ALANINE DIPEPTIDE AND DISCRIMINATION OF MISFOLDED STRUCTURES OF PROTEINS WITH THE SCREENED COULOMB POTENTIAL CONTINUUM MODEL

Sergio A. Hassan and Ernest L. Mehler, Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York, New York 10029

The Screened Coulomb Potential-Implicit Solvent Model (SCP-ISM) is used to study the alanine dipeptide in aqueous solution and the discrimination of native protein structures from misfolded conformations. For the alanine dipeptide the free energy surface in the $\phi-\psi$ space is calculated and compared with recently reported results of a detailed Molecular Dynamics simulation using an explicit solvent representation, and with other available data. The study showed that the free energy surface obtained with the SCP-ISM is closer to that of the explicit water calculation than most other continuum models. Both transition states and energy minima show a high correlation ($r>0.98$) with the results obtained from the explicit water simulation. The study of the misfolded structures of proteins included the analysis of two standard decoy sets, i.e., the EMBL and the Park and Levitt (PL) sets. In both cases the SCP-ISM correctly discriminated the native from the misfolded structures. Analysis of the EMBL showed how the electrostatic and solvation contributions in the SCP-ISM gave the correct energy ordering between native and misfolded structures. For the PL set the SCP-ISM was found able to identify energetically, structures with small main chain RMSD from the native structure.

INTRODUCTION

In contrast to other continuum electrostatic approaches (see Figs. 1A and 1B), the SCP-ISM is derived from a microscopic starting point (Fig.1C) and the continuum description is obtained using the theory of dipolar solvation and statistical averaging over microscopic states.

The SCP-ISM does not require either a solute/solvent boundary or a definition of the so-called internal dielectric constant. The resulting description, shown in Fig.1D, consists of the macromolecule embedded in a dielectric that permeates all of space and is completely characterized by a dielectric function $\varepsilon(r)$.

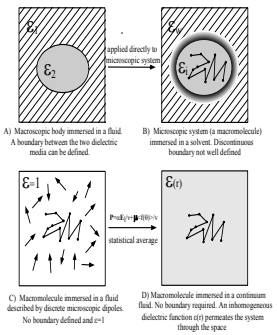

Figure 1: Schematic diagram illustrating two approaches for obtaining a continuum description of a macromolecule immersed in a polar medium. **A**: standard view of a macroscopic body immersed in a homogeneous medium. Both regions are treated as continuous dielectrics and a boundary between them can be defined as long as the uncertainty in its location can be neglected. **B**: direct application of the approach in A to a macromolecule immersed in water. Consistent with a strictly macroscopic description, the macromolecule is assigned a low dielectric constant and the water a high dielectric constant. However, at the microscopic/mesoscopic level of size scale neither the dielectric constants nor the boundary are well defined. **C**: the microscopic view of the system, where the molecule is immersed in a solvent of discrete particles with dipole moment $\boldsymbol{\mu}$ and polarizability α. **D**: the passage from the microscopic description in C to the continuum view in the SCP-ISM is formally carried out as described by the theory of polar liquids. The dielectric properties of the system are described by an effective non-homogeneous, distance-dependent dielectric function $\varepsilon(r)$ that permeates all of space; moreover, no boundary appears.

THE SCREENED COULOMB POTENTIAL-IMPLICIT SOLVENT MODEL

The SCP-ISM was reported earlier (see Refs.[1]) and has the following characteristics:

1) it is derived from a general thermodynamic cycle

2) it uses a non-homogeneous, distance-dependent dielectric function of sigmoidal form as predicted from theory and suggested by experiments

3) it makes use only of the Born approximation that is generalized further to account for the inhomogeneities of the dielectric properties in the macromolecule

4) it incorporates hydrogen bonding corrections in a fast and effective way

The total electrostatic energy is given by:

$$E_T = \frac{1}{2}\sum_{i\neq j}^{N}\frac{q_i q_j}{D(r_{ij})r_{ij}} + \frac{1}{2}\sum_{i=1}^{N}\frac{q_i^2}{R_{i,B}}\left[\frac{1}{D(R_{i,B})}-1\right]$$

ALANINE DIPEPTIDE

The alanine dipeptide is the standard system to study the energetics of backbone of amino acids. The two main degree of freedom correspond to the dihedral angle ϕ and ψ (see Figure 2)

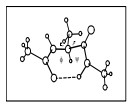

Figure 2: alanine dipeptide showing the two main degree of freedom in the most stable conformation C_7^{eq}

The free energy surface of the alanine dipeptide was calculated using the SCP-ISM, on a grid in the ϕ-ψ dihedral space. The results were compared with recently reported results from a detailed Molecular Dynamics studies with explicit solvent (see Ref.[2]).

Figure 3 shows the contour maps of the surface obtained with the SCP-ISM and the explict water calculation. Figure 4 shows the correlation of the values of minima and transition states between the SCP-ISM and explicit water calculation. Table I shows the values and locations of the minima obtained with the SCP-ISM and other available theoretical calculations.

