

Systèmes modèles pour la réplication d'une information génétique.

Jeroen Verhage

▶ **To cite this version:**

Jeroen Verhage. Systèmes modèles pour la réplication d'une information génétique.. Chimie. Ecole Polytechnique X, 2005. Français. NNT : . pastel-00001350

HAL Id: pastel-00001350
https://pastel.hal.science/pastel-00001350

Submitted on 29 Jul 2005

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

présentée à

L'Institut Européen de Chimie et Biologie

2 Rue Robert Escarpit
33607 PESSAC CEDEX

ÉCOLE DOCTORALE DE L'ÉCOLE POLYTECHNIQUE

par Jeroen VERHAGE

POUR OBTENIR LE GRADE DE

DOCTEUR

SPÉCIALITÉ: **CHIMIE ORGANIQUE**

****

Vers des systèmes modèles de la réparation de sites abasiques de l'ADN par réplication non-enzymatique de la base opposée

****

Towards artificial DNA abasic site repair systems based on the non-enzymatic replication of the opposite base

****

Soutenue le: 1 Avril 2005

Après avis de:

Bernard Rayner	(Rapporteur)	Professeur de l'Université de Bordeaux II
Sijbren Otto	(Rapporteur)	Docteur à l'Université de Cambridge

Devant la commission d'examen formée de :

Joelle Prunet	(Présidente)	Professeur chargée de cours à l'Ecole Polytechnique
Bernard Rayner	(Rapporteur)	Professeur de l'Université de Bordeaux II
Sijbren Otto	(Rapporteur)	Docteur à l'Université de Cambridge
Ivan Huc	(Directeur de Thèse)	Directeur de recherche CNRS

All molecules are equal, but some molecules are more equal than others.

Free after George Orwell

Table of contents

Summary 5
Sommaire 9
Samenvatting 14
1 Introduction 19
2 Literature background 21
2.1 Introduction 21
2.2 Templates for synthetic purposes 21
2.2.1 Macrocycles 23
2.2.2 Enzyme mimics 24
2.2.3 DNA based systems 28
2.2.4 Ribozymes 34
2.3 Dynamic combinatorial chemistry and molecular imprinting 36
2.3.1 Dynamic combinatorial chemistry 36
2.3.2 Irreversible reactions 41
2.3.3 Molecular imprinting 42
2.4 Replication model systems 44
2.4.1 Replicating systems 44
2.4.2 Self replication 50
2.4.3 Cross catalytic systems 57
3 Design of a natural abasic site repair system 60
3.1 Introduction and goal 60
3.2 Design of the system 63
3.2.1 Biological relevance 63
3.2.2 Choice of the reaction 64
3.3 Possible behavior of the system 70
3.3.1 Introduction 70
3.3.2 Conformation of the abasic site 71
3.3.3 Molecular modeling 75
3.4 Synthesis of aminobases and hydrazides 84
3.5 Model reactions with ribose and deoxyribose 87
3.5.1 Test reactions with ribose 87

3.5.2 Test reactions with deoxyribose 98
3.5.3 Conclusion 107
3.6 Design of an oligonucleotide sequence 108
3.7 Synthesis 110
3.7.1 Introduction 110
3.7.2 Synthesis of the phosphoramidite and the oligonucleotide 111
3.7.3 Purification of the oligonucleotide 112
3.8 Conclusion 117
3.9 Experimental section 118
4 Design of an artificial abasic site repair system 127
4.1 Introduction and goal 127
4.2 Design of the amines and the aldehydes 128
4.2.1 Choice of the reaction 128
4.2.2 Aldehyde design 129
4.2.3 Molecular Modeling 131
4.3 Aldehydes protected as 1,3-dioxolanes 141
4.3.1 Retro-synthesis 141
4.3.2 Synthesis 143
4.3.3 Deprotection tests 147
4.4 Aldehydes masked as 1,2-diols 152
4.5 Conclusion 155
4.6 Experimental section 155
5 Perspectives 167
5.1 Short term research on natural abasic site repair systems 167
5.2 Short term research on artificial abasic site repair systems 169
5.3 Long term research on artificial abasic site repair systems 172
6 Glossary 175
7 Literature references 177
Word of Thanks / Remerciément / Dankwoord 190

Summary

The goal of this research project is to develop an artificial replicating model system to 'repair' abasic sites within DNA double helices. We will not restore a natural base pair, but introduce a chemical modification that restores duplex stability according to the base facing the abasic site. The opposite base will thus serve to select one "repair reagent" among several possibilities. We amount here to a "information transfer system", or a "replicating system", which is an interesting topic by itself.

Transfer of information stored in DNA is only known with the help of enzymes, but no artificial replicating system able to copy DNA with high fidelity is known. So far, some artificial systems have been shown to transfer information about the length of an oligonucleotide, but not about the sequence of the bases.

We would like our system to make use of reactions that are reversible on the time scale of the experiment. That will allow error correction in the system. If a repair unit not complementary to the opposite base is incorporated, it may then be cleaved off and replaced by a better matching repair unit.

We would also like the repair units to react with the DNA backbone without having to cleave it, as in the case of the enzymatic replacement of a full nucleotide. Strand cleavage has stronger consequences on duplex stability than a simple missing base. We thus believe that avoiding strand cleavage will make the system much more reliable and will increase the chances of successful information transfer.

For our system we chose the reaction of aldehydes with amine type compounds to form imine type products. We chose this group of reactions because they can be made reversible in water at pH 7.

Since a natural abasic site, a deoxyribose, is in equilibrium between its closed hemiacetal and its open aldehyde form, it was a logical choice to use this aldehyde group as an anchor point for our reagents. We thus explored the possibility of chemically 'repairing' such natural abasic sites with amines resembling natural bases. We also explored the possibility to design and incorporate into DNA 'artificial abasic sites' comprising a more reactive aldehyde function, and to restore DNA duplex stability at these sites. These experiments amount to the 'repair' of a non-natural abasic site.

The work is divided in three main chapters; in chapter two an overview of relevant literature is given. Chapter three involves the work we did on the repair of natural abasic sites. Chapter four involves the work we did on the repair of artificial abasic sites.

In chapter two an overview will be given about the literature in the field of template directed synthesis (paragraph 2.2). The formation of macrocycles is discussed (paragraph 2.2.1), as well as DNA based systems that show template effects (paragraph 2.2.3). Dynamic combinatorial chemistry makes use of templated reactions in a different way compared to templated reactions for synthetic purposes. Dynamic combinatorial chemistry will be discussed in paragraph 2.3. The last part of the chapter (paragraph 2.4) involves the development of artificial replicating systems, based on DNA or on synthetic models.
In chapter three the design of a repair model system for natural abasic sites is discussed step by step. We designed (paragraph 3.2) and synthesized (paragraph 3.4) six possible repair units (Figure 0.1), which mimic the natural nucleobases. Compounds **1-4** react reversibly with aliphatic aldehydes in water. Compound **5** mimics a thymine and its hydrazide group is more nucleophilic than the aminobases **1-4**. Compound **6** is expected to have similar reactivity as **5** and the difluorobenzene group was designed as a universal base mimic.

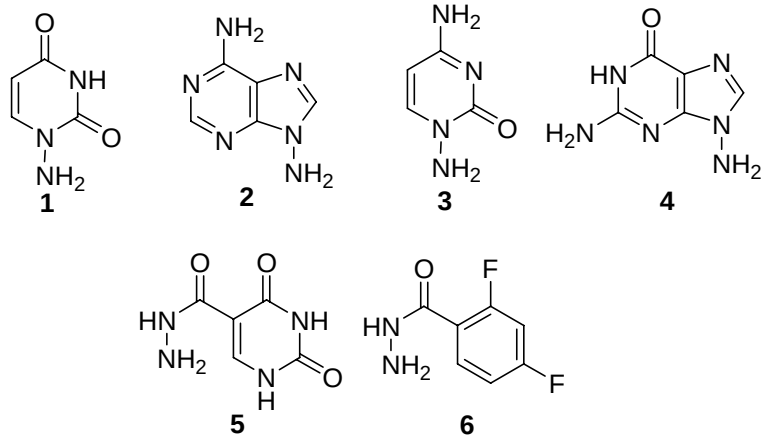


Figure 0.1. Nucleobase mimics.

Paragraph 3.3 describes the molecular modeling experiments we did, using the products of the reaction between a natural abasic site and compounds **1, 5** and **6**. We performed energy minimizations of the hydrazones and hemi-aminals that are possible reaction products. We found positions where both hydrogen bonding and π–π-stacking are possible for the hydrazones formed by the abasic site and compounds **1** and **5** and for the hemi-aminal formed by the abasic site and **5**. No stable conformation for hydrogen bonding and π–π-stacking was found for the hydrazone or the hemi-aminal of the abasic site and **6**, presumably because the

force field AMBER* underestimates the attractive interactions between difluorobenzene and an opposite adenine.

In paragraph 3.5 and 3.6 test reactions between compounds **1-6** and ribose or deoxyribose are discussed. These reactions were set up to find conditions under which compounds **1-6** would react with an RNA or DNA abasic site. Compounds **1-4** were not able to open the hemiacetal of the sugar and did not react with either ribose or deoxyribose. The reactions between ribose or deoxyribose and **5** or **6** were successful. Analysis of the reactions mixtures by ^{1}H and ^{19}F NMR spectroscopy, RP-HPLC and mass spectrometry led to the conclusions that the reaction takes about 24h to reach a final product distribution (water, pH = 6, room temperature). With 20 equivalents of ribose or deoxyribose about 80% of the starting hydrazide is converted into two or three different products respectively after 24h. The hydrolysis of the products takes place with approximately the same reaction rates. The reaction will thus not be reversible in about 30 minutes, the typical time scale of a binding experiment that we foresee with the abasic site incorporated in a DNA oligonucleotide.

