

Diss. ETH No. 12916

Chemistry and Enzymology of an Expanded Genetic Alphabet

A Dissertation Submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZÜRICH

presented by
János Tibor Kodra

Prof. Dr. G. Folkers, examiner
Prof. Dr. S. A. Benner, co-examiner

Zürich 1998

Summary

Nature uses two base pairs to store and express genetic information. In DNA, these bases pair following the rules of complementarity (purines pair with pyrimidines) as formulated by Watson and Crick (Watson and Crick, 1953)

This dissertation focuses on the chemistry and enzymology of an additional base pair between isoguanine (isoG) and isocytosine (isoC) (Fig.I).

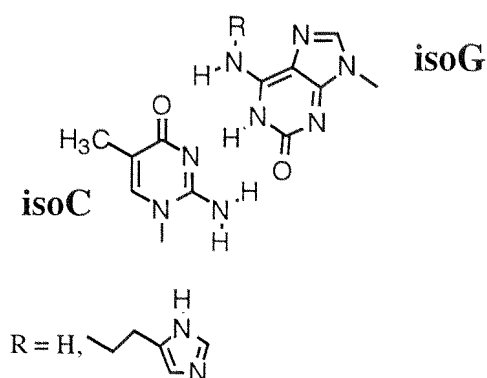
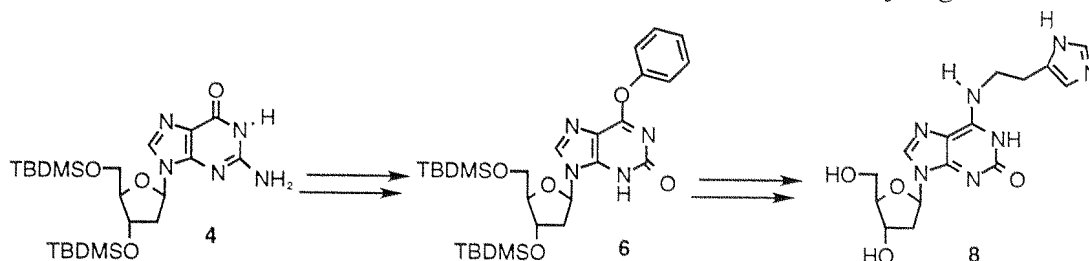


Figure I. The nonstandard base pair isoC-isoG and the imidazole carrying isoG

A synthetic method for making the nucleoside analog 6-*N*-[2-(1H-imidazole-4-yl)-ethyl] 2'-deoxyisoguanosine (dHisoG) **8** as well as its triphosphate dHisoGTP was developed starting from the sugar protected derivative of 2'-deoxyguanosine **4** (Scheme I). These compounds represent the first 2-*N*-alkylated derivatives of 2'-deoxyisoguanosine.



Scheme I. The synthesis of 6-*N*-[2-(1H-imidazole-4-yl)-ethyl] 2'-deoxyisoguanosine (dHisoG)

The synthesis of oligonucleotides containing 2'-deoxy-5-methylisocytidine (d^{Me}isoC) and 2'-deoxyisoguanosine (disoG) using phosphoramidite chemistry in solid-phase oligonucleotide synthesis is described. Supporting previous observations, the *N,N*-diisobutylformamidinium moiety was found to be a far superior protecting group than the *N*-benzoyl for 2'-deoxy-5-methylisocytidine. A formamidinium protected phosphoramidite of d^{Me}isoC **9j** (Fig. II) was incorporated several times successfully at consecutive positions during a standard automated synthesis protocol with a coupling efficiency >99%, according to DMT release. Extending the coupling times of the standard protocol for the phosphoramidite of disoG **13** (Fig. II) led to successful incorporation of multiple consecutive disoG nucleotides with a coupling efficiency >97%.

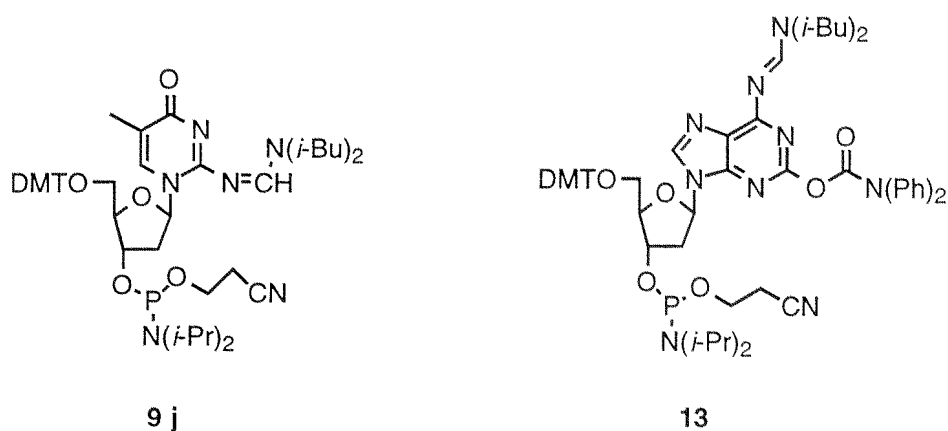


Figure II. The building blocks for the chemical synthesis of oligonucleotides containing d^{Me}isoC and disoG

