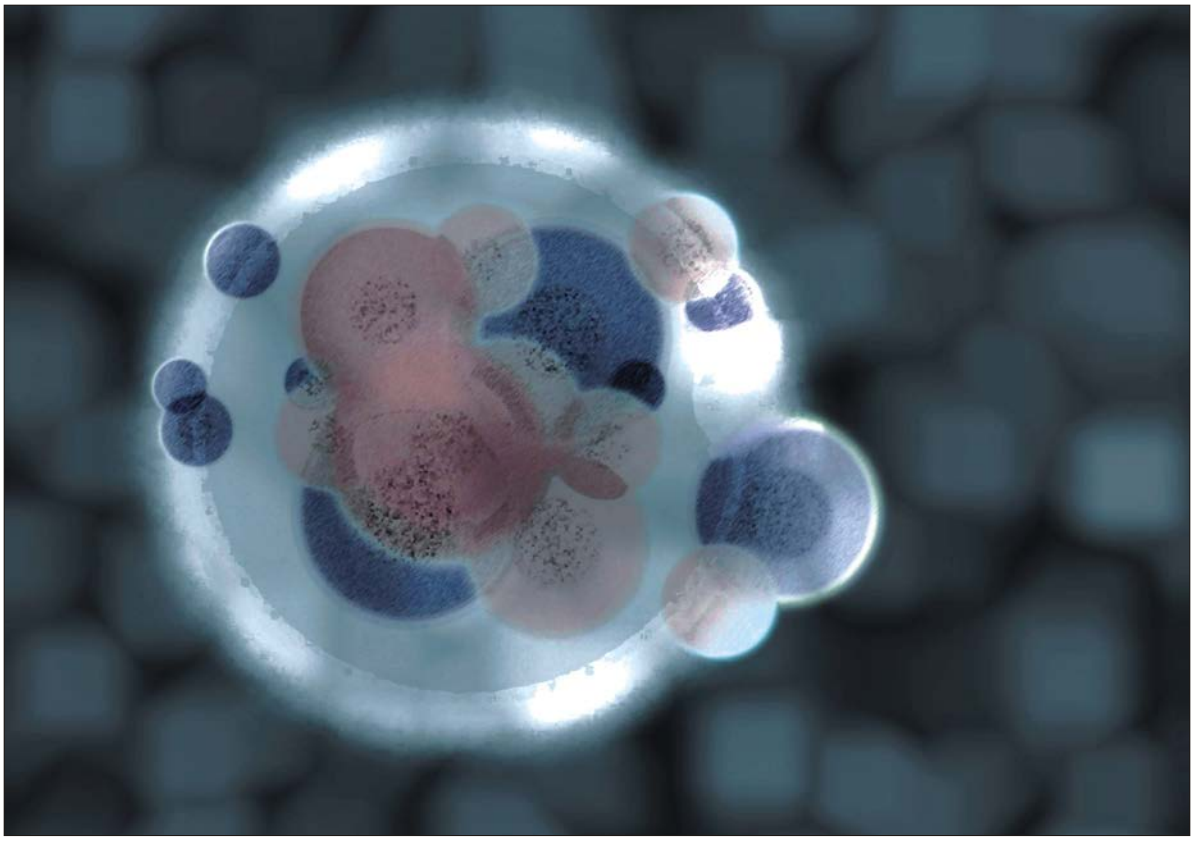

Journal of Biological Sciences

ISSN 1727-3048

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Biological Sciences 6 (3): 532-539, 2006
ISSN 1727-3048

Is RNA Synthesis Possible Without Transcription Bubble Formation?

Yu. P. Petrov
Department of Cell Culture, Institute of Cytology,
Russian Academy of Sciences, St. Petersburg, Russia

Abstract: A computer model for transcription along the major groove of DNA without formation of the transcription bubble is presented. The model is based on the following assumptions. (1) Transcription should occur without alteration of the DNA B-form. (2) The DNA template strand and the transcribed RNA are antiparallel. (3) Specificity of the transcribed RNA is exclusively determined by hydrogen bonds; the number of the DNA-RNA hydrogen bonds is one per triplet. (4) The distance from the sugar-phosphate backbone to the DNA axis does not dependent on the ribonucleotide size. (5) Such structure is stereochemically possible only provided that the transcribed RNA fragment contains ribonucleotide diphosphates. As the customary RNA with a monophosphatic backbone cannot be synthesized under such conditions, formation of an intermediate referred to as the RNA transcription fragment (tf-RNA) is assumed. Advantages of the presented model over the established hypothesis of the local DNA melting during transcription are discussed.