Albeit the reaction not being reversible, we wanted to test whether compounds **5** and **6** are able to repair a DNA abasic site. A discussion about how the oligonucleotide was designed is presented in paragraph 3.6. Paragraph 3.7 concludes the chapter with the synthesis and purification of the oligonucleotides and the problems encountered at that stage.

Our initial target was to design a system in which error correction was possible. This was not feasible using a natural abasic site, because of the slow reactions of the base mimics with the only small amount of available aldehyde. We knew that aminobases **1-4** react reversibly with aliphatic aldehydes and we would like to make use of that reaction. In chapter 4 we present the design and synthesis of aldehydes that may be incorporated in DNA as artificial abasic sites and that are not able to form a hemiacetal. The condensation products of the reaction between the repair units and these aldehydes should be in a favorable position for π–π-stacking and hydrogen bonding.

These aldehydes had to be protected during oligonucleotide solid phase synthesis. After DNA synthesis a mild deprotection method should yield the deprotected aldehydes. However, most of the known protecting groups for aldehydes necessitate cleavage under strong acidic conditions. In 1999 Marko *et al*. were the first to publish a method to deprotect 1,3-dioxolanes under neutral conditions using Cerium(IV)ammoniumnitrate (CAN) in water/acetonitrile using a buffer at pH 8. We therefore investigated the synthesis of dioxolane derivatives **7-9** of our aldehydes and this work is presented in paragraph 4.3.2 (Figure 0.2).

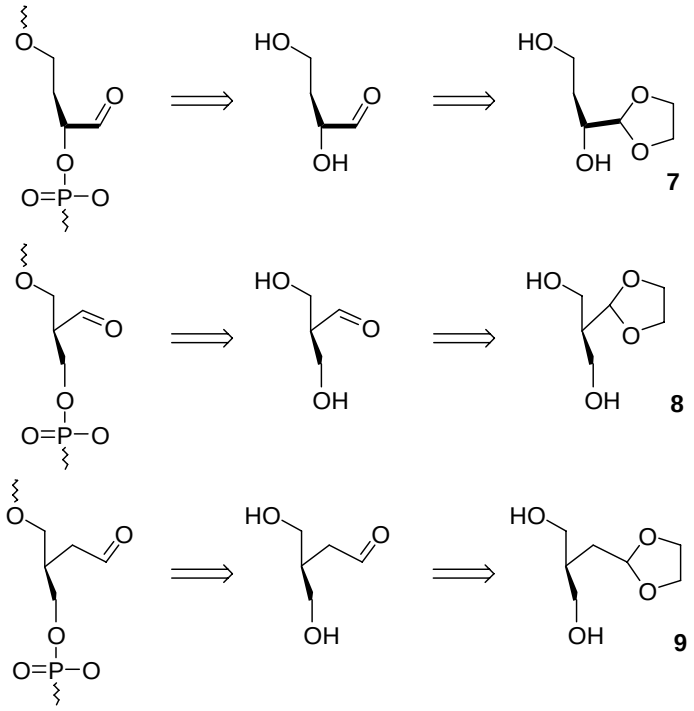


Figure 0.2. Aldehydes as artificial abasic sites.

However, deprotection attempts of various dioxolane model compounds gave mixed results as discussed in paragraph 4.3.3. In fact the reaction did not work the way described by Marko *et al.*, as revealed by Manzo *et al.* three years later. The reaction did not proceed at pH 8 as Marko claimed, but only below pH 4. At this pH, CAN is a very good cleaving agent for DNA and the method is thus not suitable to incorporate aldehydes in a DNA strand. A new method had to be found.

In 2001 an aldehyde was incorporated into a DNA strand for the first time by the group of Zatsepin *et al.*. They used a 1,2-diol, which they protected with benzoyl groups before oligonucleotide synthesis. After incorporation in a DNA strand, the benzoyl groups were cleaved during the standard treatment with ammonia to give the corresponding 1,2-diol. Mild oxidation of the 1,2-diol using $NaIO_4$ yielded the aldehyde.

We designed protected 1,2-diols similar to the dioxolane units shown above, as discussed in paragraph 4.3.4. Their synthesis involved multiple steps to discriminate between the four OH groups and was not completed. A fourth aldehyde was designed that was easier to synthesize and that can be incorporated at the end of a strand. The synthesis of this compound and its phosphoramidite was achieved.

Sommaire

Les recherches décrites dans cette thèse ont pour but le développement de systèmes modèles pour 'réparer' des sites abasiques dans des doubles hélices d'ADN. Nous ne pensons pas rétablir la base naturelle manquante, mais introduirons une modification chimique du site abasique pour restaurer la stabilité du duplex. La base opposée au site abasique sera utilisée pour sélectionner un "réactif de réparation" complémentaire au sens de l'appariement de type Watson-Crick, parmi plusieurs réactifs possibles. Cette étape de sélection devrait donc constituer une sorte de "transfert d'information" ou de "réplication" de la base opposée vers le site abasique, et elle devrait être intéressante a ce titre.

Le transfert de l'information contenu par l'ADN ne peut se faire a priori seulement avec l'aide d'enzymes. Un système artificiel capable de copier l'ADN avec une bonne fiabilité n'est pas encore connu. Seule l'information concernant la longueur d'un segment d'ADN a été transférée vers un brin complémentaire en l'absence d'enzymes, pas l'information contenue dans la séquence.

Nous voudrions que notre système de réparation de sites abasiques utilise des réactions qui soient réversibles à l'échelle de l'expérience. Cela devrait permettre une correction *in situ* des erreurs commises par le système. Si une unité de réparation est incorporée dans le brin d'ADN qui n'est pas complémentaire avec la base opposée, elle pourrait être coupée et remplacée par une unité qui stabilise le duplex davantage.

Nous voudrions également que les unités de réparation réagissent avec le squelette de l'ADN, mais sans le couper, comme c'est le cas lors du remplacement enzymatique d'un nucléotide entier. Une coupure du brin aurait des conséquences plus graves pour la stabilité du duplex, qu'une simple base manquante. Nous pensons qu'éviter la coupure du brin rendra notre système beaucoup plus fiable et augmentera la probabilité de réussite du transfert d'information.

Nous avons choisi la réaction entre des aldéhydes et des dérivés d'amines pour former des dérivés d'imines. Ces réactifs peuvent réagir de manière réversible dans l'eau à pH 7.

Parce que le site abasique naturel, un désoxyribose, est en équilibre entre sa forme hémiacetal cyclique et sa forme aldéhyde ouverte, il semble logique d'utiliser cet aldéhyde comme point d'ancrage pour nos réactifs. Nous avons exploré la possibilité de réparer ces sites abasiques naturels de façon chimique avec des amines qui ressemblent aux bases naturelles. Nous avons également exploré la possibilité de concevoir et d'incorporer des 'sites abasiques artificiels' dans un brin d'ADN. Ces sites abasiques artificiels sont constitués d'une fonction aldéhyde

plus réactive et nous avons essayé de restaurer la stabilité des duplex contenant un tel site abasique. Ces expériences représentent la 'réparation' de sites abasiques artificiels.

Cette thèse est divisée en trois chapitres principaux; dans le chapitre deux une étude bibliographique de la littérature sur laquelle nos travaux sont basés est présentée. Le chapitre trois contient les travaux sur la réparation des sites abasiques naturels. Le chapitre quatre concerne les travaux sur la réparation des sites abasiques artificiels. Le contenue ces trois chapitres est brièvement résume ci-après.

Dans le chapitre deux, un résumé concernant les travaux effectués sur la synthèse par effet gabarit (en anglais: template) est donné (paragraphe 2.2). La formation de macrocycles est discutée (paragraphe 2.2.1), ainsi que des systèmes basés sur l'ADN qui montrent un effet gabarit (paragraphe 2.2.3). La chimie combinatoire dynamique utilise les réactions gabarit dans des mélanges complexes. Elle est présentée dans le paragraphe 2.3. La dernière partie du chapitre 2 (paragraphe 2.4) contient le développement de systèmes de réplication artificiels, basés sur l'ADN ou sur des modèles synthétiques.
Dans le chapitre trois on discutera le système de réparation des sites abasiques naturels que nous avons conçu. Nous avons conçu (paragraphe 3.2) et synthétisé (paragraphe 3.4) six molécules de réparations possibles (Figure 0.1) qui miment les nucléobases naturelles. Les composés **1-4** réagissent de manière réversible avec des aldéhydes aliphatiques dans l'eau. Le composé **5** mime une thymine et son groupement hydrazide est plus nucléophile que les aminobases **1-4**. Le composé **6** devrait avoir une réactivité similaire à **5** et son groupement difluorobenzoique est conçu comme un mime de base universelle.

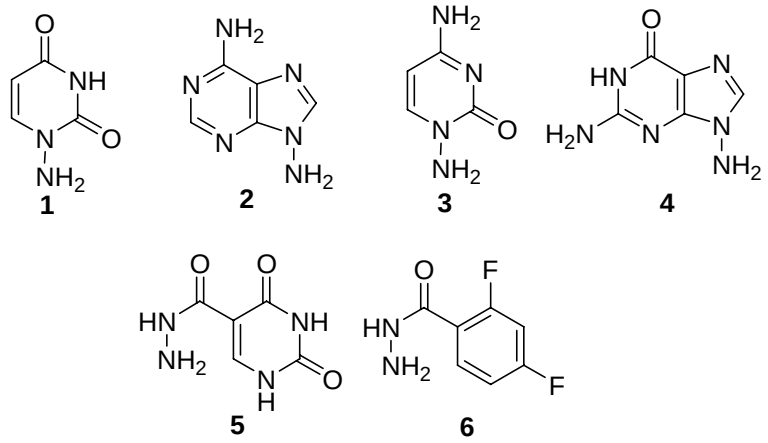


Figure 0.1. Mimes de nucleobases.

