ELSEVIER

Computer Physics Communications 91 (1995) 291–303

Computer Physics Communications

Dielectric properties of proteins from simulations: tools and techniques

Thomas Simonson [a], David Perahia [b]

[a] *Laboratoire de Biologie Structurale, C.N.R.S., I.G.B.M.C., 1 rue Laurent Fries, BP 163, 67404 Illkirch, France*

[b] *Laboratoire d'Enzymologie Physico-Chimique et Moléculaire, Université de Paris-Sud, 91405 Orsay, France*

Received 17 October 1994

Abstract

Tools and techniques to analyze the dielectric properties of proteins are described. Microscopic dielectric properties are determined by a susceptibility tensor of order $3n$, where n is the number of protein atoms. For perturbing charges not too close to the protein, the dielectric relaxation free energy is directly related to the dipole–dipole correlation matrix of the unperturbed protein, or equivalently to the covariance matrix of its atomic displacements. These are straightforward to obtain from existing molecular dynamics packages such as CHARMM or X-PLOR.

Macroscopic dielectric properties can be derived from the dipolar fluctuations of the protein, by idealizing the protein as one or more spherical media. The dipolar fluctuations are again directly related to the covariance matrix of the atomic displacements. An interesting consequence is that the quasiharmonic approximation, which by definition exactly reproduces this covariance matrix, gives the protein dielectric constant exactly.

Finally a technique is reviewed to obtain normal or quasinormal modes of vibration of symmetric protein assemblies. Using elementary group theory, and eliminating the high-frequency modes of vibration of each monomer, the limiting step in terms of memory and computation is finding the normal modes of a single monomer, with the other monomers held fixed. This technique was used to study the dielectric properties of the Tobacco Mosaic Virus protein disk.

Keywords: Protein simulations; Dielectric properties

1. Introduction

Dielectric properties of proteins play a crucial role in folding and stability, binding and recognition, as well as charge transfer and enzyme kinetics [1–3]. Theoretical tools and computational techniques have been developed and tested for computing these properties, using simple linear response theory to analyze molecular dynamics simulations of proteins. Macroscopic and microscopic properties have been analyzed, including the protein dielectric constant and its spatial variation within the protein [4], the microscopic susceptibility in response to perturbing test charges, deviations from the linear response (i.e. dielectric saturation), and the separate contributions of electronic and dipolar relaxation to the dielectric response [2,3].

We briefly review the theory that underlies the calculation of protein dielectric properties, and the computational techniques that are used. The first section deals with microscopic dielectric properties. The second section deals with macroscopic dielectric properties, primarily the protein dielectric constant and its spatial variation. The third section reviews a technique to calculate the harmonic or quasiharmonic modes of

SSDI 0010-4655(95)00054-2

vibration of symmetric assemblies of proteins, which in turn give access to some of the dielectric properties of these assemblies.

2. Microscopic dielectric properties

2.1. Theory

Consider a set of nondiffusive charges such as a folded protein *in vacuo*, and a fixed perturbing charge density ρ. The perturbing Hamiltonian V_{tot} contains a 'static' term V_{static} and a relaxation term V_{rlx},

$$V_{tot} = V_{static} + V_{rlx}. \tag{1}$$

The first term is associated with introducing the perturbation while constraining the system to retain its unperturbed structure. The second term is associated with the structural relaxation after the constraints are removed. The perturbation free energy also has a static and a relaxation component,

$$A_{tot} = A_{static} + A_{rlx}. \tag{2}$$

If ρ is sufficiently small, we are in the linear response regime, and V_{rlx} is proportional to the perturbing field,

$$V_{rlx} = -\underline{x} \cdot \underline{f}. \tag{3}$$

$\underline{f}$ denotes the $3n$-vector,

$$\underline{f} = (\underline{f_1}, \underline{f_2}, \ldots, \underline{f_n}), \tag{4}$$

where $\underline{f_i}$ is the perturbing field on atom i, and n is the number of atoms in the system; $\underline{x}$ is a $3n$-vector, and measures the structural relaxation of the system. This relaxation can be characterized by a generalized susceptibility operator $\underline{\underline{\alpha}}$,

$$\langle \underline{x} \rangle = \underline{\underline{\alpha}}\, \underline{f}. \tag{5}$$

The brackets represent an ensemble average. Thus

$$\langle V_{rlx} \rangle = -{}^t\underline{f}\, \underline{\underline{\alpha}}\, \underline{f}. \tag{6}$$

The raised t indicates vector transposition. Using a simple charging process it can be shown [3] that

$$A_{rlx} = -\tfrac{1}{2}{}^t\underline{f}\, \underline{\underline{\alpha}}\, \underline{f}. \tag{7}$$

If the perturbing charge density ρ is distant from the protein atoms, then the susceptibility operator has the simple form [2]

$$\underline{\underline{\alpha}} = \frac{1}{kT}\underline{\underline{M}}, \tag{8}$$

where k is Boltzmann's constant, T is the temperature, and $\underline{\underline{M}}$ is the dipole–dipole correlation matrix of the unperturbed protein,

$$M_{i\alpha,j\beta} = q_i q_j \langle \delta u_i^\alpha \delta u_j^\beta \rangle_0. \tag{9}$$

The brackets $\langle\ \rangle_0$ indicate an ensemble average in the absence of the perturbation, q_i is the partial charge on atom i, and δu_i^α is its instantaneous displacement from its average position along the Cartesian axis α (= x, y or z). A sufficient condition for (8) to hold is that the minimum distance d between ρ and the protein is large with respect to the fluctuations of the protein atoms. Corrections to (8) are of the second order with respect to

$$\langle \delta \underline{u}_i^2 \rangle_0 / d^2.$$

We refer to this limiting situation as the *distant-ρ* limit. In this limit, the structural relaxation $\underline{x}$ is simply

$$\underline{x}_i = q_i \delta \underline{u}_i. \tag{10}$$

Thus $\underline{x}$ is the vector of instantaneous dipoles on each atom. When $\underline{f}$ is applied, the average of these instantaneous dipoles becomes non-zero and proportional to $\underline{f}$.

