Protonic Quantum Correlations in the H-Bond Dynamics of Nucleic Acids

Part II

Correlations along the Helical Axis of Protein-Coding DNA of Living Organisms

by C. **Aris Chatzidimitriou-Dreismann**[a][b]* and **Daniel Seifert**[a][c]

[a]) I. N. Stranski-Institut für Physikalische und Theoretische Chemie, Technische Universität Berlin, Straße des 17. Juni 112, D-10623 Berlin

[b]) Institute for Basic Research, Castello Principe Pignatelli, I-86075 Monteroduni (IS)

[c]) Institut für Nichtmetallische Werkstoffe, Fachgebiet Polymertechnik, Technische Universität Berlin, Fasanenstr. 90, D-10623 Berlin

Due to their small mass, adjacent protons (or H-atoms) of molecular systems may exhibit quantum *entanglement* (or quantum correlations), even at ambient conditions. The considerable thermal disturbance and/or many-body interactions of condensed matter and the associated *decoherence* effect, however, cause this protonic entanglement to be restricted in space and time. Some aspects of entanglement and decoherence are mentioned. Extending our previous theoretical work, in the present paper the focus is on the possible existence of entangled protons belonging to the H-bonds of *adjacent* base pairs of B-type DNA. Based on the 'working hypothesis' that this effect does really exist, the most probable 'positions' for the appearance of protonic entanglement in DNA sequences are qualitatively determined. Furthermore, these 'positions' appear to correspond uniquely to dimers of adjacent base pairs of DNA. As a consequence, one can straightforwardly search for an enhanced appearance of such entangled H-bonds in DNA sequences of living organisms, using the existing DNA databases. A quantitative analysis of protein-coding DNA sequences of various organisms has been performed, the results of which provide strong evidence for the existence of the considered effect. The most striking finding may be summarized as follows: Quantum entanglement appears preferably between the third base of a codon and the first base of the following one. Quantitative estimates of this and further obtained results are presented. It is also shown that quantum-chemical considerations of stacking energies cannot account for the results. The new findings provide first evidence for the biological significance of entangled H-bonds.

1. Introduction. – Hydrogen bonds appear in many physical, chemical, and biological systems. This kind of bonding, especially, is of primary importance for biomolecular systems of living organisms, since the latter always contain a significant number of H-bonding functional groups. This is surely related with the well-known facts that many biomolecules exhibit very specific H-bonding interactions with other molecules and with liquid water, and that the evolution of life appeared in an aqueous medium. Furthermore, the discovery of the double-helical structure of DNA [1] made also clear that genetic information on our planet is only possible with H-bonds, since the *Watson-Crick* patterns of H-bonds between base pairs are essential for the DNA structure (and dynamics) of living organisms.

In the present paper, the possibility of quantum correlations between H-bonds of adjacent base pairs of the biologically relevant B-type of DNA is considered. These investigations have been started with our previous paper [2] (Part I of this series), studying the quantum correlated H-bonds in the A-T and C-G base pairs of DNA. That

work was motivated by the novel achievements of *Benner* and coworkers [3] on the extension the genetic alphabet of Nature with the aid of a new base pair of *Watson-Crick* type, called κ-π. Within the frame of a simple quantum-mechanical model, we showed that the H-bond pattern of C-G may exhibit a specific quantum-correlation effect between its H-bonds, in contrast to κ-π where such correlations do not occur [2]. As a consequence, the following question arose: *Is it possible that the considered quantum-correlation effect has some biological significance and/or provides some 'evolutionary advantage'?*

Searching for methods to handle this question, we also noted the basic work of *Eschenmoser* and coworkers [4] on novel structural forms of nucleic acids (*i.e.*, homo-DNA and p-RNA) and their possible consequences for prebiotic biomolecular evolution. The secondary structures of the novel homo-DNA form does not exhibit the well-known helical structure of the natural B-type DNA [4]. In the light of these results, the preceding question can be extended and/or precised in the following way: *Could it be that the quantum-correlation effect between protons also affects H-bonds belonging to adjacent base pairs along the helical axis of B-type DNA?*

The crucial point is now that this question is less speculative than the preceding one, since, in principle, it can be tested 'experimentally', *i.e.*, it can be confronted with experimental data. Namely, if one assumes that the considered quantum-correlation effect is really connected with some kind of evolutionary pressure (whatever 'small' it may be), then one has to expect that Nature has already made some 'use' of it, after about three billion years of evolution of DNA molecules. But then this 'use' should be manifested in specific 'features' of, or 'patterns' appearing in, DNA nucleotide sequences of present-day living organisms. A large number of the latter are already available in DNA databases.

In the present paper, direct evidence for the existence of the considered quantum-correlation effect is provided for the first time. Extending our previous work [2], we will qualitatively show in which specific patterns of the B-type DNA such correlations may be present. These patterns are defining the aforementioned specific 'features' of DNA sequences of living organisms we are searching for. Furthermore, a quantitative statistical analysis of a sufficiently large number of natural DNA sequences coding for proteins will be provided. The associated results and their discussion provide strong evidence for the appearance and the biological significance of the quantum-correlation effect under consideration.

It may be appropriate to mention already here the most striking finding of this analysis, which essentially can be stated in simple terms as follows: The well-known degeneracy of the genetic code, *i.e.*, the fact that many amino acids are specified by more than one codon, concerns mainly the third base (denoted by B_3) of each codon (being given by the three bases $B_1B_2B_3$). But, surprisingly, the choice of B_3 in natural coding DNA appers not to be as 'arbitrary' as assumed. More specifically, our results show the following striking feature: *In many cases, Nature seems to choose selectively B_3 in such a way that quantum correlations between H-bonds of B_3 and of the adjacent B_1*, i.e., *the first base of the next codon, become possible.*

The present paper is organized as follows: *Sect. 2* deals qualitatively with the concepts of 'quantum correlations' or 'entanglement', and 'decoherence', which are of particular importance for the theoretical frame of the paper. *Sect. 3* makes a short

contact with [2], *i.e.*, Part I of this series. Information concerning stacking energies between DNA base pairs is given in *Sect. 4*; these considerations are done in the framework of conventional quantum chemistry. *Sect. 5* provides a qualitative description of the possible protonic quantum correlations between adjacent base pairs of B-type DNA. Here, the specific patterns of base pairs containing quantum correlated H-bonds are presented and qualitatively discussed, leading to the explicit formulation of our 'working hypothesis' underlying the quantitative statistical analysis of DNA sequences of living organisms presented in *Sect. 6*. In particular, *Sect. 6.2* specifies the organisms whose DNA has been analyzed, and *Sect. 6.3* contains the results of that analysis, *e.g.*, the aforementioned surprising finding (concerning the enhanced appearance of protonic quantum correlations between the third base pair of one codon and the first base pair of the following codon). Finally, *Sect. 7* provides a short discussion of the obtained quantitative results, as well as short consideration of their possible biological relevance.

Since the theoretical basis of the presented work is connected with certain modern aspects of quantum theory (like quantum entanglement and decoherence), it may be appropriate to note that 'formal derivations' and/or 'abstract discussions' are not within the scope of the present paper. Rather, the focus is on the applicability and biomolecular relevance of the theoretical results.

2. Quantum Entanglement. – Quantum correlation effects (like, *e.g.*, quantum interference, quantum entanglement, *Einstein-Podolsky-Rosen* correlations, quantum coherence, *Bose* condensation, *etc.*) are intrinsically related with the superposition principle of quantum mechanics. Although quantum interference of photons and neutrons is known since long time and constitutes parts of textbooks of physics, the study of similar effects on atomic and molecular matter has been achieved just very recently. For instance, quantum delocalization, in the range of some micrometers, and self-interference of single atoms (*e.g.*, He, Na) and molecules (Na_2) has been demonstrated some years ago [5]. Moreover, the creation of macroscopic quantum-coherent matter waves of some millions of Na-atoms was achieved [6][7], and their ability to exhibit quantum interference was demonstrated very recently [7], receiving considerable interest [8].