Key words: Transcription, computer model, RNA

INTRODUCTION

The basic unit of DNA is a polymer of four different monomeric subunits, deoxyribonucleotides, arranged in a precise linear sequence. Two polymeric strands are twisted around each other to form the DNA double helix, in which every subunit in one strand pairs specifically with the complementary subunit in the opposite strand. The DNA structure is found in several forms. For a random-sequence DNA molecule under physiological conditions the B form is a most stable structure. The purine and pyrimidine bases of both strands are stacked inside the double helix, with their hydrophobic and nearly planar ring structures very close together and perpendicular to the long axis of the helix. Each base of one strand is paired in the same plane with a base of the other strand. The spatial relationship between these strands creates a major groove and a minor groove between the two strands.

This elegant and stable structure is disturbed during DNA replication which is impossible to represent without DNA melting. In analogy to replication, the idea of local melting of DNA during transcription had appeared. This idea assumes that RNA is synthesized as a part of so-called transcription bubble (Coulombe and Burton, 1999; Nikolov and Burley, 1997; Spitalny and Thomm, 2003). Today, this is an established opinion.

The idea of the transcription bubble is widely accepted and confirmed by a large body of experimental data. There are, however, the data that are hardly consistent with this concept or cannot be interpreted with its help. Some of these problems are listed below.

- During transcription thymine is used instead of uracil. Is this replacement of any functional relevance?
- Under physiological conditions DNA is in an energetically favourable B-form. The DNA-RNA hybrid cannot adopt the B-form because of the presence of the OH-group at the C2'-atom of ribose. Therefore the «RNA-readout DNA strand» pair of transcription bubble is in the A-form in which the number of base pairs per one turn of helix will be 11 rather than 10. This should increase tension in the DNA-RNA transcription complex. Moreover, the non-readout DNA strand should be placed aside and for stereochemical reasons become less coiled. This should increase the molecular tensions in DNA, too.
- What is the function of non-readout strand during transcription? What is a spatial organization of this strand in the transcription complex?
- Why is not the second strand readout simultaneously?
- It is known that width of the RNA polymerase cleft, in which the transcription bubble is presumably located, is 25 nm (Ebright, 2000). However, the bubble dimension (three strands) is of 30 nm.
- Classical transcription patterns of ribosomal genes (Morgan *et al*., 1983) show up to 120 RNA transcribed simultaneously. Such DNA segment

contains about 5000 bp. Therefore, one transcription complex contains approximately 40 pairs of nucleotides. It is known that in the transcription bubble the DNA-RNA complexes are formed simultaneously by10-12 bp. If one would assume that 5-6 bp should be melted spatially upstream and downstream from these complexes at the same time, then about half of DNA should be melted in ribosomal gene region. This suggests a high input of energy to break hydrogen bonds and an additional stress between the transcription bubbles.

- Although hypothesis about molecular structure of the transcription bubble is widely accepted and the potent computer programs are available the author could not find any data on molecular modelling of the transcription bubble.

These facts show that the mechanism of RNA synthesis via the transcription bubble formation is far of being comprehensive. It is, therefore, reasonable to assume that transcription can occur via a more saved and energetically favourable mechanism. These considerations have induced the author to construct an alternative model of transcription. As the presented computer model of transcription has no special experimental proof the author considers it as an invitation for discussion.

MATERIALS AND METHODS

The computer models for DNA and RNA fragments were taken from the Nucleic Acid Database (NDB, The IMB Jena Image Library of Biological Macromolecules-http://www.imb-jena.de/IMAGE.html). The canonical B-form of a double helix of DNA formed by the decamer oligonucleotide 5'-d(ACAAGAACTA)-3' was used. The model was constructed manually by using the program for molecular modeling and analysis of 3D-images-«Chem3D Ultra» (Cambridge Soft Corporation, USA).

RESULTS

The presented model is based on the following general principles:

- The transcription occurs without DNA melting.
- RNA is transcribed in the major groove of the DNA double helix.
- Formation of the transcription complex should not disturb B-form of DNA.
- The model should give satisfactory answers to the questions put above.
- The model should not contradict the experimental data that confirm the transcription bubble model.

Design of triplets: During transcription a correct readout along the major groove of B-DNA is possible only if the third integrated strand (RNA strand) has the same nucleotide sequence as the nontemplate DNA strand. Stability of the B-DNA is determined by the base-stacking interactions. The stacking interactions in which bases are positioned so that the planes of their rings are parallel (similar to a stack of coins) represent an important mode of interaction between two bases. The stacking involves a combination of van der Waals and dipole-dipole interactions between the bases. These base-stacking interactions are very important in stabilizing the three-dimentional structure of nucleic acids. However, the specificity of a given base sequence in each DNA strand is maintained entirely by hydrogen bonding between the base pairs. How many of such bonds are necessary for recognition? A minimum number is probably one bond per one base pair. During transcription, it is desirable that the number of hydrogen bonds of the third helix (RNA) should be less than the number of such bonds between the DNA strands. This would be important to avoid a certain energy barrier at detachment of the transcribed RNA strand from DNA. These assumptions are the key point for choice of specificity of hydrogen bonds during transcription.

