

Long-Range Electrostatic Forces

H.J.C. Berendsen and W.G.J. Hol, Groningen

(Biomolecular Study Centre, Chemical Labs. of the University of Groningen)

Long-range electrostatic forces play an important role in the interactions within and between biological macromolecules. The molecular structure of several proteins seems to be designed to exert a specific electrostatic long-range effect with functional significance. The efficient evaluation of such forces is not straightforward and poses an interesting problem in computational physics. The computational solution of this problem is urgently needed for the rapidly growing biophysical applications of *Molecular Dynamics* (MD), a computer simulation method that mimics the dynamics of complex molecular systems. The MD method has matured to the stage that it includes the realistic evaluation of free energies and the prediction of binding properties between complex molecules. This opens the way for computer predictions in such practical applications as drug and vaccine design, as well as enzyme design and protein engineering. Of the many problems to be overcome in these computational techniques, the correct evaluation of electrostatic interactions is central and challenging.

Source and Range of Electrostatic Interactions

Most atoms that compose biological macromolecules are partially charged and hence produce local dipoles. In addition some of the molecular groups, when solvated by water, are ionized and carry a full positive or negative electronic charge. In both cases electric fields result. However, if dipoles are more or less randomly oriented, their net long-range effect is not very important, while the effect of solvated charges is limited because of the screening influence of the solvent. The most important effect is to be expected from isolated charges and from correlated dipoles in a hydrophobic environment where solvent screening is absent. The correlation between dipoles causes a spatial charge separation: a linear sequence of dipoles (a line dipole) is electrically equivalent to two charges at the ends of the line. Thus correlation between dipoles causes electrostatic effects to become long-ranged.

Only in the last decade has it been recognized that such correlated dipoles play a functional role in many biological macromolecules. The most important example of a correlated dipolar structure is the α -helix, occurring as a structural element in many proteins (Fig. 1). The dipolar nature resides in the backbone structure; each peptide unit has a dipole of $1.2 - 1.5 \times 10^{-29}$ C m (depending on how much mutual polarisation occurs). With a density of 7×10^9 peptide units per metre along the helix, the linear dipole density is $0.8 - 1 \times 10^{-19}$ Coulomb. Considering the helix as a linear line dipole of homogeneous density, its electric field is equivalent to that of two charges of 0.5 to 0.6 electronic charge units at the ends of the helix. As can be seen in Fig. 2, the line dipole approximation is quite good. Thus a helix produces a non-solvated, buried positive half electronic charge at its amino terminal, and a similar negative charge at its carboxyl terminal.

The usual method of computing electrostatic forces by summation of Coulomb terms within a certain cut-off range can cause grave errors. For example, in Fig. 3 the potential and field are

plotted on the axis of a helix, for various lengths of the helix. If a cut-off of 2 nm is used (which is quite respectable; usual cut-offs do not exceed 1 nm!), the computed values deviate considerably from the true values. Moreover, the field is no longer a derivative of the potential and consistency is lost in molecular dynamics simulations.

Another problem arises from the inhomogeneity of the dielectric properties in a hydrated macromolecule: the solvent presents a dielectric environment that is totally different from the one inside the macromolecule. Unless all degrees of freedom of the solvent are explicitly taken into account and a sufficiently thick layer of solvent is considered, this inhomogeneity must be incorporated into the computation of electric fields. We shall return to this problem.

Functional Significance of Long-range Fields

As was mentioned above, the main source for long-range fields in proteins are α -helices. They serve to produce a strong electric field in a hydrophobic environment and play a well-documented role in the binding of charged small molecules (*substrates*) and, for enzymes, in the enhancement of chemical reactivity. In addition, the helix fields will contribute to the factors directing the folding of a protein; they may also serve to attract and properly orient substrates before they bind. Since these effects have been described extensively in the literature¹⁾ it will suffice to give some examples.

