# Energetics of insertion of soluble proteins into membrane

**Abhishek Mathur[1], Anurag Sethi[1], Vishwanath Jogini[1], Yogesh Bhargava[1], B. L. Tembe[2] and Anil K. Lala[1,*]**

[1]Biomembrane Research Laboratory and [2]Molecular Dynamics Laboratory, Department of Chemistry and School of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai 400 076, India

**A number of soluble proteins that insert into membranes show a characteristic hairpin bent helix–helix pair, which is presumed to play an important role in insertion into the lipid membrane. However, little is known about the energetics of membrane insertion of such proteins. It is predicted that the spontaneous insertion of hydrophobic helical proteins into lipid bilayers is mainly driven by a hydrophobic free-energy and interaction of polar and charged side chains of protein with the solvent. With a consideration of hairpin bent α-helix as an 'elliptical frustrum' we have demonstrated that insertion of such bent α-helices into the lipid bilayers is an outcome of various free energies like solvation, lipid perturbation, protein immobilization in lipid and packing energy. Different physical origins and different magnitudes of various free energies satisfy this model for insertion into the membranes.**

CELLS exchange protein chains between their membrane-limited compartments in order to survive. Insertion of protein segments of certain proteins into the membranes is accomplished either by the formation of a helical hairpin structure or by a β-barrel structure. However, very little is known about the energetics of interaction between lipids and proteins. A detailed characterization of the energetics of insertion of α-helices into membranes is an important step to gain an insight into the mechanism of insertion. It would also be useful to predict from either the 1D sequence or 3D structure of a protein whether a given peptide sequence will be in a position to insert into the membrane or not. The measurement of free-energy change associated with a single helix binding to bilayers has been attempted in a number of systems like melittin[1], cytochrome $b_5$ (ref. 2) and polyalanine[3,4]. However a very small number of proteins have been studied for a hairpin bend α-helical pair insertion. It is believed that the α-helical domains of diphtheria toxin and colicin A which have a buried pair of long apolar helices, belong to this category.

The mechanism of insertion of helical hairpins has been predicted to follow the helical hairpin hypothesis[5] according to which the spontaneous insertion of a hydrophobic helical protein into a lipid bilayer is mainly driven by a hydrophobic free-energy of about 60 kcal/mole. The helical hairpin hypothesis has been further extended by Jacobs and White[6], who emphasized the role of lipid–protein interface in the helix formation and insertion into the lipid bilayer. While making the hairpin hypothesis, only hydrophobic forces and interactions of polar and charged side chains with the solvent, were taken into account. We consider here the physical origins as well as the magnitudes of the various free-energy contributions to the transfer of hairpin bend segments of some of the proteins like Diphtheria toxin (DT), *Bacillus thuringiensis* Cry1Aa, Cry3A (whose solution structures are known) from aqueous phase to the nonpolar, lipid environment.

It has been predicted that these 'hydrophobic hairpins' become unshielded through conformational rearrangements triggered by environmental conditions and/or intermolecular contacts, allowing the hairpins to insert into their respective target bilayers[7]. This can be seen as a two-step process (Figure 1). In the first step, the helical hairpin undergoes a conformational change from the stable native state in water ($N_w$) to an open state ($I_i$) on change in local environment, e.g. lower pH found in endosomes after receptor-mediated endocytosis. This conformation is a metastable state as it resides on a local minimum in the free-energy diagram. Hence, $I_i$ can insert into the membrane into its thermodynamically stable transmembrane state ($N_l$), which can subsequently oligomerize.

The free-energy change associated with the transfer of a protein or protein fragment from an aqueous solution into a lipid membrane involves several contributions of different origin. The standard transfer free-energy, $\Delta G^0$, is often written as the sum of a solvation free-energy, $\Delta G^0_{solv}$, a lipid perturbation free-energy, $\Delta G^0_{lip}$, a term which results from immobilization of the protein in the membrane environment, $\Delta G^0_{imm}$, and the packing energy term, $\Delta G^0_{packing}$, which takes into account the free-energy change that takes place due to change in the interactions between the two helices and due to the conformational changes[4–6,8]. The magnitude of the contribution of the various terms to $\Delta G^0$ depends on the detailed molecular structure of the protein and the lipid membrane. However, very little is known about membrane–protein interactions and hence a detailed study of transmembrane proteins has not been

*For correspondence. (e-mail: anillala@iitb.ac.in)