Specificity of pairing bases via hydrogen bonds is not enough, however, for construction of the transcription model along the major groove of DNA. It should be taken into account that the sugar-phosphate backbone of the third strand should be located at the same distance from the DNA axis. Building of this structure is not trivial. Four pairs of the triplets used in modelling of the DNA-RNA triple helix are shown in Fig. 1. It should be noted that in all four variants only one hydrogen bond was used for recognition. This bond is formed between a keto form of oxygen and nitrogen. Formation of the same hydrogen bond determines an equiprobable (in the energy sense) selection of the bases of the third strand. Therefore, the probability of a specific pairing of the bases should increase. Ultimately, this gives 4 independent combinations. According to the calculated length, the hydrogen bond for all bases of the third helix is the same, 2.98±0.01 Å. This value is in a perfect agreement with one known for the hydrogen bond of this type (O-H…O-2.7 Å, O-H…N-2.9 Å, N-H…O-3.0 Å, N-H…N-3.1 Å).

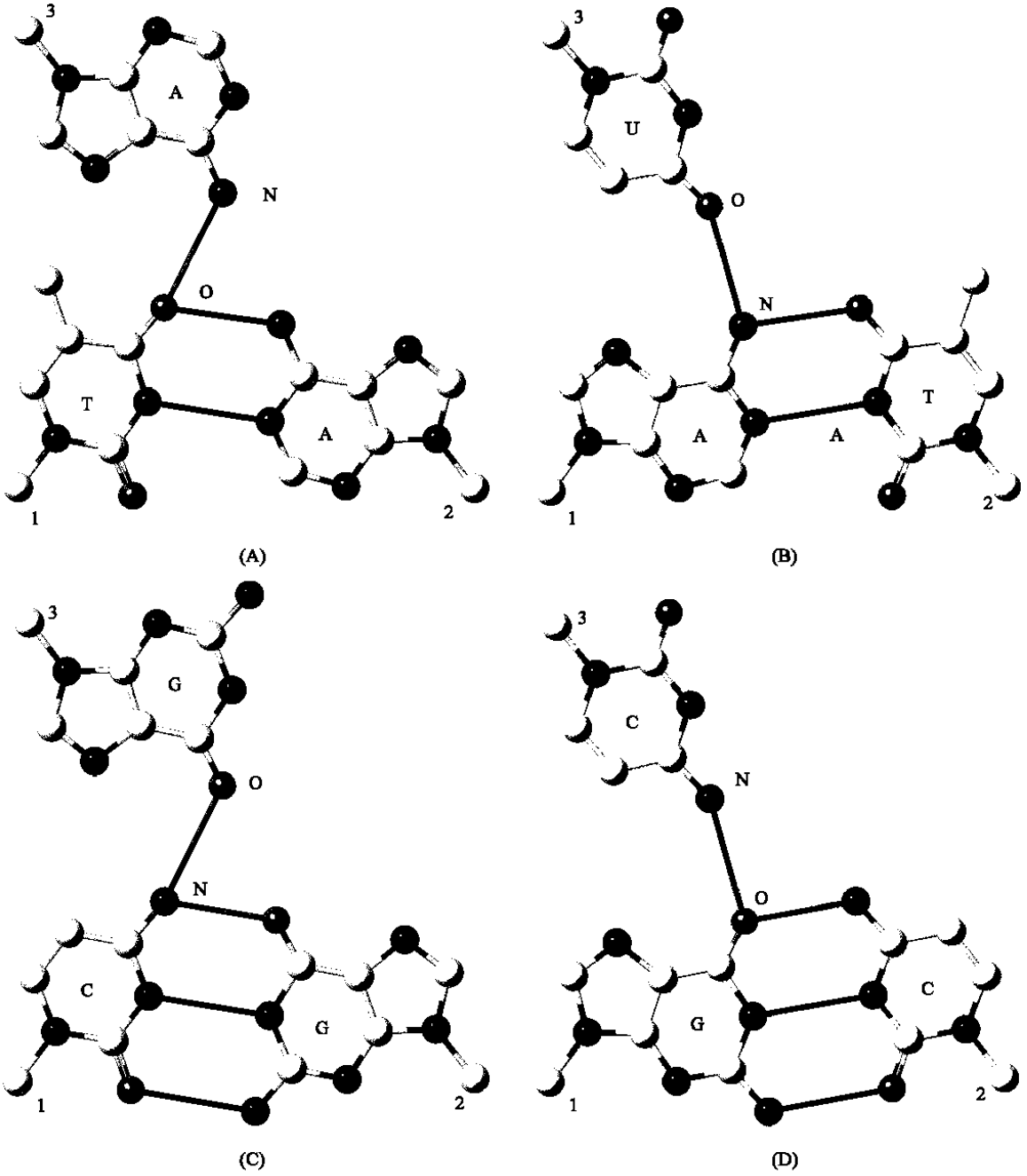

Fig. 1: The assumed DNA – tfRNA triplets (A, B, C, D). A, T, U, G, C – adenine, thymine, uracil, guanine, cytosine, respectively. O and N – oxygen and nitrogen forming hydrogen bonds between DNA and tf – RNA. 1 and 2 – the carbon atoms of desoxyribose of nontemplate and template strands of DNA, respectively. 3 – the carbon atom participating in formation of the suger – phosphate backbone of ribose. For further explanations, see the text

Despite coupling of ribonucleotides by a single hydrogen bond, the principle of complementarily (or specificity) of the template strand and the transcription product is preserved. These strands are antiparallel. Therefore, the nucleotide sequence of the RNA strand is identical to that of the DNA nontemplate strand. Thus, the same information transfer pattern is supported both by a transcription bubble and by the present model.

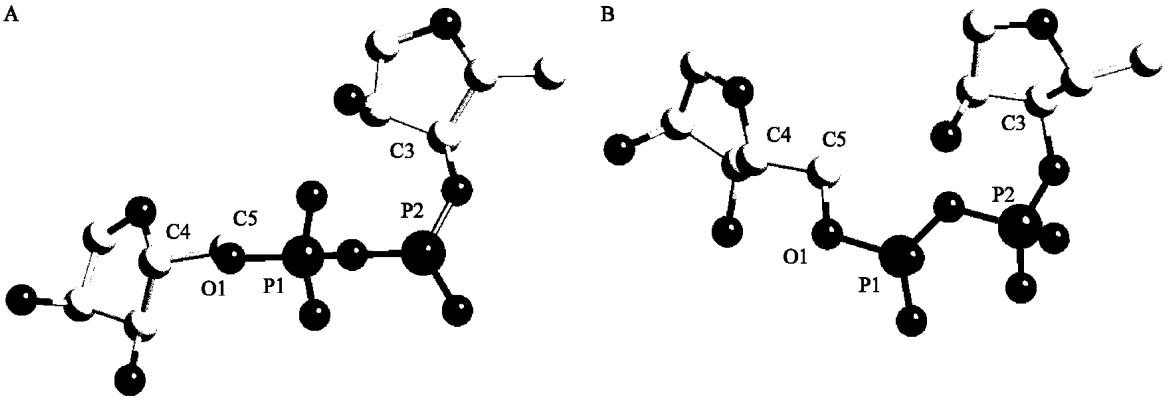

Fig. 2: A part of tf — RNA with diphosphate fitted between two ribose residues. A — a side view, B — a side view and a view from above. For further explanations, see the text

