenaar, H. I. Ingólfsson, M. Prieß, S. J. Marrink and L. V. Schaefer, *J. Phys. Chem. B*, 2013, **117**, 3516 – 3530.
58 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *J. Comp. Chem.*, 1992, **13**, 1011 – 1021.
59 D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701–1718.
60 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. D. Nola and J. R. Haak, *J. Comp. Phys.*, 1984, **81**, 3684–3690.
61 C. A. Athale, A. Dinarina, F. Nedelec and E. Karsenti, *Phys. Biol.*, 2014, **11**, 016008.
62 W. F. vanGunsteren, X. Daura and A. E. Mark, *Hel. Chem. Acta*, 2002, **85**, 3113–3129.
63 M. N. Costa, K. Radhakrishnan and J. S. Edwards, *J Biotechnol.*, 2011, **151**, 261–270.
64 D. A. Lauffenburger and J. J. Linderman, *Receptors: Models for Binding, Trafficking, and Signalling*, Oxford University Press, 1993.

8 Tables

Table 1 Parameters used for the spatial Monte Carlo simulations

Sr.no	Parameters	Description	Values
1	δt	Time (μs)	10
2	R	Lattice size (nm)	1
3	D_{mono}	Monomer Diffusion Coefficient (μm^2/s)	0.09
4	D_{dimer}	Dimer Diffusion Coefficient (μm^2/s)	0.02
5	P_{on}	Association Probability	0.9
6	k_{on}	Association rate constant (s^{-1})	90000
7	k_{off}	Dissociation rate constant (s^{-1})	20000

9 Figures

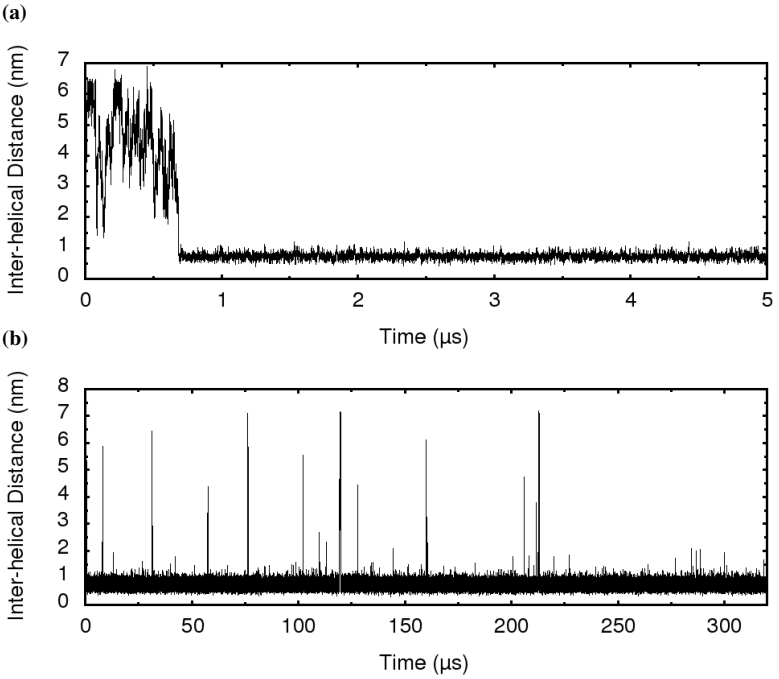

Fig. 1 Association and dissociation events of polyalanine transmembrane helices. The inter-helical distances are plotted during the course of the simulation for a) 0-5 μs and b) 0-320 μs. Helix-Helix distances greater than 1 correspond to the monomer regime and those less than 1 correspond to the dimer regime. The dissociation events are seen as spikes in the plot due to the μs time-scale lifetimes of the monomers. The times reported in the manuscript are simulation times, i.e. when multiplied by a factor of four, approximately accounts for the speed-up of coarse-grained dynamics resulting from the neglect of friction associated with the atomistic degrees of freedom.