In the general case, ρ may be close to the protein, or even located on protein atoms. Then the analytical approximation (8) to the susceptibility operator no longer holds. The expression of the relaxation free energy is given by [2]

$$A_{rlx} = -kT \ln[\langle \exp(-V_{tot}/kT) \rangle_0] - \langle V_{tot} \rangle_0. \tag{11}$$

This is a general expression, valid beyond the linear response regime. Equating this to (7) gives a functional relationship between ρ and $\underline{\underline{\alpha}}$. This relationship can be inverted to give the complete susceptibility operator. In practice, one can consider a perturbing point charge q, and consider $3n$ distinct locations of this charge. For each location one obtains an equation relating V_{tot} to the $3n$ components of $\underline{\underline{\alpha}}$. The resulting system of equations is linear and can be inverted to obtain $\underline{\underline{\alpha}}$.

A simpler, partial, characterization of the dielectric properties is obtained by considering a single perturbing point charge q, and calculating the relaxation free energy as a function of its position. Dividing the relaxation free energy by the square of the perturbing field, gives a *scalar* susceptibility $\alpha(\underline{r}_q)$,

$$A_{\mathrm{rlx}} = -\tfrac{1}{2}\alpha(\underline{r}_q)\underline{f}^2. \tag{12}$$

The scalar susceptibility is a function of the position $\underline{r}_q$ of the perturbing charge. It represents a one-dimensional contraction of the full susceptibility operator.

The relaxation free energy can always be expanded with respect to V_{tot}, giving

$$A_{\mathrm{rlx}} = -\frac{1}{2kT}(\langle V_{\mathrm{tot}}^2\rangle_0 - \langle V_{\mathrm{tot}}\rangle_0^2) + O([V_{\mathrm{tot}}/kT]^3). \tag{13}$$

Terms of order 3 and higher correspond to a nonlinear response of the system, and therefore measure dielectric saturation.

Two particular situations are of interest. In the first, the 'protein' is actually a dielectric continuum, with fixed partial charges imbedded at discrete positions, surrounded by a vacuum. When ρ is introduced, the protein atoms remain fixed, but the dielectric relaxation is expressed by rearrangement of induced volume and surface charge. The relaxation free energy can be calculated by considering two processes: in the first, the perturbing charge distribution ρ is introduced into the system, giving a free energy change $\Delta A_{\mathrm{tot}}(\rho)$. In the second, the same perturbing charge distribution, but with the opposite sign, is introduced into the (unperturbed) system, giving a free energy change $\Delta A_{\mathrm{tot}}(-\rho)$. The relaxation free energy is simply

$$A_{\mathrm{rlx}} = \tfrac{1}{2}(\Delta A_{\mathrm{tot}}(\rho) + \Delta A_{\mathrm{tot}}(-\rho)). \tag{14}$$

This can be rewritten in terms of the induced charge at the protein surface. If $\delta\sigma_{\mathrm{ind}}$ is the shift in the induced surface charge due to the perturbation, then

$$A_{\mathrm{rlx}} = \tfrac{1}{2}\oint \delta\sigma_{\mathrm{ind}} V(\rho)\, d^2r. \tag{15}$$

The integral is over the protein surface, and $V(\rho)$ is the perturbing potential at the surface element d^2r [1].

[1] Ref. [2] gave the relaxation *energy* (p. 871), but with the incorrect sign due to a typographical error.

In the second particular situation, the protein atoms are fixed, but bear point polarizabilities, which could represent the electronic polarizability of the system [2]. When ρ is introduced, dielectric relaxation corresponds to a change in the induced dipoles on each atom. To calculate the relaxation free energy, we use the same two processes as in the previous paragraph. The initial energy is

$$U_0 = \sum_{i<j}\frac{q_iq_j}{r_{ij}} - \tfrac{1}{2}\sum_i \underline{m}_i\cdot\underline{E}_i^p, \tag{16}$$

where $\underline{m}_i$ is the induced dipole on atom i, and $\underline{E}_i^p$ is the field of the protein partial charges at atom i. In the presence of $+\rho$ the energy is

$$U_+ = \sum_{i<j}\frac{q_iq_j}{r_{ij}} - \tfrac{1}{2}\sum_i(\underline{m}_i+\delta\underline{m}_i)\cdot\underline{E}_i^p + \sum_i q_iV_i(\rho) - \sum_i(\underline{m}_i+\delta\underline{m}_i)\cdot\underline{f}_i, \tag{17}$$

where $\delta\underline{m}_i$ is the shift in the induced dipole on atom i due to ρ, and $V_i(\rho)$ is the perturbing potential at atom i. In the presence of $-\rho$ the energy is

$$U_- = \sum_{i<j}\frac{q_iq_j}{r_{ij}} - \tfrac{1}{2}\sum_i(\underline{m}_i-\delta\underline{m}_i)\cdot\underline{E}_i^p - \sum_i q_iV_i(\rho) + \sum_i(\underline{m}_i-\delta\underline{m}_i)\cdot\underline{f}_i. \tag{18}$$

The relaxation energy is given by

$$2\langle V_{\mathrm{rlx}}\rangle = U_+ + U_- - 2U_0 = -2\sum_i \delta\underline{m}_i\cdot\underline{f}_i. \tag{19}$$

From Eqs. (6),(7), the relaxation free energy is $\langle V_{\mathrm{rlx}}\rangle/2$. Thus we have finally

$$A_{\mathrm{rlx}} = -\tfrac{1}{2}\sum_i \delta\underline{m}_i\cdot\underline{f}_i. \tag{20}$$

In this case, the quantity $\underline{x}$ conjugate to the perturbing field (Eq. (3)) is just the list of induced dipole shifts, $\{\delta\underline{m}_i\}$.

If the protein atoms are mobile, and simultaneously carry a point polarizability, then electronic polarization and dipolar polarization (i.e. polarization due to atomic motions) coexist. A coupling then arises between the two. An analytical expression is available for the coupling energy [2] in the form of a series expansion with respect to the dipole–dipole tensor of

the system. Because this tensor is small, the coupling represents a corrective effect compared to the main electronic and dipolar contributions.

The well-known quasiharmonic approximation to the full protein dynamics views the protein as vibrating along effective, 'quasinormal', modes [5,6]. These modes are determined by requiring that they reproduce, collectively, the exact covariance matrix $\underline{\underline{\sigma}}$ of atomic displacements of the system. They are obtained in practice by diagonalizing an effective Hessian matrix $\underline{\underline{H}}$, given by

$$\underline{\underline{H}} = kT\underline{\underline{\sigma}}^{-1}. \tag{21}$$

$\underline{\underline{\sigma}}$ is usually taken from a molecular dynamics simulation. A correlary of this definition is that the quasinormal modes reproduce the exact dipole–dipole correlation matrix $\underline{\underline{M}}$ of the system. In the distant-ρ limit defined above, this matrix gives the generalized susceptibility of the system. This means that when considering perturbing charges at a sufficient distance d from the protein, such that

$$\forall i, \langle \delta \underline{u}_i^2 \rangle / d^2 \ll 1,$$

the quasiharmonic approximation gives an essentially *exact* description of the protein's dielectric response.