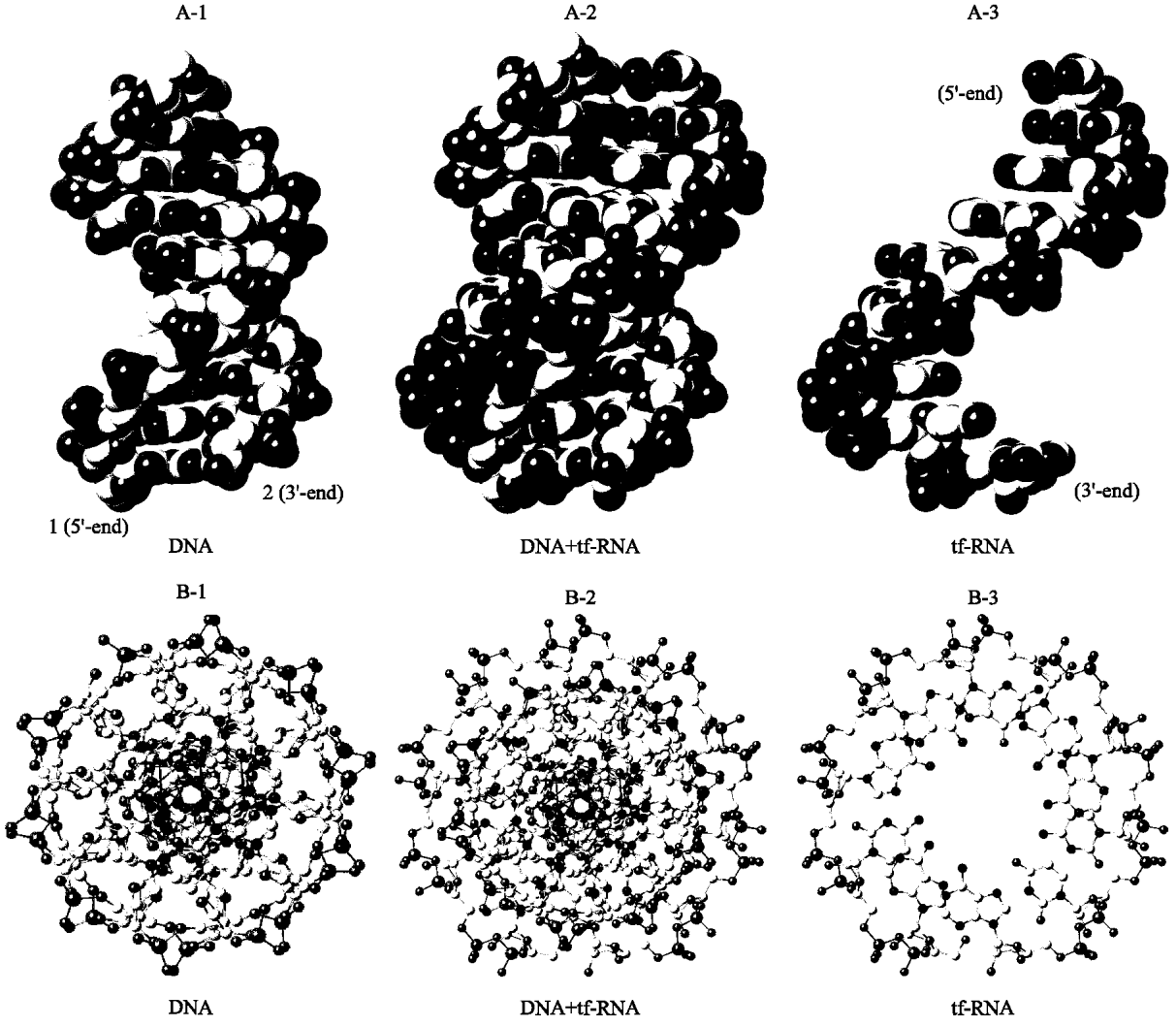

Fig. 3: A snapshot of the DNA — tfRNA model. A — a side view, B — an end view. 1 — the nontemplate strand of DNA, 2 — the template strand of DNA. For further explanations, see the text

As shown in Fig. 1, the directions of the hydrogen bonds for pyrimidine and purine bases of the third strand relative to the helix axis are different. The directions of the C-N bond relative to ribose and the bases are always the same. This means that the RNA strand has the same stereochemical position of the sugar-phosphate groups relative to each other. According to the presented model, the distance between the C1'-atom of ribose and the analogous atom (Fig. 1, 1) of desoxyribose of the DNA template strand is 11.183±0.002 Å, and for the same atom (Fig. 1 and 2) of the nontemplate strand this distance is 15.204±0.001 Å.

Design of the triple helix: Fitting of the third strand in the major groove without alteration of the B-form of the DNA-RNA hybrid has specific features. Stability of the DNA-RNA B-form should be determined by the base-stacking interactions. The distance between the C5'-atom and the appropriate OH-group at the C3'-atom of the adjacent ribose is determined by a constant steric layout of ribose irrespective of the base size. In the present model this distance is constant (as the same distance in DNA). However, since the C1-atom of ribose is located further from the helix axis than the same atom of deoxyribose it is impossible to build one phosphate group into this space. This distance is too large. On the other hand, this niche is suitable for diphosphate. Then, it is reasonable to suppose that during transcription the sugar-phosphate backbone of the transcript is formed by diphosphates rather than by monophosphates. This assumption implies that an atypical transcript is formed. I suggest that it be called the RNA transcription fragment (tf-RNA). In the presented model such tf-RNA is formed by the 10-mer oligoribonucleotide 5'-AppCppAppAppGppAppAppCppTppA-3' (Fig. 3).

To examine this hypothesis, fitting of diphosphate into the niche between C5' and C3' atoms of the adjacent riboses was performed. Figure 2 shows the fragment with the diphosphate excised from the model shown in Fig. 3. The C4-C5-O1 angle is 112°. The C5-O1-P1 angle is 114°. It should be noted that in intermolecular interactions the angle may vary over a rather wide range (Jones *et al*., 1999).

In the model of ribonucleotides the direction of the OH-group of ribose in the 3'-position differs from the direction of the same group in deoxyribonucleotides. For this reason, it is difficult to compare the P2-O2-C3 angle of the presented model (~138°) and the same angle in B-DNA (~120°). One more feature of this structure is orientation of phosphates relative to each other. On modeling a so-called staggered form of diphosphate turned out to be the most appropriate one. It permits optimal positioning of atoms and should promote reduction of tension in tf-RNA. The 2'-OH group in the ribose sterically prevents helical regions of RNA strand from adopting the B-form helical arrangement. In the presented model of tf-RNA this problem is absent.