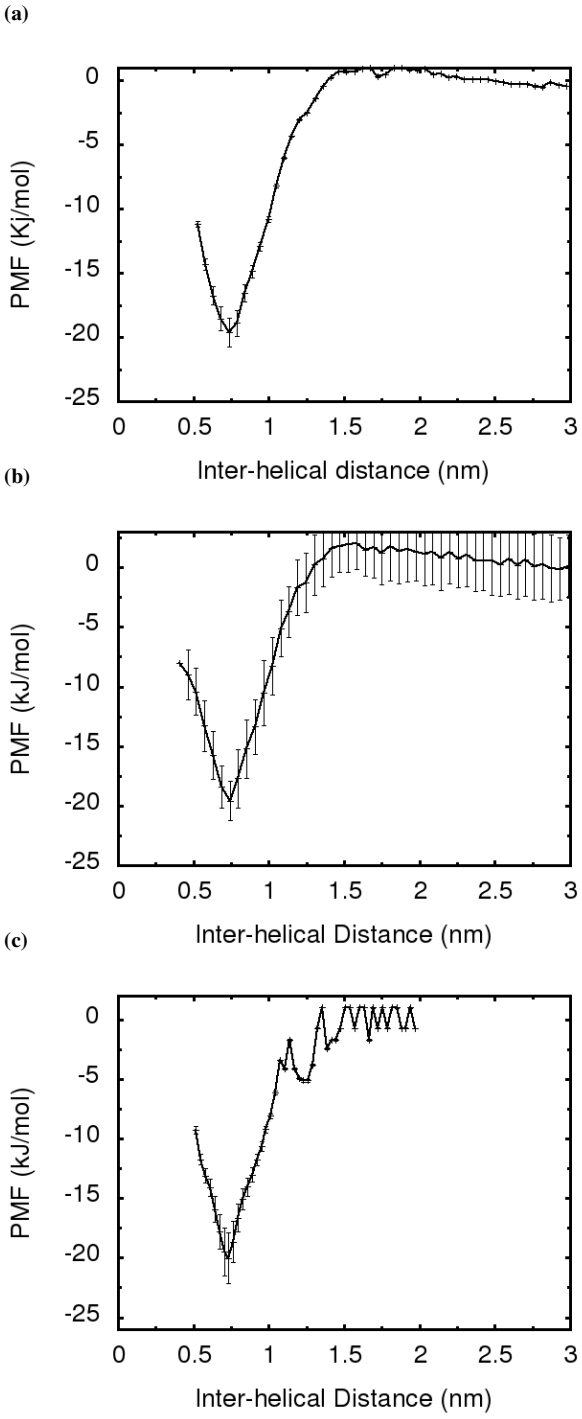

Fig. 2 The potential of mean force (PMF) of dimerization of polyalanine transmembrane helices along the inter-helical distance reaction coordinate. The PMFs have been calculated from (a) long *ms*-timescale unbiased simulations (b) biased umbrella sampling calculations and (c) multiple short μs-timescale unbiased simulations.