Within the quasiharmonic approximation, the elements of the atomic covariance matrix take the form of a sum over the normal modes,

$$\langle \delta u_i^\alpha \delta u_j^\beta \rangle = 2kT \sum_k \omega_k^{-2} \delta u_i^\alpha(k) \cdot \delta u_j^\beta(k), \tag{22}$$

where $\delta u_i^\alpha(k)$ is the α component ($= x$, y or z) corresponding to atom i in the normalized displacement vector of normal mode k, and ω_k is the frequency of the mode. Thus the susceptibility operator, in the distant-ρ limit, also takes the form of a sum over the normal modes.

2.2. Computational procedures

2.2.1. Susceptibilities from molecular dynamics simulations

It is straightforward to calculate the scalar susceptibility from a molecular dynamics simulation. The distant-ρ approximation is obtained directly from the dipole–dipole correlation matrix (Eq. (8)), while the exact susceptibility is obtained from the time-series of the perturbation energy (Eq. (11)). Fig. 1 shows a script file for the program X-PLOR that reads a trajectory file and writes out the latter time series.

The susceptibilities can be also decomposed over the quasinormal modes of the system (Eq. (22)). These can be obtained directly from the programs CHARMM [7] or X-PLOR [8].

2.2.2. Case of a macroscopic medium

In the particular case where the 'protein' is considered to be a macroscopic continuum with embedded point charges, the relaxation free energies can be obtained in two different ways. The shift in surface charge induced by the perturbation ρ can be calculated directly, or the free energies for introducing and subtracting ρ can be calculated and averaged (Eq. (14)). We consider each of these methods below.

The shift in surface charge due to ρ is given by [9]

$$(\underline{\underline{I}} - c\underline{\underline{K}})\delta\underline{\sigma} = c\underline{f}. \tag{23}$$

This equation assumes the protein surface has been broken down into n small, discrete elements. $\delta\underline{\sigma}$ is an n-vector giving the surface charge shift at each element; $\underline{f}$ is an n-vector giving the perturbing field at each element; $\underline{\underline{I}}$ is the unit tensor of order n, and the elements of $\underline{\underline{K}}$ are

$$K_{ij} = \delta S_j (\underline{r}_i - \underline{r}_j) \cdot \underline{n}_i / (\underline{r}_i - \underline{r}_j)^3. \tag{24}$$

δS_j is the surface area of element j, $\underline{r}_j$ its position, and $\underline{n}_i$ is a unit vector normal to surface element i. Finally c is a constant, $c = (1-\epsilon)/2\pi(1+\epsilon)$, where ϵ is the protein dielectric constant. To obtain the surface charge shift, we must invert (23). Two methods can be used: direct matrix inversion [9], or a self-consistent iterative method [2]. The iterative method is equivalent to summing the series $\sum_{n=0}^{\infty}(c\underline{\underline{K}})^n$, which converges if the eigenvalues of $c\underline{\underline{K}}$ are strictly smaller than 1 in magnitude. While this holds for some cases, it does not hold for large molecules with modest numbers of surface elements (Simonson & Perahia, unpublished results). For a sphere of radius R, for example, assuming the N surface elements all have the same area, the elements of $\underline{\underline{K}}$ are just

$$K_{ij} = 2\pi c R / N r_{ij}, \tag{25}$$

```
remarks    Interaction between a test charge on the Calpha of
remarks     residue 1 with the rest of the protein.

{read structural information, including atom types, charges,
 and covalent connectivity}
structure @<name of structure file> end

{read empirical energy parameters and specify electrostatic cutoff}
parameters
   @<name of parameter file
   nbonds
      cdie eps=1. e14fac=1. shift
      cutnb=12. ctonnb 10. ctofnb 11.5
   end
end

{set charge of Calpha of residue 1 to one}
vector do (charge=1.) (name ca and resid 1)

{only compute interactions between Calpha 1 and the rest of protein}
constraints
interaction=(name ca and resid 1) (all)
weights * 0. elec 1. end
end

{initialize loop for reading trajectory}
evaluate ($M=1)
evaluate ($status="READ")

{loop over trajectory frames}
while ($status = "READ") loop main !----------

if ($M=1) then                          {read 1st trajecroy frame}
read dynamics
     ascii=true
     input=<name of trajectory file>
     begin=100 skip=100 stop=120000
end
evaluate ($M=$M+1)
else                                    {read other trajectory frames}
read dynamics next end
end if

energy end              {calculate perturbation energy for current frame}

display $ELEC                           {write out perturbation energy}

end loop main   !----------------------------

stop
```

Fig. 1. X-PLOR script file to extract the time series of perturbation energies from a trajectory, for a perturbing test charge on one of the C_α's.

r_{ij} being the distance between surface elements i and j. The largest matrix elements are between neighboring surface elements, such that $r_{ii+1} \sim 2\pi R/N^{1/2}$, and

$$K_{ii+1} \sim c/N^{1/2}. \tag{26}$$

Thus when N increases, the largest matrix elements scale as $1/N^{1/2}$. This may not be sufficient to ensure convergence in all cases.

An analytical continuation of the series $\sum_{n=0}^{\infty}(c\underline{\underline{K}})^n$ can be used to extend convergence to cases where the eigenvalues of $c\underline{\underline{K}}$ are larger than 1 in absolute magnitude. Consider first the numerical case: let $g(x) = 1/(1-x)$, where x is a real number. The usual series form of g can be generalized: given a real number a,

$$\left|\frac{x+a}{1+a}\right| < 1 \Rightarrow g(x) = \frac{1}{1+a}\sum_{n=0}^{\infty}\left(\frac{x+a}{1+a}\right)^n. \tag{27}$$

Choosing a large positive a, we can shift the convergence of the series form towards negative x; choosing a large negative a, we can shift the convergence towards positive x. The same idea can be applied to the matrix series above. We simply add to $\underline{\underline{K}}$ the unit tensor $\underline{\underline{I}}$, multiplied by a real number a. Since $\underline{\underline{K}}$ and $\underline{\underline{I}}$ commute, we have the same series expansion as in the scalar case,

$$(\underline{\underline{I}} - c\underline{\underline{K}})^{-1} = \frac{1}{1+a}\sum_{n=0}^{\infty}(c\underline{\underline{K}} + a\underline{\underline{I}})^n/(1+a)^n, \tag{28}$$

as long as the eigenvalues of $c\underline{\underline{K}} + a\underline{\underline{I}}$ are smaller than $|1+a|$ in absolute magnitude. Choosing a by trial and error, we can apply the iterative method to systems that would not otherwise converge.

