Prepared for the International Symposium on Structure and Dynamics of Nucleic Acid, Proteins and Membranes, Rome April 1984 Adenine Press

Further Theoretical Studies of (Nonlinear) Conformational Motions in Double-Helix DNA

J. A. Krumhansl, G. M. Wysin†*, D. M. Alexander, A. Garcia
Laboratory for Atomic and Solid State Physics
Cornell University
Ithaca, NY 14853 U.S.A.

and

P. S. Lomdahl††, Scott P. Layne††
Los Alamos National Laboratory
Los Alamos, NM 87545 U.S.A.

Abstract

We have pursued further the earlier work of Krumhansl and Alexander (KA)[a] in the direction of developing a rationale for the use of a reduced set of variables for describing the large amplitude conformational dynamics and static structure of DNA. Guided by the structure studies of Fratini, et al.[b] the computer experiments of Keepers, et al.[c] NMR data, [d] we have used correlation data and other constraints to argue for the use of a set of conformationally significant variables (CSV), of the order of three per base. Equations of motion have been programmed at Los Alamos for segments up to 200 b.p. in length, and can accommodate arbitrary sequences. Using this program for a model dodecamer we can reproduce and [e] generalize Calladine's results somewhat. For longer chains with suitable parameters we can generate ("soliton") junctions between A and B conformations; whether the parameters used are realistic cannot yet be determined. Moreover, the effects of water and counterions must be added. We comment speculatively on implications in cancer research.

---

† And Center for Nonlinear Studies, Los Alamos National Laboratory.

* Work supported by National Foundation for Cancer Research.
†† Work supported by the U.S. Department of Energy.

(a) J. A. Krumhansl and D. M. Alexander, "Structure and Dynamics: Nucleic Acid and Proteins," Eds. E. Clementi and R. H. Sarma, Academic Press, NY (1983) p. 61.
(b) A. V. Fratini, M. L. Kopka, H. R. Drew and R. E. Dickerson, J. Biol. Chem. 257, 14686 (1982).
(c) J. W. Keepers, P. A. Kollman, P. K. Weiner, T. L. James; Proc. Nat. Acad. Sci. USA 79, 5537 (1982) Biophysics.
(d) See T. L. James; Bull. Magnetic Resonance 4, 119 (1982-83).
(e) C. R. Calladine; J. Mol. Biol. 149, 761 (1981).

Introduction

We are concerned with the static structure or the dynamics of low frequency conformational changes in DNA which can lead to significant topological changes (e.g. transitions among A, B, Z structures). Our objective is to develop a descriptive theoretical understanding which can not only provide a general phenomenological framework for understanding that class of biomolecular dynamics, but also can be utilitarian for modeling and exploring specific problems (such as opening, site selective activity, drug intercalation, radiation effects, promoter and repressor sequences, and radiation effects.) In a more general context, though as yet farther from quantitative molecular biology, we have thought for sometime that long range effects and solitons might provide a substantially new ingredient to consider in the understanding of oncogenes and sequence recognition in DNA research.

We follow the viewpoint [Lomonossoff, Butler and Klug (1981),[1] Dickerson, et al (1981), (1982), (1983)[2]] that conformational structural specificity and dynamics are of significant importance in the function of biopolymers, in addition to chemical sequence specificity.

With respect to the structure and dynamics of biopolymers a traditional approach [e.g. CHARMM, Brooks, et al;[3] AMBER, Wiener and Kollman[4]] which has become widespread with the advent of large computational facilities is to simulate the interaction potentials between all the atoms in a large model molecule and: (a) attempt to derive the equilibrium structure through energy and energy gradient minimization algorithms, (b) study the dynamical excitations about equilibrium structures, i.e. molecular dynamics. However, even with the largest computers this "first principle" [or "brute force"] approach soon reaches its limitations for molecules containing say

more than 100 atoms. Exactly that kind of limitation was found a long time ago in the physics of (condensed) many-body systems. There the solution to this difficulty has been to search (by physical intuition or experimental insight) for a few functionally significant variables (e.g. the "order parameter" introduced by Landau) which relate particularly to a specific class of behaviors which one seeks to describe (e.g. the onset of long range order in a ferromagnet). Formally, there is a procedure called renormalization group theory which has been developed to average over irrelevant degrees of freedom in order to justify the reduction of a many (ca. 10^{23}) body Hamiltonian to an effective Hamiltonian involving only a small number of order parameters (i.e. significant variables). That, and the simpler method of constraint identification together provide in principle the conceptual justification of the order parameter idea.

