

## The solvation concept approached by lipophily

M.-H. Van Eyck<sup>1</sup>, P. Ducarme<sup>1</sup>, N. Benhabiles<sup>1</sup>, A. Thomas<sup>2</sup> and R. Brasseur<sup>1</sup>

<sup>1</sup>Faculté Universitaire des Sciences Agronomiques de Gembloux, Centre de Biophysique Moléculaire Numérique, Passage des Déportés 2, B-5030-Gembloux, Belgique

<sup>2</sup>INSERM U10 de Gastro-Entérologie, Hôpital Cl. Bichat, 170 Bd. Ney, 75018 Paris, France

### Introduction

The solvent of biological media is water and interactions between water and solutes are major to explain membrane and protein structures. Therefore, mimicking solvation effects is a challenge to compute native structures of biological molecules. Lipophily is an experimental approach of solvation since it measures the partition of a solute between two solvents, the reference being water. In the last decade, different attempts were made to extract parameters from lipophily that will enable to describe solvation of complex molecules. We discuss here some analysis made with the atomic transfer energy and the atomic surface parameter in the study of protein folding and protein insertion in membranes.

The structure of a molecule depends on intramolecular and extramolecular interactions, extramolecular interactions being only with the solvent in dilute systems. In early approaches, conformations of molecules were calculated from intramolecular interactions only, as if molecules were in vacuum. However, influences of solvents are far from negligible. The overall shapes of soluble proteins illustrate that point: Kauzmann in the late fifties, highlighted that most globular proteins have an hydrophilic envelope surrounding an hydrophobic core [1]. This implicates that, in the time course of folding, some domains of proteins are burying inside to avoid interactions with the solvent. Simulating the folding of proteins in vacuum will never lead to such structures. Formation of biological membranes and micelles are similar kinds of process that emphasises the crucial role of water. Therefore it should be of interest to measure solvation energy if direct measures were possible. This is why

many approaches sprinkle to simulate solvation by theoretical calculations that use empirical functions based on experimental approaches.

If the role of water is obvious, understanding its role is not simple [2]. Effects of solvents result from a multitude of different interactions between the solute and the solvent and between solvent molecule themselves. At best, all atoms of a system should be taken into account and each force between couples of atoms calculated, disregarding whether atoms are from the solute or the solvent. Clearly, calculations get rapidly enormous and even if computing capacities increase fantastically, too much is too much and simplifications of calculations are required. Therefore, empirical treatments of solvation are compromises between the size of the calculated system, the resolution accepted and expected and, the calculation times [3].

At the moment, attempts to modelise solvation effects follow two main approaches: explicit and implicit representations of solvent [4]. In the first approach, the energetic contribution of thousands of solvent molecules is calculated. In the second one, the solvent is described as a continuous medium. Explicit representation is the most precise and should give the best results. A very clear success is obtained in mimicking occluded molecules of water. When they are not simulated, the corresponding holes are collapsed distorting the solute structure [3]. However, in explicit simulations, calculation times increase rapidly with the size of the system and thus are rapidly impossible. Actually, in large biological systems explicit representations of solvent are made when the solvent structure is analysed in response to solute. In most other cases, implicit representations of water are used. Molecules of water are replaced by a continuous medium that is defined to have physical and chemical properties of water at best. Those approaches are interesting to mimic the major solvent properties but not to mimic short distance interactions with solute.

In this review, we are interested in how water acts in the folding of globular proteins and in the insertion of proteins in membranes. Therefore, due to the large size of the biological models, simple implicit representations of water are used. We analyse the properties of solute-solvent interactions from lipophilicity and explain here different aspects of this approach.

### Forces implicated in solute-solvent interactions

As a pre-requisite to the following, we shall briefly evoke forces currently involved in solute-solvent interactions. All enthalpic effects of intra and inter-molecular interactions are electrostatic because they involve electronic interactions. Nevertheless, they are described under different terms: hydrogen bonds, polar interactions, electrostatic interactions and Van der Waals interactions. The entropic effects are, of course, not electrostatic. Since all terms describe different

aspects of a continuous electronic phenomenon, care must be taken not to overestimate one effect, as hydrogen bonds for instance, by overlapping its description in two different terms.

#### Enthalpic interactions: Electrostatic forces

The charge interactions are screened by the solvent molecules between the two charges. This screen effect is described by the dielectric constant that strongly varies with solvents. Variations are due to two effects: the polarisability of solvent molecules and their orientation in response to that polarisation. Both effects act to dissipate the electric field and thus to decrease the energy of interactions between the charges. Furthermore, the screen effect implicates many molecules of solvent, not only those between the charges. This makes the accurate description of an implicit solvent extremely difficult because the surrounding space can be filled with several kinds of molecules, heterogeneously distributed.

a) **Hydrogen bonds** play a major role when the solvent is water. The hydrogen bonds have a major role because liquid water is a continuous network of hydrogen bonds. Individual molecules of water do almost not exist but they are linked together by a mean of 3.4 hydrogen bonds per molecule [5]. Then, when you drop a solute in water, its solubility will depend on its capacity to substitute for the hydrogen bonds of water [6]. When the solute cannot make hydrogen bonds with water or, more accurately according to recent publications, when the solute makes hydrogen bonds of low energy with water [7], the molecules of water reorganise. They increase the mean number of their hydrogen bonds to minimise the loss of energy. This enthalpic gain comes with an entropic loss of energy because the mobility of water is decreased. This entropic effect is often called hydrophobicity. The whole process is energetically unfavourable [8].