The positioning of tf-RNA in the major groove relative to the DNA strands is asymmetric. However, its symmetry relative to the nontemplate strand of DNA should be noted (Fig. 1). In spite of the different size of ribonucleotides an end view shows that the oxygen and nitrogen atoms participating in formation of specific hydrogen bonds are at the same distance from the centre (i.e., from the helix axis). The side view reveals a planar construction of ribonucleotides (the B-form of tf-RNA).

DISCUSSION

A basis for the presented model is the idea of transcription along the DNA major groove without alteration of its B-configuration. It is supposed that the number of hydrogen bonds in the transcribed RNA should be less than the number of the bonds in the DNA. In this case, energy consumption for peeling RNA from DNA should be reduced. I assume also that the RNA fragment formed during transcription should have the sugar-diphosphate backbone. Although any particular assumption concerning formation of tf-RNA (the B-form of triple helix, single hydrogen bond, diphosphates) is hypothetical, every of them can be supported by specific evidence.

After the DNA double helix was discovered, the opportunity to understand mechanisms of transcription at the molecular level appeared. It seemed logical that RNA could be synthesized along the major groove of DNA. However, such models of transcription were not developed. Nevertheless, it was revealed that under artificial conditions it is possible to obtain different variants of triple helical DNA. Such polymers are obtained from the oligonucleotides formed either by pyrimidines or purines (Dadarlat and Saxena, 1998; Escude *et al*., 1999; Frank-Kamenetskii and Mirkin, 1995). As the same bases belong to each strand, their sugar-phosphate backbone is at the same distance from the helix axis. Such triplexes can be locally formed at certain DNA sites and carry out regulatory functions (Besch *et al*., 2002; Escude *et al*., 1998; Jiang and Russu, 2002; Pasternack *et al*., 2002; Potaman and Sinden, 1998; Potaman and Bissler, 1999; Rocher *et al*., 2001). In contrast to the DNA double helix, such triple helixes do not form the B-form but are similar to the A-form. Exotic triple helical structures are formed

by morpholino oligodeoxyribonucleotides (Basye *et al*., 2001). Thus, the triple helixes can really exist. Interaction between strands is maintained by hydrogen bonds.

In the presented model the number of hydrogen bonds per base pair is one. Pairing of the bases of the third strand via one hydrogen bond is possible (Parel and Leumann, 2001; Rana and Ganesh, 2000). Specifically, the 5-methyl-group of thymine, in addition to a single hydrogen bond of adenine, can orient the latter in space. This group may play the role of a stopper, limiting angular movement of adenine of tf-RNA in the direction of the complementary strand of DNA. It is known that in many cases RNA synthesis begins with adenine. This is probably determined by a special position of adenine associated with its orientation both by hydrogen bond and by stopper effect of the methyl-group. This could explain the presence in DNA of thymine instead of uracil.

The other suggestion of the model is the linkage of ribonucleotides via diphosphates. Fitting of diphosphates instead of monophosphates into such molecular construction is determined by steric reasons. Despite the fact that the diphosphates were built rather easily into tf-RNA, the angles and the bond lengths cannot be fitted manually precisely enough. Besides, including diphosphates into tf-RNA further suggests that RNA polymerases should be able to transform the diphosphate backbone into the phosphate backbone. For example, polynucleotide phosphorylase is known to synthesize polyribonucleotides from nucleoside-5'-diphosphates with release of phosphate (Li *et al*., 1998; Mohanty and Kushner, 2000; Trip and Smith, 1978). Whether there is any RNA polymerase with such properties it is not known.

Undoubtedly, the assumptions regarding the RNA transcription fragment require experimental confirmation. However, I have not found any data contradicting to presented concept. The data confirming formation of the transcription bubble unambiguously do not exist either. As a rule, experimental data based on the transcription bubble hypothesis may be explained by using the tf-RNA hypothesis and the works in which artificial transcription bubble is produced (Aiyar *et al*., 1994; Armache *et al*., 2005; Choder and Aloni, 1988; Daube and von Hippel, 1999; Kettenberger *et al*., 2004; Keene and Luse, 1999; Kireeva *et al*., 2000; Murakami *et al*., 2002; Wilson *et al*., 1999) cannot exclude transcription along the non-melted DNA. In particular, no three-dimensional structural information on the transcription bubble is yet available (Temiakov *et al*., 2002; Zaychikov *et al*., 1997).