We follow an analogous philosophy, though we have taken a much more pragmatic approach than a totally deductive theoretical procedure to reduce the many body form of molecular dynamics to significantly fewer variables for the specific purpose of studying metastable states and low frequency global vs. bond dynamics. Instead, we use experimental results to guide a heuristic choice of a few functionally important effective variables -- conformationally significant variables (CSV). This approach, as is usually the case in the use of "order parameters" in condensed matter physics, is partly empirical and lies between full blown molecular dynamics and continuum elastic modeling of DNA; it is less general than the former but more general than the latter. In particular, elastic models cannot deal with strongly nonlinear excitations whereas we can deal with such (e.g. kinks, nucleation, and solitons) relatively easily, formally or computationally, in terms of reduced effective equations of motion. At the same

time the results of static elastic models [Calladine[5] (1982)] may be achieved as a limiting case.

In this conference report our aim is simply to outline the main ideas and initial computer simulation results to illustrate the kind of applications possible. More detailed descriptions of the equation of motion and the parameters therein as used in the simulations will be discussed elsewhere.

Method

We outline: (1) The use of experimental data to reduce general equations of motion; (2) The nature of the reduced equations of motion. The approximate molecular equations of motion sought are to be used to describe asymptotic static variations from an ideal Watson-Crick structure or possible dynamic excitations from those displaced states which may lead to large conformational changes.

Fratini, et al[6] (1982) conclude: "The main elements of flexibility in double-helical DNA are sugar puckering, phosphate orientation, and propellor twist...." There have been numerous NMR studies of the motions in DNA. These are reviewed by James,[7] and a discussion particularly related to backbone motions has been given by Keepers and James.[8] Relaxation rates cover the range from picoseconds to microseconds, but there are particularly important motions in the nanosecond range which cannot be related to helix twisting and bending, tumbling, or the usual small amplitude harmonic motions. It is notable, as we will discuss in further detail in the next section, that the conformational pattern seems to correlate strongly with the static patterns found by Dickerson, et al.

Qualitatively similar behavior is found in computer simulations by Keepers, et al.[9]; by subjecting a simulated structure to various non-ideal imposed conditions and studying the response of the system to these imposed constraints it is found that only a few of the backbone torsional angles are particularly "responsive" (i.e. mobile) in conformational rearrangements.

With only Dickerson's earlier results in mind Krumhansl and Alexander[7] (1983) had postulated a phenomenological (classical) effective Hamiltonian for use in describing the slow and large conformational motions; the choice of our variables represented a qualitative embodyment of the available structural and computer studies. Now, with the above experimental information the rationale for those equations can be made much more systematic as illustrated below in the case of MPD-7. Simply put, in the static and slow motion regime a displacement-displacement correlation analysis gives defined experimental relationships between atom positions; and as such obviously provides constraints which reduce the number of independent degrees of freedom in the equations of motion. In fact, as noted above, the main flexibility of DNA seems to come from only three degrees of motion per base-strand unit. Thus, in a real sense most of the degrees of freedom are either frozen or "slave" to a few significant variables. Because of the non-independence of the variables the choice of responsive coordinates is not unique; in fact, the K-A variables chosen were sugar puckering, backbone extension and base orientation (twist about the helix axis), again three "free" variables per unit. In view of the high degree of correlation between, for example χ and δ, we may hope that the K-A and Dickerson "free" variables represent alternative but equally acceptable choices. In any case the data of Fratini, et al[6] provides further validation of the K-A

equations,[10] which remain the same in form except for one additional term which couples stretching and twisting of the helix, and terms to simulate cross-strand purine interference or the presence of attached complexes.

Rather than lay out the equations in full mathematical detail in this report we outline them functionally:

Hamiltonian = Sum (over base-ribose-phosphate; strands 1, 2) of: (sugar pucker, including possible double well potential) + (backbone stretch; phosphate rotation and base pair separation implied) + (base orientation (twist) rotation about helix axis) + (pucker-stretch coupling) + (pucker-twist coupling) + (stretch-twist coupling) + (steric forces, e.g. Calladine (1982)) + (external forces, e.g. models of intercalants) + (Langevin damping from high frequency modes or temperature bath).

These equations have been programmed generally for oligomers up to 200 base pairs at Los Alamos, for either static or dynamic simulations in response to various initial conditions or steric forces and may accept arbitrary sequences. Temperature effects are present only through damping.