For instance, the free energy of transfer of a molecule of butane, from butane to water at 25 °C is described as follows [9]:

$$\Delta G = \Delta H - T\Delta S = -4.3 + 28.7 = +24.5 \text{ kJ mol}^{-1}.$$

Therefore, the system will tend to minimise the surface of the water/butane interface: butane is hydrophobic. Energies of transfer of series of small molecules including amino acids were measured and correlated to the surface of the apolar atoms [5,10].

Reorganisation of hydrogen bonds around a solute is more complex than previously evoked because most molecules contain both polar and apolar groups, able and unable to make high-energy hydrogen bonds. Therefore precise behaviour of water is difficult to mimic [9].

b) **Monopole-monopole and monopole-dipole interactions** between solute and water have several origins. Charged atoms have strong interactions (monopole/monopole) and interact with water molecules (monopole/dipole). Moreover, electrostatic potential generated by the

solute-solvent system polarises the solvent (cf screen effect) but also the solute.

c) **Dipole/dipole interactions** are frequent in water, water being a dipole and proteins carrying many of them, especially the peptidic bonds.

d) **induced dipole-induced dipole** (van der Waals interactions) between water and solutes have a minor role since water molecules are small and mobile and will reorganise to cover the accessible solute surface at best. The resulting van der Waals energy will not vary much because of the absence of steric hindrance.

### Entropic effects

Movements of solutes in water are easy because the viscosity of water network is low. Diffusion of solutes in membranes is more restricted because the cohesion of acyl chains is high and the bilayer organisation restricts the space to two dimensions. Therefore, in membranes the diversity of solute conformations will be smaller and the native structure entropically favoured [9].

## Simulation of protein-solvent interactions by lipophily

In most of our studies, solutes are proteins. The nature of solvents influences the protein interactions. In water, two hydrophobic proteins seem to collapse [9]. In fact the protein-water system tends to minimise energy by decreasing the number of unfavourable hydrogen bonds and the easiest way to do this is to decrease the polar/apolar interfaces. This process results in protein insertion in membranes, protein folding but also in membrane formation. A simple empirical approach of problems of solvation in proteins is derived from lipophily.

Lipophily ( $\log P$ ) is used in the description of major properties of drugs. It is the partition of a solute between a polar (water) and apolar medium. The partition ( $P$ ) is measured at equilibrium as the ratio of solute concentrations in both solvents and is expressed as follows:

$$\log P = \log \frac{C_{\text{apolar}}}{C_{\text{polar}}}$$

Several organic phases have been tested, but only three were largely used: octanol, cyclohexane and water vapors. The partition coefficients are either measured or calculated [11].

### Which interactions are involved in lipophily?

Partition coefficient is the macroscopic description of the energy differences of a solute dissolved in an apolar *versus* a polar solvent. The energy of transfer  $\Delta G$  of molecules is derived from lipophily as follows

$$\Delta G = -2.3 RT \log P.$$

If the temperature is constant,  $\log P$  and  $\Delta G$  are linearly related.

Therefore, lipophily accounts for the differences of molecular interactions in both media. The differences can be on solute-solvent and solvent-solvent interactions. They can also be on intramolecular interactions if the solute changes conformation while switching solvent for instance.

The intermolecular changes can concern hydrogen bonds, monopole-monopole, monopole-dipole, dipole-dipole interactions, and induced dipole-induced dipole interactions. The major changes should be on hydrogen bonds, largely different in water and apolar media. Among those usually called apolar, some solvents actually make hydrogen bonds (octanol, ethanol), others do not (cyclohexane, hexadecane), so, some are strictly apolar, others are not. This could explain the differences in the partition coefficients of polar amino acids in octanol-water and cyclohexane-water. Residues making hydrogen bonds are more soluble in octanol than in cyclohexane [12].

As we evoked before, beside intermolecular interactions, the lipophily can also account for important changes in intramolecular interactions of the solute. Indeed, in water, the solute will tend to bury the apolar atoms and will turn inside-out in apolar solvent. Thus, important differences of solute structures can exist in both media, especially if the solute is a large and flexible molecule.

Moreover, a solute will affect the solvent conformational and rotational entropy. The magnitude of those changes depend on the solvent itself. For instance, alkyl chains of octanol have some degree of organisation and should reorganise largely to fit the solute surface. This reorganisation is not as important in globular solvents as ethanol and water [13].