Le paragraphe 3.3 décrit les expériences de modélisation moléculaire que nous avons faites, pour évaluer la stabilité des produits des réactions entre un site abasique naturel et les composés **1**, **5** et **6**. Nous avons accompli des minimisations d'énergie des hydrazones et des hémi-aminals possibles. Nous avons cherché des conformations dans lesquelles la formation des liaisons hydrogène avec la base opposée et de l'empilement π–π avec les bases voisines étaient possibles. Nous avons trouvé ces positions favorables pour les hydrazones formées par la réaction entre le site abasique et **1** ou **5,** et pour la forme hémi-aminal du composé formé par le site abasique et **5**.

Une conformation favorable pour la formation des liaisons hydrogène et de l'empilement π–π avec les bases voisines n'a pas été trouvée pour les produits de la réaction entre le site abasique et **6**, vraisemblablement parce que le champ de force AMBER* sous-estime les interactions attractives entre le difluorobenzene et l'adénine opposée.

Dans les paragraphes 3.5 et 3.6 des réactions tests entre les composés **1-6** et le ribose et le désoxyribose sont décrites. Ces réactions ont été réalisées dans le but de trouver des conditions dans lesquelles **1-6** pourront réagir avec des sites abasiques naturels dans un brin d'ARN ou ADN. Les composés **1-4** ne sont pas capables d'ouvrir l'hémiacetal du sucre et ne réagissent pas avec le ribose ou le désoxyribose. Les réactions entre le ribose ou le désoxyribose et **5** ou **6** ont bien marché. L'analyse des mélanges de réactifs par spectroscopie RMN du ^{1}H et du ^{19}F, par HPLC en phase inverse et par spectrométrie de masse, ont conduit à la conclusion que la réaction a besoin d'environ 24h pour arriver à une distribution de produits stable (eau, pH 6, température ambiante). Avec 20 équivalents de ribose ou de désoxyribose, environ 80% de l'hydrazide est converti en respectivement deux ou trois produits après 24h. La cinétique de la réaction inverse, c'est à dire l'hydrolyse des produits, a lieu dans la même gamme de temps.

La réaction n'est donc pas réversible en environ 30 minutes, l'échelle de temps typique prévue pour une expérience de mesure de la stabilité d'un duplex réparé. Malgré le fait que les réactions ne sont pratiquement pas réversibles, nous avons souhaité tester si les composés **5** et **6** sont capables de réparer un site abasique d'ADN. Le dessin d'un oligonucleotide contenant un site abasique est présenté dans le paragraphe 3.6. Le paragraphe 3.7 clôt le chapitre avec la synthèse et la purification de l'oligonucléotide et fait l'analyse des problèmes de stabilité rencontrés à cette étape.

Notre but initial était de concevoir un système dans lequel la correction d'erreurs était possible. Ce n'était pas faisable avec les sites abasiques naturels, parce que les réactions des

mimes des bases avec la petite fraction d'aldéhyde des sites abasiques sont trop lentes. Nous savions que les aminobases **1-4** réagissent de manière réversible avec des aldéhydes aliphatiques et nous voudrions utiliser cette réaction. Dans le chapitre 4 nous présentons le dessin et la synthèse d'aldéhydes qui peuvent être incorporés dans l'ADN comme des sites abasiques artificiels. Ces aldéhydes ne peuvent pas former des hémiacétals. Les produits des réactions entre les sites abasiques artificiels et les aminobases **1-4** sont conçus de façon à ce qu'ils trouvent facilement une position favorable pour former des liaisons hydrogène et pour l'empilement π–π.

Les aldéhydes doivent être protégés pendant la synthèse d'oligonucléotides en phase solide. Après la synthèse d'ADN, l'application d'une méthode de déprotection douce doit donner les aldéhydes non-protégés. Malheureusement, presque tous les groupes protecteurs d'aldéhydes nécessitent une déprotection en milieu acide fort. En 1999 Marko *et al*. étaient les premiers à publier une méthode de déprotection de 1,3-dioxolanes dans des conditions neutres en utilisant le nitrate d'ammonium et de cérium (IV) (CAN) dans de l'eau / acétonitrile avec un tampon à pH 8. Nous avons donc exploré la possibilité de synthétiser des dérivés de dioxolanes **7-9** comme précurseurs de nos aldéhydes désirés. Ces travaux sont présentés dans le paragraphe 4.3.2 (Figure 0.2).

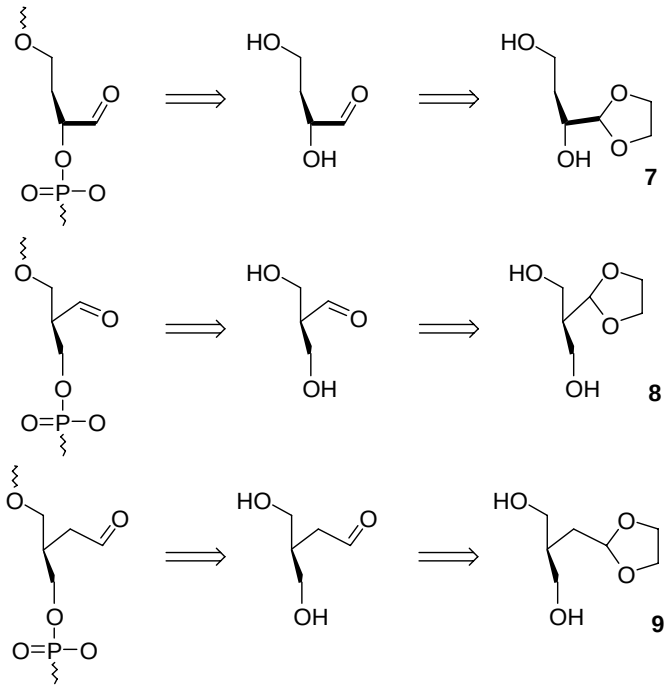


Figure 0.2. Des aldéhydes comme sites abasiques artificiels.

Malheureusement, les essais de déprotection des dioxolanes modèles variés ont donné des résultats mixtes, comme discuté dans le paragraphe 4.3.3. La réaction ne marche pas comme

décrit par Marko *et al.*, ainsi que Manzo *et al.*l'a souligné trois ans plus tard. La réaction n'a pas lieu à pH 8, comme revendiqué par Marko, mais seulement en dessous du pH 4. A ce pH acide, le CAN est un très bon agent de coupure d'ADN, cette méthode n'est donc pas valable pour incorporer des aldéhydes dans un brin d'ADN. Une autre méthode doit être trouvée.

En 2001, un aldéhyde était incorporé dans un brin d'ADN pour la première fois par le groupe de Zatsepin *et al.*. Ils utilisaient un 1,2-diol, protégé par des groupes benzoyls pendant la synthèse d'oligonucléotides. Après incorporation dans un brin d'ADN, les groupements benzoyls étaient coupés par le traitement standard avec de l'ammoniaque, pour donner le 1,2-diol correspondant. L'oxydation douce du 1,2-diol avec $NaIO_4$ donnait l'aldéhyde.

Nous avons conçu des 1,2-diols, similaires aux dioxolanes dessinés ci-dessus, comme discuté dans le paragraphe 4.3.4. La synthèse nécessite des étapes multiples pour discriminer les quatre groupements hydroxyles, et n'est pas achevée. Un quatrième aldéhyde a ete conçu, plus facile à synthétiser et qui peut être incorporé au bout d'un brin d'ADN. La synthèse de ce composé et de son phosphoramidite est présentée.

Samenvatting

Het doel van dit promotieonderzoek is het ontwikkelen van modelsystemen voor het repareren van ‘abasic sites’ in dubbele DNA helices. We zullen niet de natuurlijke base herstellen, maar we willen een chemische modificatie inbrengen, welke de stabiliteit in de dubbele helix herstelt. Welk reagens ingebouwd wordt hangt af van de base die zich tegenover de ‘abasic site’ in de complementaire keten bevindt. Deze base zal gebruikt worden om een ‘reparatie reagens’ te selecteren, complementair volgens Watson-Crick paring. Deze selectie kan worden gezien als het kopiëren van informatie in de ‘abasic site’.

Het kopiëren van informatie van de ene DNA-keten naar de andere, is tot op heden slechts mogelijk met behulp van enzymen. Er is geen enkel synthetisch kopieersysteem bekend dat met goede betrouwbaarheid DNA kan kopiëren. Er zijn enkele kunstmatige systemen bekend die de informatie over de lengte van een DNA keten kunnen overbrengen, maar het kopiëren van de informatie over de volgorde van de basen is tot op heden niet mogelijk gebleken.

We willen ons systeem zo ontwikkelen dat het reversibel is op de tijdschaal van onze experimenten. Dit zal het mogelijk maken om fouten direct door het systeem zelf te laten corrigeren. Als een molecuul wordt ingebouwd dat niet complementair is met de base in de andere DNA keten, kan deze worden gehydrolyseerd en worden vervangen door een beter passende verbinding.

We willen dat onze reacties plaatsvinden met de ‘backbone’ van de DNA-keten, zonder de keten in tweeën te splitsen, zoals by enzymatische reparatie het geval is. Het in tweeën splitsen van de DNA-keten zal grotere gevolgen hebben voor de stabiliteit van de helix dan wanneer slechts één base in de keten mist. In onze ogen zal het vermijden van het in tweeën splitsen van de keten een veel betrouwbaarder systeem opleveren, wat de kans op het succesvol kopiëren van informatie vergroot.

We kozen voor ons systeem de reacties tussen aldehydes en amines, onder de vorming van imines. We kozen deze reacties omdat ze reversibel gemaakt kunnen worden in water met een neutrale pH.

Aangezien een natuurlijke ‘abasic site’, deoxyribose, in evenwicht is tussen zijn gesloten half-acetaal vorm en zijn open aldehyde vorm, was het logisch deze aldehyde als aangrijpingspunt te nemen voor onze reacties. We hebben onderzoek gedaan naar de mogelijkheid deze ‘abasic sites’ te ‘repareren’ met amines die lijken op natuurlijke basen. Bovendien hebben we onderzocht of het mogelijk was synthetische ‘abasic sites’ te ontwerpen en te synthetiseren en de stabiliteit in een dubbele DNA-helix te herstellen. Deze kunstmatige ‘abasic sites’ hebben

een aldehyde-groep welke reactiever is dan de natuurlijke ‘abasic sites’, zodat mildere condities mogelijk zijn.