Application and Results

We report initial results of simulations on the dodecamer MPD-7 and on a 50 b.p. model system; the modified K-A model produces the kind of static structural variations along the helix found by Dickerson on MPD-7 and it simulates a B-A junction on the 50 b.p. model.

(a) MPD-7

Obvious constraints are: Bond lengths constant, bond angles constant. Data shows that several backbone angles are essentially constant; approximately, $\alpha \approx -65°$, $\beta \approx 175°$, $\gamma \approx 175°$, $\varepsilon \approx -170°$ over a wide range of phosphate and ribose configurations. Several variables are strongly correlated: δ and glycosyl χ_0; δ and backbone angle ζ; rise "h" and global twist "t". A statement that δ varies is now widely accepted as equivalent to sugar pucker variation. Then the correlation between δ and ζ is equivalent to the conclusion of Keepers and James[8] that the ^{31}P NMR results can be explained by simultaneous pucker and rotational jumps about the C3'-O3' bond. It is noteworthy that the experimental constraint relationships found by Dickerson in static X-ray measurements seem to be identical to those holding in the nanosecond range, an important proviso to our use of CSV for both static and dynamic variations.

The above refer to backbone and pucker within one ribose-phosphodiester segment; in addition we introduce an angle ϕ which measures the rotation of this segment and its attached base about the helix axis; finally we introduce the coordinate z which measures displacement of the segment parallel to the axis. These are effectively the K-A CSV. As noted below it may eventually be necessary to introduce a fourth variable to describe the B-Z transition.

We imposed cross strand purine-purine forces in the same manner as Calladine. The computer program allows us also to vary parameters such as torsional stiffness at the ends, and to introduce additional forces (due for example to Br on the MPD-7). Both static and dynamic simulations were carried out. Initially the dodecamer was placed in the ideal Watson-Crick structure; small random initial displacements were applied and the system allowed to run through about 50-100 time units, with a damping of 10 inverse time units to simulate viscous or temperature bath effects. Figure 1(a) shows the dynamic history of the interbase twist "DEL-PHI" (similar to Dickerson's t_i) vs. base pair number [Time units are the transit time for sound to travel from one base to its neighbor]. Figure 1(b) shows the asymptotic static limit. An attempt to model the Br on cytosine is illustrated. In this run the torsional stiffness at the ends was reduced to half that in the mid-region. The result closely resembles that found by Calladine's rules; this is not surprising because the reduced equations can simulate elastic behavior. These runs took only a few seconds. We are studying the effects of bromine and differences between A-A and G-G interactions, and eventually will explore the conformation of arbitrary sequences.

(b) <u>50 b.p. Model</u>

In this series of runs the principal parameters in the Hamiltonian were the same as for the MPD-7 run, but no steric interactions were assumed. The pucker potential (or equivalently for δ) was a double well with barrier height ≈ 3K cal; one well represents the A-form, the other B, and in the run shown they are taken to have the same energy. This is the condition for B→A transition. Figure 2(a) shows the time history of a B-A domain boundary (i.e. soliton) imposed on the system at t=0 (Antiperiodic

boundary conditions). The main result of this run is to show that although various transient oscillations are set up the B-A interface is asymptotiently stable within our equations of motion which include damping. Figure 2(a) depicts the sugar pucker and Figure 2(b) shows how the twist varies with time: It is quite clear that there are two distinctly different motions: (a) a torsional-stretching oscillation of the entire chain (low frequency but dynamic); and (b) a long lived "kink" in which sugar pucker, twist and stretch maintain a very characteristic pattern, which is not removed by damping. This independence of normal mode damped harmonic-like motions and solitary long-lived excitations is generic in nonlinear soliton bearing systems. There may also be an oscillatory "breathing" component within the kink region.

In either case, the runs took only a small amount of computer time, a few seconds for the dodecamer and no more than a minute for the longer helix, including graphic output.

Discussion

In the most general sense what these considerations mean is that significant features of DNA conformations and their relation to sequence are related in a systematic way to physical considerations and may be determined by suitable equations of motion. Every new structure need not be determined by X-ray methods if this method is successful.

It should be clear also that this theoretically approximate method goes considerably beyond elastic models of polynucleotides, yet for the purposes for which it is designed it should be a cheap and quick tool for analyzing and planning experiments. We envisage that a library of effective parameters for the reduced equations might be built up empirically in a systematic joint experimental-analytical (formal theory or simulation) program involving X-ray, NMR, Raman, IR and other studies on prototypical oligomers, enzymes, or other biomolecules.