The question of whether a suitable choice of a can always be made remains open. The constraint that the relaxation free energy in response to ρ is always negative, however, constrains the structure of the matrix $\underline{\underline{I}} - c\underline{\underline{K}}$. It appears that there is a universal upper bound for the eigenvalues of this matrix. This would imply that a suitable value of a can be found in all cases, and the iterative method can always be used.

The second approach to calculating susceptibilities for a macroscopic system is to calculate the free energies for introducing and subtracting ρ (Eq. (14)). This can be done using the finite difference method to solve the Poisson equation, as implemented in programs such as Delphi [10]. A complicated protocol must be used in order to subtract out numerically the lattice energy associated with distributing the perturbing charge density ρ over a three-dimensional grid [11]. This protocol is illustrated in Fig. 2.

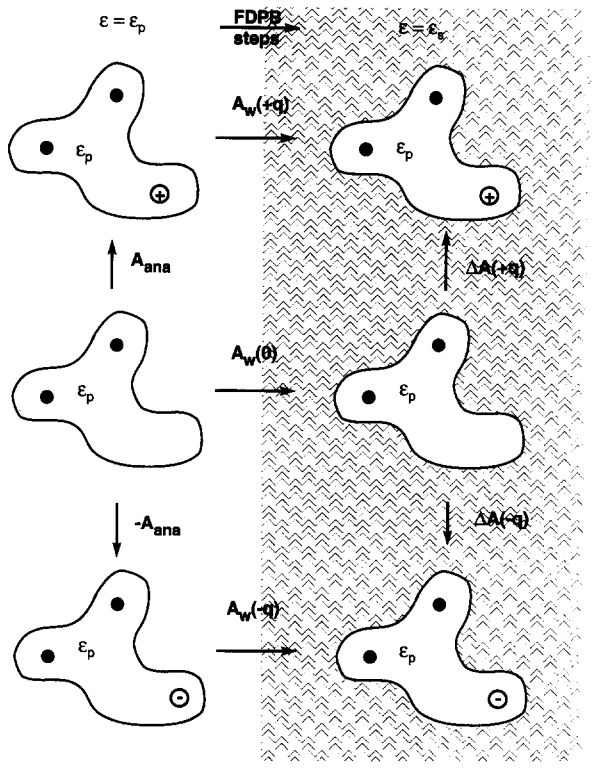

Fig. 2. Thermodynamic cycle to calculate the relaxation free energy for adding a single perturbing charge to a macroscopic system. The relaxation free energy is $A = \frac{1}{2}(\Delta A(+q) + \Delta A(-q)) = \frac{1}{2}(A_w(+q) + A_w(-q) - 2A_w(0))$. The left-hand medium has the same dielectric constant as the protein, $\epsilon = \epsilon_p$. The right-hand medium has a different dielectric constant, corresponding to a surrounding vacuum or solvent. The left-to-right transfer processes can be calculated with a finite difference Poisson–Boltzmann package [10]. The vertical processes on the left cancel out.

2.2.3. *Atomic point polarizabilities and induced dipoles*

The shift in induced dipoles due to the perturbing charge density ρ is given by [2]

$$(\underline{\underline{I}} - \underline{\underline{\alpha_{at}}}\,\underline{\underline{T}})\,\delta\underline{m} = \underline{\underline{\alpha_{at}}}\,\underline{f}, \tag{29}$$

where $\underline{\underline{\alpha_{at}}}$ is the diagonal matrix of order $3n$ formed by the list of atomic polarizabilities (each repeated 3 times), $\underline{\underline{T}}$ is the dipole–dipole matrix of the system, $\underline{f}$ is the $3n$-vector giving the perturbing field at each atom, and $\delta\underline{m}$ is the list of induced dipole shifts at each atom. As in the previous section, this equation can be solved by direct matrix inversion, or by an iterative procedure, which is equivalent to summing the matrix series $\sum_{i=0}^{\infty}(\underline{\underline{\alpha_{at}}}\,\underline{\underline{T}})^i$. This sum converges if $\underline{\underline{\alpha_{at}}}\,\underline{\underline{T}}$ has no eigenvalues greater than 1 in absolute magnitude. For the moderate point polarizabilities present in proteins, this appears to be usually the case. This is because the elements of $\underline{\underline{T}}$ are of order $1/r_{ij}^3$, and decrease rapidly with atomic separations. If interactions between bonded atoms are included in $\underline{\underline{T}}$, however, the series will sometimes diverge [2]. Damping of the dipole–dipole interaction at short range can be used to ensure convergence [12].

3. Macroscopic dielectric properties

The macroscopic dielectric properties are determined by the protein dielectric constant. For an isotropic medium this is a number; for an anisotropic medium it is a 3×3 matrix. For an inhomogeneous medium, the dielectric constant may vary throughout the system. The definition of a dielectric constant assumes that a large portion of the protein responds as a homogeneous medium. Thus spatial variation of the dielectric constant is only meaningful on a scale that is large compared to the atom–atom separation. We calculate the protein dielectric constant from Fröhlich–Kirkwood theory [13], by idealizing the protein as a homogeneous spherical medium. We summarize below the general theory derived in [2] and [4].

3.1. Theory

Consider a system made of n concentric spherical regions, of dielectric constants $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$. The outer radius of region i is r_i; the outermost region n is infinite.

Following Fröhlich [13], we can derive an expression for the dielectric constant of the innermost region (1) by introducing an applied field $\underline{E}$, uniform far from the center of symmetry. A uniform cavity field $\underline{F}$ then reigns in region (1). $\underline{F}$ is obtained from elementary electrostatics [13], and has the form [14]

$$\underline{F} = f(\epsilon_1, \epsilon_2, \ldots, \epsilon_n)\underline{E}. \tag{30}$$

The average polarization $\langle \Delta \underline{M} \rangle$ in region (1) is along $\underline{E}$, and is given by

$$\langle \Delta \underline{M} \rangle / V = \frac{\epsilon_1 - 1}{4\pi}\underline{F} = \frac{\epsilon_1 - 1}{4\pi} f(\epsilon_1, \epsilon_2, \ldots, \epsilon_n)\underline{E}, \tag{31}$$

where V is the volume of region (1).

