Journal of Mathematical Chemistry 15 (1994) 77–92

A hierarchical approach to force field calculations through spline approximations

M. Butzlaff [a], W. Dahmen [b], S. Diekmann [a], A. Dress [c], E. Schmitt [a] and E. von Kitzing [d]

[a] *Max-Planck-Institut für Biophysikalische Chemie, Postfach 2841, D-37083 Göttingen, Germany*

[b] *Lehrstuhl für Mathematik und Institut für Geometrie und Praktische Mathematik, RWTH Aachen, Templergraben 55, 5100 Aachen, Germany*

[c] *Fakultät für Mathematik, Universität Bielefeld, Postfach 8640, 4800 Bielefeld, Germany*

[d] *Max-Planck-Institut für Medizinische Forschung, Postfach 103820, 6900 Heidelberg, Germany*

Received 22 October 1992

The objective of this paper is to outline a new approach to analyzing the geometry of macro-molecules and investigating important physical properties by means of simulations. The classical method of force field calculations requires minimizing the energy as a function of the Cartesian coordinates of all atoms. Due to the large number of variables this method is limited to relatively small molecules. We describe an approach to overcome this difficulty. On the one hand, the number of free variables is effectively reduced by assembling certain groups of atoms into configurational structures with considerably less degrees of freedom. In this way we build up a whole hierarchy of coordinate spaces with decreasing dimensions. On the other hand, approximations to the energy function with respect to these variables are constructed using methods from the theory of splines and radial basis functions. The hierarchical features of wavelet decompositions are utilized to exploit the physical importance of the different force field constants on the biological function of the macro-molecule.

1. Introduction

In recent years, consistent force field calculations have proved to be a powerful means to compute the 3D-structure of biological macro-molecules. The structure of a molecule in space is determined by the interaction of its n atoms having the Cartesian coordinates $x_i \in \mathbb{R}^3$, $1 \leqslant i \leqslant n$. Interaction energies are described by analytical functions derived theoretically from quantum mechanics and empirically from experiments. The energy functions depend on distances between atoms as well as on bond and torsion angles which are computed from the coordinates x_i. For the

total internal energy E_{total} of the molecule typical force fields are given by (see e.g. [1–3])

$$E = E_{\text{total}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}} + E_{\text{NB}}, \tag{1}$$

where for example

$$E_{\text{bond}} = \sum_{(i,j) \in I_{\text{bond}}} K_{Bij}(\|\boldsymbol{x}_i - \boldsymbol{x}_j\| - R_{0ij})^2$$

denotes the energy contribution of all pairs of bonded atoms and

$$E_{\text{torsion}} = \sum_{(i,j,k,l) \in I_{\text{torsion}}} \left(E_{\text{NB1-4}} + \sum_{\nu \in J_{ijkl}} K_{\phi ijkl\nu}(1 + \cos(\nu\phi_{ijkl} + \delta_{\nu ijkl})) \right)$$

denotes the energy contributions arising from torsion angles ϕ_{ijkl} formed by all sequences of four consecutively bonded atoms i, j, k, and l. Here $\delta_{\nu ijkl}$ is a phase factor and $K_{\phi ijkl\nu}$ is a force constant as defined by the force field. All the constants depend on the atoms involved as well as on their structural neighbours. For the energies of bonding angles between three atoms connected by two bonds similar formulas hold. The non-bonding energie E_{NB} including hydrogen bonds, van der Waals forces and the long range Coulomb forces is a sum over all pairs of non-bonded atoms. It should also be mentioned that especially for E_{NB} there exist different approximative descriptions [3,4]. In specific examples we always refer to the implementation of consistent force field calculations as done in the AMBER program package [3]. Although the approach described below is very general in nature we will refer mainly to DNA, as a first specific example.

A stable three dimensional structure of the molecule corresponds to a minimum of the total internal energy E_{total}. In consistent force field calculations, the minimum of the total internal energy is usually computed iteratively by means of gradient methods. Describing the location of atoms by their Cartesian coordinates requires minimizing functions of $3n$ variables:

$$E_{\text{total}} : \mathbb{R}^{3n} \to \mathbb{R},$$

but, of course, not all configurations are feasible. Within the minimization step the function and its partial derivatives have to be evaluated and a new improved estimate for the actual configuration with a lower total internal energy is calculated.

A major drawback of the whole method is the high dimension (approximately 100–10 000) of the space of free variables. It is not surprising that this leads to severe problems in the minimization process as observed e.g. in [2]. However, as we will point out, the large number of degrees of freedom can be reduced significantly.

The main goal of this paper is to outline new theoretical and practical methods which facilitate computing the structure of biological macro-molecules of realistic

size in three dimensional space by means of energy minimization. The motion of atoms in biological macro-molecules cover a wide range of time scales. The fastest vibrations consist in the motion of pairs of chemically bonded atoms oscillating around their average distance. The periods of these vibrations are in the order of 10^{-15} seconds. This type of motion is highly localized and largely independent of the global conformation of the molecule. On the other hand, the folding of proteins range in the time scale of 10^{-3}–10^{2} seconds or even longer. The biological function is mostly related to the slower type of motions covering the time scales from 10^{-9} seconds and above, where always many atoms are involved. For any conformation $\mathbf{R} \in R^{3n}$, one can identify directions related to fast or slow motions thus decomposing the space of motion, the tangent space $\mathcal{T}$ to $\mathbf{R}$, into "fast" ($\mathcal{T}_F$) and "slow" ($\mathcal{T}_S$) subspaces:

$$\mathcal{T} = \mathcal{T}_F \oplus \mathcal{T}_S .$$

The method proposed in this paper takes advantage of the fact that for large molecules the dimension of $\mathcal{T}_S$ is considerably smaller compared to the dimension of $\mathcal{T}_F$. It is designed to reduce the number of degrees of freedom by taking into account only those which are essential for the biological function of the molecule. As a first step the atoms of the molecule are grouped into relatively rigid substructures. In the case of DNA, such substructures may consist of the bases, the ribose and the phosphate atoms (see fig. 3B). These substructures can again be grouped to define a hierarchy of substructures. This allows us to adapt the accuracy of each part of the computation to the particular need of the problem under consideration.