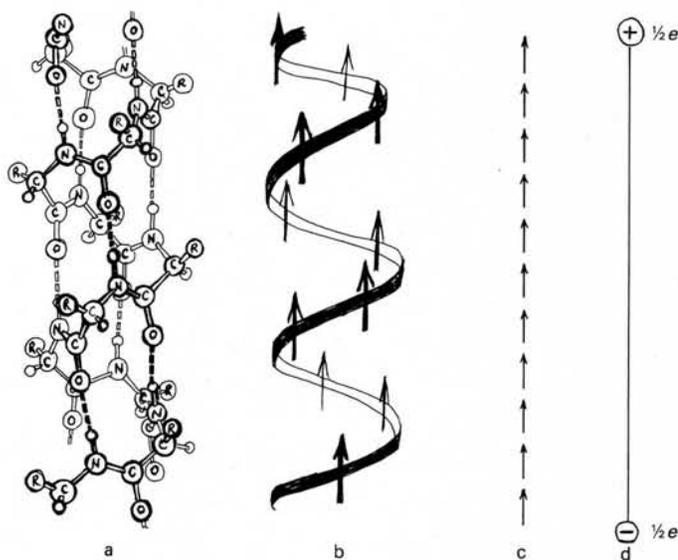


Fig. 1 — The backbone atoms of the α -helix (a), a frequently occurring structural element in proteins, represent parallel dipole moments (b), which can be approximated by a line dipole (c). The field of this line dipole is equivalent to that of two opposite charges at the ends of the helix with a magnitude of about 0.5 electronic charges (d) [from Ref. 2].

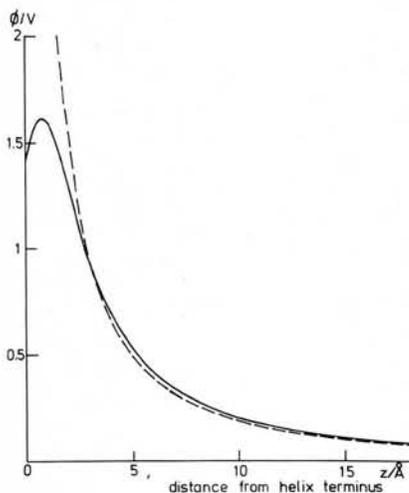


Fig. 2 — The potential of a helix is well represented by its equivalent of two charges. Full curve: potential on helix axis due to seven peptide dipoles of 1.2×10^{-29} Cm on a helix at 0.19 nm from the axis with 0.15 nm translation and 100° rotation per dipole. Dashed curve: potential due to two charges of 8×10^{-20} C at the ends of the helix.

Many enzymes bind substrates or co-enzymes that contain one or more charged phosphate groups, often bound to one or between two nucleotides. Very often the charged phosphate is situated in its binding site at the amino terminal of an α -helix. Comparison of very different and evolutionary dissimilar phosphate-binding proteins reveals that these functional helices are very similar as if Nature has invented the functional helix several times over.

A functional role in enzymatic reactions is quite clear in several cases where a proton transfer is involved in the reaction. Serine esterases (as *subtilisin*) and cysteine esterases (as *papain*) are examples. These are enzymes that cleave ester or peptide bonds in a series of reaction steps, starting with the transfer of a proton from an -OH or -SH group to a histidine. Normally this transfer would not take place because the

The Department of Physics at the University of Nijmegen has an opening for a

full professor of experimental physics

His task will include research in solid state physics and the daily responsibility for a well-equipped laboratory (low temperature facilities, very high magnetic fields, infrared spectroscopy, tunnel-microscopy etc.). The know-how of the staff and the infrastructure warrant, besides the continuation of existing research, a quick start of new research activities. Material science is co-ordinated by an institute RIM (Research Institute for Materials).

Applicants should have extensive research experience connected with recent progress in the field of solid state physics. Collaboration with theoretical groups serves as recommendation.

Didactical qualities are prerequisite. The applicant will participate in the teaching activities of the Department of Physics at undergraduate and graduate level.

A contribution to the management of the Department and the Faculty is expected. Applications should include Curriculum Vitae, a list of publications and names of references. Suggestions for potential candidates are welcome.

Prof.dr. J. Reuss, Physics Department, Katholieke Universiteit, Toernooiveld, 6525 ED Nijmegen, The Netherlands, telephone no. The Netherlands 80-558833, telex 48228 winat.nl.