We now view region (1) microscopically. The cavity field in region (1) is now $\underline{F} = f(1, \epsilon_2, \ldots, \epsilon_n)\underline{E}$. The microscopic degrees of freedom $\{X\}$ give rise to an instantaneous dipole moment $\Delta \underline{M}(X)$. The interaction with $\underline{E}$ adds a term $\Delta \underline{M}(X) \cdot \underline{F}$ to the potential energy. The Boltzmann average of $\Delta \underline{M}(X)$, for small $\underline{E}$, turns out to be

$$\langle \Delta M \rangle = \frac{\langle \Delta M^2 \rangle_0}{3kT} f(1, \epsilon_2, \ldots, \epsilon_n) E. \tag{32}$$

$\langle\ \rangle_0$ indicates a Boltzmann average with $E = 0$. In all that follows we drop the subscript '0' for simplicity. From (31–32) we obtain finally

$$\frac{\langle \Delta M^2 \rangle}{kTr_1^3} = \frac{f(\epsilon_1, \epsilon_2, \ldots, \epsilon_n)(\epsilon_1 - 1)}{f(1, \epsilon_2, \ldots, \epsilon_n)}. \tag{33}$$

If the inner region (1) contains permanent charges, the previous derivation is only slightly modified. ΔM has simply to be interpreted as the deviation of the dipole moment from its mean.

A slightly more general derivation allows for the possibility that region (1) may have electronic polarizability, which can be treated as an underlying continuum. A fluctuation formula is obtained that gives the dielectric constant as a function of the fluctuations of only the low frequency (non-electronic) degrees of freedom [4].

$\langle \Delta M^2 \rangle$ is determined by the correlations between all pairs i, j of protein atoms,

$$\langle \Delta M^2 \rangle = \sum_{i,j} q_i q_j \langle \delta \underline{u}_i \cdot \delta \underline{u}_j \rangle \tag{34}$$

(q_i is the partial charge of atom i, $\delta \underline{u}_i$ its instantaneous displacement from its mean position). It follows that the quasiharmonic approximation, which preserves the atomic covariance matrix, exactly reproduces the mean square dipole moment $\langle \Delta M^2 \rangle$. This means that the mean square dipole moment can be rigorously decomposed into a sum over the quasinormal modes, and the contributions of individual modes or frequency ranges rigorously defined and calculated.

3.2. Practical calculation of the dielectric constant

The first ingredient required in the calculation is the analytical expression of the cavity field in region (1), or equivalently, the function f (Eq. (30)). Its calculation is as follows. The potential in region i, in spherical coordinates, has the form [13]

$$\Phi_i = -(A_i/r^3 + B_i r)\cos\theta, \tag{35}$$

where A_i and B_i are constants determined by the boundary conditions. The boundary conditions include the continuity of $E_\theta = -(1/r)\partial\Phi/\partial\theta$, the continuity of $D_r = \epsilon\partial\Phi/\partial r$, and the conditions $A_1 = 0$ and $B_n = E$. Together these give a linear system of $n - 2$ equations, which can be written in matrix form,

Table 1
Boundary equations for n spherical media in a uniform field

	1	2	3	4	...	$2i$-3	$2i$-2	$2i$-1	$2i$	$2i$+1	$2i$+2	...	$2n$-5	$2n$-4	$2n$-3	$2n$-2
$\underline{X}$	B_1	A_2	B_2	A_3	...	B_{i-1}	A_i	B_i	A_{i+1}	B_{i+1}	A_{i+2}	...	B_{n-2}	A_{n-1}	B_{n-1}	A_n
$\underline{V}$	0	0	0	0				...					0	0	E	$\epsilon_n E$
$\underline{L}_1$	1	$-1/r_1^3$	-1	0				...					0	0	0	0
$\underline{L}_2$	ϵ_1	$2\epsilon_1/r_1^3$	$-\epsilon_2$	0				...					0	0	0	0
⋮																
$\underline{L}_{2i-1}$	0	0	0	0	...	0	$1/r_i^3$	1	$-1/r_i^3$	-1	0	...	0	0	0	0
$\underline{L}_{2i}$	0	0	0	0	...	0	$-2\epsilon_i/r_i^3$	ϵ_i	$2\epsilon_{i+1}/r_i^3$	$-\epsilon_{i+1}$	0	...	0	0	0	0
⋮																
$\underline{L}_{2n-3}$	0	0	0	0				...					0	$1/r_{n-1}^3$	1	$-1/r_{n-1}^3$
$\underline{L}_{2n-2}$	0	0	0	0				...					0	$-2\epsilon_{n-1}/r_{n-1}^3$	ϵ_{n-1}	$2\epsilon_n/r_{n-1}^3$

$\underline{L}_i$ is the ith line of the matrix $\underline{\underline{L}}$.

$$\underline{\underline{L}}\,\underline{X} = \underline{V}, \tag{36}$$

where

$$\underline{X} = (B_1, A_2, B_2, A_3, B_3, \ldots, A_{n-1}, B_{n-1}, A_n), \tag{37}$$

$$\underline{V} = (0, 0, \ldots, 0, E, \epsilon_n E), \tag{38}$$

and $\underline{\underline{L}}$ is an $(n-2)\times(n-2)$ matrix listed in Table 1. The formal matrix inversion of $\underline{\underline{L}}$ is easily done with a symbolic mathematics program such as Mathematica [15].

The cavity field in region (1) is uniform, and equal to B_1. The function f (Eq. (30)) is just

$$f(\epsilon_1, \epsilon_2, \ldots, \epsilon_n) = B_1/E = X_1(E=1). \tag{39}$$

In the case $n = 3$, for example, we obtain

$$f(\epsilon_1, \epsilon_2, \epsilon_3) = \frac{9\epsilon_2\epsilon_3}{(\epsilon_1+2\epsilon_2)(\epsilon_2+2\epsilon_3)-2(r_1/r_2)^3(\epsilon_3-\epsilon_2)(\epsilon_1-\epsilon_2)}. \tag{40}$$

Having determined f, we make an assumption about the protein radius r_1, which could be chosen such that the sphere of radius r_1 has the same radius of gyration as the actual protein. Fig. 3 illustrates the spherical idealization of a protein, cytochrome c. There are $n = 3$ dielectric regions in this particular application: the protein forms region 1, a shell of water forms region 2, the surrounding vacuum forms region 3. Finally, the mean square dipole moment $\langle M^2\rangle$ is estimated from a molecular dynamics simulation, and the protein dielectric constant ϵ_1 is calculated from Eq. (33). Some results are summarized in Table 2. Fig. 4 is a Mathematica script to analyze the sensitivity of the calculated dielectric constant to the exact value assumed for the protein radius r_1.