### Calculation of lipophily

Most approaches of lipophily are based on plots of energies of transfer of molecules as a function of their number of carbons. Data from McAuliffe [16] et Bohon and Clausen [15] are reported in figure 1. They lead to conclude that:

1. Energies of transfer of two molecules are equal as long as they have the same number of carbons and of unsaturated bonds: this is true for ethyl benzene and xylene (8C and 4 unsaturated bonds) and methylcyclopentane and cyclohexane (6C and one unsaturated bond) (Fig. 1).
2. For an equal number of unsaturated bonds, the energy of transfer of molecules is linearly correlated to their number of carbons (Fig. 1).
3. Since both plots of figure 1 are parallel, the drop of  $E_{tr}$  due to the 3 more unsaturated bonds is constant.

However molecules of figure 1 are some how similar: they are all small, rigid and have no net charge. From there, it is suggested that the  $E_{tr}$  of small rigid molecules are the additive functions of the  $E_{tr}$  of their constitutive elements.

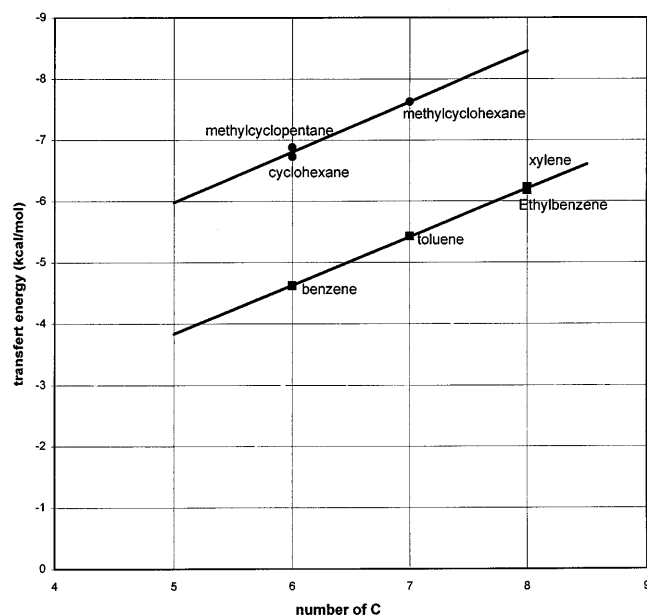


Figure 1. Plot of experimental transfer energies versus the number of carbons for several cyclic derivatives.

Several methods of calculation of  $E_{tr}$  have been developed in the literature [16–20]. The authors estimate the partition coefficients of molecules as the sum of atomic or fragmental function constants.

Where  $f_i$  is the lipophilic contribution of the molecular fragment or atom  $i$  and,  $a_i$ , the number of fragments  $i$ .

### Calculated versus experimental lipophilicity

The different authors are rapidly faced with problems of divergence between experimental and calculated values. Several reasons are suggested: proximity of charges, flexibility of solutes which change conformations, large solutes folding to create polar or apolar core according to the solvent, etc. Actually, besides the changes of solute-solvent interactions, these authors are facing intramolecular reorganisations.

In order to improve calculations, authors propose corrective factors [18,19] or define more atomic species [16,17].

### Using lipophilicity as an empirical approach of solvation

In the sixties, Tanford and collaborators attempted to define solvation as a quantifiable and measurable phenomenon [21]. One approach of solvent effects on solutes was lipophilicity. Indeed, the energy of transfer of a solute between apolar and polar solvents is measuring the variation of energy between both systems: it is an indirect approach of solvation energy. But deriving solvation from lipophilicity is not that simple. First, as we previously wrote, lipophilicity can also account for variations of intramolecular interactions. Those

are not solute-solvent interactions. This explains why to define solvation parameters from lipophilicity it is better to use small molecules as models for which variations of intramolecular interactions are negligible. Second, the partition experiments using polymeric solvents give entropic variations that are not significant of solute-solvent interactions. The theory of Flory-Huggins [22] introduces a parameter that describes the contribution of the entropic variation of the solvent in solute partition. This parameter is diversely used in the literature and, Chan and Dill in a recent review suggest that entropic correction must be restricted to polymeric solvents [13].

### Atomic solvation

Extracting solvation parameters from lipophilicity resumes therefore in calculating the contribution of solvation from solute partition. As in lipophilicity, the postulate that atomic contributions can be summed is also taken into account for solvation. Two approaches are made, whole atoms (atomic  $E_{tr}$ ) and the accessible surface of atoms, the atomic surface parameter (ASP).

### Atomic energy of transfer ( $E_{tr}$ )

In 1991, Brasseur [23] proposes a scale of atomic  $E_{tr}$  that he extracts from the  $E_{tr}$  of series of small molecules (Tab. I). These values are used to calculate molecule  $E_{tr}$  and molecular hydrophobicity potentials [3]. The scale was derived from Tanford's data [24] and proposes 7 atomic species. Those result from the resolution of a system of  $m$  equations with  $n$  unknown parameters,  $m$  being the number of molecules whose lipophilicity was measured hence the number of equations and  $n$  the number of different species of atomic required to resolve the equations.

Atomic  $E_{tr}$  were used to add a solvation term to the classical van der Waals, Coulomb, torsion terms of molecular energy [25]. In a recent study, we compared several energy fields in the calculation of KAD-129, ([2S]-2-benzyl-3-5-cis-hexahydro-2-isoindolinylcarbonyl) structure. It was found that, in the absence of the solvation term, the introduction of NMR constraints are required to reach a correct structure while no constraint is required if the solvation term is added. The energy-minimised structure has good homology with the NMR structure.