These experiments demonstrated that not only the electrons, but also the nuclei represent quantum objects. In other words, quantum phenomena involving nuclear degrees of freedom explicitly are nowadays experimentally well-established. Obviously, these findings are, at least partially, in 'conceptual conflict' with the common viewpoint of traditional quantum chemistry, as applied to standard quantum-chemical calculations of molecules, where the nuclei are treated as classical mass points.

However, all aforementioned experiments took place under extreme conditions: the considered quantum systems are quasi-isolated and at very low temperature (micro- or nano-Kelvin). These conditions guarantee that the quantum entanglement survives long enough, in order to be experimentally detected. Usually, the detection process typically requires 'macroscopic' time intervals (say, milliseconds).

On the other hand, dynamical processes in condensed matter at ambient temperatures are associated with very short *decoherence* times [9], a fact that thus far has prevented the study of quantum entanglement in such systems. In these systems, it is widely believed that the thermal motion destroys completely every quantum correlation effect within every time interval being experimentally accessible. It may be noted that the

study of quantum decoherence constitutes one of the most actual topics of modern theoretical research (*cf.* [10]).

The possibility of a specific kind of relatively 'long-lived', experimentally accessible quantum entanglement between protons, being the lightest nuclei, in condensed matter at ambient conditions was theoretically investigated by *C.-Dreismann* and *Brändas* [11]. Further work motivated several novel experiments which were proposed and carried out [12][13]. The three recent light and deep-inelastic neutron scattering experiments on H_2O/D_2O mixtures at room temperature [13] provided strong evidence for the theoretically predicted quantum-entangled protonic (and deuteronic) domains in the liquids (for a recent review, see [14]).

It should be mentioned that these scattering processes appear to be fast enough with respect to the relevant decoherence time so that the considered quantum-correlation effect becomes experimentally accessible. In particular, the most recent deep-inelastic neutron experiment [13d] has provided, for the first time, direct evidence for nuclear quantum entanglement in condensed systems at ambient conditions.

3. On Protonic Quantum Correlations in Single DNA Base Pairs. – In our previous paper [2], the possibility of quantum-correlated protons of the H-bonds in the natural G-C and *Benner*'s κ-π [3] DNA base pairs was theoretically considered. That work was motivated by certain specific aspects of the distribution of the three H-bonds of the two base pairs mentioned above. The theoretical considerations were based on *1) Löwdin*'s hypothesis on 'coupled double proton transfer' [15] and the appearance of tautomeric forms in these base pairs, as well as on *2)* the modern aspects of quantum-tunnelling dynamics as revealed in the frame of the so-called Complex Scaling Method (CSM) [11a] [11c]; *3)* the possibility of short-time quantum entanglement of light particles, especially protons, in condensed systems, even at ambient temperatures [11c][14].

One main finding was that, apparently, the G-C base pair appears to exhibit an enhanced 'quantum stability' (caused by quantum correlations, or phase stiffness, between protons) which is clearly missing in the case of the H-bond patterns of the κ-π and A-T pairs. This was demonstrated by using a simple 2×2 model Hamiltonian matrix which may become complex symmetric under the CSM transformation. The aforementioned 'enhanced quantum stability' was associated with the possible appearance of a *Jordan*-block structure in that Hamiltonian (see [2] for details).

In the present context, some related experimental findings may be of interest. For example, the surprising results of inelastic neutron scattering experiments on polyglycine of *Fillaux* and coworkers [16] should be mentioned: Crystals consisting of β-sheets have been investigated. These experiments revealed the striking feature that the protons of the H-bonds 'experience' a completely symmetric potential between N (to which they are bonded covalently) and O (which participate to the H-bonds) atoms. This effect has found thus far no conventional interpretation. However, preliminary theoretical work indicates that the aforementioned symmetric potential may be due to the quantum entanglement of adjacent H-bonds connecting polyglycine strands. The same theoretical framework has been successfully applied to the interpretation of the 'anomalous' dependence of T_1 (spin-lattice) relaxation of benzoic acid dimers on H/D isotopic composition [12a][14].

4. Stacking Energy Considerations. – Conventional quantum-chemical results of energies caused by stacked base pairs in a standard B-type DNA [17] double strand are shortly reviewed in the following.

There are ten possible combinations of base-pair dimers of the spatial structure

$$\begin{matrix} 5'\text{–}X\ Y\text{–}3' \\ \vdots\ \vdots \\ 3'\text{–}X'Y'\text{–}5' \end{matrix} \qquad \text{or in short} \qquad XY/Y'X' \tag{1}$$

denoting that on the one strand the base X is followed by the base Y in the 5′ → 3′ direction of the sugar-phosphate backbone and on the other strand the base Y′ is followed by the base X′, in the 5′ → 3′ direction. The dotted lines represent schematically the H-bonds between the complementary bases X and X′ as well as Y and Y′.

The ten base-pair dimers are: AA/TT, AT/AT, TA/TA, CC/GG, GC/GC, CG/CG, AC/GT, AG/CT, TC/GA, and TG/CA. (In this context, note that, *e.g.*, AG/CT and CT/AG represent the same pair dimer.) The stacking energies of these dimers have been calculated with several quantum-mechanical methods, *e.g.*, force-field calculations, semiempirical calculations and combinations of these methods [18–23] (see also [23][24] for reviews).

To achieve comparable results, a base-plane distance of 3.4 Å and a rotational angle of 36° was used, according to the geometry of the B-type of DNA. All these calculations assume base-pair dimers in the so-called 'frozen core approximation', *i.e.*, the geometries of the molecules are assumed to be rigid. The in-plane energy (*e.g.*, the energy of H-bonds) does not contribute to the stacking energy and, therefore, was not considered in those calculations.

The main contribution of stacking energy is caused by electrostatic forces between the considered stacked base pairs, *i.e.*, interaction between the electron-rich heteroatoms (N and O) or aromatic rings of one base and the electron-poor protons of the one of the other bases (stacked 'above' or 'under' one base on the other strand), and the interaction between the stacked aromatic rings (*cf. Fig. 1*). Thus, in a thermodynamical sense, some of the ten base-pair dimers are more stable than other ones (see *Table 1*). Conferring to the results of *Ornstein et al.* [23] base-pair dimers of the type 5′-purine-pyrimidine-3′/5′-purine-pyrimidine-3′ are more stable than the dimers of the type 5′-pyrimidine-purine-3′/5′-pyrimidine-purine-3′. For instance, AC/GT dimer is by *ca.* 4 kcal/mol more stable than the oppositional dimer CA/TG (*cf. Table 1*). The results of the cited calculations are in qualitative agreement with some associated melting experiments on oligonucleotides [23].

Aida [25a] carried out *ab initio* studies of base dimers in two conformations of DNA, the A- and the B-type, assuming that the stacked bases are parallel with 3.46 Å between them. It should be noted that all four bases of a base-pair dimer are not treated as one system, instead the system is divided in four subsystems:

$$\begin{matrix} \overrightarrow{X\ Y} \\ \underleftarrow{X'\ Y'} \end{matrix} = \begin{matrix} \overrightarrow{X\ Y} \\ \underleftarrow{\quad\quad} \end{matrix} + \begin{matrix} \overrightarrow{\quad\quad} \\ \underleftarrow{X'\ Y'} \end{matrix} + \begin{matrix} \overrightarrow{\quad\quad} \\ \underleftarrow{X'\quad} \end{matrix} + \begin{matrix} \overrightarrow{Y\quad X} \\ \underleftarrow{\quad\quad Y'} \end{matrix} \tag{2}$$

where XY denotes a stacked pair with a base X followed by a base Y in the 5′ → 3′ direction; X′ is the complementary base of X, and Y′ the complementary base of Y. It

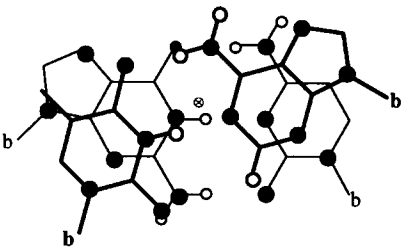

Fig. 1. *Base-pair dimer GT/AC as an example of electrostatic interactions between the stacked base pairs.* The 'upper' base pair is represented with thick lines, whereas the pair 'below' is drawn with thin lines. The H-atoms between each pair of bases are shown as small, open circles, the heteroatoms are shown as large, solid circles. The C=C and C=N bonds have been omitted for clarity. The five- and six-membered rings of the nucleotides are drawn as equilateral polygons. The view is along the helical axis (denoted as ⊕) while 'b' denotes the bond to the sugar-phosphate backbone. A turn angle of 36° per nucleotide residue is used.