In summary, our approach facilitates calculating the 3D-structure of large, biologically relevant macro-molecules. This should help to understand their interactions and biological functions.

2. Reduction of the degrees of freedom

Before describing the general principle of reducing the number of the degrees of freedom we mention two cases where hierarchical principles were successfully applied for the calculation of 3D-structures.

Example 1

First trials to calculate the structure of a 40 base pair DNA molecule resulted in a straight helix axis although experiments strongly suggested that the helix axis is curved [5]. Inspection of the resulting structure revealed significant local perturbations. In spite of optimizing all degrees of freedom of this large molecule only small deviations from the initial structure were observed. So it seems that the method could not escape from a local minimum. So instead, a hierarchical procedure was applied in [6]. The structures of smaller parts of the total sequence were calculated

separately, followed by optimizing the transitions from one part to the other. The global structure was constructed from these parts and used as the starting configuration for the optimization of all parameters. Now the resulting global structure exhibited strong curvature of the helix axis which is in agreement with the experimental results [7,8]. Further theoretical predictions deduced from this structure were verified experimentally [9–11] (see fig. 1).

Example 2

Holliday junctions are transition structures during recombination of DNA double strands. In this structure individual DNA single strands are exchanged between different double strands. The three dimensional structure depends on environmen-

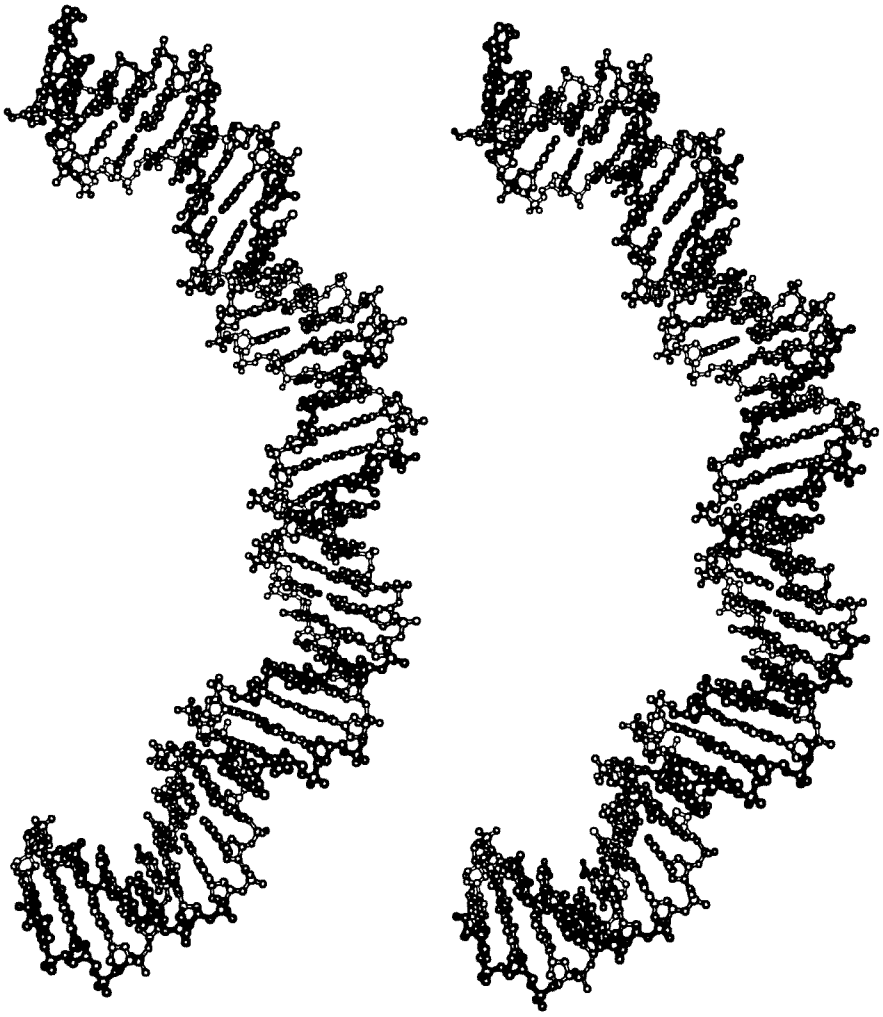

Fig. 1. The optimized structure of the DNA sequence d[$(GCTCGAAAAA)_4 \cdot (TTTTTCGAGC)_4$] in stereo [6].

tal conditions. To compute this configuration, the structure of the double helices of the four arms was kept fixed, however, their relative orientation was optimized with respect to variable torsion angles in the phosphate chains connecting the helix arms. The optimal configuration was computed for two different salt concentrations. For low salt concentration the helix arms form a planar quadratic structure with an open centre at the crossing point, whereas for high salt concentration the base pairs at the crossing point stack on one another thereby forming an X-like structure with two continuous double helices stacking through the crossing point [12] (see fig. 2). The corresponding DNA sequences were synthesized and the structural predictions deduced from the theoretical model could be accurately verified experimentally [13–15]. The DNA 4-way junction structure is not in agreement with that presented in textbooks and suggests revising the current model of recombination.

To reduce the number of free variables in general we assemble certain groups of atoms into one configurational structure which can be described by fewer variables than the number of all atom coordinates involved. For instance, a rigid substructure of several atoms can be characterized by the Cartesian coordinates of one atom and three additional variables defining the relative position of the whole rigid structure in space within a global coordinate frame. The parameter space for a rigid subunit is $\mathbb{R}^3 \times SO(3)$, the three Euler angles being a very convenient parametrization for the compact manifold SO(3).