At the same time the formulation arrived at here has the features of a number of models developed in theoretical condensed matter physics over the past decade, suggesting that a useful interdisciplinary transfer of concepts may take place. For example, a model for the production of Z regions in B due to negative supercoiling falls naturally out of an analogy with strain-induced discommensurations in epitaxial layers. Another concept is that of a generalized order parameter which may have several components. It seems to us that many such parallelisms can be found and profitably exploited.

We note briefly some preliminary thoughts on the B-Z transition. The most significant structural feature of Z-DNA is that the bases have been rolled over from the normal **syn** to **anti** configurations, and the basic unit is a dinucleoside. Earlier in the discussion we noted the possible need

for an additional fourth coordinate for each backbone segment. Perhaps this should be the glycosyl angle, and physically the effective potential for that motion might be something like a double well (corresponding to **syn** or **anti**). Ordinarily, the hump is too difficult to surmount, but salt and strong (G-C) sequence forces might drive it. If this is the case we have a most intriguing dynamic and static problem: two anharmonic bistable systems, ribose and glycosyl, coupled to backbone twist and helix extension. It is not at all beyond the capability of the present method to model this situation; we plan to do so and see whether a B-Z junction can be found in this manner.

Finally, in view of our sponsorship by the National Foundation for Cancer Research we offer a few speculations about possible implications regarding cancer initiation at the molecular level. Fundamentally, what our work suggests is that in addition to the more conventional modes of atomic vibrations there may be both anharmonically localized conformations and long range effects due to the motions of solitary (soliton) conformational excitations. These might, for example, be initiated by a carcinogen at one point on the DNA helix, diffuse or be driven chemically or by coil stress to another region where they might substantially conspire with an oncogenic sequence to direct cancer cell production. Possibly more immediately accessible is the better understanding of the role of anti-tumor drugs. However, much before one pursues such ideas too far we must plod along developing experience on simpler, well defined molecular problems.

We wish to acknowledge a number of stimulating discussions in the course of developing these ideas with: C. R. Calladine; S. W. Englander, N. R. Kallenbach, and A. J. Heeger; A. Klug; A. C. Scott; H. M. Sobell.

Institutional support has been received from the British Science Research Council, the National Foundation for Cancer Research, and Los Alamos National Laboratory.

References

1. G. P. Lomonossoff, P. J. G. Butler, A. Klug; J. Mol. Biol. 149, 745 (1981).

2. R. E. Dickerson, and H. R. Drew; J. Mol. Biol. 149, 761 (1981).

   R. E. Dickerson, H. R. Drew, B. N. Cooper, R. M. Wing, A. V. Fratini and M. L. Kopka; Science 216, 475 (1982).

   R. E. Dickerson; Scientific American, p. 94, December (1983).

3. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathanand M. Karplus; J. Comput. Chem. 4, 187 (1983).

4. P. K. Wiener and P. A. Kollman; J. Comput. Chem. 2, 287 (1983).

5. C. R. Calladine; J. Mol. Biol. 161, 343 (1982).

6. A. V. Fratini, M. L. Kopka, H. R. Drew and R. E. Dickerson; J. Biol. Chem. 257 14686 (1982).

7. ref. (d) of Abstract

8. J. W. Keepers and T. L. James, J. Am. Chem. Soc. 104, 929 (1982).

9. ref. (c) of Abstract

10. J. A. Krumhansl and D. M. Alexander; "Structure and Dynamics: Nucleic Acid and Proteins," Eds. E. Clementi and R. H. Sarma, Academic Press, NY (1983), p. 61.

Figure Captions

Figure 1: (a) Time evolution of a twist pattern driven by purine-purine forces from a uniform helical twist (DEL-PHI=0) at t=0 toward a Calladine-like pattern at 40 time units.

(b) The static limit of the variations in twist along the chain. In both cases the model represents MPD-7; in one case only purine-purine forces were included, then the effect of Br was simulated by adding additional forces.

Figure 2: (a) Pucker; and

(b) Twist variation along a uniform 50 b.p. model helix due to the introduction of a kink (i.e. B-A boundary) at t=0. Note the oscillatory transits which eventually damp out while the kink remains essentially unaffected.