Table 1. *Comparison of Stacking Energies* (kcal/mol/dimer) *of the Possible Dimers*

Dimer	[a])	[b])	[c])	[d])	[e])	[f])	[g])	[h])
CG/CG	−4.99	−15.0	−9.68	−3.41	−1.98	−9.69	−11.64	−8.14
GC/GC	−9.80	−15.5	−15.34	−5.88	−13.32	−14.59	−9.62	−7.56
CA/TG	−3.50	−12.3	−6.75	−2.61	−0.94	−6.57	−10.52	−7.50
TA/TA	−2.52	−11.7	−5.26	−1.68	−0.78	−3.82	−9.57	−7.24
GA/TC	−6.04	−13.6	−10.47	−4.20	−5.72	−9.81	−8.21	−6.68
AA/TT	−4.08	−12.8	−6.09	−3.00	−3.06	−5.37	−9.17	−5.85
AC/GT	−7.25	−13.8	−11.25	−5.03	−9.50	−10.51	−8.30	−5.40
AG/CT	−2.41	−12.1	−6.62	−2.24	−0.20	−6.78	−8.21	−4.98
GG/CC	−1.26	−9.5	−8.38	−3.09	−1.32	−8.26	−7.06	−4.58
AT/AT	−4.48	−11.9	−7.21	−4.01	−5.43	−6.57	−7.18	−4.29

The calculations of [a]) *Rein et al.* [18], [b]) *Poltev* and *Sukhorukov* [19], [c]) *Rein* [20], [d]) *Fujita et al.* [21], [e]) *Kudritskaya* and *Danilov* [22], [f]) *Ornstein et al.* [23], [g]) *Aida* [25a], and [h]) *Alhambra et al.* [25b]. For the notation of the dimers, *cf. Eqn. 1.*

should be noted, however, that adding energy eigenvalues derived from Hamiltonians on different incomplete *Hilbert* spaces is affected by the well-known *b*asis *s*et *s*uperposition *e*rror (BSSE).

The most recent *ab initio* study of *Alhambra et al.* [25b] does account for the BSSE. Forced by the limitations of current computer resources, the interstrand stacking and cross-interaction energies, similar to *Eqn. 2*, were derived from MP2/6-31G(d) single-point energies. The geometries of the investigated dimers were composed of single base pairs optimized at the HF/6-31G(d) level of theory, which were subsequently assembled with a fixed base-pair distance and base-pair twist or turn angle. For the same reason, the backbone was not accounted for. Both *ab initio* investigations [25a][25b] are in qualitative agreement. Investigating the stability of dimers of the types 5′-purine-pyrimidine-3′/5′-purine-pyrimidine-3′ and 5′-pyrimidine-purine-3′/5′-pyrimidine-purine-3′, the latter is predicted to be more stable, which is in clear contradiction to the semiempirical calculations of [23].

It should be mentioned that these calculatory results cannot be compared conclusively with experimentally determined melting temperatures T_{melt} of oligomers. This is due to the fact that the numerical value of the parameter T_{melt} is mainly determined by the number of H-bonds (or, equivalently, the percentage of C and G [24][26]) existing in a DNA oligomer, rather than by the stacking energies of its base pairs.

5. Protonic Quantum Entanglement in Adjacent Base Pairs of B-DNA. – Since the protons or H-atoms of interest are participating to the formation of H-bonds, it is obvious that the considered quantum correlations concern protonic *and* electronic degrees of freedom at the same time, even if the electron density involved is typically low due to the characteristics of the H-bond.

Certain electronic and nuclear degrees of freedom appear to be always entangled, in condensed matter problems. In illustrative (and, therefore, necessarily classical) terms, one may say that a certain electronic cloud always 'accompanies' each proton during its motion. Furthermore, the overlapping of the wavefunctions of the considered protons is insignificant. This implies that the protonic entanglement is mainly caused by, or mediated through, the (mainly electrostatic) interactions of the electronic charges 'between' the protons. This context is important to bear in mind, even if we often speak, for the sake of brevity, of quantum correlations between protons, only.

Another important point, in the present context, concerns the theory of the spatial extension and the lifetime of the considered quantum entanglement in *condensed* (*i.e.*, strongly interacting) systems which is still in its infancy (*cf.*, *e.g.*, the associated comments in [9c]). This is clearly in contrast to the situation concerning quantum-entangled systems in the 'ideal gas phase', *i.e.*, for systems being isolated from their environment (*cf. Sects. 1* and *2*). As a consequence, the following point cannot be overemphasized. The protonic quantum correlations under consideration depend on the mutual distance between two protons in an *extremely sensitive* way (see *Sect. 7* and [9a][9c] for more details). Mainly protons (or H atoms) being as close to each other as possible may be expected to exhibit such quantum correlations, since, in this case, the environmental disturbances of their entanglement are smaller than otherwise. Recall that the microscopic mechanism causing the protonic quantum entanglement is given by the direct interactions of the electronic charges 'surrounding' the considered adjacent protons. But, in condensed matter, the same electronic charges also interact with other particles of their environment being subject to the thermal motion, thus leading to the decoherence [9] of the considered protonic entanglement.

In the present section, being motivated by our previous theoretical [11] and experimental [13][14] work, we shall take for granted that the (thus far unknown) decoherence time of possible quantum correlations between closely lying protons (3.4 Å) may be sufficiently 'long' so that it can have certain detectable consequences for the primary DNA structure of living organisms. This constitutes our *working hypothesis*, the immediate consequences of which are reported in the following.

The mentioned motivations and physical hypothesis have led us to consider first the spatial structure of all ten possible base-pair dimers in the well-known B-type [17] of DNA. (Other DNA forms like Z or A, which are of less biological significance, are not considered in this paper.) *Fig. 2* shows schematically the spatial structure of each of the ten dimers, as seen 'along' the helical axis of the DNA double strand. As an approximation, the bases are considered to be, and represented as, planar structures.