The ability of various polymerases to accept the isoG-isoC base pair, as well as their ability to incorporate the new base dHisoG base as its triphosphate into DNA opposite its complementary base isoC, was studied. It was found that the Klenow fragment of polymerase I, HIV-1 RT, and the *Tth*, *Tfl* and *Bst*. DNA polymerase incorporated the dHisoGTP opposite the complementary base d^{Me}isoC. This represents an enzymatic method for synthesizing DNA molecules that contain one important functionalities (an

imidazole ring) found in proteins. This result paves the way for the generation of a novel type of biopolymer that combines the catalytic efficiency of proteins with the replicability of nucleic acids.

The polymerases that accepted dHisoG/disoG nucleobases all belong to the polymerase A family. Two polymerase A members were found (the Hot *Tub* DNA polymerase and the *Taq* DNA polymerase) that accepted disoGTP but not dHisoGTP. Various members of the polymerase B family were also investigated. In all cases, these polymerases rejected the disoG/dHisoG nucleobase analogs. This difference in the polymerase families ability to incorporate disoGTP opposite d^{Me}isoC suggests that minor groove scanning is quantitatively different in the two classes of polymerases, and may be qualitatively different as well. If indeed the disoG-disoC base pair was part of an early form of life (Rich, 1962; Benner *et al.*, 1987; Switzer *et al.*, 1993), then these results suggest that polymerase A family share a closer affinity with the earliest polymerases than the polymerase B family. Based on these findings, primer-extension experiments with disoGTP and a d^{Me}isoC-containing template would allow a rapid classification of new polymerases.

Polymerases were also examined for ability to incorporate d^{Me}isoCTP opposite disoG in a template. The Klenow fragment of polymerase I, HIV-1 RT, and the *Tth* DNA polymerase incorporated d^{Me}isoCTP opposite disoG, making them promising leads for site-directed mutagenesis in order to develop polymerases that are able to copy DNA built from an expanded genetic alphabet.

Zusammenfassung

Die Natur verwendet zwei Basenpaare zur Speicherung und Expression der genetischen Information. Wie von Watson und Crick formuliert wurde, paaren diese Basen in DNA entsprechend Komplementaritätsregeln, das heisst Purine paaren mit Pyrimidinen (Watson and Crick, 1953).

In dieser Dissertation wurden die Chemie und Enzymologie des neuen Basenpaars isoG-isoC (Abb. I) untersucht.

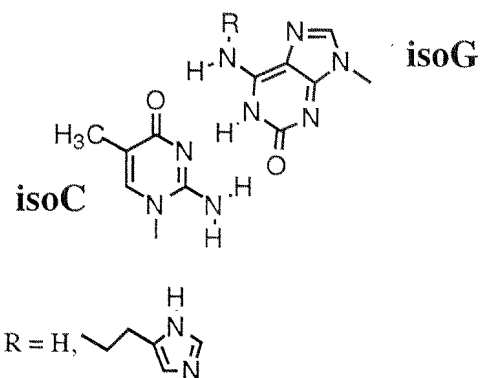
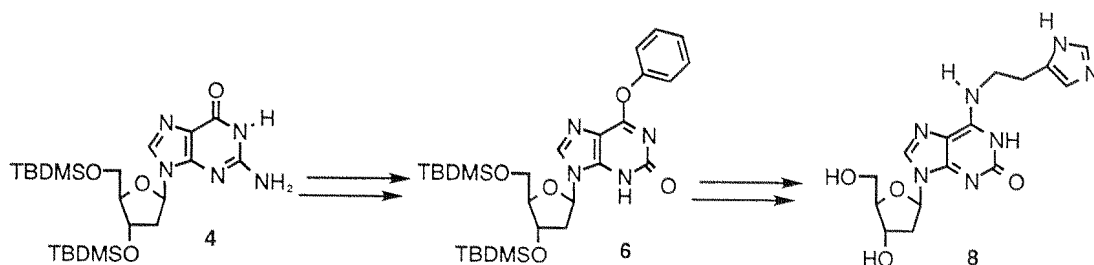


Abb. I. Das neue Basenpaar isoG-isoC

Eine Synthesemethode für die Darstellung sowohl des Nucleosidanalogons 6-*N*-[2-(1H-Imidazole-4-yl)-ethyl]-2'-desoxyisoguanosin (dHisoG) **8** als auch des Triphosphats dHisoGTP wurde entwickelt (Schema I). Diese Verbindungen sind die ersten beschriebenen 2-*N* alkylierten Derivate von 2'-Desoxyisoguanosin.