In the case of DNA the sequential structure [16] permits a division into three types of natural substructures: ribose, base and phosphate (see fig. 3). If, for example, guanine is described as a rigid subunit, the number of free parameters is reduced from 33 to 6. Experimental data and results of theoretical computations

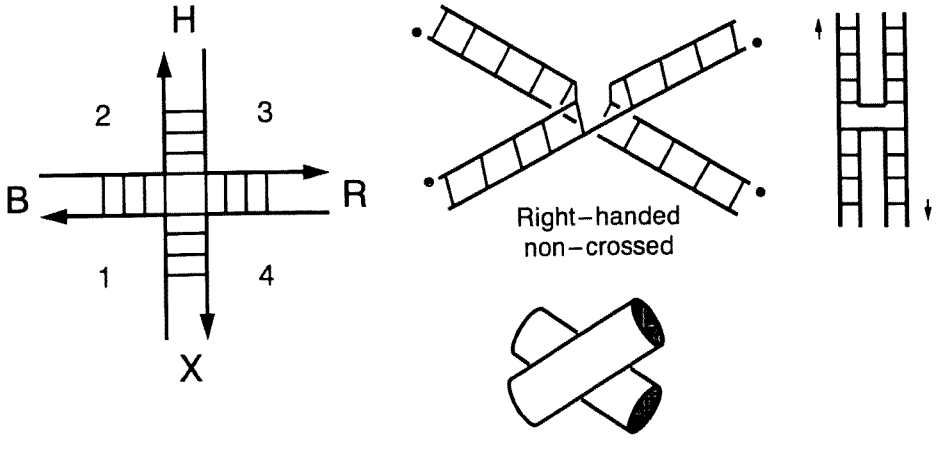

Fig. 2(A). Schematic drawing of the open (left) and folded (right) structure of a DNA 4-way junction. The letters H, R, X, and B denote the DNA arms. The arrows indicate the strand direction. The 4-way junction is a right-handed non- crossed structure with antiparallel strands

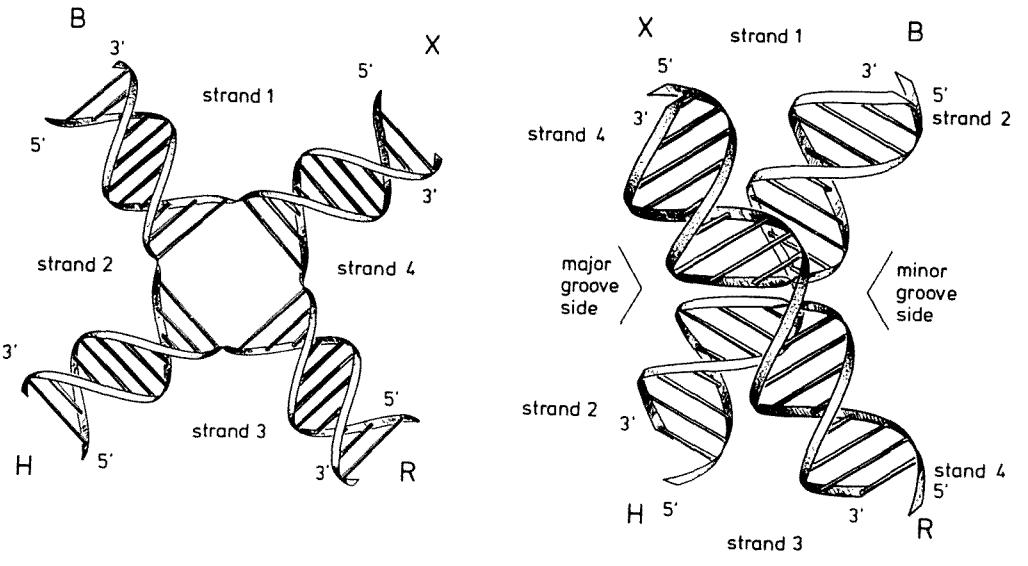

Fig. 2(B). Helical drawing of the open (left) and folded (right) structure of a DNA 4-way junction. The strands are numbered 1 to 4, their direction indicated by the chemical nature of their ends. The arms are labelled with the letters B, X, R, and H. In both structures the major grooves lie on one, the minor grooves on the other side of the structure offering different surfaces to proteins.

show that, at least in a first approximation, the four different bases can be viewed as rigid. The same is true for the phosphate group including the two neighbouring O-atoms in the strand of the DNA molecule. All rigid subunits are characterized by six variables from $\mathbb{R}^3 \times \mathrm{SO}(3)$.

In contrast, one cannot regard the ribose group as a rigid substructure [17]. It is known that the configuration of the ribose part can be described as a function of a periodic variable, the so-called pseudo-rotation angle $p \in \mathrm{S}^1$ (cf. e.g. [18–22]). So the state of the structure is parametrized by $\mathbb{R}^3 \times \mathrm{SO}(3) \times \mathrm{S}^1$, a manifold of dimension 7. Algorithms to convert the coordinates of DNA subunits mentioned above into atom coordinates and vice versa have been implemented in FORTRAN [23].

The parameters of a subunit determine the coordinates of all its atoms. If we denote the space of all parameters of all subunits by U, then we have a map

$$\rho : U \to \mathbb{R}^{3n},$$

which assigns to every point of the parameter space U the sequence of the coordinates of all n atoms of the configuration. For instance, for DNA molecules with N bases, U has the form

$$U = V_1 \times \ldots \times V_{3N}, \tag{2}$$

where V_{3k}, V_{3k+1}, and V_{3k+2} denote the parameter spaces of the three parts (phosphate, ribose, and base respectively) of base k so that