### Atomic surface parameter (ASP)

In 1986, Eisenberg and Mc Lachlan developed a method to simulate solvation [10]. The method relies on two hypotheses:

First, free energy of interactions of a solute with water is the sum of the energy of interaction of all its atoms.

Second, the solvation energy of each solute atom is proportional to its water-accessible surface [26,27]. The water-accessible surface is obtained as the surface described by the centre of a sphere rolling over the molecule surface (Fig. 2, see page 38). The sphere radius is 1.4 Angströms [28].

Table I. Atomic solvation parameters (ASP).

	<i>protein folding</i>		<i>protein insertion in membrane</i>		
	<i>Eisenberg &amp; McLachlan</i> <sup>(1)</sup>	<i>Nolde et al. (o/w)</i> <sup>(2)</sup>	<i>Brasseur</i> <sup>(3)</sup>	<i>Ducarme et al.</i> <sup>(4)</sup>	<i>Nolde et al. (c/w)</i> <sup>(2)</sup>
Csp3	67	109	87	105	130
Csp2	67	121	54	13.4	105
Cheter	/ <sup>(6)</sup>	50	/ <sup>(6)</sup>	/ <sup>(6)</sup>	17
O	-25	-63	-112	-40.3	-359
O <sup>-</sup>	-100	-59 <sup>(5)</sup>	-112	-40.3	-451.44 <sup>(5)</sup>
N	-25	-33	-116	-11.2	-339
N <sup>+</sup>	-109	-84	-116	-11.2	-736
S	87.8	171	89	108	142
H( <i>q</i> = 0)	/ <sup>(7)</sup>	/ <sup>(7)</sup>	26	39.7	/ <sup>(7)</sup>
H( <i>q</i> /0)	/ <sup>(7)</sup>	/ <sup>(7)</sup>	-51	-36.2	/ <sup>(7)</sup>

The first column is the atom type. Csp3, single-bonded carbons; Csp2 double-bonded and aromatic carbons; C heter, carbons bonded to heteroatoms; N, neutral nitrogen; N<sup>+</sup>, charged nitrogen; O, neutral oxygen; O<sup>-</sup>, charged oxygen; S, sulfur; H(*q* = 0), noncharged hydrogen (bound to C); H(*q*/0), hydrogen bound to heteroatoms.

The following columns list the ASPs (J mol<sup>-1</sup> Å<sup>-2</sup>).

(1) Eisenberg and McLachlan, 1986 [10].

(2) Nolde et al., 1997 [31]. The table specifies which type of transfer free energies were used to derive the ASP set: v + vacuum; o = *n*-octanol; w = water.

(3) Brasseur, 1991 [23]. The sets of atomic parameters are used for soluble and membrane protein. The literature values were divided by the surface of atoms to be compared with other sets.

(4) Ducarme et al., 1998 [32].

(5) Half charge is attributed to each oxygen atoms of the carboxyl groups.

(6) These sets of ASPs, carbons bonded to heteroatoms are not considered.

(7) These sets of ASPs, hydrogens are ignored.

Calculation of accessible surface can be achieved by numeric or analytic methods [29].

Thus, the energy of solvation of a protein is the sum of all multiplication products of each atom accessible surface per the corresponding atomic solvation parameter ( $\Delta g_i$ ):

$$\Delta G_{u \rightarrow f, s}^w = \Delta G_s = \sum_i \Delta g_i \cdot \Delta Sacc_i$$

Eisenberg and McLachlan [10] calculate 5 parameters of ASP by solving the system of *m* equations by the least square root method. The *m* equations were obtained from the data of Faucher and Pliska [30] who measured lipophilicity of amino acids. The values are given in table I. In that table, we compare the atomic  $E_{tr}$  of Brasseur [23] to the ASP values of Eisenberg and McLachlan [10], Nolde et al. [31] and, Ducarme et al. [32]. For sake of comparison, the values of the 1991 paper by Brasseur were corrected by the accessible surface of each kind of atoms.

As previously written the ASP should be a global parameter to describe solvation effects. How are ASP used?

We can legitimately think that apolar interactions are proportional to the accessible surfaces of atoms. Different studies reporting the linearity of the plot between the number of carbons and their accessible surface area are a good demonstration of that idea [27].

By contrast electrostatic interactions are not proportional to atomic surfaces. The ASP is the same whether the atom is buried or not. However, electrostatic free energies depend upon the composition and the thickness of the medium between two charges [33].

In conclusion, ASP mimic hydrophobic effects and can be used as global parameters but for more accurate solutions, more terms accounting for the previously restrictions should be introduced.

Both techniques (ASP and atomic  $E_{tr}$ ) are interesting since their carry out are simple and do not cost much computing time. The interest of ASP with respect to atomic  $E_{tr}$  is that atom accessible surfaces should partially compensate the artefacts due to intramolecular reorganisation because, buried atoms will not be taken into account.