Schema I. Synthese der neuen Purinbase 6-*N*-[2-(1H-Imidazol-4-yl)-ethyl]-2'-desoxyisoguanosin (dHisoG)

Festphasen-DNA-Synthesebedingungen für die Darstellung von Oligonukleotiden, die 2'-Desoxy-5-methylisocytidin ($d^{Me}isoC$) und 2'-Desoxyisoguanosin ($disoG$) enthalten, wurde untersucht. Dabei wurde *N,N*-Diisobutylformamidin als Schutzgruppe für die exozyklische Aminogruppe von $d^{Me}isoC$ benutzt. Unter Verwendungen des Phosphoramidits **9j** (Abb. II) wurden Oligonukleotide mit aufeinanderfolgenden $d^{Me}isoCs$ mit Kopplungsausbeuten über 99 % hergestellt. Für die Synthese von $disoG$ enthaltenden Oligonukleotiden wurde das Phosphoramidit **13** (Abb. II) eingesetzt. Durch Verlängerung der Kopplungszeit konnten für Oligonukleotide mit mehreren aufeinanderfolgenden $disoGs$ Kopplungsausbeuten über 97 % erreicht werden. Mit diesen Methoden wurden DNA Template, die $d^{Me}isoC/disog$ enthalten, für enzymatische Untersuchungen hergestellt.

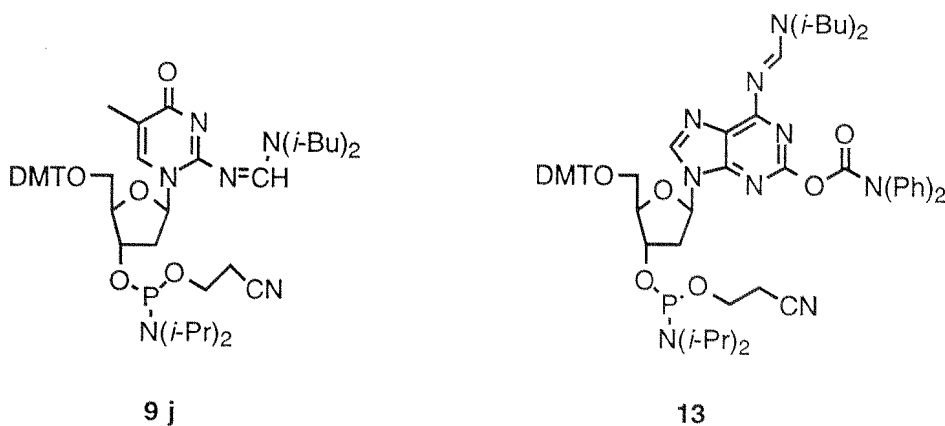


Abb. II. Die für die Festphasen-DNA-Synthese benutzten Phosphoramidite.

Der enzymatische Einbau des Basenpaars $isoG-isoC$ sowie des neuen Purinanalogs $dHisoGTP$ gegenüber $isoC$ mit verschiedenen Polymerasen wurde untersucht.

Mit Klenow fragment, HIV-1 RT, *Tth*, *Tfl* und *Bst*.DNA Polymerase gelang die Inkorporation von $dHisoGTP$ gegenüber $d^{Me}isoC$. Damit wird es möglich, mit enzymatischen Methoden DNA zu synthetisieren, die eine der wichtigsten funktionellen Gruppen von Proteinen trägt. Diese Kombination von Replizierbarkeit und Funktionalität

könnte zu einem "neuen" Typ von Biopolymer führen, der die Protein:RNA/DNA-Welt ersetzen kann.

Die Polymerasen, die dHisoGTP als Substrat akzeptieren, gehören zur Polymerase A Familie. Diese Polymerasen haben auch die nicht alkylierten Derivate disoGTP gegenüber $d^{Me}isoC$ in DNA eingebaut. Zwei weitere Polymerasen der Klasse A wurden untersucht (*Hot Tub* und *Taq*), die beide zwar disoGTP eingebaut haben, aber nicht dHisoGTP. Diejenigen Vertreter der Polymerase B-Familie, die untersucht wurden, akzeptieren disoGTP nicht als Substrat. Diese unterschiedliche Fähigkeit der Polymerasefamilien, disoGTP gegenüber $d^{Me}isoC$ einzubauen, könnte bedeuten, dass die beiden Polymeraseklassen auf unterschiedliche Art mit der kleinen Furche wechselwirken. Falls das disoG-disoC Paar in den frühesten Lebensformen vorkam, dann sollten folgerichtig Polymerasen der A-Familie die ältesten Polymerasen sein. Primer-extension Experimente mit disoGTP und einem $d^{Me}isoC$ - enthaltenden Templat erlauben eine schnelle Klassifizierung neuer Polymerasen.

Die Inkorporation in der Richtung $d^{Me}isoCTP$ gegenüber disoG wurde ebenfalls untersucht. Klenow fragment, HIV-1 RT und *Tth* DNA Polymerase haben $d^{Me}isoCTP$ gegenüber disoG eingebaut. Damit sind sie ein vielversprechender Ausgangspunkt, um mit Hilfe von site-directed mutagenesis Polymerasen zu entwickeln, die auf einem erweiterten genetischen Alphabet basierende DNA amplifizieren.