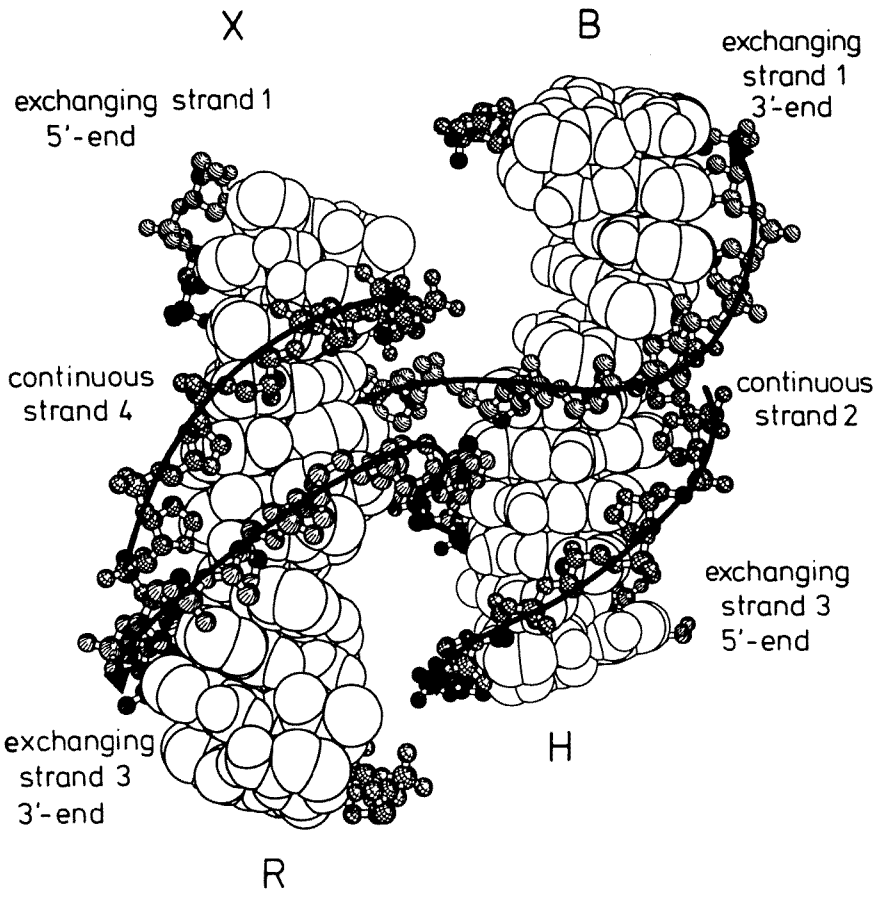

Fig. 2(C). Model of a DNA 4-way junction, bases with full van der Waals radius, sugar-phosphate backbones as ball- stick model. The strands are numbered 1 to 4, their direction indicated by the chemical nature of their ends. The arms are labelled by B, X, R, and H.

$$V_{3k} = \mathbb{R}^3 \times \mathrm{SO}(3) = V_{3k+2}, \quad V_{3k+1} = \mathbb{R}^3 \times \mathrm{SO}(3) \times S^1 .$$

Thus ρ itself is a Cartesian product of mappings $\rho^{(i)}$ acting on the respective components V_i. As the above example of guanine indicates the image $\rho(U)$ is a subset of $\mathbb{R}^{3n}$ of significantly lower dimension which, in effect, is a proper reduction of degrees of freedom (see [23]). This will be important for the evaluation of $E(\rho(u))$.

The process of introducing spaces of coordinates can be iterated to establish a whole hierarchy of coordinate sets U_i satisfying

$$\rho_m(U_m) \subset \rho_{m-1}(U_{m-1}) \subset \ldots \subset \rho_0(U_0) = U_0 = \mathbb{R}^{3n}$$

and $\dim(U_i) < \dim(U_{i-1})$ for $1 \leqslant i \leqslant m$.

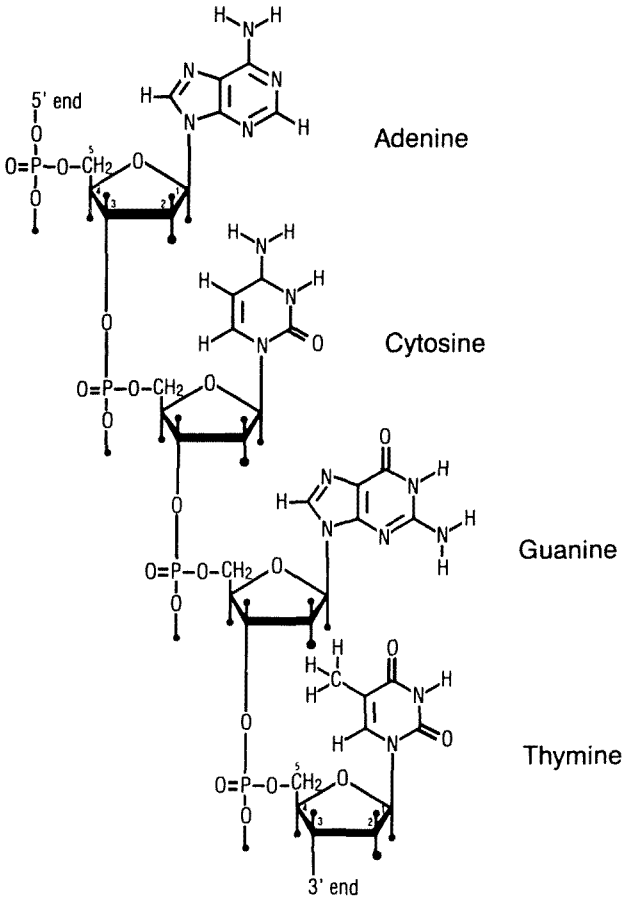

Fig. 3(A). Chemical presentation of a DNA strand of the base sequence dACGT.

In the case of DNA molecules the next level of such a hierarchy could be defined by joining related phosphate, ribose, and base groups to form a new substructure, for example a base pair.