### Simulation of solvation for the insertion of proteins in membranes

Protein insertion in membranes can be considered as a partition of a protein between two solvents, water and lipids. However membranes are complex solvents and membrane/protein interactions more difficult to simulate than

water/protein solvation. Membranes are “polymeric” constructions that are anisotropic mixes of several kinds of amphiphilic molecules. Their properties strongly differ from one end to the other and membrane/lipid interactions are directional: apolar at one end, polar at the other end. The bilayer structure is not only an intrinsic property of lipids, it also results from the contact with the surrounding water. The core is hydrophobic and the surface more hydrophilic. The surface is polar heads of lipids (Fig. 3, see page 38). The hydrophobic effect induces lipids to interact while the solvation effect stabilizes the conformation. Therefore, solvation energy will account for the protein interactions with membrane surface heads, and hydrophobic interactions will account for intramembraneous residue interactions.

Due to its own properties, simulation of membrane structure, even in the absence of protein is a major challenge. Until now, we are limited to simulate small patches of a few nanometers for a few nanoseconds [34].

Beside, it appears that:

- a) membranes do not behave like fluid mosaics [34].
- b) Proteins influence the bilayer structure, for instance when they induce fusion [35].
- c) Specific lipids are required for the activity of some proteins, and thus proteins may have preferential interactions with some molecules of the “solvent” [36].

Thus, it is clear that explicit simulation of membrane lipids and water should be necessary to discover the subtle mechanism of membrane activity. Unfortunately, molecular dynamics of such systems require very large calculation times and those investigations of the folding/insertion process has not been possible yet [37].

Surprisingly according to the previous considerations,  $E_{tr}$  calculations have been quite useful in this field. For instance, they are unable to describe 5 classes of alpha helices interacting differently with membranes [23]. To do so, the partition of hydrophobicity in the helix structure was analysed. The procedure is based on the calculation of hydrophilic and hydrophobic centres [38] given by:

$$C_{pho} = \frac{\sum E_{tri} w_i}{\sum E_{tri}}$$

where  $C_{pho}$  is the coordinate of the hydrophobic or hydrophilic centre and  $E_{tri}$  the transfer energy of  $i^{th}$  hydrophobic or hydrophilic atom whether the hydrophobic or hydrophilic centres are calculated.  $w_i$  are the coordinates of atoms  $i$ . Then, the hydrophobicity interface is defined as a plane perpendicular to the segment joining both centres so that the plane/segment intersection splits the segment proportionally to the hydrophobic and hydrophilic potentials respectively.

These analysis led to the concept of tilted peptides, i.e. peptides whose hydrophobicity is asymmetrically distributed along the helix axis. This particularity was shown to be essential for fusogenic properties especially for fusion pro-

teins of viruses [35, 39–42]. This is a good example of how lipophilicity can help in understanding peptide/membrane interactions: however, the mechanism of fusion cannot be explained. It is the fusogenic capacity of a peptide that can be predicted and this is a very important aspect.

Using amino acid  $E_{tr}$  is also useful for predicting membrane spans. Since the first methods of hydrophobicity plots [43], many progresses have been accomplished and recent techniques now claim to predict transmembrane helices with 95% accuracy [44].

However, this does not help us in understanding the mechanism of protein insertion in membranes.

The folding of alpha helices has however been studied by means of simplified models of peptides and of a continuous membrane description [45]. Starting from random structures in water, helices as complex as pf1 coat protein (a transmembrane and an adsorbed segments) were successfully inserted in membranes where they folded.

We use a similar approach to model the peptide/membrane interaction by Monte Carlo optimisation [32]. Because the final aim of this work is to optimise bundle of helices, a full atomic description of the peptide structure was used instead of the centroid models used in the previous study. This kind of approach was previously used by others [46, 47]. Two simple hypotheses were made:

- a) to simulate the solvation effect, we have to maximise the hydrophobic and to minimise the hydrophilic accessible surfaces of proteins that insert in membranes.

- b) to take into account unfavourable van der Waals interactions between peptides and lipids (as compared to lipids/lipids), we have to minimise the surface of contact between peptides and lipids.

The function has no energy terms and does not describe actual physical interactions, it is just an empirical function and its minima correspond to native-like conformations. In a recent assay, we tested the insertion of helices; the helical backbone structure was given as starting point and the procedure was sufficient to predict the correct position of alpha helices.

Conversely to classical hydrophobicity plots of sequences like Eisenberg's which predicts membrane spans [10], our approach tests whether the fragments do insert in membrane or not and calculates the dynamical behaviour in membranes. It was shown that melittin has several distinct local minima, suggesting that previous static classifications are too simplistic [32]. This hypothesis is in line with recent simulations of synthetic peptides (Ducarme, unpublished results) whose calculated conformations were in good agreement with solid-state NMR studies (Bechinger, personal communication).

### Atomic $E_{tr}$ and ASP in protein folding

If one considers the folding of proteins, the free energy of folding can be described as follows [48]:

$$\Delta G_{u \rightarrow f}^w = \Delta G_{u \rightarrow f, I}^w + \Delta G_{u \rightarrow f, S}^w$$

$u$  and  $f$  are unfolded and folded states respectively.  $w$  means that the phenomenon occurs in water.