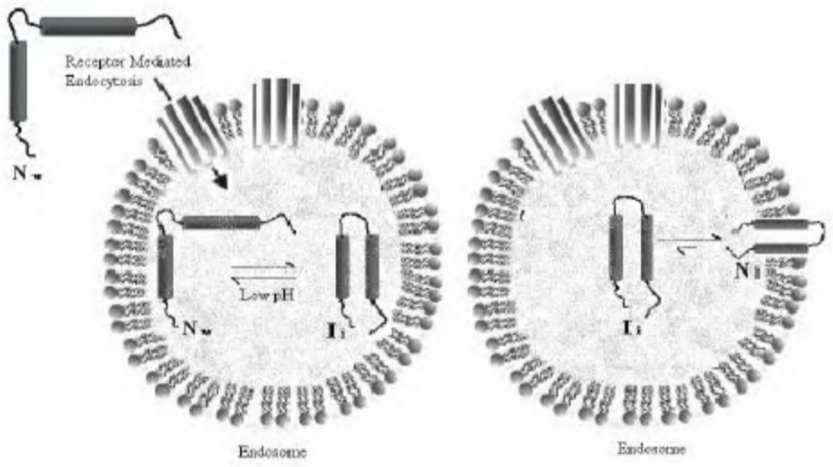


**Figure 1.** Two-step process of helical hairpin insertion in lipid bilayer.

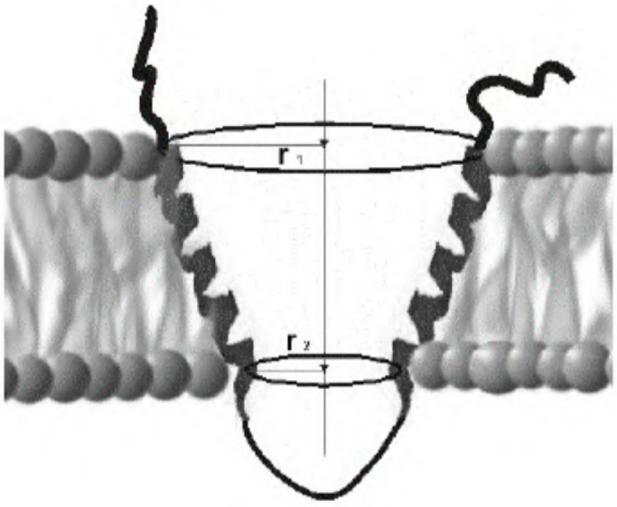

**Figure 2.** Lipid bilayer inserted hairpin bent protein is similar to 'Elliptical-frustum'.

possible. Nevertheless, some general insights into the mechanisms governing protein insertion into membranes can be gained by considering simple model systems. In this paper, we present a model for the energetics of insertion of the helical hairpin. Our treatment demonstrates how each of the free-energy terms contribute to the feasibility of protein insertion process, and helps to predict the portions of the protein that tend to be present in membrane as transmembrane helices in the form of a hairpin bend. In agreement with previous work, the hydrophobic effect is found to provide the major driving force for helix insertion[5]. However, an opposing effect of comparable magnitude is also identified and attributed to the large free-energy penalty associated with the relocation of peptide hydrogen bonds on transfer to the low dielectric environment of the bilayer. Lipid perturbation effects as well as the entropy loss associated with helix immobilization in the bilayer are also evaluated by a method previously published[9]. However, the most influential contribution is expected to come from the packing energy term, which forms the major factor in determining the feasibility of helix insertion.

We have calculated the net free-energy change for each of our bi-helical systems and on this basis we have predicted which of the systems can insert into the membrane in the form of hairpin bend.

## Model

We have put forward a model to simulate the insertion of a helix–turn–helix motif (hairpin bend) protein into membranes. We have assumed the membrane to be a dielectric slab of total area $A$ and a hydrophobic core of thickness $d_L$. The membrane is embedded in an aqueous solution of total volume $V$. Let us assume that the hairpin bend protein has a shape similar to 'elliptical frustum', with two different radii at each face and the length $d_P$ (Figure 2). We are assuming each helix of the system to be a uniform cylinder of radius $r = 5$ Å and length $d$. The angle between the two helices in the lipid environment was obtained as discussed in methodology.

We then proceed to replace this elliptical frustrum by an equivalent cylinder (Figure 3) as follows. The length $d_P$ of the polypeptide is taken to be equal to the average length of the two helices. The radius $r_P$ of the equivalent cylinder is based on surface area as well as the volume calculations. The basis of this assumption is that, free-energy is a surface and volume property and not a geometrical property. So, we replace the elliptical frustrum by a cylinder of equal volume or surface volume and whose length is the average of the two individual helices. The values of $d_P$ and $r_P$ based on surface area and vol-