Het proefschrift is verdeeld in drie hoofdstukken; in hoofstuk twee wordt een overzicht gegeven van de relevante literatuur. In hoofdstuk drie wordt de reparatie van natuurlijke ‘abasic sites’ behandeld. Hoofdstuk vier gaat over de reparatie van synthetische ‘abasic sites’. Deze drie hoofdstukken worden nu kort besproken.

In hoofdstuk twee wordt een overzicht gegeven van de literatuur over ‘template’ gestuurde synthese (paragraaf 2.2). Dit houdt onder andere de vorming van grote ringsystemen (paragraaf 2.2.1) in. Ook op DNA gebaseerde systemen die een ‘template’ effect laten zien worden behandeld (paragraaf 2.2.3). ‘Dynamic combinatorial chemistry’ maakt gebruik van ‘template’ gestuurde reacties, maar op een andere manier dan de reacties voor synthetische doeleinden. Dit zal worden besproken in paragraaf 2.3. Het laatste deel van hoofdstuk twee (paragraaf 2.4) behandelt de ontwikkeling van kunstmatige kopieersystemen, gebaseerd op DNA of synthetische modellen.

In hoofdstuk drie wordt het ontwerp van het reparatiesysteem voor natuurlijke ‘abasic sites’, stap voor stap besproken. We ontworpen (paragraaf 3.2) en synthetiseerden (paragraaf 3.4) zes mogelijke reagentia (Figure 0.1), die de natuurlijke basen nabootsen. Aliphatische aldehydes reageren reversibel met **1-4** in water. Product **5** bootst een thymine na, en zijn hydrazide-groep is reactiever dan de amines van de aminobasen **1-4**. Product **6** is ongeveer net zo reactief als **5**, en de difluorobenzeen-groep was ontworpen als universele base.

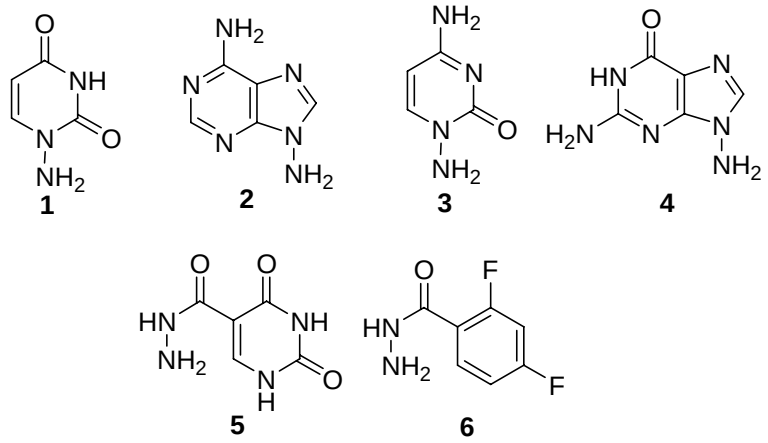


Figure 0.1. Reagentia voor de reparatie van natuurlijke ‘abasic sites’.

De modelisatieexperimenten die we uitgevoerd hebben met de aldehyde en verbindingen **1**, **5** en **6** worden besproken in paragraaf 3.3. We modelleerden de hydrazonen en de half-aminals die mogelijk zijn door de natuurlijke 'abasic site' te reageren met **1, 5** of **6**. Posities waarin π–π-overlap met de naburige basen en waterstofbruggen met de complementaire base mogelijk waren, werden gevonden voor de hydrazonen van **1** en **5**. Ook voor de hemi-aminal vorm van de reactie tussen de natuurlijke 'abasic site' en **5** werd een goede positie gevonden. Voor de producten van **6** met de abasic site werd geen goede positie gevonden voor waterstofbruggen of π–π-overlap. Dit komt zeer waarschijnlijk doordat het modelisatiekrachtveld AMBER* de attractieve interacties tussen difluorobenzeen en de complementaire adenine onderschat.

In paragraaf 3.5 en 3.6 worden testreacties tussen **1-6** en ribose of deoxyribose besproken. Deze reacties zijn uitgevoerd om condities te vinden onder welke **1-6** met een 'abasic site' zouden kunnen reageren. **1-4** bleken niet in staat de half-acetaal van de suikers te openen en reageerden niet met ribose, noch met deoxyribose. De reacties tussen deoxyribose of ribose en **5** of **6** waren wel succesvol. Analyse van de reactiemengsels met behulp van ^{1}H en ^{19}F spectroscopie, inverse fase HPLC en massaspectrometrie, leidde tot de conclusie dat de reactie ongeveer 24 uur nodig heeft om tot een definitieve productverdeling te komen (water, pH = 6, kamertemperatuur). Als 20 equivalenten ribose of deoxyribose worden toegevoegd aan **5** of **6**, wordt in 24 uur ongeveer 80% van **5** of **6** omgezet in respectievelijk twee of drie producten. Hydrolyse van de producten gebeurt met vergelijkbare reactiesnelheden. De reactie zal dus niet reversibel zijn tijdens de geplande bindingsexperimenten, welke ongeveer een half uur per experiment zullen duren.

Alhoewel de reactie niet reversibel is, willen we toch **5** en **6** testen en kijken of het mogelijk is een natuurlijke DNA 'abasic site' te repareren. Het ontwerp van de DNA-keten dat voor deze tests nodig is, wordt besproken in paragraaf 3.6. Paragraaf 3.7 besluit het hoofdstuk met de synthese en purificatie van de oligonucleotides en de problemen die we in deze fase zijn tegen gekomen.

Ons initiële doel was het ontwikkelen van een systeem dat automatisch zijn eigen fouten herstelt. Dit bleek niet haalbaar met natuurlijke 'abasic sites', vanwege de langzame reacties tussen **5** en **6** en de slechts kleine hoeveelheid aanwezige aldehyde. Het was ons bekend dat aminobasen **1-4** reversibel met aliphatische aldehydes reageren. We wilden daarvan graag gebruik maken en in hoofdstuk 4 wordt daarom het ontwerp en de synthese van aldehydes besproken, die ingebouwd kunnen worden in een DNA-keten. Deze aldehydes kunnen geen

half-acetalen vormen, en zijn daarom reactiever dan de natuurlijke 'abasic site'. De aldehydes zijn zo ontworpen dat na reactie met de aminobasen, de producten een goede positie kunnen vinden waarin π–π-overlap met de naburige basen kan plaatsvinden, alsmede waterstofbruggen met de complementaire keten gevormd kunnen worden.

Tijdens synthese van de oligonucleotides moeten de aldehydes beschermd worden. Na de synthese moet deze beschermende groep vervolgens gemakkelijk verwijderd kunnen worden. De meeste bekende beschermende groepen worden echter ontschermd onder sterk zure condities. In 1999 werd door Marko *et al*. de eerste methode gepubliceerd die 1,3-dioxolaan beschermde aldehydes kon ontschermen met behulp van Cerium(IV)ammoniumnitraat (CAN) in water/acetonitril met een buffer (pH 8). We hebben daarom de synthese van dioxolanes **7-9** onderzocht en dit wordt besproken in paragraaf 4.3.2 (Figure 0.2).

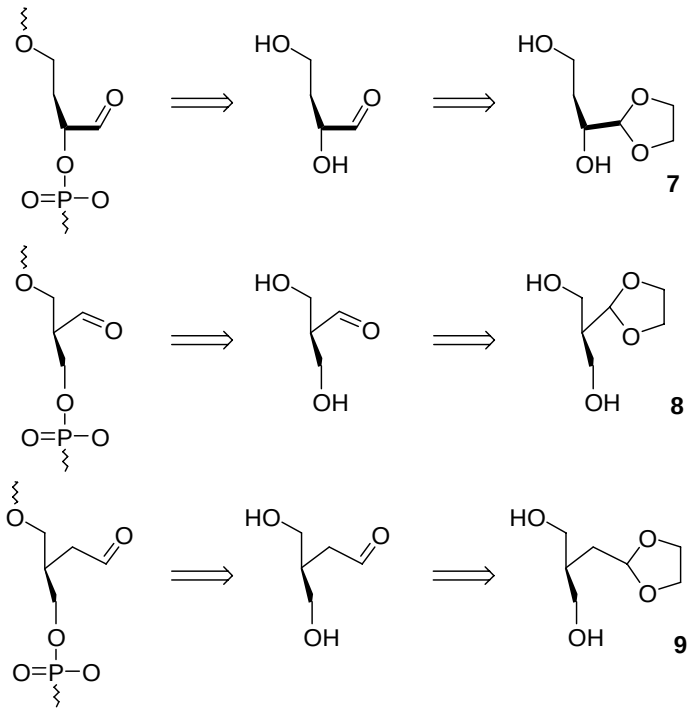


Figure 0.2. Aldehydes als synthetische 'abasic sites'.

Ontscherming van verschillende dioxolane-modellen gaf echter geen goede resultaten, zoals wordt besproken in paragraaf 4.3.3. Het bleek dat de reactie niet werkte onder de condities zoals beschreven door Marko. Drie jaar na het verschijnen van het artikel van Marko werd door een andere groep gepubliceerd dat de reactie niet bij pH 8 werkt, maar alleen beneden pH 4. DNA is echter niet stabiel met CAN onder zure pH, de fosfaatbindingen worden met CAN gehydrolyseerd. Een moest dus een andere methode worden gevonden.

In 2001 werd voor het eerst een aldehyde in een DNA-keten ingebouwd door de groep van Zatsepin *et al*.. Zij synthetiseerden een 1,2-diol die zij beschermden met benzoylgroepen voor

DNA-synthese. De 1,2-diol wordt ontschermd tijdens de standaardbehandeling met ammonia. Milde oxidatie met $NaIO_4$ geeft vervolgens de aldehyde.