3. Approximating energy functions

We now consider one particular coordinate space U representing N groups of atoms and hence having the form (2). In (1) the total internal energy E was defined as a sum of functions depending on the cartesian coordinates of two, three, or four atoms [1]. As all the atoms belong to exactly one of the N groups, the terms in the expression (1) for E can be rearranged so that E has the form

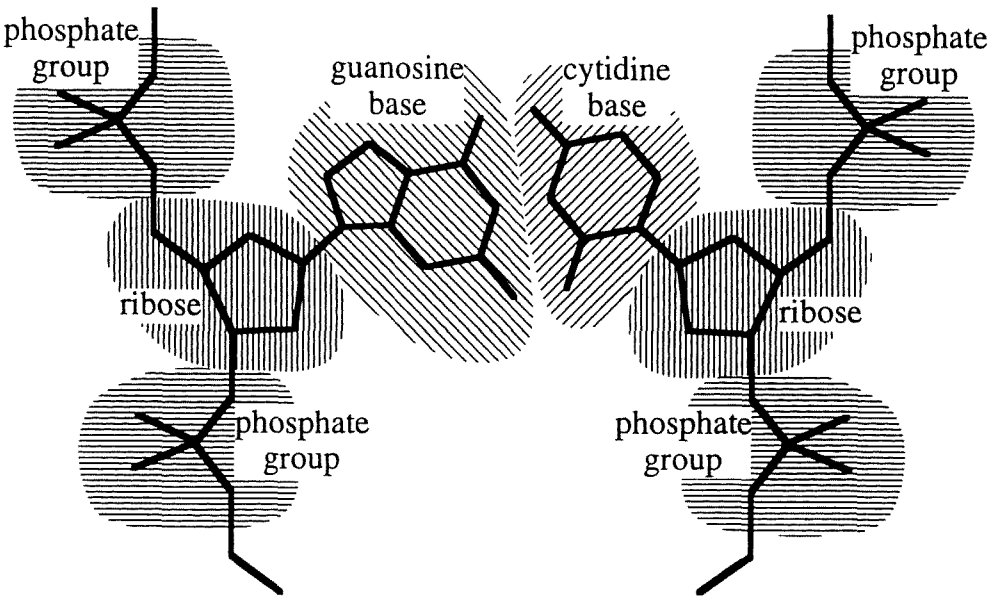

Fig. 3(B). Division of the DNA structure into 3 types of substructures: phosphate group, ribose, and base (here: guanosine and cytidine). Most interactions in the force field (1) concern only pairs of atoms. But there are also interactions involving three and four atoms. The three atom interactions are found between atoms (i,j,k), if atoms (i,j) and (j,k) are chemically bonded; the four atom interaction (i,j,k,l) requires bonds between pairs (i,j), (j,k), and (k,l). The substructures are defined in such a way that despite the four center interactions at the atomic level, only interactions between pairs of substructures have to be taken into account.

$$E = \sum_{i \in I_1} f_i(v_i) + \sum_{(i,j) \in I_2} f_{(i,j)}(v_i, v_j) + \sum_{(i,j,k) \in I_3} f_{(i,j,k)}(v_i, v_j, v_k) + \sum_{(i,j,k,l) \in I_4} f_{(i,j,k,l)}(v_i, v_j, v_k, v_l), \quad v_i \in V_i, \tag{3}$$

where I_s denotes the set of all s-tuples of distinct groups which contribute to one of the terms in the energy function, and we write briefly i instead of (i) when $s = 1$. Here, for any tuple $t \in I_s, f_t$ is the corresponding energy distribution. More precisely, f_i is the internal energy of group i whereas $f_{i,j}$ contains all energy contributions of 2-, 3-, or 4-tuples of atoms with atoms belonging to groups i and j, similarly for $f_{i,j,k}$ and $f_{i,j,k,l}$. We should remark at this point that most force field approximations consider only energy contributions depending on up to four atoms, so that there are no terms in our energy expansion (3) depending on more than four groups.

In the case of DNA molecules the groups in U_1 are defined such that all energies arising from distance or angle interactions can be expressed by the relative coordinates of only pairs of groups (see fig. 3B). So I_3 and I_4 are in fact empty. We will

restrict our further discussion to this case, although the generalization is straightforward.

Expression (3) for E can be evaluated at any point of the coordinate space U, using formula (1) and the map ρ. Recall that our ultimate central objective is to combine a reduction of degrees of freedom with an efficient evaluation of the energy function E. While the reduction is taken care of by the application of ρ, our approach to speeding up the evaluation is based on approximating the components $f_t, t \in I_s$, in (3) by suitable functions which, on the one hand can be evaluated at low cost and whose representation on the other hand exploits invariances of the energy components f_t. Moreover, the special type of approximating function has to be selected so as to minimize the effort of optimization.

As an example, we mention the following facts used in our implementation [24]. As the position of subunits within the global coordinates is described by a point in $\mathbb{R}^3 \times \mathrm{SO}(3)$, the relative position of two groups is again a parameter from $\mathbb{R}^3 \times \mathrm{SO}(3)$. Thus we can regard f_{ij} as depending rather on $V_i \times V_j$ modulo the space of Eucledian motions than on $V_i \times V_j$ itself. For rigid subunits this space is $\mathbb{R}^3 \times \mathrm{SO}(3)$. For ribose groups of course we have an additional parameter in S^1. In the case of two interacting base groups, for example, we have a parameter space of dimension six, while two interacting ribose groups define a parameter space of dimension eight. Altogether, we have 28 bonding or non-bonding interacting pairs of groups with energies parametrized by six to eight parameters (see table 1). The ribose group is the only group with non-constant inner energy which depends on the pseudo-rotation angle in S^1.

The energy between two groups has to be expressed in these new variables. This can be done on a relatively fine grid in corresponding parameter spaces of six to eight dimensions by computing the usual Cartesian coordinates of all atoms using ρ and then evaluating the analytical expression (1) for the total energy of the two groups under consideration. As a result, the exact value of the total internal energy is known at the grid points. To evaluate the energy at arbitrary points of the para-

Table 1
Number of different pairs of interacting groups and dimension of parameter spaces describing the state space of interaction. *C, G, A, T* denote bases, *R* ribose, and *P* phosphate. Note that *P* and *R* are bounded in two different ways in the strand of a DNA molecule.