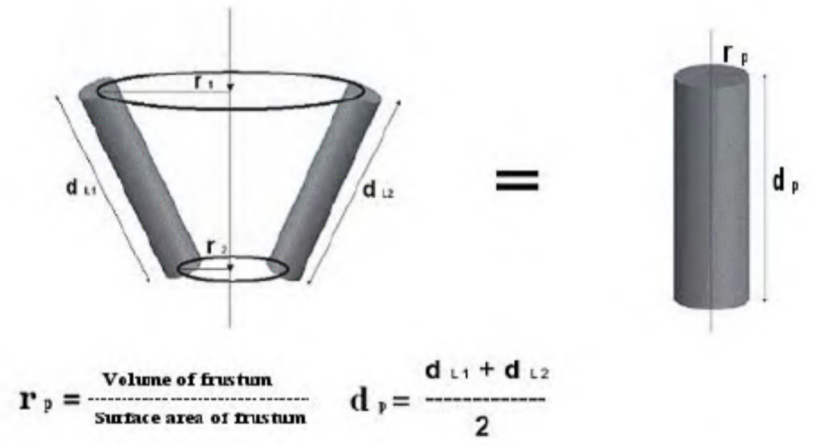


**Figure 3.** Replacement of elliptical frustum by an equivalent cylinder.

**Table 1.** Measurement parameters for different systems

| System | Sequence | Vol. rad, $r_P$ (Å) | Equivalent cylinder length, $d_P$ (Å) | Width of membrane slab, $d_L$ (Å) | Angle between two helices (Å) |
|---|---|---|---|---|---|
| DT TH8 TH9 | 326–347, 358–376 | 6.87 | 31.23 | 30 | 15 |
| DT TH1 TH2 | 206–222, 224–232 | 6.97 | 18.64 | 30 | 34* |
| Bcl-2 H5 H6 | 143–165, 168–191 | 7.18 | 34.54 | 30 | 23 |
| Bcl-xl H5 H6 | 137–156, 160–170 | 7.00 | 27.98 | 30 | 24* |
| Cry1Aa | 124–148, 154–178 | 7.14 | 33.07 | 30 | 23* |
| Cry3A | 160–182, 187–215 | 6.94 | 38.71 | 30 | 14* |
| Colicin A H1 H2 | 8–30, 32–46 | 6.89 | 28.17 | 30 | 17.5* |
| Colicin A H2 H3 | 32–46, 56–67 | 7.05 | 19.15 | 30 | 38* |
| Colicin A H3 H4 | 56–67, 76–87 | 7.16 | 15.96 | 30 | 55* |
| Colicin A H4 H5 | 76–87, 91–101 | 7.72 | 10.85 | 30 | 102* |
| Colicin A H5 H6 | 91–101, 110–127 | 7.89 | 15.91 | 30 | 86* |
| Colicin A H6 H7 | 110–127, 131–142 | 7.06 | 21.57 | 30 | 33* |
| Colicin A H7 H8 | 131–142, 147–164 | 6.85 | 22.12 | 30 | 21* |
| Colicin A H8 H9 | 147–164, 169–185 | 6.90 | 25.85 | 30 | 20 |
| Colicin A H9 H10 | 169–185, 190–199 | 6.92 | 19.60 | 30 | 29* |
| Haemoglobin H15 H16 | 99–117, 123–143 | 6.93 | 29.63 | 30 | 18* |

ume, for our test protein systems are listed in Table 1. Since values of 'surface radii' and 'volume radii' are almost equal (see Table 1), we are reporting the free-energy values for calculations based on 'volume radii'.

The protein will be considered as 'incorporated', if its center of mass is within the hydrophobic core of the membrane. Furthermore, we assume that in the minimum free-energy configuration of the helical hairpin in the membrane, the axis of the equivalent cylinder is perpendicular to the membrane plane, with its center of mass located at the mid-plane between the two lipid leaflets. This model is similar to the model proposed earlier[9], but with the equivalent radius and length of a helical hairpin segment. In the current model we have equated our helical hairpin to a solid cylinder. Both the $\Delta G^0_{\text{lip}}$ and $\Delta G^0_{\text{imm}}$ for the incorporation of a solid cylinder into the membrane have been calculated as explained in the methodology section. The equilibrium distribution of proteins between the membrane and the solution is determined by the equality of the protein chemical potentials in the two phases, which in turn depend upon the net free-energy change that takes place when the polypeptide gets transferred from the aqueous phase to the non-polar phase.

The loop joining the two helices either crosses to the other side of the membrane or it remains in the exterior medium, depending upon the position of hairpin bend. In both the configurations, it remains in the aqueous phase. So, in considering the free-energy changes of the hairpin bend polypeptide, we do not consider any net free-energy change associated with the loop.

## Methodology

First the elliptical frustrum was converted to an equivalent cylinder. For this purpose, we obtain the angle between the two helices by first minimizing the structure of the bi-helical system with the constraints that the backbone remains as a helical hairpin, using 1000 conjugate gradient iterations and the CVFF force-field, in *n*-octanol solvent, using the DISCOVER module of INSIGHTII software.

As explained in references 5, 6 and 8, the total free-energy difference for insertion into the membrane ($\Delta G^0$) can be decomposed into the terms connoting solvation free-energy ($\Delta G^0_{\text{solv}}$), lipid perturbation effects ($\Delta G^0_{\text{lip}}$) and peptide immobilization effects ($\Delta G^0_{\text{imm}}$), along with an additional contribution from energy changes due to peptide conformational changes. In the case of hairpin bend system with two helices, it refers to the energy due to the difference in the interactions between the two helices in the two media. We will call this 'conformational and interaction energy' term as the 'packing energy' term. It can be denoted by $\Delta G^0_{\text{packing}}$. Thus, we have

$$\Delta G^0 = \Delta G^0_{\text{solv}} + \Delta G^0_{\text{lip}} + \Delta G^0_{\text{imm}} + \Delta G^0_{\text{packing}}. \quad (1)$$

Solvation free-energy $\Delta G^0_{\text{solv}}$ is by itself composed of the non-polar interactions ($\Delta G^0_{\text{np}}$) and electrostatic effects ($\Delta G^0_{\text{ele}}$), that is,

$$\Delta G^0_{\text{solv}} = \Delta G^0_{\text{np}} + \Delta G^0_{\text{ele}}. \quad (2)$$

$\Delta G^0_{\text{solv}}$ is the free-energy of transfer of a solute from water to a bulk hydrocarbon phase. It accounts for electrostatic contributions resulting from changes in the solvent dielectric constant and partitioning the hydrogen bonds from the aqueous to the membrane phase as well as for the van der Waal's interactions and solvent structure effects, which are grouped in the nonpolar term and together define the classical hydrophobic effect. This term was estimated[5] for the hairpin insertion, to be –60 kcal/mole. In principle, lipid perturbation term should also be included in this because, lipid perturbation effects have much in common with the hydrophobic effect in that both involve disruption of the solvent structure owing to the presence of solute[4]. Nevertheless, we have preferred to define $\Delta G^0_{\text{solv}}$ as a separate term because it is calculated from water to *n*-octanol phase transfer data, whereas $\Delta G^0_{\text{lip}}$ is estimated separately.

Calculations of the $\Delta G^0_{\text{solv}}$ were based on the transfer free energies of the amino acid residues from water to octanol. Such values of transfer energies were obtained from the data regarding the partitioning of N-acetyl-amino acid-amides between water and octanol[10]. These transfer energies take into account the change in dielectric medium, change in hydrophobic and electrostatic interactions with the solvent, as well as the extent of intermolecular hydrogen bonds with the solvent molecules. These values contain the contributions from the nonpolar as well as electrostatic interactions, in them.