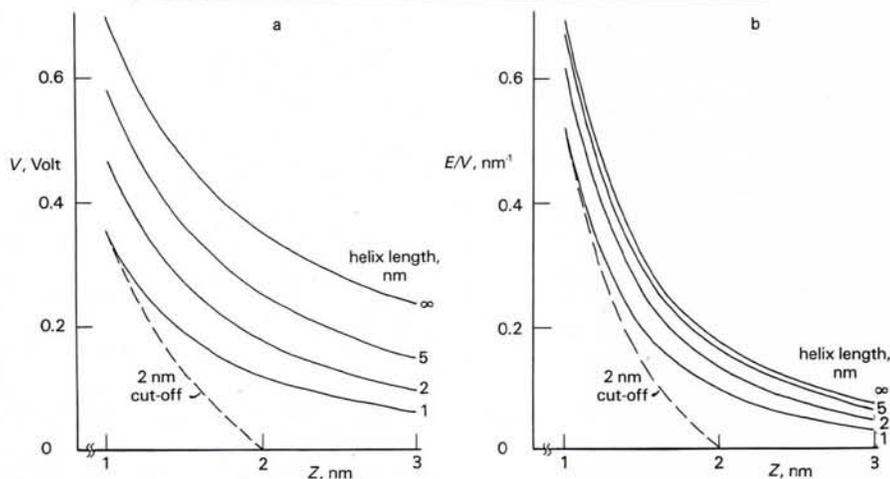


Fig. 3 — Potential (a) and field (b) on the axis on an α -helix of various lengths, as a function of the distance Z from the helix terminal. Dashed curves: the same if computed with a 2 nm cut-off radius; the dashed curves are valid for any helix longer than 1 nm. Dielectric constant was taken as 1 [from Ref. 2].

proton binds more favourably to the oxygen or sulphur than to the histidine. But the helix field is directed in such a way that the transfer becomes much facilitated. Elaborate quantum mechanical calculations have shown this effect for the enzyme papain. The field appears to originate essentially from the dipole moments of the α -helix.

How to Compute Long-range Fields?

Although the evaluation of electric fields resulting from a known charge distribution is a standard problem in electrostatics, its efficient computation is not a trivial matter²). The size of the problem is appreciable: a typical protein contains 2000-3000 atoms, most of which carry partial or full charges. In addition some 5000 water molecules, with three charges each, must be incorporated. In MD simulations the time step used to integrate the equations of motion is 1-2 fs, and a typical simulation requires 50000 steps to obtain statistically valid data. The correct electric field is required at each of these time steps, and its computation must be carried out in times well below one second (on a supercomputer) in order not to dominate the total computational effort.

Fortunately, the long-range fields fluctuate on a much slower time scale than the short-range interactions. This makes it possible to compute these fields only at certain intervals in a simulation, and limit the fast components and corrections to short ranges. Computationally such long-range updates can be efficiently combined with the construction of neighbour lists needed for the short-range interactions. Thus the adverse effects of short cut-off radii are avoided.

The straightforward summation of Coulomb terms is not correct for a macromolecule in an infinite solvent, even if part of the solvent molecules are treated explicitly. This is due to the reaction field produced by the external dielectric (and sometimes also conducting) medium. For the simple geometry of a sphere (with internal dielectric constant ϵ_1 and radius a), the reaction field due to a medium of high dielectric constant ϵ_2 can be calculated using image charges outside the sphere: each internal charge q at radius r has an image q_{im} at a radius r_{im} , where

$$r_{im} = a^2/r, \quad \text{and} \\ q_{im} = -(\epsilon_2 - \epsilon_1) / (\epsilon_2 + \epsilon_1) (a/r)q$$