Net als voor de 1,3-dioxolanes werden drie 1,2-diolen ontworpen die na DNA-synthese en ontscherming de gewenste aldehydes moesten opleveren. De synthese van deze 1,2-diolen is echter een meerstapssynthese om onderscheid te maken tussen de vier hydroxy groepen en de syntheses zijn daarom niet voltooid. Een vierde aldehyde is ontworpen die makkelijker te synthetiseren is en kan worden ingebouwd aan het eind van een DNA-keten. De synthese van deze aldehyde en zijn fosforamidite is uitgevoerd.

1 Introduction

The development of life has something mysterious. How did life develop the way it did? Is there a chance that on other planets life has developed? Will those life forms be similar to life on earth, or completely different? What is important in the development, which are the critical stages?

A lot of far reaching questions that are going far beyond the scope of one thesis. This thesis will address one part of one small subject only; how might replicating systems have developed on the early earth? A reliable replicating system is one of the prerequisites of life. The development of membranes and systems to store and convert energy are two examples of other vital steps in the development of life, but we will not comment on these topics here.

Many people already addressed questions about the development of replicating systems. RNA was proposed as a possible early replicating system from which DNA developed [1]. Others [2] do not agree and think RNA and DNA both had their function and developed simultaneously. But was RNA the *first* replicating system? It is a very complicated molecule to synthesize out of the blue, without any resembling predecessors. An inorganic 'replicating clay' was therefore hypothesized by Cairns-Smith [3]. In this clay some organic compounds might have come together, reacted, and started replicating themselves. Through evolution the most successful replicator became the most abundant.

Many DNA analogues have been synthesized over the years and tested for their ability to transfer information of the nucleobases of which they consist. No reliable replication of DNA or any DNA analogues has so far been reported though. However, small steps towards knowing more about how early replicating systems looked like were achieved. The synthesis of the nucleobases and even nucleosides under 'prebiotic' conditions was successful. How those nucleosides then polymerized is not yet known.

The replicating model system we intend to develop in this thesis is different from the natural systems in the sense that it does not involve linking recognition units (nucleotides) together to assemble an oligomeric sequence, but instead involves linking recognition units at specific sites of a pre-formed and stable backbone (Figure 1.1). We will not copy full nucleotides as in DNA, but only replace the bases that are absent in a blank backbone that is already present. We think that such a system may lead to a reliable replication of a DNA sequence and would certainly be more efficient than letting nucleotides polymerize in the absence of enzymes.

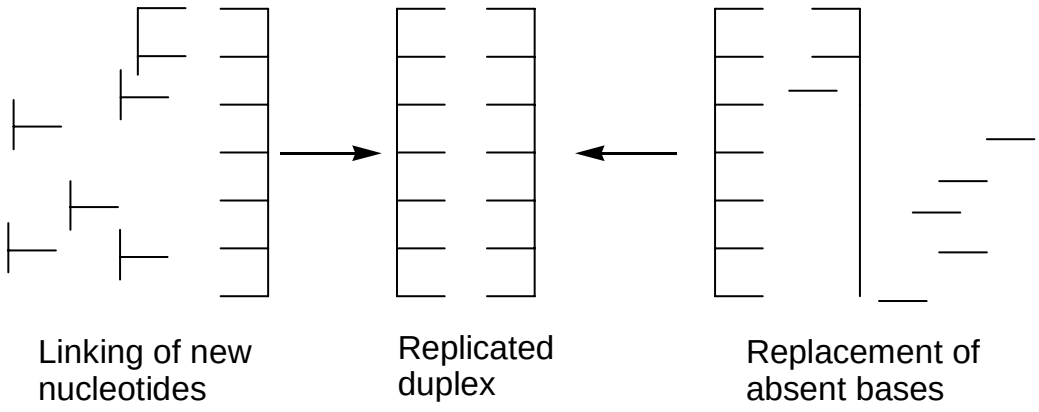


Figure 1.1. Enzymatic DNA replication compared to base replacement in a present backbone.

An important feature of our design is the introduction of automatic error correction. In cells, enzymes constantly repair damaged or wrongly incorporated nucleotides. In our opinion, a wrongly incorporated nucleotide will severely hinder successful replication of the rest of the oligonucleotide. We will use reversible reactions, allowing mismatching bases to be replaced by a better match. In most replicating systems that were so far developed, error correction was not possible because of the irreversible reactions that were used [2, 4, 5].

As a start for the design of our replicating systems, we made use of natural abasic sites. These are deoxyribose units that miss the nucleobase. We will try to replicate the base facing the abasic site into a complementary unit replacing the abasic site, which locally restores base pairing, without the use of enzymes. We called this 'abasic site repair'. In a second stage we designed artificial abasic sites and tried to develop a reliable information transfer system that is able to correct its own errors

Error correction and the use of a stable backbone should, in our opinion, provide better chances of enzyme free replication. Such a system may remain very far from the replicating systems that have developed at the early stages of life. However, we hope that the ideas presented here are a source of inspiration for others to discover which replicating systems were the basis for life as we know it.

This thesis is divided in three main chapters. Chapter two contains the relevant literature. In chapter three a natural abasic site repair system will be presented. In chapter four an artificial abasic site repair system will be designed.

2 Literature background

2.1 Introduction

In this chapter a brief literature overview is given on various aspects of templated reactions, and in particular on templated reactions involving DNA. I will give some selected illustrative examples but I do not intend to be exhaustive.

I will show various examples of the development of templated reactions, for instance the formation of macrocycles, which sometimes does not take place without a proper template. Templating is not a new, artificial notion, but nature developed the principle of templated reaction millions of years ago. Enzymes bring reagents in close proximity. Inspired by this, various authors have designed “synthetic enzyme mimics”, that I will briefly describe. Complementary DNA strands have the ability to recognize each other and bind in a specific and predictable manner. They are thus excellent moieties to bring reagents to predefined positions and can be excellent templates. Optimizing catalysts based on DNA can give rise to very good templating with major increases in reaction rates.

Most of the reactions described in the first part of the chapter deal with the ability of templates to bring reagents in close proximity, thus accelerating the reaction. The second part of this chapter will deal with a second property of templates, the ability to enhance the formation of a product among a mixture of many undesired side products. This is called dynamic combinatorial chemistry (DCC). In principle, DCC involves reversible reactions, but the concept has been extended to irreversible reactions as well. Templating the formation of a desired receptor cavity by polymerization is well known as molecular imprinting.

The last part of the chapter will deal with systems capable of information transfer. DNA is the obvious example when thinking of information transfer. A few templates will be discussed that are able to produce copies of themselves. Mind that DNA replication does not yield a copy of the initial strand, but a complementary strand. Replication of the complementary strand gives a copy of the initial DNA strand. Models of this phenomenon are called 'cross-catalytic systems'.

2.2 Templates for synthetic purposes

"Templates preorganize assemblies of atoms into a particular geometric shape, in order to achieve a particular linking of these atoms" [6]. A template provides instructions for the

formation of one single product from one or more substrates, which, in absence of the template, have the potential to assemble and react in a variety of ways.
A DNA strand is for instance templated by its complementary strand. Genetic copying of a trinucleotide like e.g. $^{5'}$ACG$^{3'}$ yields another trinucleotide with the sequence $^{5'}$CGT$^{3'}$. If a trinucleotide is formed randomly from the four possible base pairs, $4^3 = 64$ products will be formed.
Ideally, each template should yield one unique product. Changing the template results in a different substrate-template assembly and consequently yields a different product. For example, changing the sequence from the trinucleotide above to $^{5'}$GCA$^{3'}$ would result in the formation of the sequence $^{5'}$TGC$^{3'}$.

The distinction between a template and a catalyst is that the template organizes the geometry of the reaction, but is not participating in the reaction. A catalyst is not necessary pre-organizing the reagents, but it is usually participating in the reaction. It is thereby temporarily modified and later formed back. Acid catalysis of ester formation for example, involves protonation of the acid group in a first step. In a second step the ester is formed and in the last step the proton is recovered from the formed ester. It is sometimes hard to determine whether a system shows catalysis or templating. Various systems show both effects simultaneously.
Two types of templates can be distinguished, 'thermodynamic' and 'kinetic' templates [6, 7]. 'Thermodynamic templating' occurs when the reactions are reversible. The template binds to one of the products of the reaction, thereby shifting the equilibrium towards this species. $-\Delta G$ of that specific reaction is increased. The template may or may not accelerate product formation, but in the end, all what matters is how well the product is bound. High yields can be obtained by properly designing the template.
'Kinetic' templates operate on irreversible reactions. They stabilize the transition states leading to the product. They bring the reagents in close proximity, thereby lowering $\Delta S^{\ddagger}$. This increases their apparent concentration and thus their reactivity. Product stabilization is not necessary, but 'kinetic' templates almost invariably bind the product more strongly than the starting material(s) [6]. This is interpreted as the result of 'late transition states'. The templates bind to those transition states, which generally resemble the products more than the reagents.
I will now discuss some illustrative examples of templated reactions

2.2.1 Macrocycles

Macrocyclizations are prototypes of reactions that can be templated. They are not *a priori* favored over polymerization. In addition they give rise to a cavity where a template can be positioned to favor ring closure.

Electron rich aromatics can template the formation of macrocycles containing electron deficient aromatics very efficiently. For example, the nucleophilic attack of the pyridine unit on the methyl bromide of **2** as shown in Figure 2.1 is enhanced a factor of 1900 in the presence of 20 equivalents of the crown ether template **3** [8]. The template binds to a viologen unit and serves to bring the two reactive units into close proximity and is thus an example of a 'kinetic' template. Without the template the reactive units are on opposite sites of the molecules and the formation of polymers is favored.

In this particular case, the template becomes an integral part of the structure after the nucleophilic reaction to form a rotaxane. Recovery of the template for a second reaction is thus not possible.