4. Normal modes of symmetric protein assemblies

Because the quasiharmonic approximation gives exact dielectric properties when perturbing charges are not too close to the protein (distant-ρ limit), methods to compute protein normal modes are of interest here. One method was specifically developed to treat large symmetric protein assemblies [17]. This method is summarized below.

Consider a symmetric assembly of asymmetric protein monomers. Let G be the symmetry group of the multimer. Let n_G be the order of G, which is also equal to the number of monomers. Let N be the number of degrees of freedom of one monomer. By symmetry, the normal mode calculation of the multimer can be reduced from a problem of order $n_G N$ to a series of problems of order at most $2N$, using elementary group theory [18,19].

Let $\{T_p; p = 1, p_G\}$ be the irreducible representa-

Table 2
Dielectric constant of cytochrome c from simulations

protocol	surroundings[a]	treatment of charged groups[b]	ϵ_1
normal modes[c]	vacuum		2.9
normal modes[c]	vacuum	scaled by 0.3	2.5
molecular dynamics[d]	vacuum		3.5
molecular dynamics[d]	vacuum	scaled by 0.3	2.8
molecular dynamics[e]	water		25
molecular dynamics[e]	water	omitted	3.6
experiment[f]	protein powder		3.6

[a]Region 2.
[b]Treatment of charged groups (such as carboxylate and ammonium groups) in the analysis only. No special treatment is used in the simulations.
[c]Ref. [2].
[d]Ref. [3], 120 ps of MD.
[e]Ref. [4], 1000 ps of MD with 1400 explicit waters.
[f]Ref. [16].

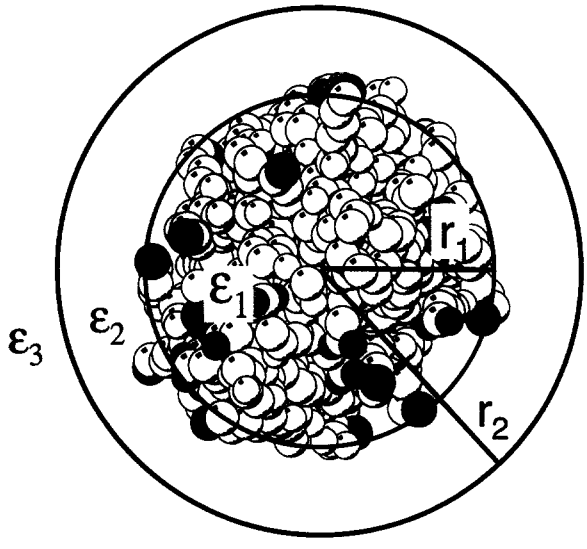

Fig. 3. Spherical idealization of cytochrome c (region 1) and surrounding water (region 2) as two concentric dielectric media surrounded by a vacuum (region 3).

tions of G of dimension s_p. For $g \in G$, let $\underline{\underline{\tau}}^p(g)$ be the (complex) matrix representing g in the representation T_p. Let $\{\underline{u}_k; k = 1, N\}$ be an orthonormal basis of the configuration space of monomer 1. A basis of symmetrical coordinates is obtained by symmetrizing the $\underline{u}_k$. Denoting g_i the symmetry element that transforms monomer 1 into monomer i, the symmetrical coordinate $\underline{q}_k^{ps}$ corresponds to a displacement of monomer i of the form

$$\underline{r}_i = \tau_{ss}^p(g_i)^* g_i \underline{u}_k. \tag{41}$$

Let $\underline{\underline{H}}$ be the mass-weighted Hessian matrix of the potential energy function of the multimer, expressed in the basis of Cartesian displacements. Let $\underline{\underline{H}}^{(ij)}$ be the block in $\underline{\underline{H}}$ corresponding to the interaction of monomers i and j. Let $\underline{\underline{\mathcal{H}}}$ be the Hessian matrix in the basis of symmetric coordinates $\underline{q}_k^{ps}$. We find [17]

$$\mathcal{H}_{k,\kappa}^{ps,\pi\sigma} = \frac{n_G}{s_p} {}^t\underline{u}_k \{\sum_{j=1}^{n_G} \tau_{ss}^p(g_j) \underline{\underline{V}}^{(1j)} g_j\} \underline{u}_\kappa \, \delta_{s\sigma}\delta_{p\pi}. \tag{42}$$

In the basis of symmetrical coordinates, the Hessian matrix is thus reduced to a series of blocks of order N. If the irreducible representation p is not one-dimensional, each line in the matrix $\underline{\underline{\tau}}^p$ contributes one block, giving $s_1 + s_2 + \ldots + s_{p_G} = n_G$ in all. The s_p blocks corresponding to $\underline{\underline{\tau}}^p$ are all identical, so that there is an s_p-fold degeneracy of the spectrum for each irreducible representation T_p.

If the irreducible representation T_p is complex, then the complex, symmetric, subspaces corresponding to T_p and T_p^* must be combined to give a physically meaningful symmetric subspace of double the dimension. The basis vectors $\underline{q}_k^{ps}$ and $\underline{q}_k^{ps*}$ are replaced by their real and imaginary parts, and the complex blocks $\underline{\underline{\mathcal{H}}}^{ps,ps}$ and $\underline{\underline{\mathcal{H}}}^{ps,ps*}$, of order N, combine to form a real block of order $2N$.

In the quasiharmonic approximation, the effective Hessian matrix is related to the covariance matrix by

```
kT = 0.59 ; e3 = 1. ;
parameters = { e2 -> 80, r2 -> 24., m2 -> 130.} ;

f[e1_,e2_,e3_] :=
 9 e2 e3 / ((e1+2e2)(2e3+e2) - 2(r1/r2)^3 (e3-e2)(e1-e2));

lhs = 332 m2 / (kT r1^3) ;

rhs = f[e1,e2,e3] (e1-1) / f[1,e2,e3];

lhsN = lhs /. parameters ;
rhsN = rhs /. parameters ;

rule = Solve[ lhsN == rhsN, e1] ;
e1 = e1 /. rule[[1]]
```

Out[167]=

$$\frac{-1.06295\ 10^{9} + 59457.1\ r1^{3} + 1.\ r1^{6}}{6.64344\ 10^{6} - 13695.4\ r1^{3} + 1.\ r1^{6}}$$

```
Plot[e1, {r1,14,18}, AxesLabel -> {"r1 (A)","e1"} ];
```