A direct estimate of $\Delta G^0_{\text{lip}}$ can be obtained based on the molecular-level model of lipid–protein interaction[11]. In this approach, the changes in the chain conformation and headgroup interaction free energies are calculated as a function of $d_{\text{P}}$ and $d_{\text{L}}$. Explicit numerical results were reported only for a cylindrical protein of cross-sectional $r_{\text{P}} \to \infty$, that is, $r_{\text{P}} \gg r_{\text{L}}$, where $r_{\text{L}}$ is the average distance between lipid head groups. In principle, the same kind of calculation can be carried out for any $r_P$ value. The numerical results obtained for the large protein case suggest that, to a good approximation the lipid–protein interaction free-energy can be expressed in the form

$$\Delta G_{\text{lip}} = \Delta G_{\text{lip},0} + 0.5 N_{\text{a}} \kappa (d_{\text{P}} - d_{\text{L}})^2. \quad (3)$$

The first term here represents the lipid perturbation free-energy for the case of zero hydrophobic mismatch and the value of this term is got by using the formula $\Delta G_{\text{lip},0} = 0.22^* L$. In this expression, '$L$' denotes the circumference of the cylinder[11]. The second term accounts for the additional free-energy cost in cases of finite mismatch. The concept of the hydrophobic mismatch plays an important role in various models of lipid–protein interactions in membranes. A representation similar to the above equation, was first suggested in the 'mattress model'[12] of lipid–protein interaction. However, their calculation of the restoring force $\kappa$ is based on a different model than the one adopted in the present paper. The numerical values corresponding to the lipid perturbation energy in the case of a flat protein wall ($r_{\text{P}} \gg r_{\text{L}}$) and a lipid bilayer in which the average cross-sectional area per chain (at the octanol–water interface is $\sim 34$ Å$^2$ (implying $r_{\text{L}} \approx 3.3$ Å) are $\Delta G_{\text{lip},0 \approx 0.22^* L}$ kcal/mole Å, and $N_{\text{a}}\kappa \approx L \times 0.09$ kcal/mole Å$^3$, where $L$ is the length of the protein (cross-sectional) circumference. For a smaller, curved inclusion such as helices of radii comparable to ours, the perturbation free-energy should be considerably smaller. As a crude curvature correction, we multiply the value of $\Delta G_{\text{lip},0}$ by the factor $(1 - (r_{\text{L}}/r_{\text{P}}))$. Using a similar curvature correction, we obtain the values of $N_{\text{a}}\kappa$. For the case of helices of unequal length in a pair, we calculate the additional term due to hydrophobic mismatch, by taking the average of modulus of the mismatch for the two helices.

In bulk solution the hairpin has three translational degrees of freedom of its center of mass and three rotational degrees of freedom. Upon insertion into the membrane, the helical hairpin retains two free translations and two free rotations, the two translations in the *xy* plane and the rotations $\psi$ (around helix axis) and $\phi$ (azimuthal angle between the helix axis projection in the *xy* plane and *x* axis), specifically $0 \le \psi \le 2\pi$ and $0 \le \phi \le 2\pi$. The confinement of the translational and rotational motion results in an entropy loss for the hairpin in the membrane compared with

its free motion in solution. This effect is responsible for the immobilization free-energy contribution. A detailed theoretical model for estimating this contribution is given by Ben-Shaul[9].

The free-energy change due to immobilization of the protein on insertion is calculated considering the isotropy of the lipid bilayer in $xy$ plane. The value for reduced partition function $q_{imm}$, was calculated[9] for a single helix to be $(2\pi)^{1/2}(kT/\kappa d_L^2)$. The relation between $q_{imm}$ and the immobilization free-energy change is given by the equation:

$$\Delta G = -RT\ln q_{imm}. \quad (4)$$

We are using our model of 'elliptical frustum' as an equivalent single cylindrical helix with the parameters enlisted in Table 1. Apart from the aforementioned contributions to the free-energy of insertion of a pair of helices, there is one more contribution, which comes from the conformational change of the hairpin on insertion. This energy emerges from the difference in interactions between the two helices in the two different media. It involves the contribution from several structural terms like the changes in the side-chain torsion angles as the peptide gets transferred to the nonpolar medium.

In the hydrophobic environment of the lipids, we would expect the hydrogen bonds of the polar peptide backbone to be satisfied, as their insertion into the membrane would be highly unfavourable if hydrogen bonds were not re-established after bilayer entry. Thus, it is important for the peptide segment to retain its bihelical structure on insertion[13]. Hence, we do not expect the individual helices to undergo any change in their secondary structure, in the process of transfer from aqueous domain to the nonpolar domain. So, the packing energy effectively includes only the changes, which take place to the extent of interactions between the helices of the pair.

This energy was acquired by calculating the packing energy, i.e. the nonbond interaction energy between two helices was calculated for two media and difference was taken to be equal to $\Delta G_{pack}$ using the CVFF forcefield. Coordinates and structure files or the aqueous medium were taken from Brookhaven Protein Data Bank as *.pdb* files. The *pdb* files used were 1MDT for *Corynebacterium diphtheriae* Diphtheria toxin[14], 1GJH for Anti apoptotic protein Bcl-2 (ref. 15), 1CIY for *Bacillus thuringiensis* pesticidal toxin Cry1Aa (ref. 16), 1DLC for *Bacillus thuringiensis* insecticidal toxin Cry3A (ref. 17), 1COL for *E. coli* Colicin A (ref. 18), 1MAZ for Anti apoptotic protein Bcl-xl (ref. 19) and 1BBB for Human blood monocarboxy Haemoglobin[20].