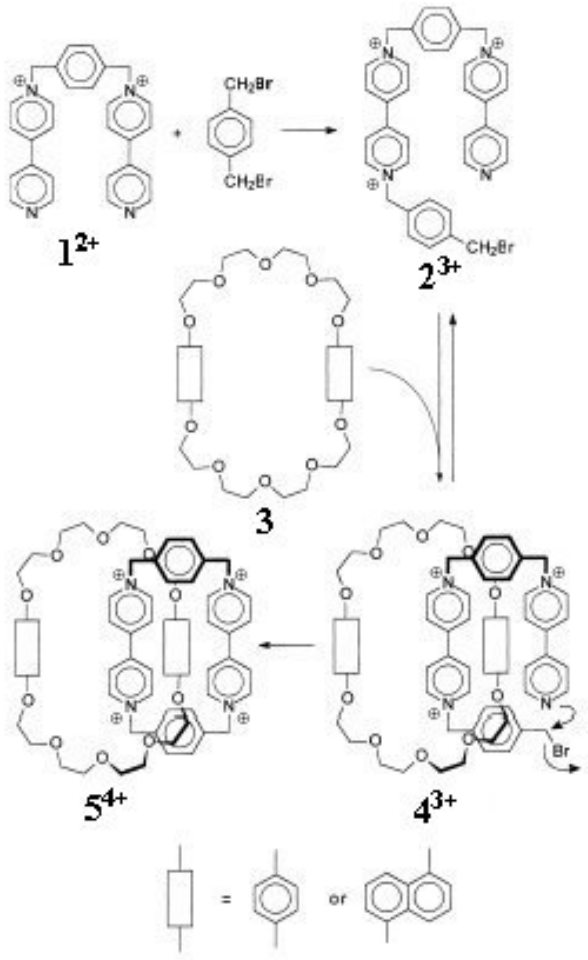


Figure 2.1. Rotaxane formation templated by a crown ether

I mentioned before that changing the template may give a different products. This was for instance shown by Hasenknopf *et al.* in 1997 [9]. Simply changing the anion in the system shown in Figure 2.2 gives a hexanuclear circular helicate instead of a pentanuclear.

The sulfate ion is too big to be contained by the pentanuclear helicate, the system is thus directed to the hexanuclear species. Selection seems to be made on size rather than on charge. Also $Fe(BF_4)_2$ and $FeSiF_6$ form hexanuclear helicates, whereas $FeBr_2$ forms both penta- and hexanuclear species. Mind that the presence of the sulfate ion in the hexanuclear species was never proven.

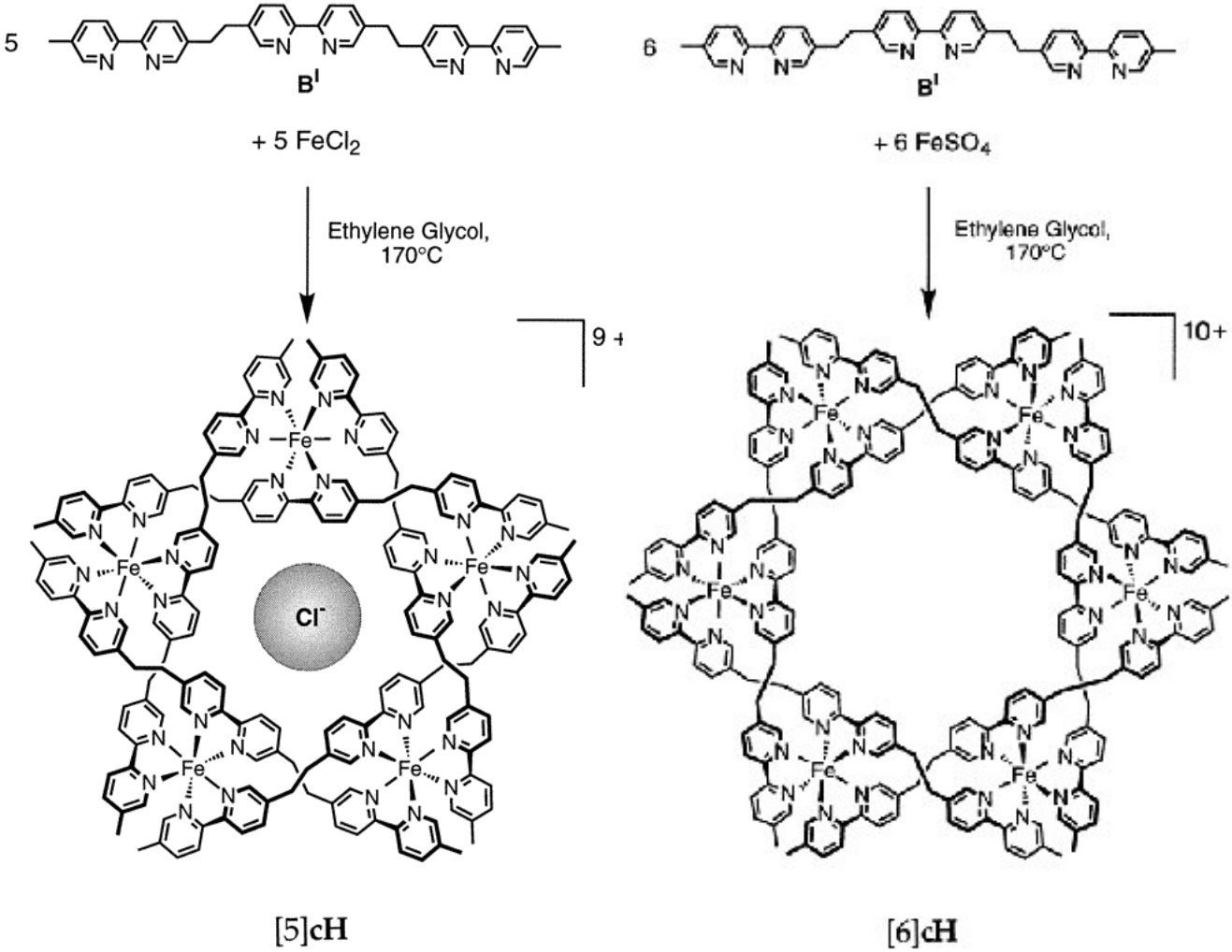


Figure 2.2. Changing template induces formation of another product.

2.2.2 Enzyme mimics

Enzymes strongly bind to their substrates and make extensive use of template effects. Reactions that do not occur under physiological conditions do occur with the help of enzymes. Reactions that are normally *inter*molecular become pseudo *intra*molecular in the enzyme,

which increases their rates significantly. Enzyme reaction kinetics are also enhanced because of transition states stabilization, but the role of bringing the reagents together is significant [10].

For every biological reaction a specific enzyme exists to favor the formation of one specific product. Reaction rate enhancements can be huge. Orotic acid monophosphate decarboxylase accelerates the decarboxylation of orotic acid monophosphate with a factor of 10^{17} [10]. Hydration of CO_2 by carbonic anhydrase is accelerated with a factor of "only" 10^7, but with a rate constant of $10^6 s^{-1}$ it is one of the fastest known enzymes [10, 11].

Enzymes served as biological examples for the design of artificial templates. Enzyme-like templates were developed to accelerate specifically a given *inter*molecular reaction by transforming it into a pseudo-*intra*molecular reaction.

This was for instance shown in the design of the enzyme mimic described in the early 1980's by Breslow *et al*. (Figure 2.3) [12, 13]. Pyridoxamine phosphate is known to be a co-enzyme for transamination reactions and it is known that this unit can also catalyze transamination reactions without the enzyme. The yields and selectivity of the reaction are lower in that case. β-cyclodextrine is known to bind various substrates in a specific manner and able to catalyze various reactions with high yields and rate accelerations. The authors speculated that linking the pyridoxamine phosphate unit to a β-cyclodextrine would yield a selective synthetic enzyme that was able to selectively form various aminoacids from the corresponding keto acids.

Two compounds were synthesized (**6** and **7**), **6** having the catalytic group on the opposite site of the cyclodextrine compared to **7**. The rates of transamination of compounds **8-10** to give compounds **11-13** templated by **6** were compared. Pyruvic acid **8** was found to react at the same rate compared to transamination with pyridoxamine which does not contain a cyclodextrine moiety. Phenylpyruvic acid **9** reacted approximately 35 times faster than the reference transamination with only pyridoxamine. Indolepyruvic acid **10** was converted into tryptophan **13** with a rate of 50 times the reference reaction. Compound **7** also shows rate accelerations for **9** and **10**, but they are about a factor 2 lower compared to **6**.

Templates **6** and **7** turned out not only to show rate accelerations, but also chiral biasing of one stereoisomer over another. Compound **6** was able to favor the formation of L-phenylalanine over D-phenylalanine 5/1. Compound **7** does not show any detectable enantiomeric preference in this case. However, in the conversion of indolepyruvic acid to tryptophan, **6** shows a 2/1 preference for the *L*-enantiomer, while **7** shows a 1.8/1 preference

for the *D*-enantiomer. This reversal of stereochemical selectivity was attributed to the helicity of the cyclodextrine relative to the catalyzing part of the molecule.