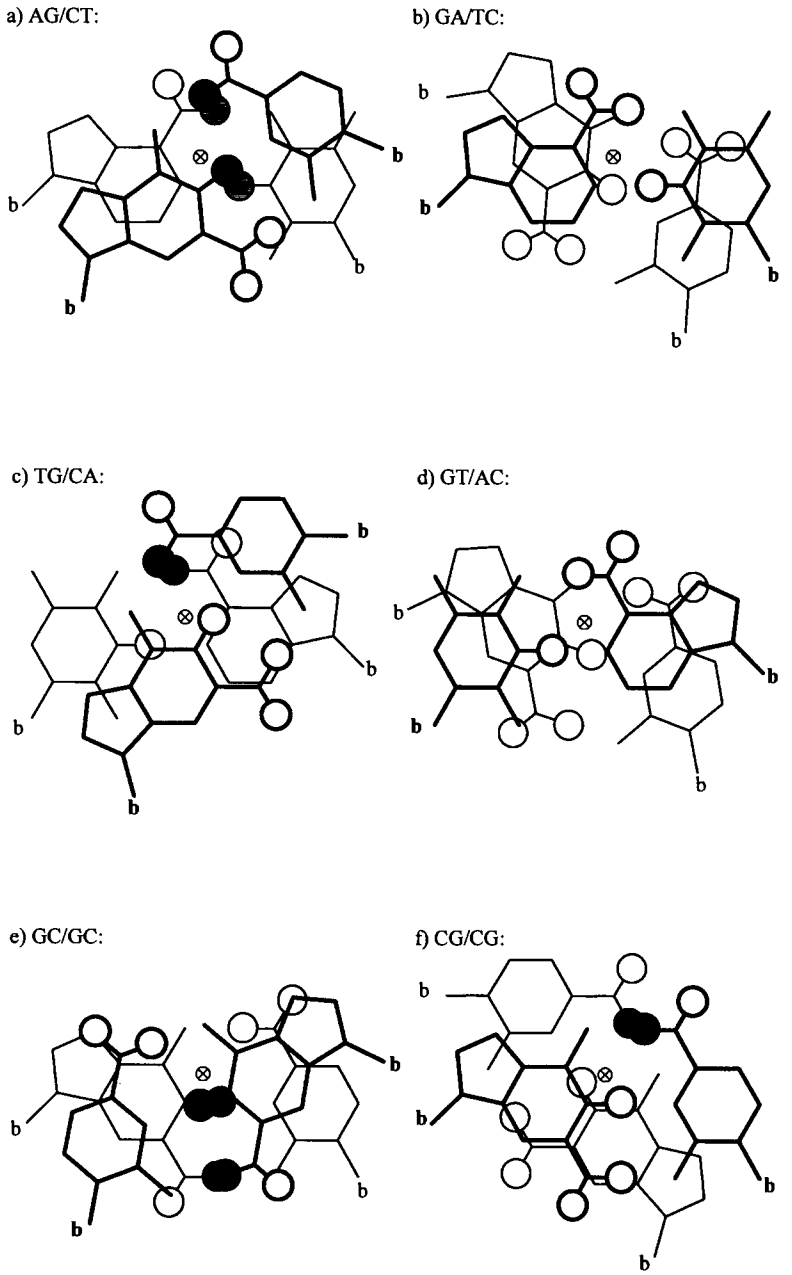

Fig. 2. *Schematic representation of DNA base-pairs dimers.* The 'upper' base pair is represented with thick lines whereas the pair 'below' is drawn with thin lines. The H-atoms between each pair of bases are shown as large circles. The C=C and C=N bonds have been omitted for clarity. The five- and six-membered rings of the nucleotides have been drawn as equilateral polygons. The view is along the helix axes (denoted with ⊕) while 'b' denotes the bond to the sugar-phosphate backbone. A turn angle of 36° per nucleotide residue is used. Note that the protons exhibiting enhanced quantum correlations are marked with hatches.

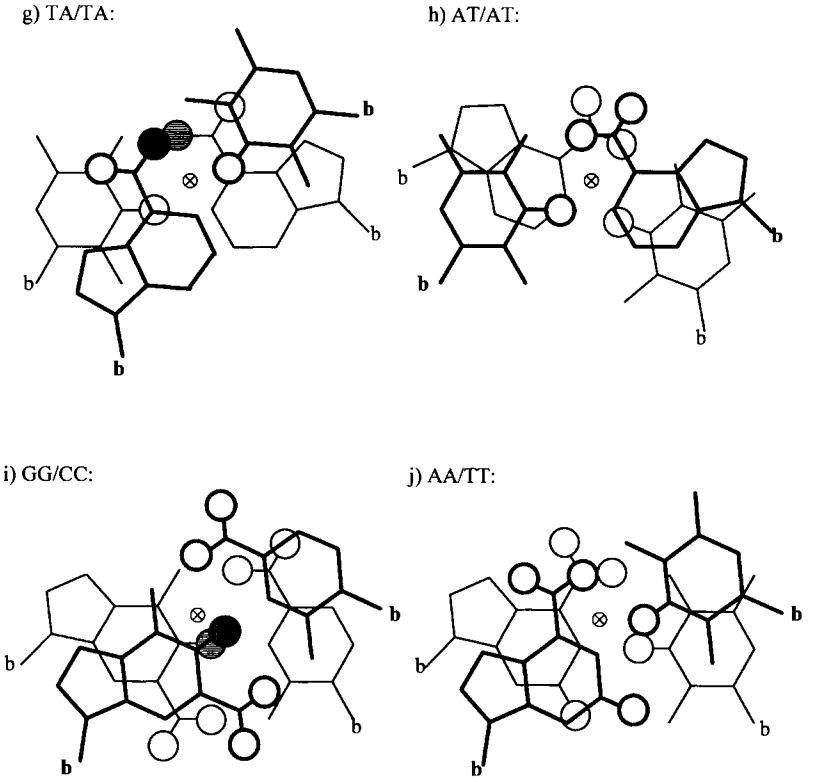

Fig. 2 (cont.)

Secondly, we look at the protons (H-atoms) of the H-bonds belonging to adjacent base pairs, with the aim to single out those XY/Y'X' dimers having the most closely lying protons. (Recall again that the aforementioned biologically significant B-type of DNA has been used.) For illustration, those H-atoms (protons) belonging to H-bonds are represented by circles as well as the second H-atoms of the amino groups. The close inspection of these figures (*Fig. 2*, *a–j*) reveals the following distinguished features.

In some dimers, two protons (H-atoms) belonging to different base pairs are able to approach each other more closely than in the reversely arranged dimers. As an example, the dimers AG/CT and GA/TC may be compared: In AG/CT, there exist two such pairs of 'approaching' protons. These are called in the following text 'quantum-correlated protons' and are represented in *Fig. 2* by circles with hatchings. On the contrary, the GA/TC contains no such protons. Another feature of these considerations is that all three dimers consisting only of guanine and cytosine (GC/GC, CG/CG, and GG/CC) own at least one pair of quantum-correlated protons.

These considerations give us the opportunity to distinguish the ten base-pair dimers into groups with two, one, and no quantum-correlated protons:

Dimers with 2 quantum correlated protons:	AG/CT and GC/GC.	
Dimers with 1 quantum correlated proton:	TG/CA, CG/CG, TA/TA and GG/CC.	(3)
Dimers without quantum correlated protons:	GA/TC, GT/AC, AT/AT and AA/TT.	

To make these observations advantageous for statistical examinations we compare two base-pair dimers with reverse arrangement (*e.g.*, GC/GC and CG/CG) in each case. Therefore, *Fig. 2* shows in one row always two dimers, each in reverse arrangement to the other. The dimers in the left column always contain more quantum-correlated protons than those in the right column. This classification provides the basis of the following statistical investigations.

6. DNA of Living Organisms: Statistical Investigations. – In this section, the possible existence of quantum-correlated protons in DNA of living organisms is quantitatively tested with the aid of statistical methods. To be as self-contained and concrete as possible, in this paper only protein coding DNA sequences are investigated.

6.1. *Definitons.* There are 16 possible nucleotide pairs of the pattern 5′-XY-3′, denoting a nucleotide X followed by a nucleotide Y in the 5′ → 3′ direction of the sugar-phosphate backbone. (It should be mentioned that the number of 16 possible nucleotide pairs is not in contradiction to the number of 10 base-pair dimers mentioned above.) Also note that a DNA sequence with a length of L *Watson-Crick* base pairs contains $L - 1$ pairs of the pattern 5′-XY-3′. Let us classify these 16 nucleotide pairs in three groups:

a) The first group, called 'Q', contains the nucleotide pairs with quantum-correlated protons.

b) Each member of the second group (which is called 'N') either does not contain any quantum-correlated protons, or it contains less quantum-correlated protons than the associated pair of the group Q. (For instance, AG contains two quantum-correlated protons, whereas GA contains none. Another example: The pair CG contains one quantum-correlated couple of protons and could, therefore, be attached to group Q. But this pair is found in group N, since the associated pair GC contains two quantum-correlated proton pairs; see *Fig. 2.*)

c) The remaining nucleotide pairs belong to the third group, called 'R'. In short:

Group Q: AG, TG, TA, GC, CA, and CT.	
Group N: GA, GT, AT, CG, AC, and TC.	(4)
Group R: AA, TT, CC and GG.	