Here, we are more concerned with those pairs of helices with length of each helix between 15 and 25 amino acids (22.5 to 37.5 Å) and with the loop joining them to be longer than 2 amino acids (> 3 Å). Helices outside this range are known to have very high values of lipid perturbation free-energy cost, which makes the net free-energy cost to be positive disallowing any spontaneous insertion. We have taken all the possible helical hairpins of Colicin A to show that this result comes naturally from our model.

## Results

### *Diphtheria toxin, DT (TH8 and TH9)*

In DT, the α-helical domain (termed the T or transmembrane domain) is located in the primary structure between the amino-terminal catalytic (C) and carboxyl-terminal receptor-binding (R) domains. After binding to its receptor via the R domain, the toxin undergoes receptor-mediated endocytosis and is conveyed to the endosomal compartment. The acidic conditions in the compartment promote conformational changes in the toxin that induce the T domain to insert into the endosomal membrane, a step required for efficient translocation of the C domain to the cytosol. It has been proposed that the channel formed by the T domain serves as a conduit through which the unfolded C domain traverses the membrane[21], but direct evidence for this is lacking. However several approaches indicate that acidic conditions within the endosomal compartment (~ pH 5.3) trigger a conformational alteration of the T domain that allows the overlying helices to separate from the buried hairpin (helices TH8 and TH9) and permits the hairpin to insert into the endosomal membrane[22–26].

The translocation domain of diphtheria toxin consists of nine helices arranged in three layers. The middle layer consists of the carboxy-terminal helices, TH8 and TH9, both of which are hydrophobic and long enough to span the membrane as shown in Figure 4. It is noteworthy that the loop connecting the helices on one face of the helix

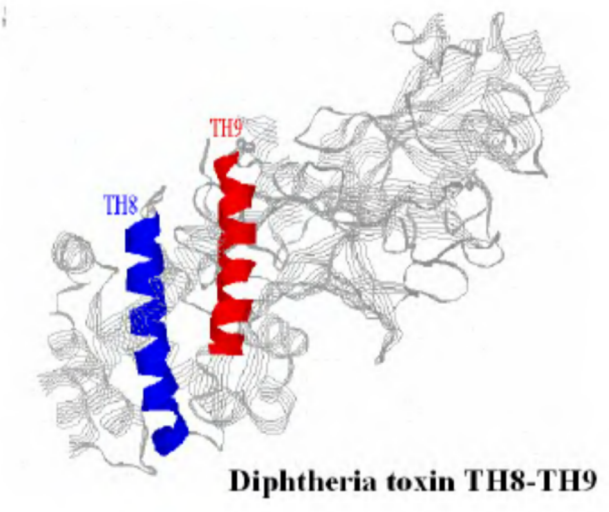


**Figure 4.** Hydrophobic translocation domains of diphtheria toxin (DT).

**Table 2.** Different contributions to total free-energy values (in kcal/mole)

| System | $\Delta G_{solv}$ | $\Delta G_{lip}$ | $\Delta G_{imm}$ | $\Delta G_{pack}$ | $\Delta G_{tot}$ |
|---|---|---|---|---|---|
| DT TH8 TH9 | –38.00 | 6.46 | 6.46 | –42.27 | –67.35 |
| DT TH1 TH2 | –2.63 | 138.91 | 6.57 | 17.28 | +160.13 |
| Bcl-2 H5 H6 | –15.96 | 5.58 | 6.65 | –46.31 | –50.04 |
| Bcl-xl H5 H6 | –18.85 | 9.40 | 6.53 | 0.00 | –2.92 |
| Cry1Aa | –33.89 | 15.57 | 6.73 | –91.16 | –102.75 |
| Cry3A | –30.65 | 83.22 | 6.54 | –151.70 | –92.59 |
| Colicin A H1 H2 | –9.14 | 8.36 | 6.48 | 0.00 | +5.70 |
| Colicin A H2 H3 | –7.23 | 130.30 | 6.65 | 56.50 | +186.22 |
| Colicin A H3 H4 | –7.16 | 220.08 | 6.75 | 47.43 | +267.10 |
| Colicin A H4 H5 | –1.26 | 464.63 | 7.26 | 59.38 | +530.01 |
| Colicin A H5 H6 | –2.90 | 264.01 | 7.39 | 125.97 | +394.47 |
| Colicin A H6 H7 | –17.64 | 80.69 | 6.65 | 146.69 | +216.39 |
| Colicin A H7 H8 | –30.78 | 67.11 | 6.44 | –14.27 | +28.5 |
| Colicin A H8 H9 | –36.08 | 22.52 | 6.50 | 0.00 | –7.06 |
| Colicin A H9 H10 | –22.59 | 115.76 | 6.52 | –88.83 | +10.86 |
| Haemoglobin H15 H16 | –13.07 | 5.15 | 6.51 | 94.23 | +92.82 |


**Figure 5.** Antiparallel arrangement of two central helices (α5 and α6) of Bcl-xL with a crossing angle of 23°.

barrel is very acidic and a large number of charged residues reside on the face opposite to it, thereby creating a large molecular dipole. TH8 and TH9 are known to form ion-selective channels after insertion. The free-energy of burying this system is well negative indicating its great propensity to lie inside the membrane (Table 2). No other helical pair in the translocation domain is of sufficient length to span the membrane. Though TH6 and TH7 are also hydrophobic in nature, their smaller size costs them a high lipid perturbation energy, $\Delta G_{lip}$, which makes the total free-energy cost of insertion, $\Delta G_{tot}$ positive, thereby rendering the spontaneous insertion impossible. Similar is the case for TH1-TH4 helices, which are hydrophilic in nature and are smaller in size.