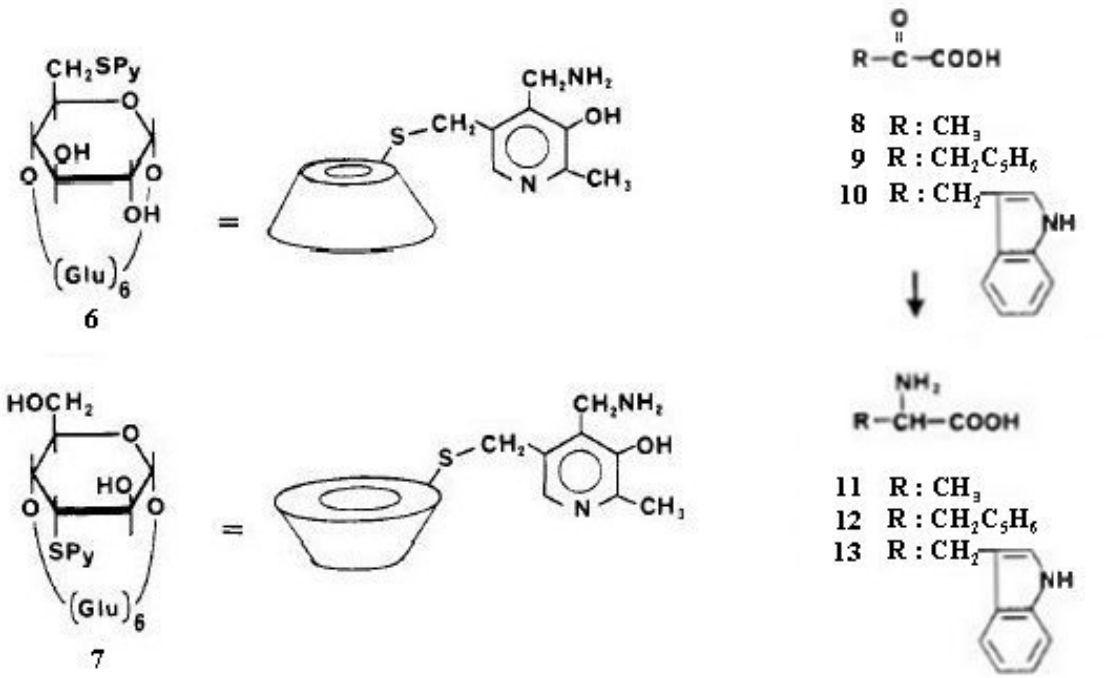


Figure 2.3. Cyclodextrine enzyme mimics for transamination reactions.

A mechanistically straightforward S_N2 alkylation of an amine by an alkyl bromide was templated by compound **14** in Figure 2.4 using hydrogen bonding [14]. Specific hydrogen bonding between the template and the two substrates preorganizes the substrates in the correct position to react. For simplicity, the authors chose to make the two binding sites identical. Of course it is preferable two make the two sites non-identical, to avoid that two identical compounds (two amines or two alkyl bromides) bind to the template and inhibit alkylation. Kinetic studies demonstrated that the template with the identical binding sites enhances the alkylation by a factor of only six (all compounds at 4.0 mM in $CDCl_3$), far from the rate accelerations obtained by enzymes.

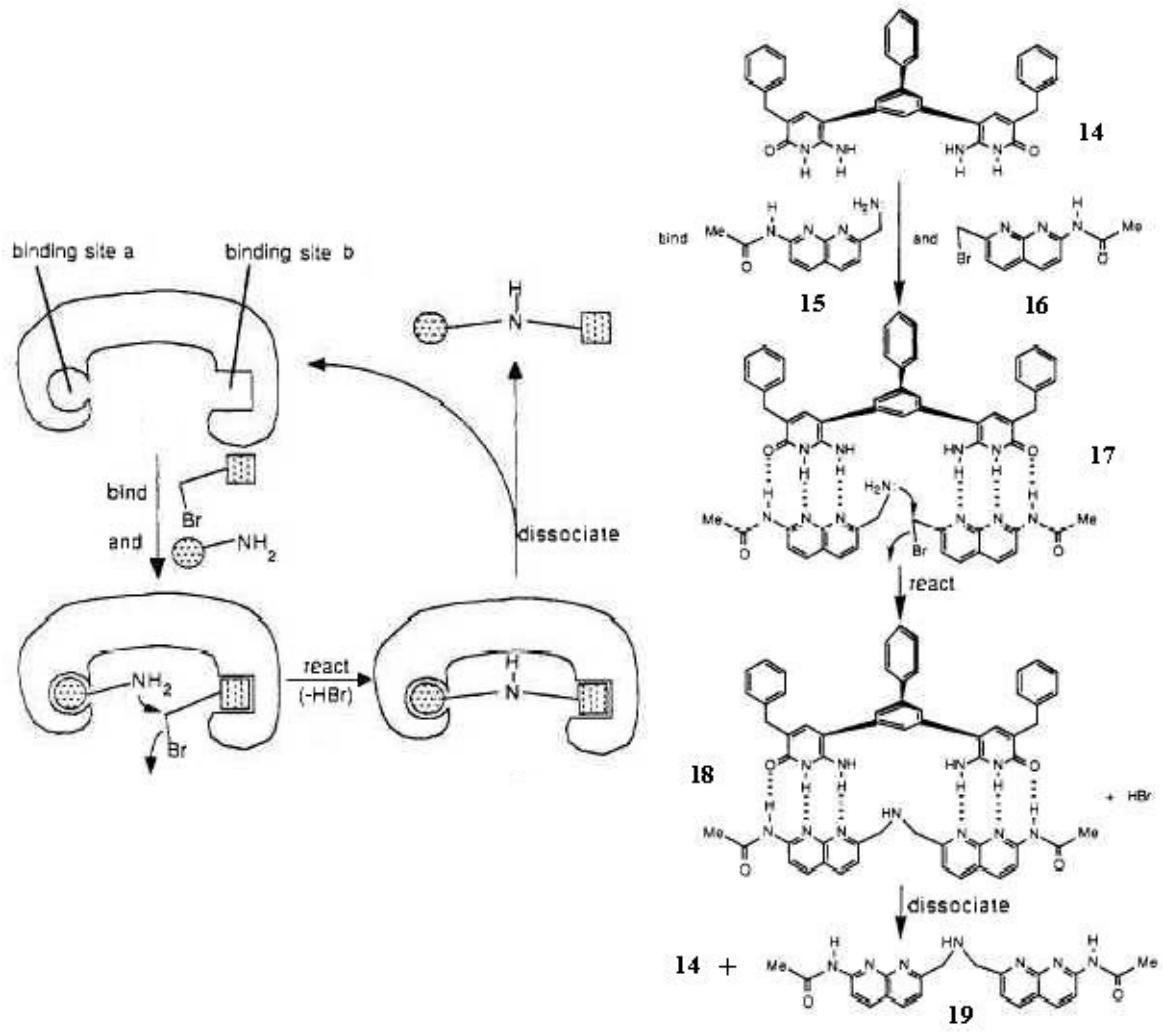


Figure 2.4. Templated S_N2 alkylation.

Another example of preorganization by hydrogen bonding was published by Huc *et al.* a few years later [15]. They used an adenine binding motif to favor aminolysis of a highly reactive *p*-nitrophenyl (PNP) ester by an aminoadenosine in chloroform (Figure 2.5).

The fact that correct alignment of the two recognition units is very important, was proven by their research on different spacers between the recognition units. Optimal preorganization of the recognition units led to rate enhancements of up to 160. If the two recognition sites are too far apart, the reaction proceeds slower, down to reaction rates equivalent to or even below that of the background reaction.

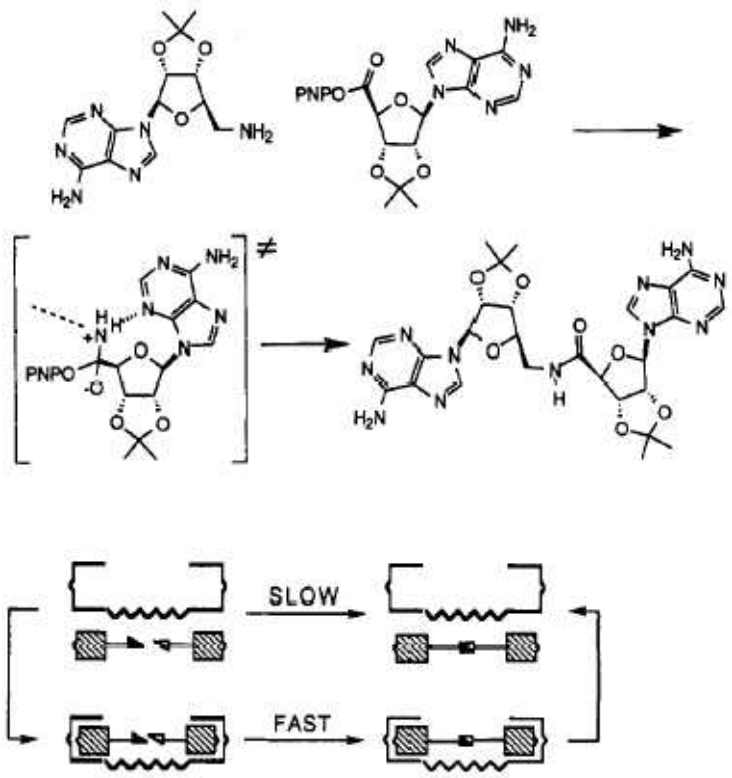


Figure 2.5. Templated aminolysis.

2.2.3 DNA based systems

Complementary DNA oligonucleotides are able to recognize each other with high fidelity. Attaching reagents to oligonucleotides at distinct places will thus enable reagents to come into close proximity of each other and react. Complementary DNA strands can thus, in principle, be excellent kinetic templates.

The most obvious example of DNA templating is the ligation of two DNA oligonucleotides. The first article briefly reporting non-enzymatic, template-directed synthesis of oligonucleotides was written by Naylor and Gilham in 1966 [16]. They mixed a decameric adenine template, and two thymine pentamers to make small amounts of decathymine strands. They activated a phosphate group of the thymine pentamer with 1-cyclohexyl-3-(3-dimethylaminopropyl) carbodiimide metho-p-toluenesulfonate (**20**), that enabled reaction with a hydroxylgroup of the other thymine pentamer (Figure 2.6).