$\Delta G_{u \rightarrow f, I}^w$  et  $\Delta G_{u \rightarrow f, S}^w$  are the respective contributions of solute intramolecular interactions and solute-solvent extramolecular interaction changes in the delta of free energy.

The empirical equation of contribution of solvation in the changes of free energy during folding is thus:

$$\Delta G_{u \rightarrow f, S}^w = \Delta G_S = \sum_i \Delta g_i \cdot \Delta Sacc_i$$

Where  $\Delta g_i$  is the atomic solvation parameter and  $Sacc_i$  the water-accessible surface of atom  $i$ . It is expected that, in most proteins, the energy of solvation is negative in the folded as compared to unfolded state. Actually, the major driving force of protein folding in water is solvation. The atomic solvation parameters should therefore mimic the preference of the different species of atoms for water or for an hydrophobic core better than the atomic  $E_{tr}$ . Thus, ideally, the experimental conditions chosen for lipophily measures should simulate the properties of the native protein core at best. Unfortunately, this core is heterogeneous. Indeed, besides the presence of apolar atoms, many core of proteins have intramolecular saline bridges. It was also described that large proteins, classically more than 100 residues have about 2% of their volume as cavity some containing occluded water molecules [49,50]. These create local hydrophilic patches. From this, it results that it is difficult to simulate the protein core as an homogenous solvent [51]. All apolar solvents mimic one or another property of the core but none gives a realistic description. Anyhow, for most authors, the best medium to simulate the protein core is octanol [52] but the debate is still open [8].

Some authors approach the problem differently [53]. They postulate that proteins fold the same way in water and in vacuum and thus that, intramolecular contributions are the same in both media. Therefore, differences in free energy should be due to solvent/protein interactions. They used the  $E_{tr}$  obtained from vapor/water partitions. Undoubtedly, the simplification is far from negligible.

Schiffer et al. [54] compared 4 sets of ASP in the optimisation of protein structure, 2 based on octanol/water partition and 2 on vapor/water partition. They conclude that the data from octanol/water are better.

The major criticism to previous techniques is that all authors do not converge toward similar values. Reasons are

multiple: first, there is no consensus on the atomic radius; second, there is not enough experimental data of lipophily. For example, the set of 20 amino acids used to describe all ASP for proteins is short. Some atomic species as sulphur for instance are too infrequently represented. More analysis should be carried out to define the different atomic species.

### Development of procedures to calculate protein structures from sequences

The folding of a nascent polypeptide chain into a compact native structure is a key-step in the translation of genetic information into molecular activity.

Polar amino acids are amphiphilic or biamphiphilic (Fig. 4) with a polar head and a hydrophobic segment [5,55]. Therefore, there is a similarity between amphiphilic characters of amino acids, lipids and detergents. All those molecules spontaneously make compact organised structures where hydrophobic domains are buried inside and thus protected against water.

A similar phenomenon to the formation of membranes is implicated in the folding of soluble proteins. The core is hydrophobic and the surface is more hydrophilic, made of polar extremities of the biamphiphilic amino acids. A protein should thus tend towards a compact form, the water-accessible surface of which is continuous and minimal [56]. Therefore, in unfolded proteins all residues are or can be water-accessible and, in folded proteins solvation energy will account for the surface of residue interactions with water, and hydrophobic interactions will account for intramolecular residue interactions.

Diverse methods of prediction of tridimensional structures of proteins were recently published [56–61]. They share the fact that, in addition to usual energetic terms describing the steric repulsion, the electrostatic interactions and the formation of hydrogen bonds, a term simulating solvation effects is added.

Srinivasan and Rose [58] simulate the energy of proteins with a very simple equation of three terms

$$E_{tot} = E_1 + E_2 + E_3$$

In which  $E_1$  is a scaled contact energy with both attractive and repulsive terms. The attractive term include pseudo-hydrophobic interactions.

$E_2$  is the backbone torsional potential.

$E_3$  simulates the hydrogen bonds of backbone.

Sun et al. [59] calculate amino acid interactions with an energy equation of two terms.

$$E_{tot} = E_1 + E_2$$

In which  $E_1$  takes into account the steric hindrance of amino acids and  $E_2$ , the hydrophobic interactions of the core using the following equation:

$$E_2 = \varepsilon \frac{1}{1 + e^{d_{ij}}}$$

$\varepsilon$  is equal to  $-1$  if both  $i$  and  $j$  residues in interaction are hydrophobic and is null in all other cases.  $D_{ij}$  is the distance between centroid centres.

In both methods, the amino acids are simulated as centroids, that is to say the backbone is explicitly described and lateral chains are simplified. They are replaced by a brief description of their steric hindrance and of their hydrophobicity character.