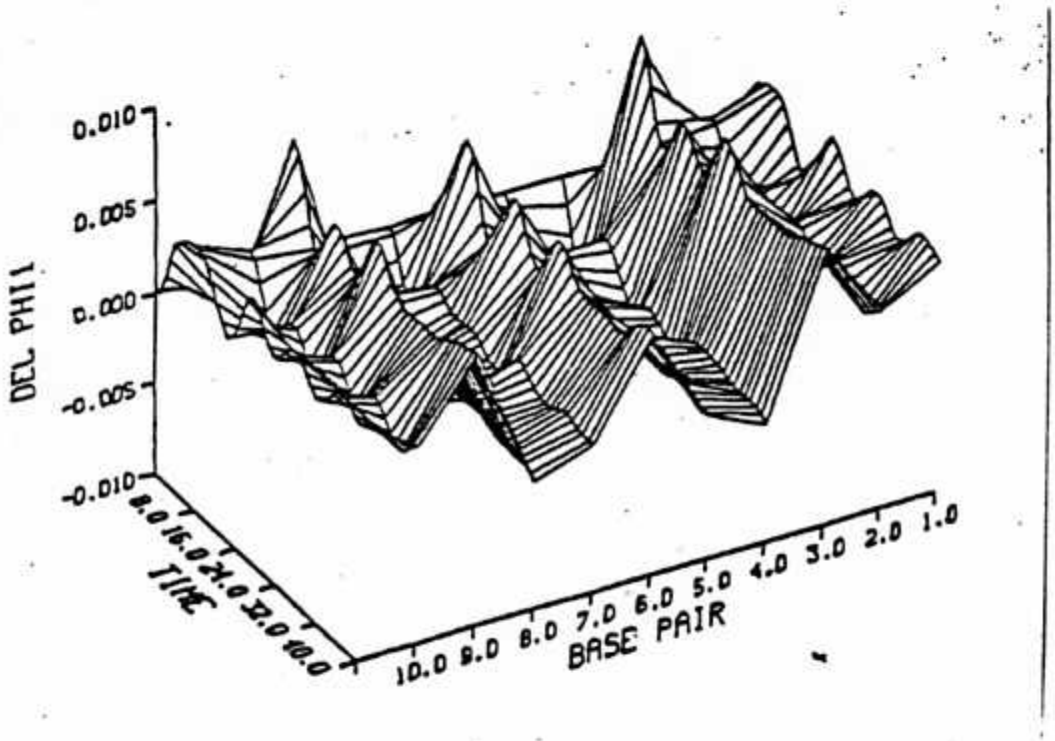
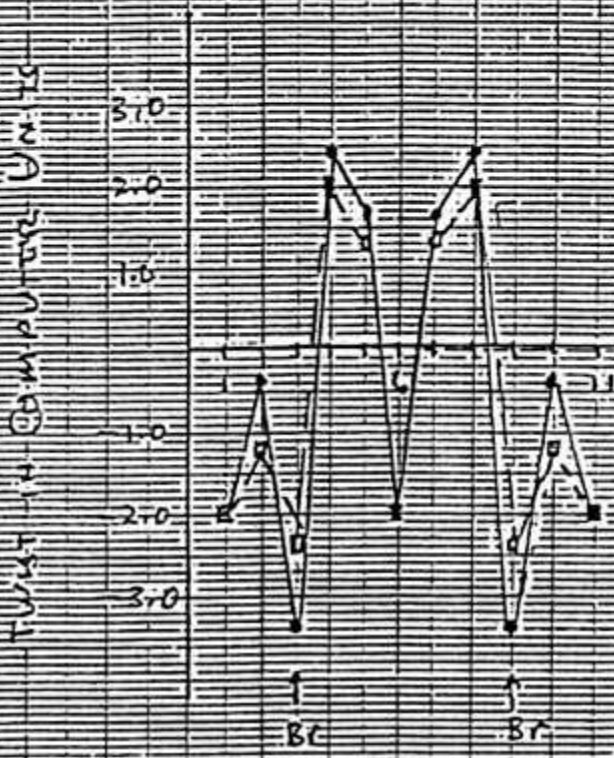

DEL PHI1
0.010
0.005
0.000
-0.005
-0.010
0.0 16.0 24.0 32.0 40.0
TIME
10.0 9.0 8.0 7.0 6.0 5.0 4.0 3.0 2.0 1.0
BASE PAIR

Fig 1 (a)

Fig 1(b)



N = BASE STEP

PURINE-FORCES ONLY
[WYSIWYG]

BROMINE-PLUS PURINE FORCES
[WYSIWYG]

Fig. 1(b)

FIG. 1 COMPUTER MODELING OF ADDING BROMINE FORCES TO PURINE STERIC INTERFERENCE IN MFD-7.