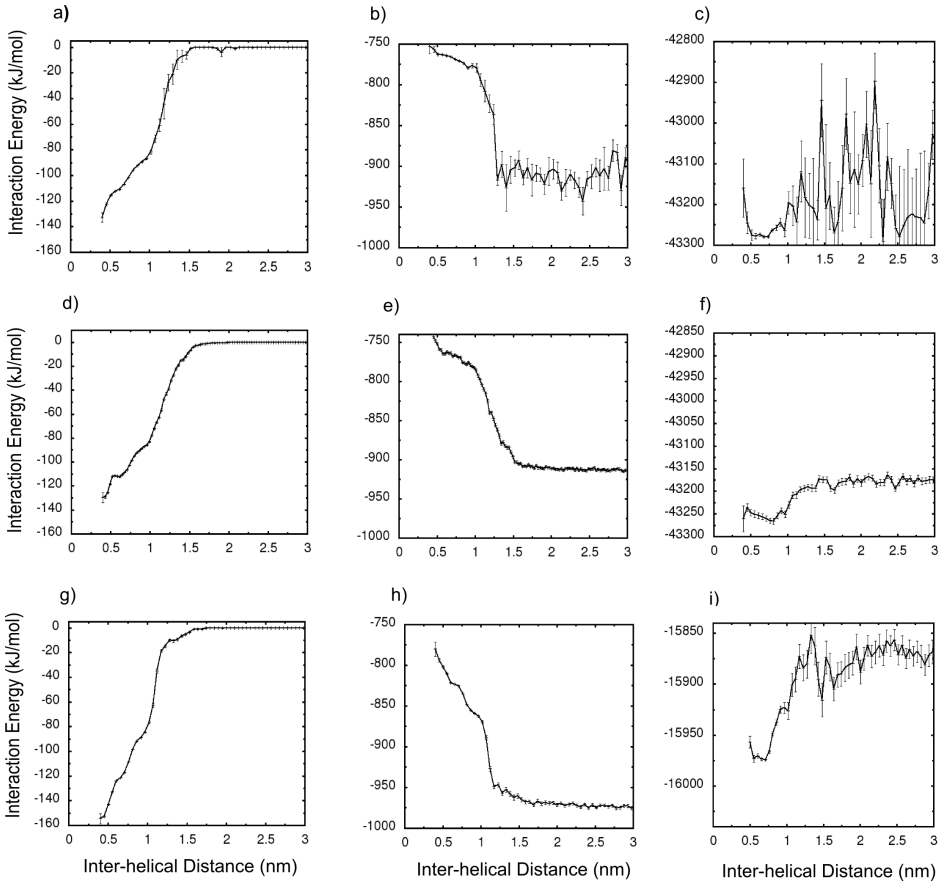

Fig. 3 The protein-protein (a, d, g), protein-lipid (b, e, h) and lipid-lipid (c, f, i) interaction energies as a function of inter-helical distance. The values are calculated from (top panel) long *ms*-timescale unbiased simulations (middle panel) biased umbrella sampling calculations and (bottom panel) multiple short μs-timescale unbiased simulations.

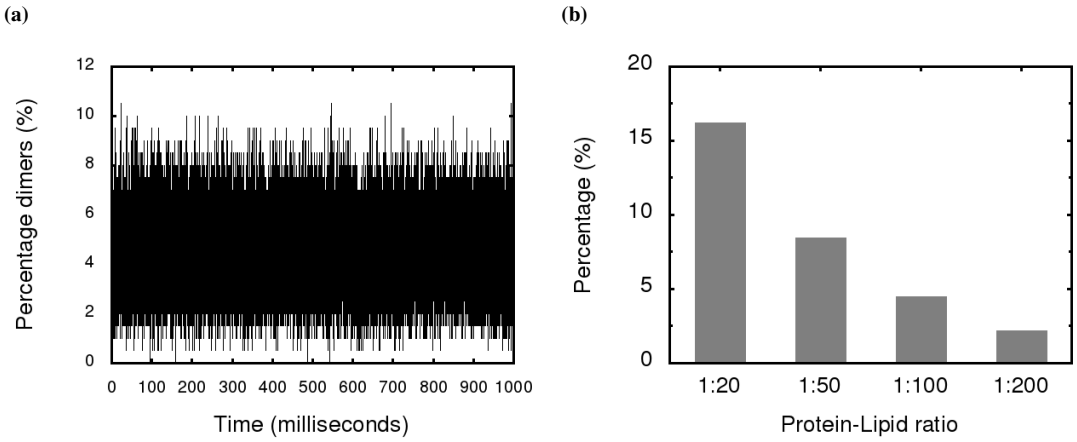

Fig. 4 (a) The time evolution of the percentage of the dimers formed in the meso-scale simulation at a protein-lipid ratio of 1:100. (b) The average dimer percentages obtained from the simulations at different protein-lipid ratios (1:20, 1:50, 1:100, 1:200).

10 Table of Contents Figure