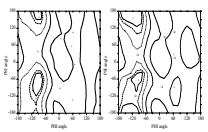

Figure 3: contour plots of the free energy surface calculated with A) the explicit water MD simulation of Ref.[2], and B) the SCP-ISM. The SCP-ISM description is very similar to the explicit water MD results. Both models found four minima (solid dots) and six transition states (open dots). Contour levels (in Kcal/mol) are as follows: short dashes 0.6; dashes 1.2; dots 2.4; dots/dashes 4.8; solid line 9.6.

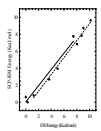

Figure 4: Scatter plot of the energy of the minima (solid dots; correlation r=0.98) and transition states (open dots; r=0.99), as calculated in the SCP-ISM and explicit water of Ref.[2].

Table I: Comparison of the minima of the free energy surface between the SCP-ISM and other solvent model calculations[a]

	C_7^{eq}	C_7^{ax}	α_R	α_L
SCP-ISM	0.00 [-115/160]	3.96 [65/-85]	0.62 [-80/-65]	7.79 [65/30]
Smith[2]	0.29 [-80/162]	5.02 [61/-133]	0.00 [-72/-56]	7.50 [59/57]
Anderson and Hermans[3]	0.00 [-110/120]	2.77 [70/-60]	1.46 [-120/-40]	2.48 [60'100]
Tobias and Brooks[4]	0.00 [-80/120]	3.61 [60/-80]	0.19 [-80/-60]	4.11 [60/60]
Apostolakis et al.[5]	0.00 [-75/136]	3.85 [57/-84]	1.41 [-76/-50]	4.37 [51/81]

[a] energy in Kcal/mol; angles in degrees

DSICRIMINATION OF MISFOLDED CONFORMATIONS OF PROTEINS

The study of the energetics of misfolded structures was performed on two standard decoy sets: the EMBL and the Park and Levitt (PL). In all cases the SCP-ISM identified the native structure as the one with the lowest energy, and a very good discrimination was obtained in all cases.

EMBL decoy set:

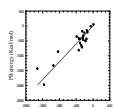

Figure 5: scatter plot of the total energy of the EMBL decoy set obtained with the SCP-ISM and with the Poisson-Boltzmann finite difference calculation reported recently [6]. The correlation is r=0.91

Park and Levitt decoy set:

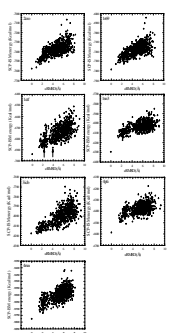

Figure 6: scatter plot of the total energy versus cRMSD (with respect to the native structure) of the Park and Levitt decoys set. The native structure is located at cRMSD=0. Table II shows the predicted structure (decoy with lowest SCP-ISM energy) for each of the proteins of the PL set.

Table II: Best-predicted conformations of the Park and Levitt decoy set

Structure id	cRMSD (Å)[a]	ΔE (Kcal/mol)[b]
1r69	0.88	56.3
2cro	0.88	12.4
1ctf	2.29	56.0
4pti	2.76	76.6
3icb	1.23	4.1
1sn3	2.21	120.9
4pxn	2.10	42.2

[a] cRMSD between predicted structure (decoy with lowest energy) and native conformation
[b] ΔE=E(predicted)-E(native)

CONCLUSIONS: The SCP-ISM is a general continuum model for calculating electrostatics in macromolecules. It is derived from a microscopic description using proper statistical averaging and a standard thermodynamic cycle. The model is based on the theory of dielectric media and the only assumption is the validity of the Born approximation that is generalized to take into account the inhomogeneous dielectric properties of the solvated macromolecule. The SCP-ISM was applied to two different systems comprising a single amino acid and whole proteins. It performed well in both applications. These results, along with others reported elsewhere [1,7] confirm the generality of the approach and show that it can be applied with confidence to molecules of biological relevance, at any size scale without requiring ad hoc modifications or parametrization.

REFERENCES

[1] S. A. Hassan, F. Guarnieri and E. L. Mehler, J. Phys. Chem. **104**, 6478 (2000); S. A. Hassan, F. Guarnieri and E. L. Mehler, J. Phys. Chem. **104**, 6490 (2000); S. A. Hassan and E. L. Mehler, Proteins **47**, 45 (2002)
[2] P. Smith, J. Chem. Phys. **111**, 5568 (1999)
[3] A. G. Anderson and J. Hermans, Proteins **3**, 262 (1988)
[4] D. J. Tobias and C. L. Brooks III, J. Phys. Chem. **96**, 3864 (1992)
[5] J. Apostolakis, P. Ferrara and A. Carlish, J. Chem. Phys. **110**, 2099 (1999)
[6] D. Petrey and B. Honig, Protein Sci. **9**, 2181 (2000)
[7] S. A. Hassan and E. L. Mehler, Int. J. Quant. Chem. **83**, 193 (2001); S. A. Hassan, E. L. Mehler and H. Weinstein, *Lecture Notes in Computational Science and Engineering*, T. Schlick, Ed., Springer, New York (2002)
M. Tartaglia, E. L. Mehler et al., *Nature Genetics* **29**, 465 (2001)