It should be emphasized that all four DNA nucleotides A, T, C, and G are present in Q as well as in N three times, *i.e.*, they are distributed with *equal* statistical weight in both groups. The introduction of the third group R preserves this equal distribution. (Parenthetically, certain poly(A) and poly(C) tracks are known to have specific biological functions in specific, noncoding DNA sequences [27], and, therefore, they should be considered more carefully in future work concerning noncoding DNA sequences, too.)

6.2. *Organisms.* Several sequences, assigned in six groups of different organisms containing 10 long coding sequences each, have been investigated. The first group contains sequences of the bacterium *Haemophilus influenzae* which is the first organism being sequenced completely [28]. Secondly, we have chosen the well-known bacterio-

phage λ. As an example of a simple eucariotic genome, we have chosen the chromosome III of *Saccharomyces cerevisiae* (yeast), the first chromosome being sequenced completely [29]. As a representative of the invertebrates, we have chosen *Caenorhabditis elegans* [30]. In the fifth group are included coding sequences of animals with vertebrae; the chicken, the rat, the mouse, and the gorilla. In the last group, we have selected some sequences of the human genome. (According to specific biological viewpoints, one may combine the last two groups into a single one, or also split them into more groups.) For details on the investigated organisms, *cf. Table 2.*

Table 2. *Fractions* F_{all}, *Fractions* F_{nm} *at Sites* nm *and Their Relative Deviations* ΔF_{nm}. For F, *cf. Eqn. 5*; for F_{12}, F_{23}, and F_{31}, *cf. Eqn. 6*; for ΔF_{12}, ΔF_{23}, and ΔF_{31}, *cf. Eqn. 7*; see text for detailed information.

Organism	Length [bp]	F_{all}	F_{12}	F_{23}	F_{31}	ΔF_{12} [%]	ΔF_{23} [%]	ΔF_{31} [%]
H. influenzae[a])	37,959	1.03	0.63	1.06	1.60	−30	−5	21
Bacteriophage λ[b])	16,923	1.03	0.79	1.05	1.38	−11	1	17
S. cerevisiae[c])	37,755	1.01	0.66	1.18	1.30	−27	4	13
C. elegans[d])	47,379	0.91	0.74	0.91	1.12	−30	−5	10
Vertebrates[e])	37,716	1.30	0.84	1.43	1.89	−12	18	29
Homo sapiens[f])	21,432	1.26	0.97	1.37	1.72	−6	14	24

The names of the investigated organisms, files, and genes or qualifiers with sequence length in brackets: [a]) *H. influenzae* (sequence file found at the database of *The National Center for Genome Resources* [28]): 1. IgA1 protease (5085 BP); 2. Cell division protein (4533 BP); 3. DNA-directed RNA polymerase beta chain (4032); 4. Formylglycineamide ribonucleotide synthetase (3963 BP); 5. Exodeoxyribonuclease V (3636 BP); 6. DNA polymerase III, alpha chain (3480 BP); 7. Transcription-repair coupling factor (trcF) (3441 BP); 8. Exodeoxyribonuclease V (3366 BP); 9. Transferrin-binding protein 1 (tbp1) (3255 BP); 10. Type I restriction enzyme ECOR124/R protein (3168 BP). [b]) Bacteriophage λ (found at the database of the *European Molecular Biology Laboratory* (*EMBL*), rel. 28). 1. J (3399 BP); 2. H (2562 BP); 3. A (1926 BP); 4. B (1602); 5. Ea59 (1578 BP); 6. C (1320 BP); 7. Ea47 (1233 BP); 8. orf401 (1206 BP); 9. E (1026 BP); 10. int (1071 BP). [c]) *S. cerevisiae* (file name SCCHRIII found at the database of *EMBL*, rel. 33): 1. YCR32w (6504 BP); 2. YCR93w (6327 BP); 3. YCL19w (4044 BP); 4. YCR33w (3681 BP); 5. YCR81w (3681 BP); 6. YCL14w (2985 BP); 7. YCR17c (2862 BP); 8. YCR37e (2772 BP); 9. YCR106w (2499 BP); 10. YCL30c (4200 BP). [d]) *C. elegans* (found at the database of *EMBL*, rel. 41): 1. C50C3.6 protein (6990 BP); 2. R10E11.1 protein (6048 BP); 3. ZK688.5 protein (5400 BP); 4. Ko4H4.1, a collagen (5235 BP); 5. F44E2.4 protein (4830 BP); 6. C29E4.3 protein (4236 BP); 7. ZK370.4 protein (4056 BP); 8. B0464.2 protein (3825 BP); 9. F10E9.8 protein (3420 BP); 10. C38C10.5 protein (3339 BP). [e]) Vertebrates' genes (all files found at the database of *EMBL*, rel. 33 or older): 1. chicken myosin heavy chain, in file GGMYHE (5823 BP); 2. chicken vitellogenin, in file GVITIIG (5550 BP); 3. chicken collagen alpha 2, in file GDCOL6A2 (3069 BP); 4. chicken lipoprotein lipase, in file GDLIPLIP (1473 BP); 5. chicken delta crystallin, in file GGCRYDS (1401 BP); 6. rat sodium channel protein II, in file RNSCPIIR (6015 BP); 7. rat sodium channel III, in file RNSCIII (5853 BP); 8. mouse sex limited protein, in file MMSLPSEX (5208 BP); 9. mouse cytochrome P-450, in file MMSLPSEX (1464 BP); gorilla alpha-fetoprotein, in file GGAFPA (1830 BP). [f]) *Homo sapiens* (all files found at the database of *EMBL*, rel. 33 or older): 1. HSCBMYHC, cardiac beta myosin heavy chain (5802 BP); 2. HSATP1A2, Na-, K-ATPase (3063 BP); 3. HSBFXIII, blood coagulation factor XIIIb (1986 BP); 4. HSTHB, Prothrombin (1869 BP); 5. HSAFPCP, alpha-fetoprotein (1830 BP); 6. HSIFNAR, interferon alpha/beta receptor (1674 BP); 7. HSG6PDGE, glucose-6-phosphate dehydrogenase (1548 BP); 8. HSFIXG, factor IX prepropeptide (1386 BP); 9. HSP53G, protein p53 (1182 BP); 10. HSADAG, adenosine deaminase (1092 BP).

6.3. *Results.* The first investigation of the mentioned coding DNA sequences is the simplest possible one: we count and classify all $(L - 1)$ pairs of the considered sequence with a length of L. Let now Q_{all} be the number of pairs of the group Q, N_{all} be the number of pairs of the group N, and R_{all} be the number of pairs of the group R, all over the

considered ten sequences in one organism group. F_{all} is the ratio

$$F_{all} = \frac{Q_{all}}{N_{all}}. \tag{5}$$

This ratio F_{all} is *independent* of the overall base composition, because the four nucleotides are evenly distributed among the numerator and the denumerator of this fraction. (The base composition only influences R_{all} which is not considered here.) If the assumed quantum correlations do not 'influence' the primary structure of DNA sequences, we may expect that F_{all} is equal to 1, within small deviations. On the contrary, if it appears that $F_{all} > 1$, then one may speculate that the enhanced quantum correlated dimers of the group Q could be related with some, thus far unknown, 'evolutionary advantage' with respect to those of the group N. Simply formulated: 'plain statistics' requires $F_{all} \approx 1$.

The results of our numerical investigations presented in *Table 2* clearly reveal strong deviations from the 'stochastic average' $F_{all} \approx 1$; while the sequences of *H. influenzae*, bacteriophage λ, and *S. cerevisiae* behave as conventionally expected, *i.e.*, having $F_{all} \approx 1$, the sequences of *C. elegans* contain more pairs belonging to group N than Q; *i.e.*, $F_{all} < 1$. On the contrary, the sequences of the vertebrates and of the human genome contain *ca.* 30% or 26%, respectively, more pairs or group Q than N.

Thus, these results show that DNA of different organisms contain a different quantity of quantum-correlated protons. The quantum correlations appear to be more frequent in DNA of 'higher' animals (vertebrates including the human genome) than in the DNA of other organisms.

