

# THE LIMITS OF TEMPLATE-DIRECTED SYNTHESIS WITH NUCLEOSIDE-5'-PHOSPHORO(2-METHYL)IMIDAZOLIDES

AUBREY R. HILL JR., and LESLIE E. ORGEL

*The Salk Institute for Biological Studies, Post Office Box 85800, San Diego, CA 92186-5800, U.S.A.*

TAIFENG WU

*Department of Chemistry, Northwestern University, Evanston, IL 60208, U.S.A.*

(Received January 11, 1993)

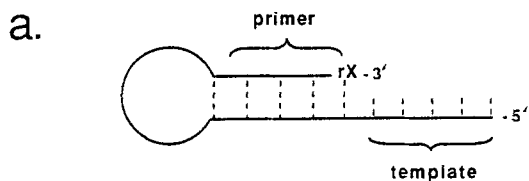
**Abstract.** In earlier work we have shown that C-rich templates containing isolated A, T or G residues and short oligo(G) sequences can be copied effectively using nucleoside-5'-phosphoro(2-methyl)imidazolides as substrates. We now show that isolated A or T residues within an oligo(G) sequence are a complete block to copying and that an isolated C residue is copied inefficiently.

Replication is possible only if there are two complementary oligonucleotides each of which acts as a template to facilitate the synthesis of the other. We emphasize the severity of the problems that need to be overcome to make possible non-enzymatic replication in homogeneous aqueous solution. We conclude that an efficient catalyst was involved in the origin of polynucleotide replication.

## 1. Introduction

Template-directed reactions of nucleoside 5'-phosphoro(2-methyl)imidazolides have been studied intensively as models of the enzymatic replication or transcription of nucleic acids (Joyce, 1987). Recently, we introduced a novel method of analysis that has greatly extended our detailed knowledge of the various steps involved (Wu and Orgel, 1992a, b, c). We synthesize oligodeoxynucleotide sequences that form stem loop secondary structures (Figure 1a). The template sequences are the 5'-terminal single-strand portions of the hairpin. The 3'-terminal segments of the oligonucleotide, which are part of the double-helical stems, act as intramolecular 'primers' for template-directed incorporation of monomers. Ribonucleotides are introduced as terminal residues at the 3'-ends of the oligomers to improve the efficiency of the reactions. The hairpins are [<sup>32</sup>P]-labelled at their 5'-termini and the products are analyzed by autoradiography of gel-electrophoretograms.

The use of the substrate oligonucleotides at nanomolar concentrations eliminates most of the problems caused by the association of two or more templates. The use of oligodeoxynucleotide hairpins also simplifies analysis of the reaction products. The method has revealed many novel features of template-directed reactions involving two or more bases. In particular, we found that efficient transcription occurs for templates in which C is the predominant base, G is present in runs of 1 to 4 or more residues, and A and T are present as isolated residues. The sequences TT, GT, and TG in the template are partial barriers to transcription and the sequences AT, TA, AA, GA and AG are almost total barriers (Wu and Orgel, 1992c).



b.

Template	Sequence
S <sub>1</sub>	<b>GGGTGGGAGTTTTTCTCC</b> rC
S <sub>2</sub>	<b>GGGAGGGAGTTTTTCTCC</b> rC
S <sub>3</sub>	<b>GGGCGGGAGTTTTTCTCC</b> rC
S <sub>4</sub>	<b>CCCGCCCTAGTTTTTCTAG</b> rG
S <sub>5</sub>	<b>GGGTGGTAGTTTTTCTAC</b> rC
S <sub>6</sub>	<b>GGGAGGTAGTTTTTCTAC</b> rC
S <sub>7</sub>	<b>GGGTGCTAGTTTTTCTAG</b> rC
S <sub>8</sub>	<b>GGGAGCTAGTTTTTCTAG</b> rC
S <sub>9</sub>	<b>GCGCGCTAGTTTTTCTAG</b> rC

Fig. 1. (a) The secondary structure of hairpin substrates. (b) The sequences of the substrates. The bases of the template sub-sequence are in bold face.

Here we explore the template activities of sequences complementary to some of the effective templates discovered in the earlier work, that is templates containing a run of G residues interrupted by a single U, C or A residue.

## 2. Materials and Methods

The sources of all reagents and all experimental procedures have been described in previous papers (Wu and Orgel, 1992a, b, c). The template sequences that we used are illustrated in Figure 1b.

### 3. Results

Replication requires a pair of templates each of which facilitates the synthesis of the other. Our program was initially based on the observation that poly(C) is a template for the synthesis of oligo(G)'s from guanosine-5'-phosphoro(2-methyl)imidazolide (2-MeImpG) (Inoue and Orgel, 1982). We have subsequently shown that isolated A, T or G residues within a sequence of C residues are copied efficiently (Wu and Orgel, 1992b, c). In the context of a sequence of C residues, sequences of G residues are the only other template sequences that are copied efficiently (Wu and Orgel, 1992b). It is clearly important to determine whether complementary sequences, particularly those in which isolated T, A or C residues are included in a sequence of G residues, can be copied.