Although the image charge method is very useful to get estimates of the magnitude of reaction fields, for complicated shapes it does not work, and Poisson's equation has to be solved for the given source terms and boundary conditions. In the most common case of an

odd-shaped macromolecule embedded in an electrolyte, the potential in the external region obeys the Poisson-Boltzmann equation

$$\nabla^2 \psi(\mathbf{r}) = \kappa^2 \psi(\mathbf{r})$$

where κ is the inverse Debye length:

$$\kappa^2 = 2IF^2/\epsilon_2 RT$$

(I is the ionic strength of the medium in mol/m³, F is the Faraday constant (96484 C) and ϵ_2 is the dielectric constant of the medium including the permittivity of vacuum.) The boundary conditions are the usual continuity of potential at the surface and dielectric displacement normal to the surface. Classical methods to solve such equations numerically can be classified as series expansions (e.g. in Legendre polynomials), Fourier methods, finite element methods, and finite difference methods. The latter have been employed to calculate fields in proteins; unfortunately they are quite time consuming. Fourier methods have been worked out for periodic homogeneous cases, but not for the problem presented here. New techniques, using Green's functions, have not yet been applied to realistic cases, but seem quite promising because they reduce a three-dimensional problem to a two-dimensional one. Briefly, the method produces a kind of surface charge density that generates the proper reaction field.

Thus far really efficient and correct methods have not been worked out. *Ad hoc* but incorrect solutions, like using a distance-dependent dielectric constant and/or a position-dependent charge reduction, have found widespread use. A much more reliable, but also time consuming method is the use of a large array of Langevin dipoles to mimic water molecules³). But the general solution awaits some clever ideas!

Future Biomolecular Simulations

Molecular dynamics simulations of biomolecules have come very close to applications that can have far-reaching consequences, both for the understanding of biochemical processes on an atomic scale, and for practical and industrial applications. Let us consider an example in the field of drug design. Certain drugs inhibit the enzyme *dihydrofolate reductase* (DHFR) which is essential for cell reproduction, by strongly binding to its active site and thus preventing the normal substrate from entering into a reaction. Inhibitors to the bacterial DHFR are antibiotics; inhibitors to human DHFR are cytostatic drugs used in the treatment of cancer. It is of great interest to find very effective and very specific inhibitors. The normal design

process of such drugs involves a great deal of chemical insight, reasoning by analogy, inspection of structures by graphic displays, and trial and error through the expensive processes of chemical synthesis and functional tests. If reliable calculations could be done to predict the binding properties of possible inhibitors, the design process could be greatly rationalized, providing effective short-cuts to new products. Another example is the prediction of the properties of a *modified* protein. If enzymes can be modified to yield catalysts for new products, or modified to become resistant to higher temperatures or different solvents, a whole new range of industrial products and processes can be envisioned. Genetic engineering techniques are now well advanced to the point where a theoretically designed protein can actually be made and produced in quantity. The next decades will undoubtedly show an unprecedented evolution in this field of biotechnology, aided by computational physics.

The problems are grave, however. Binding properties, for example, require the calculation of *free energies*, not structures or energies. Methods to calculate free energy are now available, but they involve integration over a reversible path from a reference state to the unknown state⁴). For a hydrated macromolecule with rather slow internal structural rearrangements, the required computation is enormously time consuming: hundreds of hours on a present-day supercomputer indicates the order of magnitude. Full scale applications will await the next generation of supercomputers, or — and this is more likely in a European context — the construction of very cost-effective special purpose computational machines for macromolecular simulations. Thus there is a task for computational physicists, information scientists and biophysicists: *design the instrumentation and effective computational methods for reliable and fast simulation of biomolecular processes*. A challenging and rewarding field!

LITERATURE

- Hol W.G.J., van Duijnen P.Th. and Berendsen H.J.C., *Nature* **273** (1978) 443; Hol W.G.J., *Progr. Biophys. Molec. Biol.* **45** (1985) 149.
- Berendsen H.J.C., in *Molecular Dynamics and Protein Structure*, ed. J. Hermans (Polycrystal Book Service, Western Springs, Ill) 1985, p. 18 and references quoted therein.
- Warshel A. and Russel S.T., *Quart. Revs Biophys.* **17** (1984) 283.
- Berendsen H.J.C., Postma J.P.M. and van Gunsteren W.F., in *Molecular Dynamics and Protein Structure*, ed. J. Hermans, p. 43 (see ref. 2).