As evident from the results (Table 2) the hairpin bend system TH8 (328–348) and TH9 (351–371) in diphtheria toxin, is expected to undergo spontaneous insertion into membrane by the formation of hairpin. It has been established in the literature that only this hairpin system in DT forms a transmembrane segment[27,28], in accordance with our results.

### *Bcl-2 and Bcl-xl: An anti-apoptotic protein family*

Bcl-2, first identified as an oncogenic protein involved in human lymphomas, belongs to a family of proteins that regulates apoptosis. Bcl-2 and Bcl-xl both contain a transmembrane sequence at their carboxyl terminus that anchors the proteins in intracellular membranes, primarily the endoplasmic reticulum (ER) membrane, the nuclear envelope, and the mitochondrial outer membrane. The three-dimensional structures of the cytosolic domain of Bcl-2 as well as of Bcl-xl reveal a striking similarity to the pore-forming domain of diphtheria toxin and bacterial colicins. Indeed, conductance measurements have shown that the cytosolic domains of several members of the Bcl-2 family and Bcl-xl have pore-forming activity in synthetic lipid bilayers.

The structure of Bcl-xl consists of two central α-helices (α5 and α6), which are about 30 Å in length, flanked on one side by α3 and α4 and on the other by α1, α2 and α7 (Figure 5). The two central helices contain predominantly hydrophobic residues, and are arranged in an antiparallel fashion, with a crossing angle of 23°. Helix α6 contains a kink at Pro 180 that causes a change in the direction of this helix[19].

Both the proteins have membrane insertion domains containing two central helices consisting of apolar residues, and are long enough to span a membrane. Helices α5 and α6 and three of the surrounding amphipathic heli-

ces (α1, α3 and α4) in the Bcl-xl structure align well (r.m.s.d., 2.0 Å for Cα positions) with corresponding elements of the transmembrane of diphtheria toxin.

We expect these two central hydrophobic helices to get inserted into the membrane, by assuming a hairpin structure. The free-energy cost associated with this has been calculated and found to be very much negative, indicating its tendency to assume transmembrane configuration.

### *B. thuringiensis insecticidal δ-endotoxin family (Cry1Aa and Cry3A)*

The crystal structures of two homologous toxins (Cry3A; Coleopteran-specific and Cry1Aa; Lepidopteran-specific) have been reported[29,30]. They contain three domains: Domain I (N-termini; 220 residues) is the channel-forming domain. It is a seven-helix bundle in which a central helix (helix 5) is completely surrounded by the other six helices. The six helices are amphipathic and long enough to span the 30 Å membrane bilayer. Domain II (central 200 residues) is the specificity-determining domain, which binds the receptor proteins in the insect midgut membrane. Domain III (C-termini; ~150 residues) is functionally less well defined and may transduce information from domain II (receptor binding domain) to domain I (channel-forming domain). Domain III may also regulate channel activity, stabilize the toxin, and function together with Domain II in receptor binding.

The structure of Domain I of insecticidal δ-endotoxin (Cry3A) is shown diagrammatically in Figure 6. In it, the central helix is not entirely hydrophobic as there are a few polar residues, all of which are involved in hydrogen bonds and salt bridges. It is likely that the central helix plays an important role in insertion as its residues are highly conserved throughout the δ-endotoxin family and mutations of nonpolar residues to polar residues in the helix inhibit pore formation[31].

Taking energy considerations, for Cry1Aa, it is expected that the α4 (Pro124-Leu148) and α5 (Gln154–Phe178) helical pair would insert spontaneously into the membrane. Similarly, for Cry3A, α4 (Pro160–Pro182) and α5 (Ser187–Gly215) helices are likely to form the hairpin insertion domain.

### *Human blood haemoglobin*

This is the protein, which resides in the blood of humans. It is an oxygen carrier protein. Most of the amino acids in haemoglobin form alpha helices, connected by short non-helical segments. Haemoglobin has no beta strands and no disulphide bonds. It consists of 16 α-helices. A haemoglobin molecule is composed of a protein group, known as globin, and four heme groups, each associated with an iron atom. Each haemoglobin molecule also has attached a single cysteine amino acid.

Haemoglobin is a water-soluble globular protein, which prefers to stay inside the cytosol and has no tendency to diffuse into the membrane. We have obtained the same results by inferring from the positive net free-energy change, as a pair of its helices (α15 and α16), of length