Interacting groups	Pairs	Manifold	Dim
Non-bonding:			
C, G, A, T, P	15	$\mathbb{R}^3 \times \mathrm{SO}(3)$	6
C, G, A, T, P with R	5	$\mathbb{R}^3 \times \mathrm{SO}(3) \times \mathrm{S}^1$	7
R with R	1	$\mathbb{R}^3 \times \mathrm{SO}(3) \times \mathrm{S}^1 \times \mathrm{S}^1$	8
Bonding:			
C, G, A, T with R	4	$\mathbb{R}^3 \times \mathrm{SO}(3) \times \mathrm{S}^1$	7
P with R	2	$\mathbb{R}^3 \times \mathrm{SO}(3) \times \mathrm{S}^1$	7
R (internal energy)	1	S^1	1

meter space, an approximating or interpolating function to these data has to be computed. This can be done on the whole grid or on parts of it. Due to the structure of the energy function, several choices of function types are possible. The actual choice will be determined by approximation and stability properties as well as numerical efficiency in the minimization process.

Candidates for such approximations are linear combinations of radial basis functions (see e.g. [25]). Each basis function is only distance dependent and thus has radial symmetries. Such linear combinations reproduce well some structural properties of the energy functions, especially for non-bonding interactions and larger distances. However, there are two principal difficulties. Since these functions are global, the cost of evaluation depends on the number of basis functions. While in the bivariate case multipole expansions seem to lead to more efficient evaluation schemes, the higher dimensional case which is relevant in the present context, is less understood. Secondly, a principal difficulty lies in interrelating rather different physical scales. On the one hand, local effects on atom level have to be resolved while, on the other hand, macroscopic effects should not be neglected. It should therefore be important to choose mathematical representations that reflect such hierarchies of scales in an appropriate fashion. In this regard, spline functions and wavelets are more promising than radial basis functions. Due to their local structure the costs of evaluating the approximating function at a point remains independent of the discretization level. Moreover, efficient and stable subdivision techniques facilitate local refinements. This provides a most natural framework for employing adaptive methods which are ultimately indispensible for handling problems of interesting size.

The theory of spline functions is very well developed for regular grids on Euclidean spaces of arbitrary dimensions and also for some types of parameter spaces with periodic variables. Both types of variables can be combined by tensor product constructions. Nevertheless, appropriate multiresolution setups and wavelet expansions relative to such parameter manifolds have yet to be constructed.

The following remarks are based on our numerical experience with energy functions defined on the manifolds mentioned above on the first level of the hierarchy of coordinates [26]. In order to combine the principal advantages of both types of approximations one could consider functions of the form

$$f_t(x,\phi) = \sigma(|x|)\cdot R(x) + (1-\sigma(|x|))\cdot w(x,\phi), \quad (x,\phi)\in\mathbb{R}^3\times \mathrm{SO}(3)\,.$$

Here $\sigma : \mathbb{R}_{\geqslant 0} \to [0,1]$ is some suitable sigmoidal function satisfying $\sigma(r)\to 1$ for $r\to\infty$ and $\sigma(r)\to 0$ for $r\to 0$. R should be a linear combination of just a few radial basis functions, while $w(x,\phi)$ should be an expansion of wavelet type basis functions (see e.g. [27–29]) which, in view of the form of σ, have to be determined only on essentially bounded domains. The advantages of wavelet representations lie in the following facts. The coefficients in wavelet expansions reflect the behaviour of the function relative to different scales. They are extremely suitable for adaptive data compression [28] depending on the required information. Finally, one expects

e.g. from [30] that such expansions are particularly well suited for preconditioning gradient methods for optimization purposes.

We find the possibility of combining the hierarchical organization of substructures described above with the hierarchical features of wavelet decompositions particularly intriguing and promising.

4. Splines on compact manifolds

The energy distributions $f_{(i,j)}$ are functions of variables in $\mathbb{R}^3$, SO(3), and S^1. Since spline functions on products of manifolds can easily be formed by taking tensor products, it is important to have suitable spline functions on each individual manifold. For $\mathbb{R}^3$ as well as for S^1 the theory is well developed. In fact, S^1 was for a long time the only compact manifold where explicit descriptions of splines were known. It was only in 1991 when Schumaker and Traas [31] gave a satisfactory description of spline functions on S^2, the simplest 2-dimensional compact manifold. We give a short description of their method as it can be utilized as a guideline to constructing splines on compact manifolds of similar type.

The splines on S^2 are obtained as tensor products of polynomial and trigonometric splines. In this case the sphere S^2 can be covered by only one chart

$$V := \{(\theta, \phi) : -\pi/2 \leqslant \theta \leqslant \pi/2 \quad \text{and} \quad 0 \leqslant \phi \leqslant 2\pi\}$$

with $\theta = -\pi/2$ corresponding to the south pole S of the sphere and $\theta = \pi/2$ to the north pole N. A differentiable function f on S^2 and hence on V needs to fulfill the periodicity conditions

$$f(\theta, 0) = f(\theta, 2\pi), \quad -\pi/2 \leqslant \theta \leqslant \pi/2,$$

$$f(-\pi/2, \phi) = f_S \quad \text{and } f(\pi/2, \phi) = f_N, \quad \text{for all } \phi,$$

and it is known that the following conditions on the partial derivatives of f

$$f_\phi(\theta, 0) = f_\phi(\theta, 2\pi), \quad -\pi/2 \leqslant \theta \leqslant \pi/2,$$

$$f_\theta(-\pi/2, \phi) = A_S \cos(\phi) + B_S \sin(\phi), \quad 0 \leqslant \phi \leqslant 2\pi,$$

$$f_\theta(\pi/2, \phi) = A_N \cos(\phi) + B_N \sin(\phi), \quad 0 \leqslant \phi \leqslant 2\pi,$$

are equivalent to $f \in C^1(V)$. Here, A_S, A_N, B_S, and B_N denote suitable constants.