The incorporation of A opposite T within the template sequence GGGTGGG was investigated using substrate  $S_1$ . When  $S_1$  was incubated with cytidine-5'-phosphoro(2-methyl)imidazolide (2-MeImpC), adenosine-5'-phosphoro(2-methyl)imidazolide (2-MeImpA) or a mixture of both, the yield of extension products after 1 week was very low. Analogous results were obtained for the incorporation of T opposite A in the template sequence GGGAGGG using substrate  $S_2$ . We conclude that neither T or A can be copied within a sequence of G residues.

The incorporation of G opposite C in the template sequence GGGCGGG was investigated using substrate  $S_3$  in the presence of 2-MeImpG and 2-MeImpC. After 1 week a moderate yield of a first extension product and smaller yields of products containing two or three added bases were formed (Figure 2a). Thus C is not as effective a block as A or T when incorporated into a sequence of G residues. However, the copying of a template sequence GGGCGGG ( $S_4$ ) is much less efficient than the copying of CCCGCC (Figure 2b).

We carried out numerous further experiments with substrates  $S_5$ - $S_9$  and many others (Data not shown). The results were monotonous. They confirmed that neither A nor T within a template strand composed of G residues can be copied and that a C residue within a sequence of G residues is copied at best inefficiently. They also showed that sequences such as GCGCG cannot be copied.

### 4. Discussion

We have now completed a detailed examination of the template-directed reactions of one group of activated nucleotides, the nucleoside 5'-phosphoro(2-methyl)imidazolides. Our first experiments established that 2-MeImpG polymerizes efficiently and regioselectively on poly(C) (Inoue and Orgel, 1982). Subsequently we showed that isolated G residues within a sequence of C residues efficiently direct the incorporation of C residues within a sequence of G residues. Thus sequences such as CCGCC (Inoue *et al.*, 1984) or CCGCCCCCGCCC (Acevedo and Orgel, 1987) can be copied to give their complements.

More recently we have shown that isolated A or T residues in a sequence

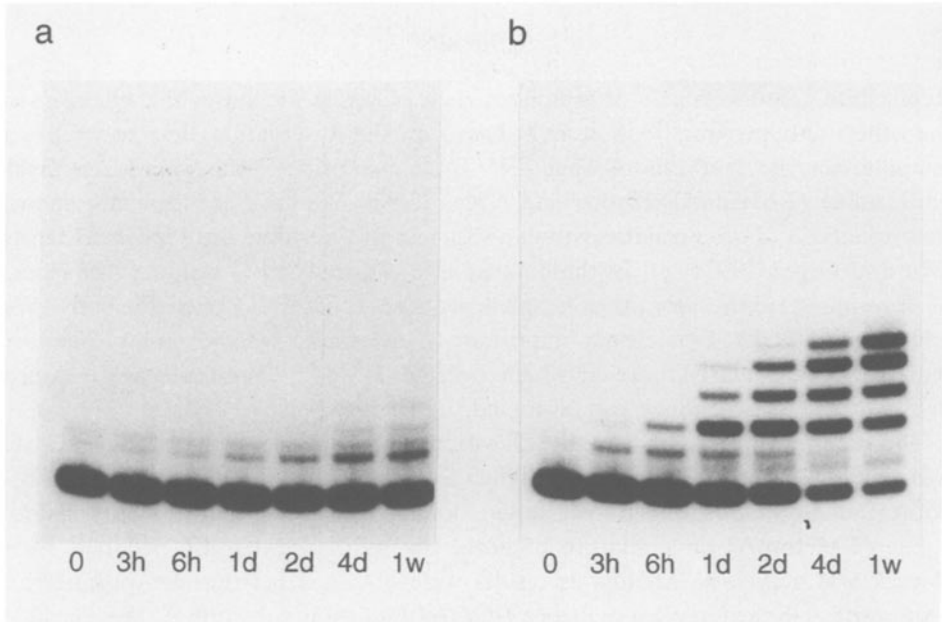


Fig. 2. Autoradiogram showing products of the reaction of a mixture of 2-MeImpG and 2-MeImpC on (a) Substrate  $S_3$ , and (b) Substrate  $S_4$ .

of C residues are also copied faithfully (Wu and Orgel, 1992c). Furthermore, sequences of G residues, provided they cannot associate to form tetraplexes, will direct the synthesis of sequences of C residues (Wu and Orgel, 1992b). We cannot detect significant yields of products from template sequences AA, AT, TA, GA and AG, and copying of GT, TG and TT sequences is inefficient (Wu and Orgel, 1992c).

These results suggest that many C-rich sequences containing isolated A and T residues and short sequences of G residues can be copied, but leaves open the question of replication. Can the G-rich sequences formed by copying C-rich sequences act as templates in turn to complete the replication cycle?