To give more evidence to this striking effect, we proceed with the following analysis, taking into account the biological purpose of a coding DNA sequence, *i.e.*, DNA-coding sequences are considered as sequences of nucleotide triplets, called codons, coding one amino acid each. Let us denote the three nucleotides of any codon by 5′-$B_1B_2B_3$-3′. We count separately the numbers of nucleotide pairs in the positions B_1B_2, B_2B_3, and B_3B_1 (always noted in the standard 5′ → 3′ direction; see *Sect. 6.1.*) belonging to the three groups Q, N, and R. To mark the site of a counted nucleotide pair, an index is used: *e.g.*, Q_{23} denotes the number of counted nucleotide pairs which consist of the bases B_2 and B_3 of a codon, and belong to Q all over the ten considered genes. Note that the pair B_3B_1 is located 'between' two adjacent codons whereas the other two sites (B_1B_2 and B_2B_3) are located within only one codon.

According to the above notations and definitions, let us denote with Q_{12} (or N_{12}) the amount of nucleotide pairs B_1B_2 belonging to the group Q (or N) all over the ten considered genes mentioned in *Sect. 6.2.* Q_{23}, Q_{31}, N_{23}, and N_{31} are defined similarly. For clarity, let us mention that, *e.g.*, $Q_{all} = Q_{12} + Q_{23} + Q_{31}$. Let us again define fractions in analogy to *Eqn. 5*:

$$F_{nm} = \frac{Q_{nm}}{N_{nm}}, \tag{6}$$

where nm = 12, 23, or 31.

According to the well-known 'degeneracy' of the genetic code (see textbooks, *e.g.*, [26][27]), the third nucleotide B_3 of a codon is not uniquely determined. In contrast, the distributions of the nucleotides B_1 and B_2 of a codon are almost exclusively determined by the amino-acid sequence. The results of our calculations are summarized in *Table 2.*

Note that *i*) all averages of F_{12} are smaller than 1, *ii*) that the averages of $F_{23} \approx 1$, except for the genes of yeast, the vertebrates, and of human genome whose averages are 1.18, 1.43, and 1.37, respectively, and that *iii*) all averages over the F_{31} values are significantly larger than 1, being between 1.12 for *C. elegans* and 1.89 for 'vertebrates' (see *Table 2*). All groups of DNA sequences exhibit, on average, larger values for F_{31} than for F_{12} or F_{23}. It is obvious that on the sites containing B_3 more entangled nucleotide pairs are found than on the site B_1B_2. Secondly, the trend of our first calculations could be mostly confirmed, although we took more biological information into account. The first calculations were made just by counting the base-pair dimers without considering the triplet-nature of codons. Now, we take this structure of triplets into account, and the observed effect remains qualitatively unchanged. So, the next step is to stay at this point of biological information but to extend the above analysis as follows.

To give more statistical evidence for the significance of these findings, one has to demonstrate that these results are not trivially caused by the base compositions of the genes. To achieve this aim, we compared (see below) the natural DNA sequences with artificial DNA sequences being 'associated' with the natural ones. These artificial sequences have been produced with the aid of random number generators by the following steps:

1) The base composition of the considered natural DNA sequence is calculated.

2) The random number generator provides a number between 0 and 1. As a random number generator, we used RAN3() which is described in the standard reference 'Numerical Recipes' [31], which is fully sufficient for these short sequences [32].

3) Respectively to the random number and the base composition of the natural sequence, one character (A, T, C, or G) is generated and added to the artificial sequence.

4) *Steps 2* and *3* are repeated L times for a natural DNA of length L under consideration. For a detailed description of this numerical procedure and different concrete applications, see [32].

To minimize statistical fluctuations, 100 different artificial sequences for each natural DNA sequence were created. We counted the amount of pairs belonging to the family Q at the three mentioned locations of each artificial sequence and calculated the average over them resulting in values called Q^*_{12}, Q^*_{23}, and Q^*_{31}. (The asterisks indicate that these quantities were obtained from artificial sequences.) In other words, these averages quantify the appearance of base pairs with the 'property Q' in artificial sequences having the same base composition with the natural DNA sequence. To quantify the 'differences' of the natural DNA sequences from the artificial ones, let us define the 'relative deviation'

$$\Delta F_{nm} = \frac{Q_{nm} - Q^*_{nm}}{Q^*_{nm}}, \tag{7}$$

where $nm = 12$, 23, or 31. See *Table 2* for the results of our calculations.

If the artificial sequences would 'behave' like the natural ones one expects that $\Delta F_{nm} \approx 0\%$. But surprisingly, we observe a large deviation between the compared sequences: first, let us consider the sites B_1B_2. All averages of ΔF_{12} are negative, with the human genes providing the least negative values (-6%). Second, some of the averages of ΔF_{23} are positive (bacteriophage λ + 1%, yeast + 4%, vertebrates' genes + 18%, and human's genes + 14%), others are negative (*H. influenzae* -5% and *C. elegans* -5%).

In other words, at codon site B_2B_3, ΔF_{23} seems to be dependent on the organism. Third, all ΔF_{31} values are positive, with the vertebrates' genes providing the highest amount (+ 29%). Summarizing, in the natural DNA sequences the quantum-correlated pairs at codon site B_1B_2 are less than randomly expected, while the largest amount of quantum correlated pairs is located at B_3B_1.

This result is consistent with our previous calculations, although we used here a statistically more relevant method. Again, this calculation shows a similar trend like our first calculations: the DNA of 'higher' organisms contains more quantum-correlated pairs than the DNA of the other investigated organisms.

7. Conclusion. – In the present paper, we provide for the first time evidence that quantum entanglement between protons belonging to adjacent H-bonds of DNA of living organisms may be effective. The existence of this new effect in liquid water at room temperature has been very recently demonstrated experimentally, using the method of deep-inelastic neutron scattering [13d].

The applied theoretical model is based on our 'working hypothesis' that this quantum effect preferably occurs in such base-pair configurations of DNA, in which two H-bond protons belonging to different base pairs can approach each other as much as possible. The reason for this is that, due to the typical many-body interactions of condensed matter at ambient conditions, the effect of decoherence becomes strongly attenuated (*i.e.*, by the factor $\exp[-\text{const}\,(\Delta x)^2]$) with increasing distance (Δx) between the two particles (see [9a][9c], for details). Recall that the considered effect does not concern 'idealized' systems like, *e.g.*, ideal gases or perfect crystals.

Based on these theoretical considerations (*cf. Sect. 5*), we were able to differentiate between specific dimers of adjacent DNA base pairs containing an enhanced number of quantum-entangled H-bonds, which define the group Q, and the 'associated' dimers containing less or no such quantum entanglement, which define the group N (*cf. Sects. 5* and *6*). For example, the dimer abbreviated with AG belongs to the group Q, whereas the 'associated' dimer GA belongs to the group N.

The crucial part of the present paper is given by the quantitative analysis of protein-coding DNA sequences of various living organisms (listed in *Sect. 6.2*), which determined quantitatively the frequencies of appearance of quantum-entangled H-bonds in the different DNA sequences. These frequencies constitute the *experimental* material which has been then analyzed in the frame of statistics. Our numerical investigations have explicitly demonstrated the following striking finding: *In protein-coding DNA sequences of living organisms, protonic quantum entanglement appears preferably between the third base of a codon and the first base of the following one.*

In short, we may say that the considered quantum entanglement appears preferably 'between' adjacent codons. In this context, recall the well-known degeneracy of the universal genetic code, which is mainly related with the 'free choice' of the third base of a codon corresponding to a specific amino acid. To be more concrete, consider the DNA strand determining the reading frame (*i.e.*, the strand whose information is transcribed to m-RNA and then translated to a protein). Our finding then implies that the third base (denoted by B_3) of a given codon is chosen in a way which tends to increase the quantum entanglement with the first base (denoted with B_1) of the following codon. This surprising finding seems to indicate that quantum entanglement may play some specific role in,

or have some biological significance for, the structure and/or dynamics of natural DNA, which is hitherto unknown.