These conditions suggest to consider spline functions of the form

$$f(\theta, \phi) := \sum_{i=1}^{M} \sum_{j=1}^{N} c_{ij} B_i(\theta) T_j(\phi), \tag{4}$$

where $B_i(\theta) = N_i^m(\theta), i = 1, \ldots, M$, are the usual normalized polynomial B-splines of order m associated with a suitable knot sequence on the θ-axis. To fulfill the conditions on f_θ at the poles it is convenient to select the functions $T_1, \ldots, T_N$ to be peri-

odic trigonometric B-splines of odd order $n = 2q + 1$. One can then show that for $m = n = 3$ the conditions for the tensor-product function f to be C^1 have the form

$$Bc = 0 \tag{5}$$

with a $(4N + 2M - 6) \times M(N - 2)$ matrix B (see [31]).

The explicit representation of spline functions on S^2 can be obtained e.g. through interpolation or least squares approximation which leads to a large sparse system of equations which has to be solved subject to the condition (5).

Once we have determined a representation of the form (4) we wish to compress it in order to facilitate efficient subsequent evaluations during the optimization process. By this we mean that we seek for a sufficiently good approximation to f which involves as few as possible coefficients relative to a suitably chosen basis [28]. Here, "suitable" means that the coefficients in such a representation should reflect the structure of f relative to different scales. Thus wavelet type bases suggest themselves as candidates.

Wavelet bases for polynomial spline functions with equidistant knots are well-known by now [29]. The idea is to view the spline space

$$\mathcal{S}_n := \operatorname{span}\{N^m(2^n \cdot - j) : j \in \mathbb{Z}\}$$

as the result of consecutive refinements of the spaces $\mathcal{S}_l$, $l \leqslant n$. Denoting by $\mathcal{W}_l$ the orthogonal complement of $\mathcal{S}_l$ in $\mathcal{S}_{l+1}$,

$$\mathcal{W}_l := S_{l+1} \ominus \mathcal{S}_l ,$$

one has

$$\mathcal{S}_n = \mathcal{S}_0 \oplus \bigoplus_{l=0}^{n-1} \mathcal{W}_l .$$

The objective is then to construct a spline $\psi \in \mathcal{S}_1$ of possibly small support such that

$$\mathcal{W}_l = \operatorname{span}\{\psi(2^l \cdot - j) : j \in \mathbb{Z}\} .$$

ψ is called wavelet. Thus, in a wavelet expansion of f, the coefficients corresponding to high values of l reflect those contributions of f which can be only seen on a fine grid.

If one could handle the trigonometric case in a similar fashion, a tensor product wavelet expansion for (4) would be readily available. However, for trigonometric B-splines the situation is slightly more complicated. Nevertheless, wavelet type bases can be constructed using information from the theory of so-called E-splines which have been thoroughly studied (cf. [32]).

In the relevant case of SO(3) it is natural to use Euler angles (e.g. [33]) for the purpose of parametrization. We prefer to use the following map τ on the 3-dimensional cube $C := [0, 2\pi] \times [0, 2\pi] \times [-\pi/2, \pi/2]$

$$\tau : C \ni (\phi, \psi, \theta) \mapsto M(\phi, \psi, \theta) \in \mathrm{SO}(3),$$

where $M(\phi, \psi, \theta)$ is defined by

$$\begin{pmatrix} \cos\psi\cos\theta & -\cos\phi\sin\psi - \sin\phi\cos\psi\sin\theta & \sin\phi\sin\psi - \cos\phi\cos\psi\sin\theta \\ \sin\phi\cos\theta & \cos\phi\cos\psi - \sin\phi\sin\psi\sin\theta & -\sin\phi\cos\psi - \cos\phi\sin\psi\sin\theta \\ \sin\theta & \sin\phi\cos\theta & \cos\phi\cos\theta \end{pmatrix}.$$

The identity in SO(3) then has a regular neighbourhood and those rotations describing the usual relative positions in the DNA double helix have well defined pre-images. In fact, in this way one can cover the whole manifold by one coordinate chart. However, the compatibility conditions for C^1 functions with respect to this parametrization turn out to be significantly more complicated. So it is obviously necessary for a continuously differentiable function $f : C \to \mathrm{SO}(3) \to \mathbb{R}$ to fulfill the following periodicity conditions:

$$\begin{aligned} f(0, \psi, \theta) &= f(2\pi, \psi, \theta), && 0 \leqslant \psi \leqslant 2\pi \quad \text{and} \quad -\pi/2 \leqslant \theta \leqslant \pi/2, \\ f(\phi, 0, \theta) &= f(\phi, 2\pi, \theta), && 0 \leqslant \phi \leqslant 2\pi \quad \text{and} \quad -\pi/2 \leqslant \theta \leqslant \pi/2, \\ f_\phi(0, \psi, \theta) &= f_\phi(2\pi, \psi, \theta), && 0 \leqslant \psi \leqslant 2\pi \quad \text{and} \quad -\pi/2 \leqslant \theta \leqslant \pi/2, \\ f_\psi(\phi, 0, \theta) &= f_\psi(\phi, 2\pi, \theta), && 0 \leqslant \phi \leqslant 2\pi \quad \text{and} \quad -\pi/2 \leqslant \theta \leqslant \pi/2. \end{aligned} \tag{6}$$

Conditions on the boundary faces $C_\pm := \{(\phi, \psi, \pm\pi/2) \in C\}$ have a more complicated structure due to the singular behaviour of τ on $C_\pm$. More precisely, one can prove the following result [34]:

THEOREM 4.1

Let $F : \mathrm{SO}(3) \to \mathbb{R}$ be a function and define f by $f := F \circ \tau : C \to \mathbb{R}$. Then F is C^1 if and only if the following conditions hold:

(1) f fulfills the periodicity conditions (6).
(2) there are realvalued 2π-periodic C^1 functions $F_\pm$ as well as 2π-periodic continuous functions $A_\pm, B_\pm$ such that

$$f(\phi, \psi, \pm\pi/2) = F_\pm(\phi \pm \psi)$$

and

$$f_\theta(\phi, \psi, \pm\pi/2) = A_\pm(\phi \pm \psi)\cos\phi + B_\pm(\phi \pm \psi)\sin\phi$$

for $0 \leqslant \phi \leqslant 2\pi$ and $0 \leqslant \psi \leqslant 2\pi$.