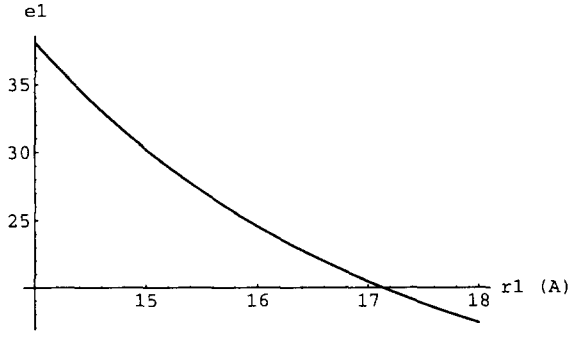

Fig. 4. Mathematica [15] file to analyze the sensitivity of the protein dielectric constant to the choice of protein radius. Numerical values are typical of cytochrome c.

$$\underline{\underline{\sigma}} = kT\underline{\underline{H}}^{-1}. \tag{43}$$

In symmetric coordinates, the Hessian $\underline{\underline{\mathcal{H}}}$ is made up of diagonal blocks $\underline{\underline{\mathcal{H}}}^{pp}$ of order N. Separating out the six external degrees of freedom, the remaining blocks are positive definite at the energy minimum, therefore invertible. The inverse $\underline{\underline{\mathcal{H}}}^{-1}$ has the same block-diagonal structure as $\underline{\underline{\mathcal{H}}}$. It follows that the transformation (42) that reduces $\underline{\underline{H}}$ to block-diagonal form reduces $\underline{\underline{H}}^{-1}$, and therefore $\underline{\underline{\sigma}}$, to the same block-diagonal form. In other words the method presented in the preceding paragraphs can be applied without modification to the quasiharmonic approximation.

To reduce the computing burden, we seek to eliminate the high-frequency modes of the system at an early stage of the calculation. The natural way to do this is to eliminate the high-frequency modes of the individual monomers. At this point, the basis $\{\underline{u}_k; k = 1, N\}$ is arbitrary. Let us now take the $\underline{u}_k$ to be the N eigenvectors of the block $\underline{\underline{H}}^{(11)}$ of the Hessian $\underline{\underline{H}}$. This means that $\underline{u}_k$ is a normal mode of monomer 1, with its neighboring monomers artificially held rigid. Let us discard the high-frequency modes from this set, retaining the n lowest-frequency modes $\{\underline{u}_k; k = 1, n\}$. Constructing the corresponding symmetric coordinates, we reduce $\underline{\underline{H}}$ to a series of diagonal blocks of order n or $2n$. For large systems such as proteins,

only the lowest fraction of the spectrum need be retained, so that $n \ll N$. Then the calculation is essentially reduced to the initial determination of the $\underline{u}_k$, a single problem of order N.

The sequence of operations to obtain the normal modes of the multimer is the following:

(i) Calculate the Hessian matrix of monomer 1 (or the covariance matrix in the quasiharmonic approach) in the presence of its rigid neighbors, $\underline{\underline{H}}^{(11)}$;
(ii) Diagonalize this matrix, retaining only the n lowest-frequency modes, $\{\underline{u}_k; k = 1, n\}$;
(iii) Calculate the other blocks of the Hessian (or the covariance matrix in the quasiharmonic approach), coupling monomer 1 to its neighbors, $\underline{\underline{H}}^{(1,j)}$;
(iv) Contract and symmetrize these blocks by applying Eq. (42), to obtain small blocks of order $\leq 2n$, $(\mathcal{H}^{ps,ps}_{k,\kappa})$;
(v) Diagonalize these small blocks to obtain the normal modes of the multimer, expressed in symmetric coordinates;
(vi) Convert these modes back to Cartesian coordinates.

Each block $\underline{\underline{H}}^{(1,j)}$ contributes separately to $\underline{\underline{\mathcal{H}}}^{(ps,ps)}$ (Eq. (42)), so that each one can be calculated and processed separately, without the need to store more than one in computer memory at any time. Since $\underline{\underline{\mathcal{H}}}^{(ps,ps)}$ is typically much smaller than $\underline{\underline{H}}^{(1,j)}$, the memory requirements are essentially the same as for a normal mode calculation of a single monomer. C.p.u. requirements are very small for any normal mode calculation (<5 minutes for a TMV protein monomer on a Fujitsu VP200 supercomputer using highly vectorized subroutines).

4.1. Computational procedures: application to the disk of TMV protein

In the case of the disk of TMV protein, the symmetry group is C_{17}. The complex, one-dimensional, irreducible representations are listed in Table 3. The real, physically irreducible, representations are

$$T_1, T_2 + T_{17}, T_3 + T_{16}, T_4 + T_{15}, \ldots, T_9 + T_{10}.$$

The complex symmetrical coordinate q^p_k has the form

Table 3
Irreducible representations of C_{17}

g	I	R	R^2	R^3	...	R^{16}
T_1	1	1	1	1	...	1
T_2	1	ω	ω^2	ω^3	...	ω^{16}
T_3	1	ω^2	ω^4	ω^6	...	ω^{32}
T_4	1	ω^3	ω^6	ω^9	...	ω^{48}
⋮	⋮	⋮	⋮	⋮	⋱	⋮
T_{17}	1	ω^{16}	ω^{32}	ω^{48}	...	$\omega^{16\times 16}$

R represents the rotation of $2\pi/17$ around the axis of the disk, and $\omega = \exp(2i\pi/17)$.

$$\underline{q}^p_k = \begin{pmatrix} \underline{u}_k \\ \omega^{(p-1)*} R\underline{u}_k \\ \omega^{2(p-1)*} R^2 \underline{u}_k \\ \vdots \\ \omega^{16(p-1)*} R^{16} \underline{u}_k \end{pmatrix}. \tag{44}$$

R represents the rotation of $2\pi/17$ around the axis of the disk, and $\omega = \exp(2i\pi/17)$. The real symmetric coordinates have the form

$$\underline{Q}^p_k = \frac{1}{\sqrt{2}}(\underline{q}^p_k + \underline{q}^{p*}_k),$$
$$\underline{R}^p_k = \frac{1}{i\sqrt{2}}(\underline{q}^p_k - \underline{q}^{p*}_k). \tag{45}$$

Routines to implement these equations for groups of the form C_n were included in the program CHARMM [7].