The results presented in the present paper make it clear that no sequence containing an A or T residue can replicate. Optimal templates, CCCTCCC and CCCACCC, for example, would generate products GGGAGGG and GGGTGGG respectively. We have seen that isolated A or T residues within a run of G residues are a complete block to copying. So the second step of the replication cycle does not occur.

The situation is more complicated for templates containing C and G residues. The template sequence CCCGCC would direct the synthesis of the product sequence GGGCGGG efficiently. This sequence, while not totally ineffective as a template is copied slowly and inefficiently. Thus it would be difficult to achieve replication in this system, but perhaps not impossible.

Pairs of sequences such as CCCCGGG and CCCGGGG are more promising.

Each, in principle, would promote the formation of the other. The same would be true for longer sequences in which blocks of C residues alternate with blocks of G residues. Unfortunately, sequences consisting of alternating blocks of C and G residues carry rather little information.

In practice, replication of even this restricted range of oligomers would present some additional difficulties. We have seen that the copying of the last C residue of an oligo(C) template is slow (Wu and Orgel, 1992a), and the copying of the last G residue of an oligo(G) sequence is very slow (Wu and Orgel, 1992b). Thus for a template  $C_nG_m$ , for example, there would be a strong tendency to truncate the sequence by one residue after each round of replication. The difficulty of dissociating the products of template-directed synthesis from the template, and the stability of tetraplexes formed by oligo(G) sequences are other well-recognized obstacles to replication (Joyce, 1987).

We conclude that it will not be possible to achieve a practical informational relicating system using 5'-phosphoro(2-methyl)imidazolides as substrates in homogeneous aqueous solution. Are the difficulties intrinsic to any self-replicating system based on mononucleotides, or do they depend on the particular group of activated nucleotides that we have employed? This question cannot be answered definitively. However, we do not think that the prospects of finding some other quartet of activated nucleotides that supports an efficient replicating system are good.

The major difficulty arises through the sequence-dependence of the conformation at the end of a growing product strand (Joyce and Orgel, 1986). Competition between internucleotide-bond formation and hydrolysis of the activated nucleotide is favorable only if the 3'-OH of the growing chain is in exactly the right position relative to the activated phosphate. Regiospecificity is observed only if the 2'-OH group fails to compete with the 3'-OH group. These stringent conditions can be met by a single set of activated mononucleotides for one or two types of additions in homopolymers, for example G to G or C to C, on oligo(C) or oligo(G) templates, respectively. However, when sequences of any generality are concerned, replication involves so many different addition reactions, each with a different substrate geometry, that a single type of activated nucleotide is unlikely to work efficiently for all of them.

We must note one minor reservation concerning our choice of substrates. We have used oligodeoxynucleotides terminated by a single ribonucleotide as substrates, rather than oligoribonucleotides. Thus we have not literally shown that oligonucleotides cannot replicate. However, much earlier work has shown that oligodeoxynucleotides and oligonucleotides react similarly as templates (Joyce, 1987), and as non-terminal constituents of primers (Wu and Orgel, 1992a, b).

The arguments presented above suggest that direct replication in homogeneous solution using mononucleotides as substrates is unlikely. The origin of polynucleotide replication probably involved catalysts that 'smoothed away' the sequence-dependent differences in configuration that occur at the end growing chains. We do not know whether the catalysts were inorganic (minerals), the products of a 'biological' system

that preceded the RNA world or, just possibly, a ribozyme formed independently of a template (Joyce and Orgel, In press).

### Acknowledgments

This work was supported by NASA Grant No. NAGW-1660. We thank Sylvia Bailey for manuscript preparation.

### References

- Acevedo, O. L. and Orgel, L. E.: 1987, *J. Mol. Biol.* **197**, 187-193.
- Inoue, T., Joyce, G. F., Grzeskowiak, K., Orgel, L. E., Brown, J. M., and Reese, C. B.: 1984, *J. Mol. Biol.* **178**, 669-676.
- Inoue, T. and Orgel, L. E.: 1982, *J. Mol. Biol.* **162**, 201-217.
- Joyce, G. F.: 1987, 'Non-enzymatic Template-directed Synthesis of Informational Macromolecules', *Cold Spring Harbor Symposia on Quantitative Biology, Volume LIII*. Cold Spring Harbor Press, Cold Spring Harbor, New York. 41-51.
- Joyce, G. F. and Orgel, L. E.: 1986, *J. Mol Biol.* **188**, 433-441.
- Joyce, G. F. and Orgel, L. E.: Origin of the RNA World. *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, (in press).
- Wu, T. and Orgel, L. E.: 1992a, *J. Chem. Soc.* **114**, 318-322.
- Wu, T. and Orgel, L. E.: 1992b, *J. Am. Chem. Soc.* **114**, 5496-5501.
- Wu, T. and Orgel, L. E.: 1992c, *J. Am. Chem. Soc.* **114**, 7963-7969.