This conclusion may also be illustrated quantitatively with the aid of the magnitude of the effect. For example, in analyzed DNA sequences of vertebrates (including human DNA), we found that the ratio $F_{31} = Q_{31}/N_{31}$ of pairs 'with' and 'without' quantum entanglement is *ca.* 1.8, *i.e.*, it appears to be by *ca.* 80% larger than one may expect on the basis of simple statistics (where one plainly would expect that $F_{31} = 1.0$). Noting that this number is based on the analysis of different sequences with a total length of *ca.* 20,000 codons (*cf. Table 2*), one immediately sees that the result is statistically highly significant.

To give further support to the significance of this finding, we also compared the considered natural DNA sequences with artificial sequences having the same length and the *same* global base-pair composition with the natural ones. (The artificial sequences were computer-generated, using random number generators; *cf. Sect. 6.3.*) The direct comparison between the natural and artificial sequences clearly showed that the natural DNA sequences are characterized by significantly more (*ca.* 25%) quantum entanglement 'between adjacent codons' than the corresponding artificial sequences. Of course, the relevance of this comparison should be not overestimated, since the artificial sequences only resemble the natural ones with respect to base-pair composition, but they do not code for the same proteins. Nevertheless, this comparison clearly demonstrates that the obtained result is not just an artefact being related with the 'uneven' or 'biased' base-pair composition of the natural DNA sequences.

Further work, being now in progress, is intended to produce artificial sequences containing equal 'information', as concerns protein synthesis, with a natural DNA sequence. (In simple terms, both artificial and natural DNA sequences lead to the same protein.) Preliminary results of comparisons between such sequences indicate that our new finding survives also this test, *i.e.*, it remains highly significant.

Another crucial point, in view of molecular biology, concerns the so-called *codon usage* [33] of natural organisms. This means, in simple terms, that all codons coding for a specific amino acid do not appear with equal probability in natural DNA sequences. (Note that, in the scientific literature, there is no unique opinion about the 'biological meaning' of this observation.) In the near future, we also intend to elucidate our finding with respect to codon usage associated with the specific natural DNA sequences being under investigation. For these investigations, suitable artificial sequences will be produced, which *a*) code for the same sequence of amino acids as the natural DNA sequence, and *b*) have on the average the codon distributions of the considered organism. First related results appear to be encouraging, too.

Analyzing the result from the viewpoint of traditional quantum chemistry, one could speculate the main result (say, $F_{all} > 1.0$ and $F_{31} > 1.0$) to be trivially caused by differences in stacking energies. One would then expect the averaged stacking energies of the base-pair dimers forming group Q to be more negative (*i.e.*, more favorable in terms of energy) than those of group N. However, testing this assumption on the base of the data of *Table 2*, a clear contradiction arises (even though the thus far available data [18–23][25] provide only a trend in energies, due to the applied limited quantum chemical methods). In almost all investigations [18–23], the nucleotide pairs of group N appear to be more favorable (or stabilized) than those of group Q by *ca.* 10 kcal/mol, and thus

one would expect F_{all} to be smaller than 1.0. This, however, is in contrast to the present findings, which revealed F_{all} to be larger than 1.0. The results in [25] suggest no preference for either of the two groups. Summarizing, differences on stacking energies appear not to be able to account for the observed finding $F_{all} > 1.0$.

At least, some speculations concerning the possible biological significance of the revealed quantum entanglement between protons (H-bonds) may be appropriate. In Part I [2], where the possible existence of entangled H-bonds in individual base pairs has been considered, it has been noticed that that effect might play some important role in biomolecular systems, *e.g.*, providing some enhanced stability (rigidity or stiffness) of H-bonded molecular systems [2]. Moreover, it was speculated that, if the effect under consideration does really cause some kind of 'evolutionary advantage', then one would expect that Nature has made already some 'use' of it. But then this ought to be manifested in specific features of DNA nucleotide sequences of living organisms. Indeed, the new findings presented above strongly indicate that such 'specific features' do really exist in natural coding DNA. Of particular biological interest seems to be the large value, as compared with statistical estimates, of F_{31}. Namely, the third base B_3 of a codon may be used to 'store' a hitherto unknown biological information; *e.g.*, *i*) it may be chosen properly in order to increase the quantum-correlation stability [2] of DNA *along* the helical axis; or *ii*) its entangled H-bonds with those of the next codon may provide an extra 'marker' which could facilitate transcription. The latter possibility would then 'explain' the well-known phenomenon of varying codon usage [33], which has found thus far no clear explanation. A further speculative thought is that the quantum entanglement between H-bonds may be used to facilitate molecular recognition between different biomolecules, *e.g.*, in the formation of highly specific DNA-protein or protein-protein complexes. Further work related with the above findings and/or speculative considerations is in progress.

We are indebted to Profs. *D. Larhammar* and *M. Ehrenberg* (Biomedical Center, Uppsala University) for many invaluable discussions. This work has been supported, in part, by the *Fonds der Chemischen Industrie*, *EMBO*, and *DAAD*.

REFERENCES

[1] J. Watson, F. H. C. Crick, *Nature (London)* **1953**, *171*, 964.

[2] C. A. Chatzidimitriou-Dreismann, *Helv. Chim. Acta* **1992**, *75*, 2252.

[3] a) J. A. Piccirilli, T. Krauch, S. E. Moroney, S. A. Benner, *Nature (London)* **1990**, *343*, 33; b) C. Switzer, S. E. Moroney, S. A. Benner, *J. Am. Chem. Soc.* **1989**, *111*, 8322; c) S. A. Benner, *Nachr. Chem. Tech. Lab.* **1990**, *38*, 442.

[4] a) A. Eschenmoser, in 'Verhandlungen der Gesellschaft Deutscher Naturforscher und Ärzte', 1991, Vol. 116, p. 135; b) A. Eschenmoser, M. Dobler, *Helv. Chim. Acta* **1992**, *75*, 218; c) I. Schlönvogt, S. Pitsch, C. Lesueur, A. Eschenmoser, B. Jaun, R. M. Wolf, *ibid.* **1996**, *79*, 2316.

[5] a) O. Carnal, J. Mlynek, *Phys. Rev. Lett.* **1991**, *66*, 2689; b) D. W. Keith, C. R. Ekstrom, Q. A. Turchette, D. E. Pritchard, *ibid.* **1991**, *66*, 2693; c) M. S. Chapman, C. R. Ekstrom, T. D. Hammond, R. A. Rubenstein, J. Schmiedmayer, S. Wehinger, D. E. Prichard, *ibid.* **1995**, *74*, 4783.

[6] M. H. Anderson, J. R. Ensher, M. R. Matthews, C. E. Wiemann, E. A. Cornell, *Science* **1995**, *269*, 198.

[7] M. R. Andrews, C. G. Townsend, H.-J. Miesner, D. S. Durfee, D. M. Kurn, W. Ketterle, *Science* **1997**, *275*, 637.

[8] K. Burnett, *Nature (London)* **1997**, *385*, 482.

[9] a) D. Giulini, E. Joos, C. Kiefer, J. Kupsch, I.-O. Stamatescu, H. D. Zeh, 'Decoherence and the Appearance of a Classical World in Quantum Theory', Springer, Berlin, 1996; b) W. H. Zurek, *Physics Today* **1991**, *44*, 36; c) R. Omnès, 'The Interpretation of Quantum Mechanics', Princeton University Press, Princeton, 1994, see Chapt. 7.