Fig. 5 shows the scalar susceptibilities of the TMV protein disk, in response to a test charge placed on a C_α, as a function of the residue number (reproduced from [17]). The susceptibilities are calculated from the normal modes of the disk. The individual contributions of the nine symmetric subspaces are shown separately.

5. Conclusion

A variety of tools have been described for the analysis of dielectric properties of proteins. The general theory of microscopic dielectric properties of proteins [2,3] has been reviewed. This theory is concerned with calculating relaxation free energies and generalized susceptibilities of proteins in response to perturbing charges. Practical calculations are usually based

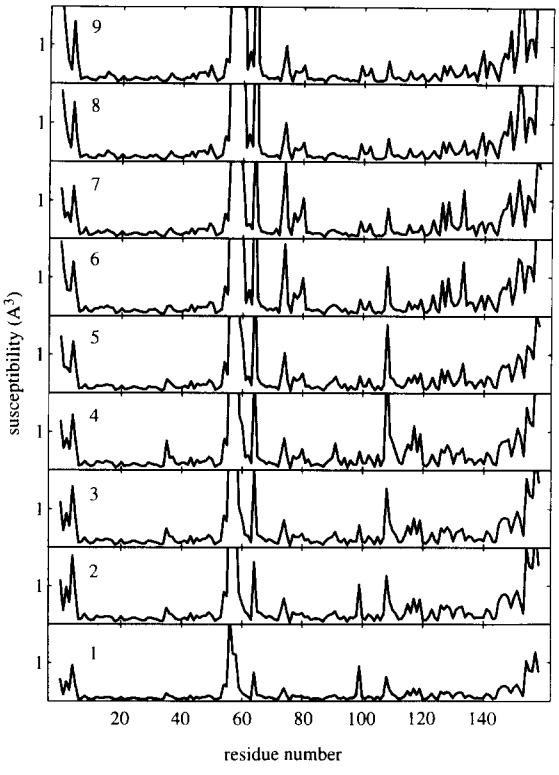

Fig. 5. TMV protein disk susceptibilities in symmetric subspaces [17]. A test charge is placed successively on each C_α of subunit 1 and the scalar susceptibility computed from the normal modes. Each panel corresponds to one symmetric subspace.

on molecular dynamics simulations. Susceptibilities are obtained numerically either from the atom–atom covariance matrix (distant-ρ limit) or from the time series of perturbing energies. Two special models are also of interest: in the first the protein is treated as a continuum with embedded point charges; in the second, the protein atoms are fixed, but bear point polarizabilities. In the continuum case, the relaxation free energy is related to the shift in induced surface charge due to the perturbation. Practical calculations can use either a self-consistent iterative procedure, or direct matrix inversion to obtain the surface charge shift. Alternatively programs can be used that solve the Poisson equation with a finite-difference method. A simple thermodynamic cycle permits one to eliminate the lattice energy associated with distributing ρ over a grid.

The macroscopic dielectric constant can be obtained from the protein dipolar fluctuations, if one idealizes the protein as a homogeneous spherical medium. Spatial variation of the dielectric constant going from the center of the protein to the outside can also be analyzed.

Finally a technique to calculate normal or quasinormal modes of symmetric protein assemblies was reviewed. Since the quasiharmonic approximation preserves the atom–atom covariance matrix, it also preserves the susceptibility matrix in the distant-ρ limit, as well as the macroscopic dielectric constant. Thus an exact decomposition of these quantities over the quasinormal modes can be made, indicating the contributions of different frequency ranges to the dielectric response.

Time-dependent dielectric properties were not treated here; they will be discussed elsewhere.

Acknowledgements

Some of the work discussed here was done in collaboration with Gérard Bricogne and Axel Brünger.

References

[1] A. Warshel and S. Russell, Calculations of electrostatic effects in biological systems and in solutions, Q. Rev. Biophys. 17 (1984) 283–342.

[2] T. Simonson, D. Perahia and G. Bricogne, Intramolecular dielectric screening in proteins, J. Mol. Biol. 218 (1991) 859–886.

[3] T. Simonson, D. Perahia and A.T. Brünger, Microscopic theory of the dielectric properties of proteins, Biophys. J. 59 (1991) 670–90.

[4] T. Simonson and D. Perahia, Internal and interfacial dielectric properties of cytochrome c from molecular dynamics simulations in aqueous solution, Proc. Natl. Acad. Sci. USA 92 (1995) 1082–1086.

[5] M. Karplus and J. Kushick, Method for estimating the configurational entropy of macromolecules, Macromolecules 14 (1981) 325–332.

[6] R. Levy, M. Karplus, J. Kushick and D. Perahia, Evaluation of the configurational entropy for proteins: application to molecular dynamics simulations of an α-helix, Macromolecules 17 (1984) 1370–1374.

[7] B. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan and M. Karplus, Charmm: a program for macromolecular energy, minimization and molecular dynamics calculations, J. Comp. Chem. 4 (1983) 187–217.

[8] A.T. Brünger, X-plor version 3.1, A System for X-ray crystallography and NMR (Yale Univ. Press, New Haven, 1992).

[9] R. Zauhar and R. Morgan, A new method for computing the macromolecular electric potential, J. Molec. Biol. 186 (1985) 815.
[10] K. Sharp, DelPhi, Version 3.0, Columbia University, New York (1988).
[11] M. Gilson, K. Sharp and B. Honig, Calculating electrostatic interactions in bio-molecules: method and error assessment, J. Comp. Chem. 9 (1988) 327–335.
[12] B. Thole, Molecular polarizabilities calculated with a modified dipole interaction, Chem. Phys. 59 (1981) 341–350.
[13] Fröhlich, H., Theory of Dielectrics (Clarendon Press, Oxford, 1949).
[14] J. Powles, R. Fowler and W. Evans, The dielectric constant of a polar liquid by the simulation of liquid drops, Chem. Phys. Lett. 107 (1984) 280–283.
[15] S. Wolfram, Mathematica. A system for doing mathematics by computer. (Addison Wesley, New York, 1991).
[16] S. Bone and R. Pethig, Dielectric studies of protein hydration and hydration-induced flexibility, J. Mol. Biol. 181 (1985) 323–326.
[17] T. Simonson and D. Perahia, Normal modes of symmetric protein assemblies. application to the tobacco mosaic virus protein disk, Biophys. J. 61 (1992) 410–427.
[18] E. Wilson, J. Decius and P. Cross, Molecular vibrations. The theory of infrared and Raman vibrational spectra (McGraw-Hill, New York, 1955).
[19] G. Lyubarskii, The application of group theory in physics (Pergamon Press, New York, 1960).