We developed a method for protein structure prediction with 6 energy terms [55–57]:

$$E_{\text{tot}} = E_1 + E_2 + E_3 + E_4 + E_5 + E_6.$$

In which  $E_1$  simulates van der Waals interactions,  $E_2$ , electrostatic interactions,  $E_3$  torsion energy,  $E_4$  hydrogen bonds and  $E_5$  hydrophobic interactions between atoms using the following equation:

$$E_4 = \delta_{ij} [ |E_{\text{tr}} f_{ij}| + |E_{\text{tr}} f_{ji}| ] \exp(-d_{ij} + r_i + r_j) / D_{\text{H}_2\text{O}}.$$

In which  $\delta_{ij} = -1$  if atoms  $i$  and  $j$  are of the same type, hydrophobic or hydrophilic, and  $= 1$  if they are different.  $E_{\text{tr}}$  and  $E_{\text{tr}}$  are the  $E_{\text{tr}}$  of  $i$  and  $j$  atoms,  $f_{ij}$ ,  $f_{ji}$  are the covering factors and  $r_i$ ,  $r_j$  the van der Waals radii.  $D_{\text{H}_2\text{O}}$  is the water diameter.  $E_5$  simulates the solvation energy using the water-accessible surface or, more precisely the number of solvent molecules directly in contact with the protein. In this method, the description of all atoms of amino acids are explicit. Water molecules are implicitly described via the water-accessible surface.

## Conclusions

We attempted to overview the complexity of solvation problems in protein folding and protein insertion in membranes with respect to the lipophilicity approach. The methods evoked in this review are simple, partial and they account only for part of the true phenomena. However their efficacy is proved. In the future, improvement of computing capacities should enable to go to explicit description of solvent molecules and to a better description of hydrogen bonds.

## Acknowledgements

R.B. is Research Director at the National Funds for Scientific Research of Belgium (FNRS). This work was supported by the “Interuniversity Poles of Attraction Programme – Belgian State, Prime Minister’s Office – Federal Office for

Scientific, Technical and Cultural Affairs” contract No. P4/03. The work of M-H.V. was supported by a grant from the Funds for Industrial and Agricultural Research (FRIA). The works of N. B. and Ph. D. were supported by a grant FNRS Televie. The collaboration between the two groups is supported by a CGRI-INSERM project (1998/0902).

## References

1. Kauzmann, W. *Adv. Protein Chem.* **1959**, *13*, 1-63.
2. Montulet, E. J. M. Molecular Force fields and implicit solvation. [http://cmcind.far.ruu.nl/montulet/solvation/solvatie\\_scriptie.html](http://cmcind.far.ruu.nl/montulet/solvation/solvatie_scriptie.html)
3. Van Gunsteren, W. F.; Luque, F. J.; Timms, D.; Torda, A. E. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 847-863.
4. Cramer, C. J.; Truhlar, D. G. *Reviews in Computational Chemistry* **1995**, *6*, 1-57.
5. Creighton, T. E. *Proteins: Structures and Molecular Properties*, W.H. Freeman and Company, New York, 1993.
6. Hoof, R. W. W.; Sander, C.; Vriend, G. *PROTEINS, Structure, Function and Genetic* **1996**, *26*, 363-376.
7. Wahl, M. C.; Sundaralingam *TIBS* **1997**, *22*, 97-102.
8. Dill, K. A. *Biochemistry* **1990**, *29*, 7133-7155.
9. Israeachvilli, J. *Intermolecular and surface forces*, Academic Press, London, 1991.
10. Eisenberg, D.; McLachlan, A. D. *Nature* **1986**, *319*, 199-203.
11. Carrupt, P. A.; Testa, B.; Gaillard, P. *Rev. Comp. Chem.* **1998**, *11*, 241-315.
12. Radzicka, A.; Wolfenden, R. *Biochemistry* **1988**, *27*, 1664-1670.
13. Chan, H. S.; Dill, K. A. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 425-459.
14. McAuliff, C. *J. Phys. Chem.* **1966**, *70*, 1267.
15. Bohon, R. L.; Claussen W. F. *J. Am. Chem. Soc.* **1951**, *73*, 1571.
16. Broto, P.; Moreau, G.; Vandycke, C. *Eur. J. Med. Chem.* **1984**, *19*, 71-78.
17. Ghose, A. K.; Crippen, G. M. *J. Comp. Chem.* **1986**, *7*, 565-577.
18. Leo, A.; Jow, P. Y. C.; Silipo, C.; Hansch, C. *J. Med. Chem.* **1975**, *18*, 865-868.
19. Rekker et al., *The hydrophobic fragmental constant*, Elsevier, New York, 1977.
20. Ben-Naim, A., Mazo, R. M. *J. Phys. Chem.* **1993**, *97*, 10829-10834.
21. Tanford, C. *J. Amer. Chem. Soc.* **1962**, *84*, 4240-4246.
22. Flory, P.J. *Principles of Polymer Chemistry*, Cornell Univ. Press, New York, 1953.
23. Brasseur, R. *J. Biol. Chem.* **1991**, *266*, 16120-16127.
24. Tanford, C. *The hydrophobic effect: formation of micelles and biological membranes*, Krieger Pub. Co., Malabar, 1991.
25. Lins, L.; Brasseur, R.; Malaisse, W. J.; Biesemans, M.; Verheyden, P.; Willem, R. *Biochem. Pharmacol.* **1996**, *52*, 1155-1168.
26. Hermann, R. B. *J. Phys. Chem.* **1972**, *76*, 2754-2759.
27. Reynolds, J. A.; Gilbert, D. B.; Tanford, C. *Proc. Nat. Acad. Sci.* **1974**, *71*, 2925-2927.
28. Lee, B.; Richards F. M. *J. Mol. Biol.* **1971**, *55*, 379-400.
29. Flower, D. R. *J. Mol. Graphics Mod.* **1997**, *15*, 238-244.