At the same time, the present model has a number of advantages in comparison with the hypothesis of the transcription bubble. Local melting of DNA requires energy expenses both for breaking hydrogen bonds and for the DNA supercoiling. Such expenses can be sharply increased by a simultaneous reading of the short tandem repeats. The hypothesis of the transcription bubble does not seem to explain, how energy expenses are regulated in such a mode of transcription. In the presented model, similar questions do not arise.

According to the hypothesis of the tf-RNA, DNA supercoiling during transcription should take place and be phenomenologically similar to supercoiling during replication. There are data confirming participation of topoisomerases in transcription (Mondal *et al*., 2003). It has been shown that in the absence of topoisomerase, transcriptional repression on chromatin depends on RNA length. Synthesis of transcripts of 100 nucleotides or shorter was unaffected by chromatin, whereas repression was apparent when the RNA transcript was 200 nucleotides or longer (Mondal *et al*., 2003). Such type of supercoiling may not have any relation to the specific supercoiling during formation of the transcription bubble. It is easy to imagine that the extent of supercoiling connected to local melting of DNA under the same length of transcription bubble should be constant. This supercoiling does not depend on the length of RNA.

In a certain sense, RNA synthesis in the transcription bubble is similar to DNA replication. However, during replication both strands of DNA are synthesized. In the transcription bubble only one DNA strand is transcribed. RNA polymerase should probably have an additional function suppressing the readout in the opposite direction. The hypothesis of the RNA transcription fragment explains easily the fact that only one DNA strand can be read out. In this case, the simultaneous readout of both strands of DNA is impossible because of stereochemical restrictions.

I think that the weakest spot of the hypothesis about the transcription bubble is absence of stereochemical models. For different reasons, such models are very difficult to construct.

Thus, the hypothesis of formation of tf-RNA has a number of advantages in comparison with the standard model of the transcription bubble. It requires, however, an appropriate experimental confirmation.

ACKNOWLEDGMENTS

I wish to thank Drs S.Yu. Khaitlina, B.A. Margulis, V.I. Vorob'ev, S.A. Borkhsenius and O.K. Glebov for discussions and help at the time of preparation of the manuscript.

REFERENCES

Aiyar, S.E., J.D. Helmann and P.L. deHaseth, 1994. A Mismatch Bubble in Double-stranded DNA Suffices to Direct Precise Transcription Initiation by *Escherichia coli* RNA Polymerase. J. Biol. Chem., 269: 13179-13184.

Armache, K.J., H. Kettenberger and P. Cramer, 2005. The dynamic machinery of mRNA elongation. Curr. Opin. Struct. Biol., 15: 197-203.

Basye, J., J.O. Trent, D. Gao and S.W. Ebbinghaus, 2001. Triplex formation by morpholino oligodeoxyribonucleotides in the HER-2/neu promoter requires the pyrimidine motif. Nucleic Acids Res., 29: 4873-4880.

Besch, R., C. Giovannangeli, C. Kammerbauer and K. Degitz, 2002. Specific inhibition of ICAM-1 expression mediated by gene targeting with triplex-forming oligonucleotides. J. Biol. Chem., 277: 32473-32479.

Choder, M. and Y. Aloni, 1988. RNA polymerase II allows unwinding and rewinding of the DNA and thus maintains a constant length of the transcription bubble. J. Biol. Chem., 263: 12994-13002.

Coulombe, B. and Z.F. Burton, 1999. DNA bending and wrapping around RNA polymerase: A revolutionary model describing transcriptional mechanisms. Microbiol. Mol. Biol., Rev., 63: 457-478.

Dadarlat, V.M. and V.K. Saxena, 1998. Stability of Triple-Helical Poly(dT)-Poly(dA)-Poly(dT) DNA with Counterions. Biophys. J., 75: 70-91.

Daube, S.S. and P.H. von Hippel, 1999. Interactions of *Escherichia coli* s70 within the transcription elongation complex. PNAS, 96: 8390-8395.

Ebright, R.H., 2000. RNA Polymerase: Structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. J. Mol. Biol., 304: 687-698.

Escude, C., C.H. Nguyen, S. Kukreti, Y. Janin, J.S. Sun, E. Bisagni, T. Garestier and C. Helene, 1998. Rational design of a triple helix-specific intercalating ligand. PNAS, 95: 3591-3596.

Escude, C., T. Garestier and C. Helene, 1999. Padlock oligonucleotides for duplex DNA based on sequence-specific triple helix formation. Proc. Natl. Acad. Sci. USA., 96: 10603-10607.

Jiang, L. and I.M. Russu, 2002. Internal dynamics in a DNA triple helix probed by 1H-15N-NMR spectroscopy. Biophys. J., 82: 3181-3185.