[10] D. M. Greenberger, A. Zeilinger, Eds., 'Fundamental Problems in Quantum Theory', Ann. N. Y. Acad. Sci. New York, 1995, Vol. 755.
[11] a) E. J. Brändas, C. A. Chatzidimitriou-Dreismann, in 'Resonances', Eds. E. J. Brändas and N. Elander, Lecture Notes in Physics, Springer, Berlin, 1989, Vol. 325, pp. 485–540; b) C. A. Chatzidimitriou-Dreismann, E. J. Brändas, *Ber. Bunsenges. Phys. Chem.* **1991**, *95*, 263; c) C. A. Chatzidimitriou-Dreismann, *Adv. Chem. Phys.* **1991**, *80*, 201.
[12] a) S. Takeda, A. Chatzidimitriou-Dreismann, *Chem. Phys. Lett.* **1992**, *198*, 316; b) H. Weingärtner, C. A. Chatzidimitriou-Dreismann, *Nature (London)* **1990**, *346*, 548.
[13] a) U. K. Krieger, F. Aurich, C. A. Chatzidimitriou-Dreismann, *Phys. Rev. A* **1995**, *52*, R1827; b) C. A. Chatzidimitriou-Dreismann, U. K. Krieger, A. Möller, M. Stern, *Phys. Rev. Lett.* **1995**, *75*, 3008; c) C. A. Chatzidimitriou-Dreismann, J. Mayers, *Physica B* **1996**, *226*, 231; d) C. A. Chatzidimitriou-Dreismann, T. Abdul Redah, R. M. F. Streffer, J. Mayers, *Phys. Rev. Lett.* **1997**, *79*, 2839.
[14] C. A. Chatzidimitriou-Dreismann, *Adv. Chem. Phys.* **1997**, *99*, 393.
[15] P.-O. Löwdin, *Rev. Mod. Phys.* **1963**, *35*, 724; P.-O. Löwdin, *Adv. Quant. Chem.* **1965**, *2*, 212.
[16] G. J. Kearley, F. Fillaux, M.-H. Baron, S. Bennington, J. Tomkinson, *Science* **1994**, *264*, 1285.
[17] S. Arnott, R. Chandrasekaran, D. W. L. Hukins, P. J. C. Smith, L. Watts, *J. Mol. Biol.* **1974**, *88*, 523; R. E. Dickerson, H. R. Drew, *ibid.* **1981**, *149*, 761; H. R. Drew, R. E. Dickerson, *ibid.* **1981**, *151*, 535.
[18] R. Rein, P. Claverie, M. Pollack, *Int. J. Quantum Chem.* **1968**, *2*, 129.
[19] V. I. Poltev, B. I. Sukhorukov, *Stud. Biophys. (Berlin)* **1970**, *24/25*, 179.
[20] R. Rein, *Adv. Quantum Chem.* **1973**, *7*, 335.
[21] H. Fujita, A. Imamura, C. Nagata, *J. Theor. Biol.* **1974**, *45*, 411.
[22] Z. G. Kudritskaya, V. I. Danilov, *J. Theor. Biol.* **1976**, *59*, 303.
[23] R. L. Ornstein, R. Rein, D. L. Breen, R. D. Macelroy, *Biopolymers* **1978**, *17*, 2341.
[24] W. Saenger, 'Principles of Nucleic Acid Structure', Springer Verlag, New York, 1984, pp. 137–140.
[25] a) M. Aida, *J. Theor. Biol.* **1988**, *130*, 327; b) C. Alhambra, F. J. Luque, F. Gago, M. Orozco, *J. Phys. Chem. B* **1997**, *101*, 3846.
[26] B. Lewin, 'Genes V', Oxford University Press, Oxford, 1994.
[27] J. D. Watson, N. H. Hopkins, J. W. Roberts, J. A. Steitz, A. M. Weiner, 'Molecular Biology of the Gene', The Benjamin/Cummings Publishing Company, Inc., Menlo Park, California, 1987.
[28] R. D. Fleischman, M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J.-F. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. Sutton, W. FitzHugh, C. A. Fields, J. D. Gocayne, J. D. Scott, R. Shirley, L.-I. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. R. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S. M. Geoghagen, C. L. Gnehm, L. A. McDonald, K. V. Small, C. M. Fraser, H. O. Smith, J. C. Venter, *Science* **1995**, *269*, 496.
[29] S. G. Oliver, Q. J. M. Van der Aart, M. L. Agostini-Carbone, M. Aigle, L. Alberghina, D. Alexandraki, G. Antoine, R. Anwar, J. P. G. Ballesta, P. Benit, G. Berben, E. Bergantino, N. Biteau, P. A. Bolle, M. Bolotin-Fukuhara, A. Brown, A. J. P. Brown, J. M. Buhler, C. Carcano, G. Carignani, H. Cederberg, R. Chanet, R. Contreras, M. Crouzet, B. Daignan-Fornier, E. Defoor, M. Delgado, J. Demolder, C. Doira, E. Dubois, B. Dujon, A. Dusterhoft, D. Erdmann, M. Esteban, F. Fabre, C. Fairhead, G. Faye, H. Feldmann, W. Fiers, M. C. Francinques-Gaillard, L. Franco, L. Frontali, H. Fukuhara, L. J. Fuller, P. Galland, M. E. Gent, D. Gigot, V. Gilliquet, N. Glansdorff, A. Goffeau, M. Grenson, P. Grisanti, L. A. Grivell, M. De Haan, M. Haasemann, D. Hatat, J. Hoenicka, J. Hegemann, C. J. Herbert, F. Hilger, S. Hohmann, C. P. Hollenberg, K. Huse, F. Iborra, K. J. Indige, K. Isono, C. Jacq, M. Jacquet, C. M. James, J. C. Jauniaux, Y. Jia, A. Jimenez, A. Kelly, U. Kleinhans, P. Kreisi, G. Lanfranchi, C. Lewis, C. G. van der Linden, G. Lucchini, K. Lutzenkirchen, M. J. Maat, L. Mallet, G. Mannhaupt, E. Martegani, A. Mathieu, C. T. C. Maurer, D. McConnell, R. A. McKee, F. Messenguy, H. W. Mewes, F. Molemans, M. A. Montague, M. Muzi Falconi, L. Navas, C. S. Newlon, D. Noone, C. Pallier, L. Panzeri, B. M. Pearson, J. Perea, P. Philippsen, A. Pierard, R. J. Planta, P. Plevani, B. Poetsch, F. Pohl, B. Purnelle, M. Ramezani Rad, S. W. Rasmussen, A. Raynal, M. Remacha, P. Richterich, A. B. Roberts, F. Rodriguez, E. Sanz, I. Schaaff-Gerstenschlager, B. Scherens, B. Schweitzer, Y. Shu, J. Skala, P. P. Slonimski, F. Sor, C. Soustelle, R. Spiegelberg, L. I. Stateva, H. Y. Steensma, S. Steiner, A. Thierry, G. Thiereos, M. Tzermia, L. A. Urrestarazu, G. Valle, L. Vetter, J. C. van Vliet-Reedijk, M. Voet, G. Volckaert, P. Vreken, H. Wang, J. R. Warmington, D. von Wettstein, B. L. Wicksteed, C. Wilson, H. Wurst, G. Xu, A. Yoshikawa, F. K. Zimmermann, J. G. Sgouros, *Nature (London)* **1992**, *357*, 38.

[30] J. Sulston, Z. Du, K. Thomas, R. Wilson, L. Hillier, R. Staden, N. Halloran, P. Green, J. Thierry-Mieg, L. Qiu, S. Dear, A. Coulson, M. Craxton, R. Durbin, M. Berks, M. Metzstein, T. Hawkins, R. Ainscough, R. Waterston, *Nature (London)* **1992**, *356*, 37.

[31] W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, 'Numerical Recipes', Cambridge University Press, Cambridge, 1986.

[32] C. A. Chatzidimitriou-Dreismann, R. M. F. Streffer, D. Larhammar, *Biochim. Biophys. Acta* **1994**, *1217*, 181; C. A. Chatzidimitriou-Dreismann, R. M. F. Streffer, D. Larhammar, *Eur. J. Biochem.* **1994**, *224*, 365; C. A. Chatzidimitriou-Dreismann, R. M. F. Streffer, D. Larhammar, *Nucleic Acids Res.* **1996**, *24*, 1676.

[33] Y. Nakamura, K. Wada, Y. Wada, H. Doi, S. Kanaya, T. Gojobori, T. Ikemura, *Nucleic Acids Res.* **1996**, *24*, 214.

Received November 7, 1997