

- [1] V.N. Bolotov, L.N. Gerdyukov, S.S. Gershtein, et al., Preprint SEF/STF/OP/SKU 68 - 56 - V, Institute of High Energy Physics, Serpukhov, 1968.
- [2] E.D. Zhizhin, Yu.P. Nikitin, G.V. Rozhnov, and E.P. Shabalin, *Yad. Fiz.* 14, 420 (1971) [*Sov. J. Nucl. Phys.* 14, 235 (1972)]. A.V. Berkov, E.D. Zhizhin, Yu.P. Nikitin, and E.P. Shabalin, *ibid.* 14, 422 (1971) [14, 236 (1972)].
- [3] V.I. Voronov, I.A. Danil'chenko, R.A. Rzaev, and A.V. Samoilov, Preprint OP/SPK 70 - 93, Institute of High Energy Physics, Serpukhov, 1970.
- [4] V.V. Solov'ev and I.S. Tsukerman, *Zh. Eksp. Teor. Fiz.* 42, 1252 (1962) [*Sov. Phys.-JETP* 15, 868 (1962)].

THE INVERSE PROBLEM FOR DNA

M.Ya. Azbel'

L.D. Landau Institute of Theoretical Physics, USSR Academy of Sciences

Submitted 19 June 1972

ZhETF Pis. Red. 16, No. 3, 183 - 186 (5 August 1972)

1. To study the character of the information recorded in DNA, it is necessary to ascertain the regularity of the sequence of the nucleotides in it. In DNA, the large surface free energy V on the boundaries between long melted and helical states (MS and HS) makes it possible to determine this regularity from the melting curve of DNA with stabilizers. At $V = \infty$, the melting of DNA would correspond to a first-order phase transition at a temperature T_{tr} near which $E_{\sigma} = \sigma \alpha_{\sigma} (-T_{tr} + T)$, where E_{σ} is the free energy per link, $\sigma = 1, -1$, and 0 correspond to the MS, HS, and the melting point, and α_{σ} and T_{tr} depend on the composition of the DNA. The term independent of σ has been omitted from E_{σ} . At large V at the beginning of the melting, when $T > T_{AT}^{1)}$ (where T_{AT} is the melting temperature of the homo-AT), the significant (long) sections are so enriched with the easy-melting AT pair, that the given value of T coincides with their melting temperature \bar{T} , at which (taking into account the surface energy) the free energies of the MS and HS are equal. (We shall call these s-sections, and their states s-states. All the remaining sections are located below "their own" melting temperatures and, accurate to within the entropy term (see below) are helical.) The probability of such an appreciable fluctuation of the composition decreases rapidly with increasing length of the section, and therefore in the principal approximation the wings of the melting curve are determined directly by sections of definite length and composition. This indeed makes it possible to determine their probability. The accuracy of such a determination, however, has a certain limit (determined by the relative distance from T_{tr} , i.e., by the value of $1/V$, and this accuracy governs the extent to which the thermodynamic quantities are independent of the detailed form of the nucleotide sequency and the solution of the inverse problem is stable.

2. We shall therefore consider only the wings of the curve (for concreteness, at $T < T_{tr}$) and only with the indicated accuracy. On the wings (see above) the change of the total free energy E' (per link) in comparison with \tilde{E} for "pure" hs DNA is connected only with the appearance of the S-section. If the probability of an S-section of length λ is w_{λ} , with $w = \sum w_{\lambda} \ll 1$, then in the principal approximation such sections are independent, do not intersect, and form on the DNA a weak "solid" solution (since the sequence of the nucleotides remains unchanged). (Since $\lambda \gg 1$, we always have $w_{\lambda} \ll 1$, even at T_{tr} .) This

¹⁾ AT and GC are pairs of nitrogenous bases, adenine-thiamine and guanine-cytosine, respectively; the stabilizers ("clips") are low-molecular impurities in the solution (see [1]).

means that when $m_\lambda w_\lambda \ll |\ln w|$ (m_λ is the characteristic correlation length on the section λ , w_λ^{-1} is the average distance between the sections λ), with relative accuracy w ,

$$E = E - \tilde{E} = \sum w_\lambda (E_\lambda^S - E_\lambda^M), \quad (1)$$

where $E_\lambda^{S,H,M}$ correspond to the surface, helical, and melted state for a surface section of length λ . The surface terms (see above) are included in E_λ^M (which enters in E_λ^S). Since the melting between the HS of a section of length λ produce (per link), besides the surface energy $-2V/\lambda$ (the dependence of which on the composition of the MS and HS can be neglected accurate to λ^{-1}), also a non-additive entropy $-S(\lambda) \sim -\ln \lambda$ (see [1], it follows that $E_\lambda^H \approx E_\lambda^M$ at a melting temperature $T_\lambda = T - 2\tilde{V}[\lambda(\alpha_+ + \alpha_-)]^{-1} \equiv T - \tau/\lambda$; $\tilde{V} = V + S(\lambda)/2$, with $\lambda \geq \lambda_0$, where $T_{\lambda_0} = T_{AT}$. The composition of the surface section is determined by $T_\lambda < \tilde{T} < T_{\lambda+1}$. Therefore, if $c(\lambda, \tilde{T})d\tilde{T}$ is the probability that the composition of a section of length λ is such as to ensure a melting temperature in the interval $d\tilde{T}$, then $w_\lambda = c(\lambda, T_\lambda)\tau/\lambda^2$. Since $w_\lambda \ll 1$, it suffices to write down E_λ^S in the zeroth approximation, i.e., as $V \rightarrow \infty$, meaning also as $\lambda \rightarrow \infty$ (this is also possible for E_λ^H , but not advantageous). As a result we obtain:

$$-E' \sim \int_0^\infty \alpha_+ r^2 c(\lambda, T_\lambda) \frac{d\lambda}{\lambda^3} \sim c(\lambda^*, T^*); \quad T^* = T - \frac{\tau(\lambda^*)}{\lambda^*};$$

$$\lambda^* = \max(\lambda_0, \tilde{\lambda}); \quad dc/d\tilde{\lambda} = 0 \quad (2)$$

The accuracy of this formula, in connection with the transition $\lambda^* \rightarrow \infty$ in E_λ^S , is determined by $\beta = m_{\lambda^*}/\lambda^*$. If $\beta \gtrsim 1$ and $c(\lambda, \tilde{T})$ decreases rapidly with increasing $\tilde{T} - T_\lambda$, then E' is determined in order of magnitude, and $\ln|E'|$ is determined with a relative accuracy $|\ln|E'|^{-1}$. In the more general case we have

$$E' \sim \max_\lambda \int_{T_\lambda}^{T_{AT}} (\tilde{T} - T_\lambda) c(\lambda, \tilde{T}) d\tilde{T} = f(\lambda_m, T_{\lambda_m})$$

(cf. (2)). Since $w \gtrsim 1$, such an accuracy is maintained for all values of T . (In the particular case of a linear dependence of E_σ on the concentration of the AT pairs, we obtain the results of [2] at $\tilde{\lambda} > \lambda_0$ for a random sequence and with the same accuracy, and $\ln E \sim T - T_{tr}$.) At $V \rightarrow \infty$ the value, as $\lambda \rightarrow \infty$ become important, corresponding to a random sequence and an exponential tendency of $c(\lambda, x_\lambda)$ to zero. It is therefore natural to put $c(\lambda, x_\lambda) = \exp[-\phi(\lambda, x_\lambda)]$; this enables us to calculate the pre-exponential factor in the formula for E' .

3. Let $E' = E'(T, \tau)$ be known. Then, taking the definition of $\tilde{\lambda}$ into account, we obtain the physically clear formula $\lambda' = -[\partial E'/\partial T][\partial E'/\partial \tau]^{-1} = h(T, \tau)$. Solving this equation in conjunction with $\tilde{T} = T - \tau(h)/h$ relative to T and τ as functions of λ and \tilde{T} , we obtain $c(\lambda, \tilde{T})$ in the zeroth approximation as a result of a solution of an algebraic problem. The higher terms of the expansion of E' will give the corresponding approximations for c . \tilde{T} itself is determined only by the MS and HS, and is expressed in a definite manner, depending on the

accuracy, in terms of the concentrations of the elements (AT and GC), the pairs of neighboring elements, etc. Therefore, at the indicated limitations, the details of the information obtained concerning the structure of DNA are limited only by the measurement accuracy. The determination of the form of the function $E'(T, \tau)$ reduces to a replacement of the variables in the experimentally obtained dependence of the free energy of the DNA on T and the concentration ξ of the low-molecular impurities (the stabilizers) in the solution, since the stabilizers, which have throughout the DNA a constant chemical potential equal to their chemical potential $\mu(T, \xi)$ in the solution, renormalize α_+ and α_- in an obvious manner; for non-interacting stabilizers

$$E_{\sigma\alpha} \rightarrow E_{\sigma\alpha} - \ln [1 + \exp(\mu - \mathcal{E}_{\sigma\alpha})/T] \quad (3)$$

($\mathcal{E}_{\sigma\alpha}$ is the energy introduced by the stabilizer into the link of type α). To obtain more information concerning the DNA structure it is necessary to take into account the dependence of the interaction energy of the neighboring pairs on their type, meaning also that it is necessary to use simultaneously a corresponding number of different stabilizers. When $\bar{\lambda} < \lambda_0$, the determination of $c(\lambda, T_{AT})$ is obvious and does not require the use of stabilizers. The method is very sensitive to the structure of the DNA. Thus, if $c(\lambda, T_{AT}) = 0$ at $\lambda > \lambda_g$, then a sharp decrease of $|E'|$ occurs at $\lambda_0 > \lambda_g$.

4. The presence of a block structure [1] can be easily taken into account by introducing the distribution function of the blocks with respect to the concentrations, and leads only to a replacement of T_λ by its smallest value in the blocks. To solve the inverse problem it is convenient, of course, first to subdivide the DNA into blocks of the same type and to perform the measurements just for these blocks, using an ensemble of identical blocks, taken from DNA of different representatives of the same biological species, genus, etc. in order to study the character of the recorded information.