30. Faucher, J. L.; Pliska, V. *Eur. J. Med. Chem.* **1983**, *18*, 369-375.
31. Nolde, D. E.; Arseniev, A. S.; Vregoten, G.; Efremov, R. G. *J. Biomol. Struct. Dyn.* **1997**, *15*, 1-18.
32. Ducarme, P.; Rahman, M.; Brasseur, R. *Proteins*, **1998**, *30*, 357-371.
33. Smith, K. C.; Honig, B. *Proteins*, **1994**, *18*, 119-132.
34. Jacobson, K.; Sheets, E. D.; Simson, R. *Sci.* **1995**, *268*, 1441-1442.
35. Brasseur, R.; Pillot, T.; Lins, L.; Vandekerckhove, J.; Rosseneu, M. *Trends. Biochem. Sci.* **1997**, *22*, 167-171.
36. de Kruijff, B. *Nat.* **1997**, *386*, 129-130.
37. Merz, K. M. J. *Curr. Opin. Struct. Biol.* **1997**, *7*, 511-517.
38. Brasseur, R.; Deleers, M.; Ruyschaert J. M.; *J. Coll. Int. Sc.* **1986**, *114*, 277-281.
39. Horth, M.; Lambrecht, B.; Khim, M. C. L.; Bex, F.; Thiriart, C.; Ruyschaert, J. M.; Burny, A.; Brasseur, R. *EMBO J.* **1991**, *10*, 2747-2755.
40. Martin, I.; Defrise-Quertain, F.; Mandieau, V.; Nielsen, N. M.; Saermark, T.; Burny, A.; Brasseur, R.; Ruyschaert, J. M.; Vandenbranden, M. *Biochem. Biophys. Res. Comm.* **1991**, *175*, 872-879.
41. Martin, I.; Dubois, M. C.; Saermark, T.; Ruyschaert, J. M. *Biochem. Biophys. Res. Comm.* **1992**, *186*, 95-101.
42. Martin, I.; Dubois, M. C.; Defrise-Quertain, F.; Saermark, T.; Burny, A.; Brasseur, R.; Ruyschaert, J. M. *J. Virol.* **1994**, *68*, 1139-1148.
43. Kyte, J.; Doolittle, R.F. *J. Mol. Biol.* **1982**, *157*, 105-132.
44. Rost, B., Casadio, R., Fariselli, P.; Sander, C. *Protein Sci.* **1995**, *4*(3), 521-533.
45. Millik, M., Skolnick, J. *Protein Structure, Function and Genetics* **1993**, *15*, 10-25.
46. Edholm, O.; Jähnig, F. *Biophys. Chem.* **1988**, *30*, 279-292.
47. Biggin, P. C.; Breed, J.; Son, H. S.; Sansom, M. S. *Biophys. J.* **1997**, *72*, 627-636.
48. Juffer, A. H.; Eisenhaber, F.; Hubbard, S. J.; Walther, D.; Argos, P. *Protein Sci.* **1995**, *4*, 2499-2509.
49. Rashin, A. A.; Iofin, M.; Honig, B. *Biochemistry* **1986**, *25*, 3619-3625.
50. Hubbard, S. J.; Argos, P. *Protein Sci.* **1994**, *3*, 2194-2206.
51. Rose, G. D.; Wolfenden, R. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 381-415.
52. Kim, A.; Szoka, F. C. *Pharm. Res.* **1992**, *9*, 504-514.
53. Wesson, L.; Eisenberg, D. *Protein Sci.* **1992**, *1*, 227-235.
54. Schiffer, C. A.; Caldwell, J. W.; Kollman, P. A.; Stroud, R. M. *Mol. Simul.* **1993**, *10*, 121-149.
55. Lins, L.; Brasseur, R. *FASEB J.* **1995**, *9*, 535-540.
56. Brasseur, R. *J. Mol. Graphics* **1995**, *13*, 312-322.
57. Benhabiles, N.; Gallet, X.; Thomas-Soumarmon, A.; Brasseur, R. *J. Comput. Bio.*, in press.
58. Srinivasan, R.; Rose, G. D. *Proteins* **1995**, *22*, 81-99.
59. Sun, S.; Thomas, P. D.; Dill, K. A. *Protein Engineering* **1995**, *8*, 769-778.
60. Avbelj, F., Moul, J. *Proteins: Structure, Function and Genetics* **1995**, *23*, 129-141.
61. Defay, T., Cohen, F. E. *PROTEINS: Structure, Function and Genetics* **1995**, *23*, 431-445.