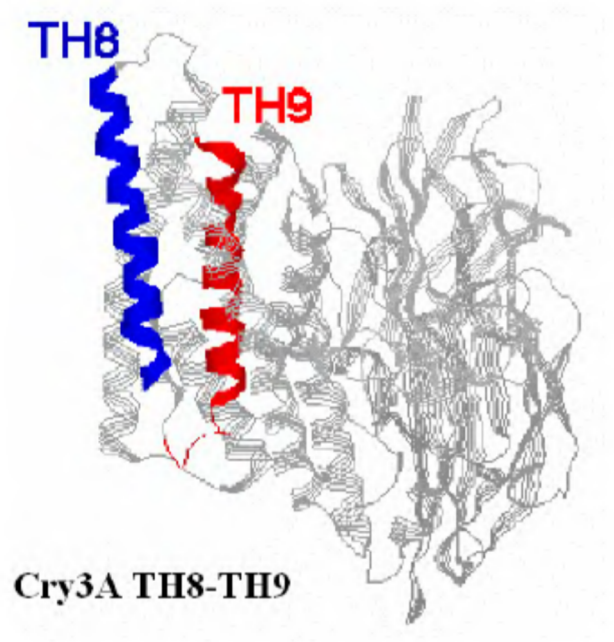


**Figure 6.** Central helices of insecticidal δ-endotoxin of Cry 3A.

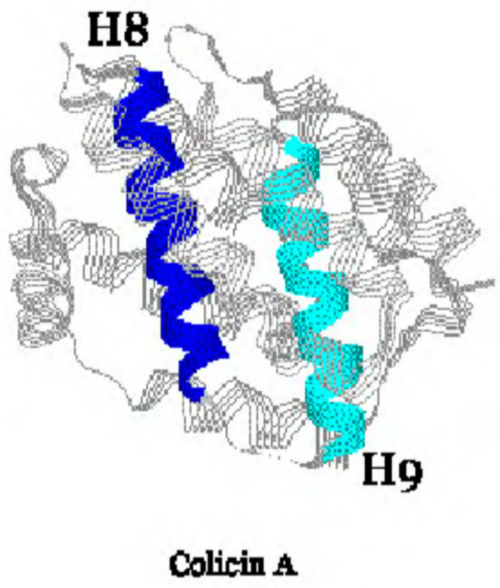


**Figure 7.** Hydrophobic bent α-helical hairpin of colicin A.

just sufficient to span the membrane, gets transferred to the nonpolar environment (octanol) from aqueous medium.

### *E. coli colicins*

The pore-forming domain of the α channel forming toxins (α CFTs) are characterized by three layers of 7–10 α-helices, in which the putative pore former, the hydrophobic helical hairpin is shielded in the middle (Figure 7). Structural changes with acidic pH are believed to expose this hairpin and allow channel formation. Increasing evidence indicates that these toxins do not insert to form a well-ordered oligomeric structure in the membrane.

In the case of Colicin A, previous work indicated a 'molten globule' intermediate for the inserted state. The 'molten globule' state is a partially unfolded but compact state characterized by a marked increase in sidechain flexibility whilst retaining a near-native secondary structure. The increased flexibility afforded by the molten globule state reduces the energy barrier for unmasking the hydrophobic hairpin.

The three-dimensional fold of the pore-forming domain of Colicin consists of ten α-helices with the two outer layers sandwiching a middle layer of three helices (helices 8, 9 and 10), the first two of which are completely buried and consist of hydrophobic residues only. Three of the helices (1, 8 and 9) are sufficiently long to span a membrane. Colicins are known to preferentially insert into negatively charged lipid bilayers at low pH. It is known to get inserted spontaneously following the interaction of hydrophobic loop of the hairpin with the lipid bilayer. The hydrophobic hairpin is known to follow a perpendicular insertion.

Free-energy of insertion has been calculated for each pair of consecutive helices in Colicin A. They have been represented by a hypothetical hairpin bend structure, with the angle between two helices in two different media to be obtained from energy minimization using DISCOVER module in INSIGHTII (see Table 1). In accordance with what we expected, smaller helices have high free-energy cost due to their higher tendency to 'perturb' the lipid bilayer (see $\Delta G_{lip}$ values in Table 3), therefore they are unlikely to enter the membrane. Only single pair of helices, α8 and α9 satisfy the negative free-energy requirement, and thus is expected to get inside in the form of a hairpin (Table 2).

## Discussion

Till now, not much work has been done on the problem of hairpin bend protein insertion into a membrane. Our goal in this paper is to determine the origin of the various contributions to the free-energy of inserting a hairpin bend α-helical peptide sequence into a lipid membrane. We have started with basic thermodynamics and calculated the different free-energy contributions by adopting an 'elliptical frustum' model for hairpin bend protein. Our calculations of 'solvation', 'lipid perturbation' and 'immobilization' free energies are not conceptually different from those given in previous studies of the protein insertion problem. Here we have dealt quite differently from the previous work in the case of 'packing free-energy' term. This term has never been accounted for in the previous studies. This has been calculated by energy minimization techniques using DISCOVER module of INSIGHTII software. Apart from this, many hairpins, which are known to form a hairpin inside the membrane, have been correctly categorized as transmembrane by the free-energy values we calculated. Then we have extended our method to identify the other transmembrane hairpins on a few other proteins like Bcl-xl and Bcl-2. The results are in accordance with the speculated data in literature.

Then we have enlisted those sequences, which are hydrophobic and of sufficient length to span the membrane but which differ in their insertion properties due to differences in their loop size and loop sequence (see Table 3). It has been observed that those hairpin proteins, which get inserted into the membrane, have their loops rich in valine, alanine, proline and leucine. As explained from the insertion properties of proteins like haemoglobin, which do not insert, histidine, lysine and threonine inhibit insertion process. This can be a potential source of information about the prediction of those helices which are hydrophobic and long enough to span a membrane but unable to make it due to the loop sequence.

By applying the above method of calculating the net free-energy change that accompanies the insertion of a hairpin bend, transmembrane helices can be predicted. By

**Table 3.** Helices of size sufficient to span a membrane

| Entity | Size (helix–loop–helix) | Loop sequence | $\Delta G_{tot}$ kcal/mole | Insertion |
|---|---|---|---|---|
| DT TH8 TH9 | 22–10–19 | IPLV-GELVDIGFAA-Y | –54.32 | Yes |
| Bcl-2 H5 H6 | 23–2–24 | VNRE-MS-PLV | –45.68 | Yes |
| Bcl-xl H5 H6 | 20–3–18 | ESVD-KEM-QVL | –3.92 | Yes |
| Cry1Aa | 25–5–25 | IPLL-AVQNY-QYP | –87.45 | Yes |
| Cry3A | 23–4–29 | SMPS-FAIS-GY | –79.27 | Yes |
| Colicin A H8 H9 | 18–4–17 | ALS-LGVP-ALAV | –7.06 | Yes |
| Haemoglobin H15 H16 | 19–5–21 | AHH-FGKEF-TPP | 92.81 | No |

applying such calculations, one can also have an insight into the mechanism of insertion process of soluble proteins, which is critical to the understanding of targeted drug delivery.