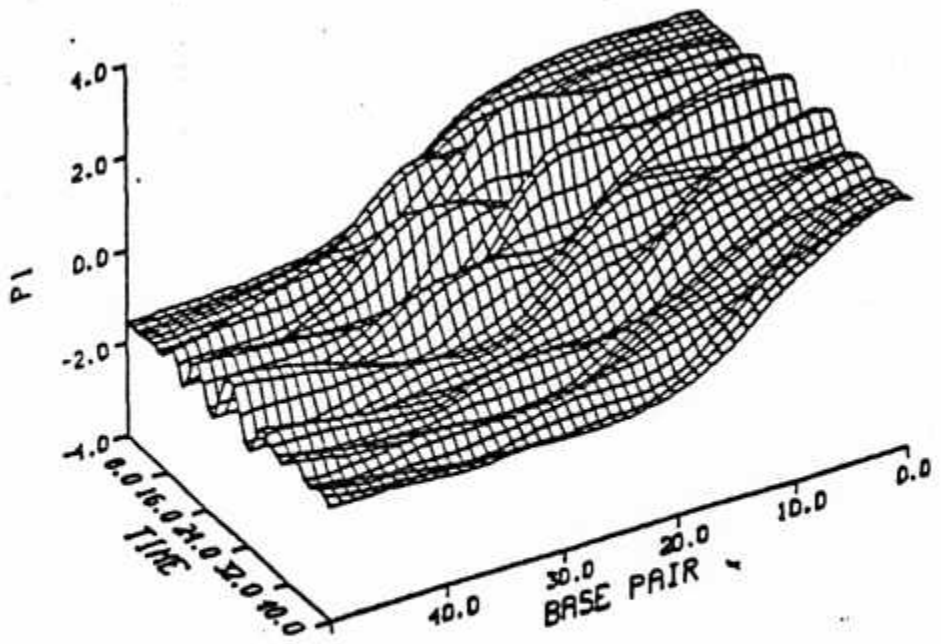

4.0
2.0
0.0
-2.0
-4.0
PI
8.0
16.0
24.0
32.0
40.0
TIME
40.0
30.0
20.0
10.0
0.0
BASE PAIR

Fig 2(a)

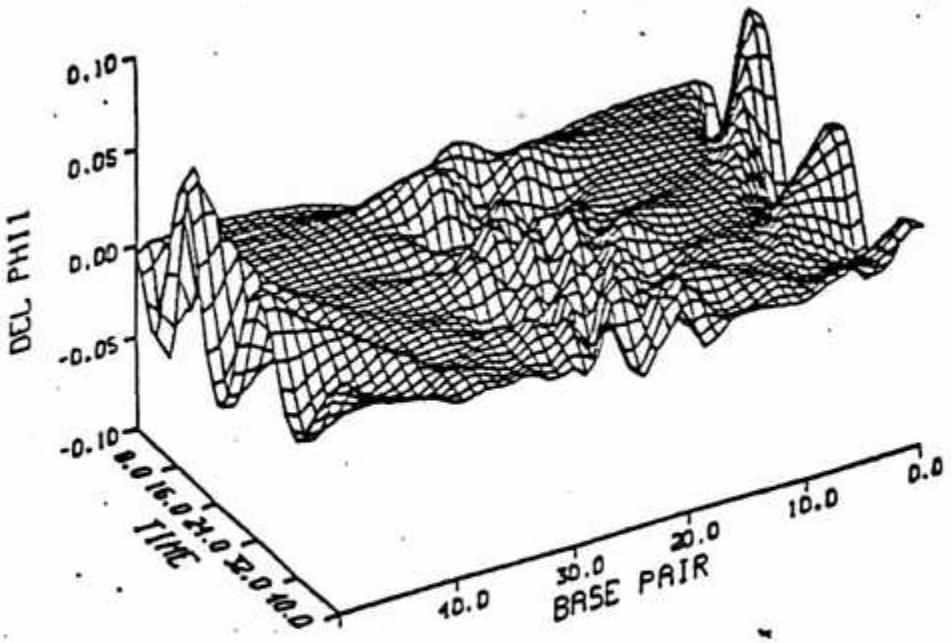

0.10
0.05
0.00
-0.05
-0.10
DEL PHI1
8.0
16.0
24.0
32.0
40.0
TIME
40.0
30.0
20.0
10.0
0.0
BASE PAIR

Fig 2(b)