The construction of corresponding C^1 wavelets based on this result on SO(3) is currently under investigation.

An alternative possibility is to tolerate the lack of regularity on the 1-dimensional submanifolds $S_\pm := \tau(\{(\phi, \psi, \pm\pi/2)\})$ of SO(3), which are essentially defined by

$$\tau(\phi, \psi, \pm\frac{\pi}{2}) = \begin{pmatrix} 0 & \mp\sin(\phi \pm \psi) & \mp\cos(\phi \pm \psi) \\ 0 & \cos(\phi \pm \psi) & -\sin(\phi \pm \psi) \\ \pm 1 & 0 & 0 \end{pmatrix},$$

and to reparametrize if necessary. More detailed information on these investigations will be given elsewhere [35].

Acknowledgements

We thank the DFG (grant Di 258/6-2) and ZiF (Bielefeld) for support.

References

[1] S.R. Niketic and K. Rasmussen, *The Consistant Force Field*, Lecture Notes in Chemistry 3 (Springer, Berlin, Heidelberg, New York, 1977).
[2] E. von Kitzing, *Molekülsimulation mit Hilfe von Kraftfeldrechnungen am Beispiel der Aggregation von Nukleinsäuren verschiedener Konformation zu einem Komplex mit Übersetzungsfunktion*, Dissertation (Rader Verlag, Aachen, Germany, 1986).
[3] S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr. and P. Weiner, J. Am. Chem. S. 106 (1984) 765.
[4] L. Nilsson and M. Karplus, J. Comput. Chem. 7 (1986) 591.
[5] S. Diekmann and E. von Kitzing, in: *Structure and Expression*, Vol. 3, eds. W.K. Olson, M.H. Sarma, R.H. Sarma and M. Sunderalingam (Adenine Press, Schenectady, NY, 1988) pp. 57–67.
[6] E. von Kitzing and S. Diekmann, Eur. Biophys. J. 15 (1987) 13.
[7] S. Diekmann, Nucl. Acids R. 15 (1987) 247.
[8] S. Diekmann, in: *Nucleic Acids and Molecular Biology*, eds. F. Eckstein and D.M.J. Lilley (Springer, Berlin, 1987) p. 138-156.
[9] S. Diekmann, EMBO J. 6 (1987) 4213.
[10] S. Diekmann and L.W. McLaughlin, J. Mol. Biol. 202 (1988) 823.
[11] S. Diekmann, E. von Kitzing, L. McLaughlin, J. Ott and F. Eckstein, P.N.A.S. US 84 (1987) 8257.
[12] E. von Kitzing, D.M.J. Lilley and S. Diekmann, Nucl. Acids R. 18 (1990) 2671.
[13] D.R. Duckett, A.I.H. Murchie, S. Diekmann, E. von Kitzing, B. Kemper and D.M.J. Lilley, Cell 55 (1988) 79.
[14] A.I.H. Murchie, R.M. Clegg, E. von Kitzing, D.R. Duckett, S. Diekmann and D.M.J. Lilley, Nature 341 (1989) 763.
[15] D.R. Duckett, A.I.H. Murchie, R.M. Clegg, E. von Kitzing, S. Diekmann and D.M.J. Lilley, in: *Structure and Methods*, Vol. 1, eds. R.H. Sarma and M.H. Sarma (Adenine Press, Schenectady, NY, 1990) pp. 157–181.
[16] W. Saenger, *Principles of Nucleic Acid Structure* (Springer, New York, Berlin, Heidelberg, 1984).
[17] C. Altona and M. Sundaralingam, J. Am. Chem. S. 94 (1972) 8205.
[18] H.P.M. de Leeuw, C.A.G. Haasnoot and C. Altona, Isr. J. Chem. 20 (1980) 108.
[19] M. Levitt and A. Warshel, J. Am. Chem. S. 100 (1978) 2607.
[20] D. Pearlman and S.-H. Kim, J. Bio. Struc. Dyn. 3 (1985) 85.

[21] D. Pearlman and S.-H. Kim, J. Bio. Struc. Dyn. 3 (1985) 99.
[22] T. Schlick, C. Peskin, S. Broyde and M. Overton, J. Comput. Chem. 8 (1987) 1199.
[23] M. Butzlaff and E. Schmitt, Coordinates of Subunits of DNA-molecules (incl. FORTRAN-programs), Technical Report 1 (1991).
[24] M. Butzlaff and E. Schmitt, Force field energies between DNA-subunits (incl. FORTRAN-programs), Technical Report 2 (1991).
[25] M.J.D. Powell, in: *Advances in Numerical Analysis*, Vol. II, ed. W. Light (Oxford Science Publications, Clarendon Press, Oxford, 1992) pp. 105–210.
[26] M. Butzlaff and E. Schmitt, Characterization of internal energy components of interacting DNA-subunits, Technical Report 1 (1992).
[27] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics 61 (SIAM Publications, Philadelphia, USA, 1992).
[28] R.A. deVore and B.J. Lucier, Acta Numerica (1992) 1.
[29] C.K. Chui, *An Introduction to Wavelets* (Academic Press, Boston, USA, 1992).
[30] W. Dahmen and A. Kunoth, Numer. Meth. 63 (1992) 315.
[31] L.L. Schumaker and C. Traas, Numer. Math. 60 (1991) 133.
[32] W. Dahmen and C.A. Micchelli, Adv. Math. 76 (1989) 33.
[33] D.M. Soumpasis and C.-S. Tung, J. Bio. Struc. Dyn. 6 (1988) 397.
[34] E. Schmitt, Differentiable functions on SO(3), Preprint (1992).
[35] E. Schmitt, Approximations of differentiable functions on SO(3), Technical Report 2 (1992).