An important finding of this work is the existence of a deep free-energy well for the hairpin insertion into apolar medium. The earlier estimate[5] was very primitive and did not take into account other different free-energy terms. We have determined that only 'hydrophobicity' does not form the key. There are other factors involved, which in some cases, are much more significant. Our model is much more refined and accounts for all energy terms of significance, thereby giving it a note of validity, as observed from the matching of results in literature. Some hairpins tend to lie strongly into apolar environment depending upon their free-energy cost. Higher negative values indicate greater propensity to lie inside the membrane.

Finally, we note that the theoretical model presented here, which was applied only for the case of hairpin bend α-helical pairs is valid for all membrane proteins and, in addition, can be readily extended to the oligomeric aggregates in the membranes. A variety of biological processes, for example, viral infection and signal transduction, involve interaction between proteins and protein fragments in lipid bilayers[32]. Targeted drug delivery is a major research area related to this. The theoretical development presented in this work provides a platform for a detailed study of the molecular events that underlie these processes.


1. Vogel, H., *FEBS Lett.*, 1981, **134**, 37–42.
2. Leto, T. L. and Holloway, P. W., *J. Biol. Chem.*, 1979, **254**, 5015–5019.
3. Moll, T. S. and Thompson, T. E., *Biochemistry*, 1994, **33**, 15469–15482.
4. Ben-Tal, N., Ben-Shaul, A., Nivholls, A. and Honig, B., *Biophys. J.*, 1996, **70**, 1803–1812.
5. Engelman, D. M. and Steitz, T. A., *Cell*, 1981, **23**, 411–422.
6. Jacobs, R. E. and White, S. H., *Biochemistry*, 1989, **28**, 3421–3437.
7. Orlandini, E., Seno, F., Banavar, J. R., Laio, A. and Maritan, A., *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 14229–14234.
8. Jahnig, F., *Proc. Natl. Acad. Sci. USA*, 1983, **88**, 3691–3695.
9. Ben-Shaul, A., Ben-Tal. N. and Honig, B., *Biophys. J.*, 1996, **71**, 130–137.
10. Fauchere, J. L. and Pliska, V., *Eur. J. Med. Chem.*, 1983, **18**, 369–375.
11. Fattal, D. R. and Ben Shaul, A., *Biophys. J.*, 1993, **65**, 1795–1809.
12. Mouritsen, O. G. and Bloom, M., *Biophys. J.*, 1984, **46**, 141–153.
13. Engelman, D. M., Steitz, T. A. and Goldman, A., *Annu. Rev. Biophys. Chem.*, 1986. **15**, 321–353.
14. Bennett, M. J. and Eisenberg, D., *Protein Sci.*, 1994, **3**, 1464–1475.
15. Petros, A. M. *et al.*, *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 3012–3017.
16. Berghuis, A. M., Guillemette, J. G., Smith, M. and Brayer, G. D., *J. Mol. Biol.*, 1994, **235**, 1326–1341.
17. Li, J. D., Carroll, J. and Ellar, D. J., *Nature*, 1991, **353**, 815–821.
18. Parker, M. W., Postma, J. P., Pattus, F., Tucker, A. D. and Tsernoglou, D., *J. Mol. Biol.*, 1992, **224**, 639–657.
19. Muchmore, S. W. *et al.*, *Nature*, 1996, **381**, 335–341.
20. Silva, M. M., Rogers, P. H. and Arnone, A., *J. Biol. Chem.*, 1992, **267**, 17248–17256.
21. Kagan, B. L., Finkelstein, A. and Colombini, M., *Proc. Natl. Acad. Sci. USA*, 1981, **78**, 4950–4954.
22. Cabiaux, V., Mindell, J. and Collier, R. J., *Infect. Immunol.*, 1993, **61**, 2200–2202.
23. Mindell, J. A., Zhan, H., Huynh, P. D., Collier, R. J. and Finkelstein, A., *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 5272–5276.
24. Mindell, J. A., Silverman, J. A., Collier, R. J. and Finkelstein, A., *J. Membr. Biol.*, 1994, **137**, 45–57.
25. Silverman, J. A., Mindell, J. A., Finkelstein, A., Shen, W. H. and Collier, R. J., *J. Biol. Chem.*, 1994, **269**, 22524–22532.
26. Zhan, H., Choe, S., Huynh, P. D., Finkelstein, A., Eisenberg, D. and Collier, R. J., *Biochemistry*, 1994, **33**, 11254–11263.
27. Lacy, D. B. and Stevens, R., *Curr. Opin. Struct. Biol.*, 1998, **8**, 778–784.
28. D'Silva, P. R. and Lala, A. K., *J. Biol. Chem.*, 2000, **277**, 11771–11777.
29. Grochulski, P., Masson, L., Borisova, S., Pusztai-Carey, M., Schwartz, J. L., Brousseau, R. and Cygler, M., *J. Mol. Biol.*, 1995, **254**, 447–464.
30. Li, J. D., Carroll, J. and Ellar, D. J., *Nature*, 1991, **353**, 815–821.
31. Wu, D. and Aaronson, A. I., *J. Biol. Chem.*, 1992, **261**, 2311–2317.
32. Shai, Y., *Trends Biochem. Sci.*, 1995, **20**, 460–464.