Jones, S., P. Heyningen, H.M. Berman and J.M. Thornton, 1999. Protein-DNA interactions: A structural analysis. J. Mol. Biol., 287: 877-896.

Frank-Kamenetskii, M.D. and S.M. Mirkin, 1995. Triplex DNA Structures. Ann. Rev. Biochem., 64: 65-95.

Keene, R.G. and D.S. Luse, 1999. Initially transcribed sequences strongly affect the extent of abortive initiation by RNA polymerase II. J. Biol. Chem., 274: 11526-11534.

Kettenberger, H., K.J. Armache and P. Cramer, 2004. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. Mol. Cell., 16: 955-965.

Kireeva, M.L., N. Komissarova and M. Kashlev, 2000. Overextended RNA: DNA Hybrid as a negative regulator of RNA polymerase II processivity. J. Mol. Biol., 299: 325-335.

Li, Q.S., J.D. Gupta and A.G. Hunt, 1998. Polynucleotide phosphorylase is a component of a novel plant poly(A) polymerase. J. Biol. Chem., 273: 17539-17543.

Mohanty, B.K. and S.R. Kushner, 2000. Polynucleotide phosphorylase functions both as a 3′→5′ exonuclease and a poly(A) polymerase in *Escherichia coli*. PNAS, 97: 11966-11971.

Mondal, N., Y. Zhang, Z. Jonsson, S.K. Dhar, M. Kannapiran and J.D. Parvin, 2003. Elongation by RNA polymerase II on chromatin templates requires topoisomerase activity. Nucleic Acids Res., 31: 5016-5024.

Morgan, G.T., R.H. Reeder and A.H. Bakken, 1983. Transcription in cloned spacers of Xenopus laevis ribosomal DNA. PNAS, 80: 6490-6494.

Murakami, K.S., S. Masuda, E.A. Campbell, O. Muzzin and S.A. Darst, 2002. Structural basis of transcription initiation: An RNA polymerase holoenzyme-DNA complex. Sci., 296: 1285-1290.

Nikolov, D.B. and S.K. Burley, 1997. RNA polymerase II transcription initiation: A structural view. PNAS, 94: 15-22.

Parel, S.P. and C.J. Leumann, 2001. Triple-helix formation in the antiparallel binding motif of oligodeoxynucleotides containing N9-and N7-2-aminopurine deoxynucleosides. Nucleic Acids Res., 29: 2260-2267.

Pasternack, L.B., S.B. Lin, T.M. Chin, W.C. Lin, D.H. Huang and L.S. Kan, 2002. Proton NMR studies of 5′-d-$(TC)_3(CT)_3(AG)_3$-3′-A paperclip triplex: The Structural Relevance of Turns. Biophys. J., 82: 3170-3180.

Potaman, V.N. and R.R. Sinden, 1998. Stabilization of intramolecular triple/single-strand structure by cationic peptides. Biochem., 37: 12952-12961.

Potaman, V.N. and J.J. Bissler, 1999. Overcoming a barrier for DNA polymerization in triplex-forming sequences. Nucleic Acids Res., 27: 1-5.

Rana, V.S. and K.N. Ganesh, 2000. Recognition of 5-aminouracil (U#) in the central strand of DNA triplex: Orientation selective binding of different third strand bases. Nucleic Acids Res., 28: 1162-1169.

Rocher, C., R. Dalibart, T. Letellier, G. Précigoux1 and P. Lestienne, 2001. Initiation of DNA replication by DNA polymerases from primers forming a triple helix. Nucleic Acids Res., 29: 3320-3326.

Spitalny, P. and M. Thomm, 2003. Analysis of the open region and of DNA-protein contacts of archaeal RNA polymerase transcription complexes during transition from initiation to elongation. J. Biol. Chem., 278: 30497-30505.

Temiakov, D., M. Anikin and W.T. McAllister, 2002. Characterization of T7 RNA polymerase transcription complexes assembled on nucleic acid scaffolds. J. Biol. Chem., 277: 47035-47043.

Trip, E.M. and M. Smith, 1978. Enzymatic synthesis of oligodeoxyribonucleotides of defined sequence. Polynucleotide phosphorylase catalysed addition of deoxyribonucleotides to primers which are good or poor acceptors. Nucleic Acids Res., 5: 1529-1538.

Wilson, K.S., C.R. Conant and P.H. von Hippel, 1999. Determinants of the stability of transcription elongation complexes: Interactions of the nascent RNA with the DNA template and the RNA polymerase. J. Mol. Biol., 289: 1179-1194.

Zaychikov, E., L. Denissova, T. Meier, M. Götte and H. Heumann, 1997. Influence of Mg^{2+} and temperature on formation of the transcription bubble. J. Biol. Chem., 272: 2259-2267.