5. We now take into account the entropy increment to E' , which is unique at $T < T_{AT}$ (and $T > T_{GC}$). It is determined (since V is large) by a small number of short melting sections. In the region where the energy gain is non-existent or negligibly small, the melting-sections collide infrequently and do not tend to combine (as in the case when $T_{AT} < T$) into long sections. Therefore the probabilities $q_{\lambda m} = c_{\lambda m} \exp[(\lambda E - E_{\lambda m})T^{-1}]$ of the melting sections of length $\lambda \geq 1$ and with concentrations $x = m/\lambda$ ($E_{\lambda m}$ corresponds to the MS) can be regarded as independent, and the probability of the helical section (which is known to be very long) can be assumed equal to $q_s = \exp(+E'/T)$. Hence $\sum q_{\lambda m} q_s = 1 - q_s$ (for a homopolymer, where there is only an entropy increment, the q_λ are always strictly independent, and the obtained formula is exact). From $E' \sim \exp(-2V)$ (cf. with Sec. 3) we can reconstruct $|\ln|E'|$ of the first coefficients $c_{\lambda m}$. In essence, it is precisely the possibility of separating the pure "entropy" and the pure "energy" regions at $V \gg 1$ which makes the solution constructive.

6. The need for stability of the linear reading of the information in the DNA probably gives rise to a considerable redundancy. This can lead to the existence of "morphemes" - segments on which maxima of $c_{\gamma_1 \gamma_2 \dots \gamma_k} / c_{\gamma_1 \dots \gamma_{k-1}}$ and $c_{\gamma_1 \dots \gamma_k} / c_{\gamma_1 \dots \gamma_k}$ are realized ($c_{\gamma_1 \dots \gamma_k}$ is the concentration of the sequence $\gamma_1 \dots \gamma_k$) at the locally largest value of k . If the redundance is sufficient, they can be separated by studying the melting curve. (In the "ideal

case" $c_{\lambda m} = c_{\lambda+1, m+1}$ or $c_{\lambda m} = c_{\lambda+1, m}$.) A study of sufficiently large λ would make it possible to determine the probability that different morphemes are neighbors. The resultant analogy with a language text have been discussed in [3]. In connection with the foregoing, it is of interest to calculate with a computer the melting point of a sequence of pairs specified by means of the text of a sufficiently long book expressed in binary form, and to investigate the inverse problem for this case.

I am deeply grateful to I.M. Lifshitz for very useful discussions, and also to A.I. Larkin and M.D. Frank-Kamenetskii.

- [1] A.A. Vedenov, A.M. Dykhne, and M.D. Frank-Kamenetskii, Usp. Fiz. Nauk 105, 479 (1971) [Sov. Phys.-Usp. 14, 715 (1972)].
- [2] A.A. Vedenov and A.M. Dykhne, Zh. Eksp. Teor. Fiz. 55, 357 (1968) [Sov. Phys.-JETP 28, 187 (1969)].
- [3] M.Ya. Azbel, Ideen des Exakten Wissens, No. 12 (1972)

LOW-FREQUENCY CYCLOTRON RESONANCE IN AN IDEAL METALLIC CRYSTAL

S.M. Cheremisin

Institute of Physics Problems, USSR Academy of Sciences

Submitted 22 June 1972

ZhETF Pis. Red. 16, No. 3, 186 - 190 (5 August 1972)

The progress attained recently in the production of perfect single crystals and in the mastery of the millimeter radio band has made it possible to observe cyclotron resonance in metals under conditions when the free path time is determined not by scattering from the impurities and defects of the crystal, but by the interaction of the resonant electrons with other quasiparticles, primarily with phonons. The most detailed studies were made on bismuth [1] and lead [2, 3].

Sher and Holstein have constructed a rigorous theory [4] that enables us to calculate the cyclotron-resonance line width determined by the electron-phonon interaction. The electromagnetic field frequency was assumed by them to be very high, $\omega = n\omega_c \sim \omega_D$ (ω_c is the cyclotron frequency and ω_D the Debye frequency). The relaxation time in this case is $\tau \sim 1/\omega$ and the influence of the Landau quantization on the scattering can be disregarded even at $T = 0$. Actually, however, the experiments are performed at $\omega \ll \omega_D$ and $\omega\tau \gg 1$, and, as will be shown below, at sufficiently low frequency the relaxation time of the resonant electrons may turn out to be much larger than expected from the theory of [4].

Let us consider a quasiclassical model of electron-phonon interaction at $kT \ll \hbar\omega$. The main process determining the relaxation time will in this case be, as is well known, the spontaneous emission of phonons by the excited resonant electrons. (The residual time of the relaxation determined by the collisions of the electrons with the impurities and defects of the crystal is not taken into account.) The electron energy near the (convex) Fermi surface is given by the formula

$$E = (N + 1/2)\hbar\omega_c + p_z^2/2m_H$$

The energy of an electron having $p_z = 0$ changes by $\Delta E = \Delta N\hbar\omega_c - q_z^2/2m_H$ following the emission of a phonon (q_z is the projection of the phonon momentum on the magnetic-field direction).