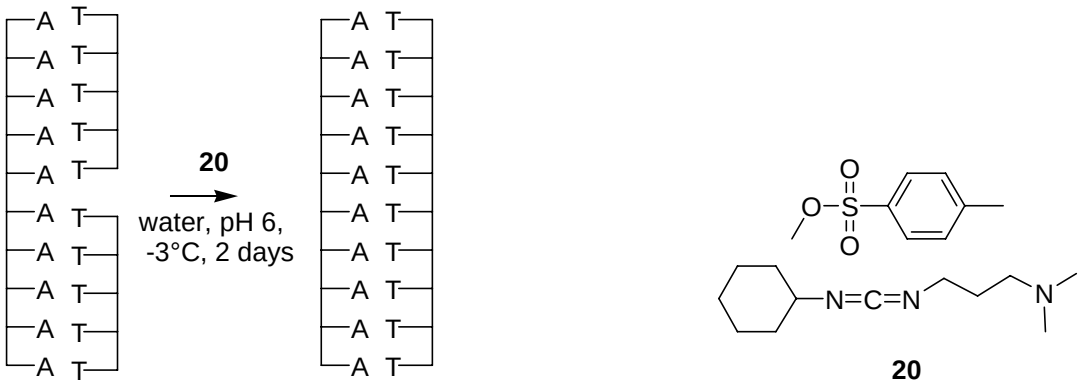


Figure 2.6. First templated DNA ligation as shown by Naylor *et al.*.

Another example of DNA ligation was published by the group of Eric Kool in 2000 [17]. They synthesized an oligonucleotide consisting of seven nucleotides bearing a 3'-phosphoroselenoate group, and another oligonucleotide consisting of 12 nucleobases modified with a 5'-iodothymidine end group (Figure 2.7). A template DNA strand was used to bring the two groups in close proximity, in order to get a nucleophilic attack of the negatively charged selenium atom on the 5'-iodothymidine.

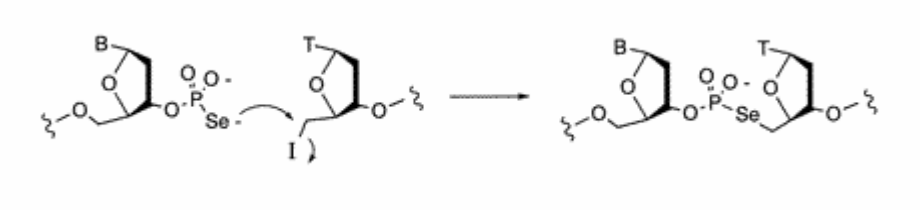


Figure 2.7. Non-enzymatic DNA coupling.

The ligation proceeded to a 27% yield in 60 minutes for the selenium case using a DNA template. If a sulfur atom was used instead of a selenium, the reaction proceeds slower. Using an RNA template instead of a DNA also has a negative effect on the reaction kinetics. To establish the template effect of the complementary DNA strand, a mismatch was incorporated in the oligonucleotides. In this case, no reaction product was observed after 1h.

Another artificial templating reaction was published by Fujimoto *et al.* in 2000 [18]. They used an oligomeric DNA strand of 12 nucleobases as a template, and two strands of 6 nucleobases as reactants. Of these last strands, one contained a cytosine moiety at the 3'-end (Figure 2.8a). An alkyl modified thymine was incorporated at the 5'-end of the other oligonucleotide.

Irradiating the mixture of oligonucleotides with light of a wavelength of 366 nm yielded the adduct depicted in Figure 2.8b, containing a butyl ring system between the cytosine and the

modified thymine. Irradiating this product with light of a lower wavelength (302 nm) showed the inverse reaction, reverting to the two initial unsaturated bonds. Continuing the cycle by irradiating this mixture with light of 366 nm a second time, yielded again the four membered ring system without significant product loss.

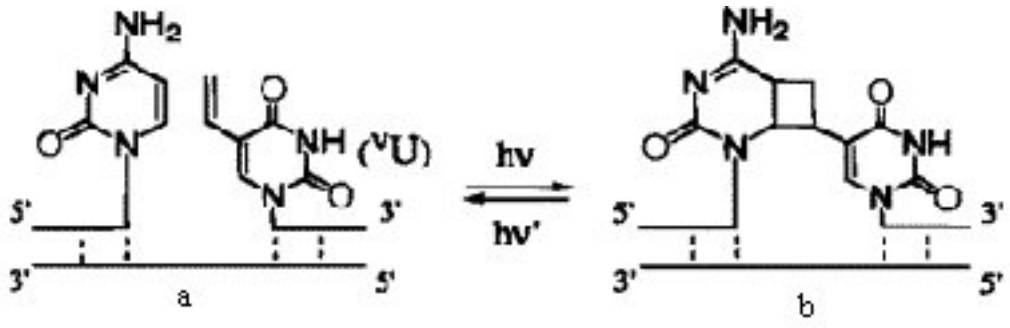


Figure 2.8. DNA templated Photoligation reaction. a) starting alkyl. b) cyclobutyl adduct.

The templated synthesis of a metallosalen compound ligated to DNA as shown by Czlapinski *et al.,* is depicted in Figure 2.9. The synthetic approach consists of modifying two DNA strands with salicylaldehydes moieties on the 3' and 5' end respectively. The two strands are aligned using a complementary DNA template to bring the two salicylaldehydes in close proximity. Addition of 1,2-diaminoethylene and an appropriate coordinating metal ion yields the metallosalen complex **21**. The reaction proceeds to a maximum yield of 65% in about 1h. This is much faster than the situation without the DNA-template, where the product is only just detectable after 8h. Metallosalen compounds serve as ligands for a broad range of metal ions. The synthesis of these metallosalen-DNA complexes may be useful since these compounds may catalyze a range of reactions in an enantioselective manner. Synthesis *via* a DNA template can be regarded as a double coordination process, with the metal ion templated in the Salens mold.

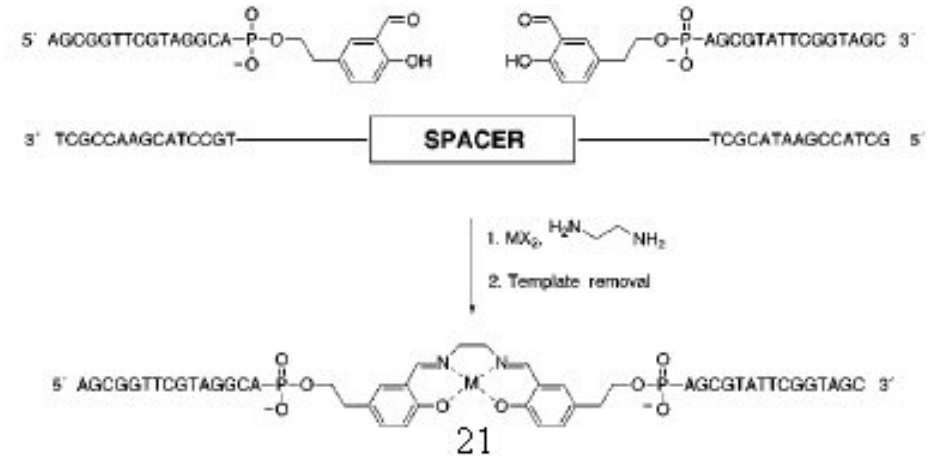


Figure 2.9. DNA templated metallosalen reaction.

Various examples of DNA catalyzed reactions were published by the group of David Liu [19-25]. The first example shown here is the formation of thioethers by nucleophilic attack of a thiol to a maleimide derivative (Figure 2.10) [20]. When the reactants are attached to one of the two template architectures shown in Figure 2.10, the reaction proceeds about 700 fold faster than in solution. The difference in reaction rates between the End-of-helix (E) and the Hairpin (H) template is minimal. Introduction of mismatches leads to a large decrease in reaction rates, in accordance with the templating reactions of the various DNA backbone analogues showed before.

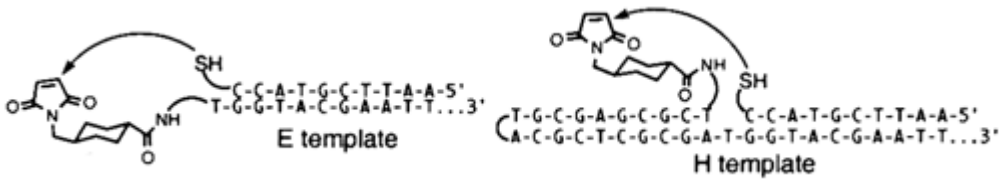


Figure 2.10. Thioether formation.

Transfer of the right-handed helicity of the DNA double helix into the stereochemistry of the products was shown by the same group a few years later [24]. Enantioselective thioether formation worked best with the End-of-helix motif. A thiol on a DNA strand was brought in proximity of a chiral bromine moiety on the complementary strand (Figure 2.11). It turned out that the bromide with S-chirality reacted four to five times faster than the R-enantiomer.

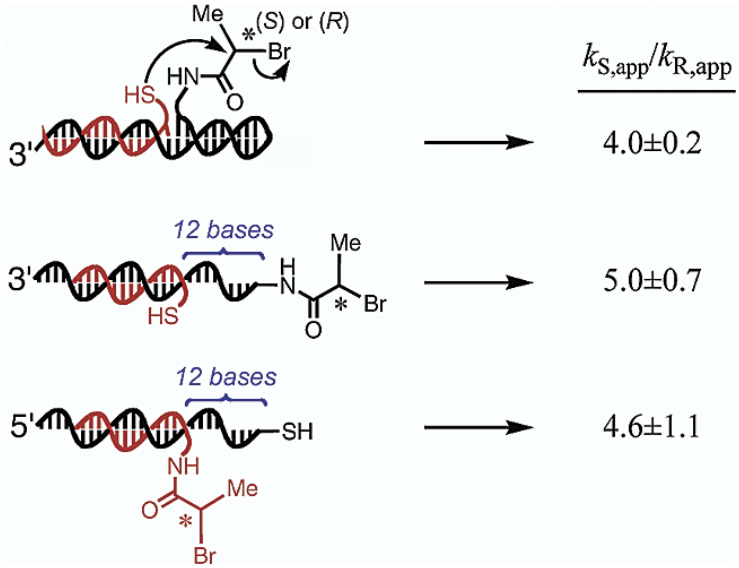


Figure 2.11. Stereoselective DNA-templated thioether formation.

To investigate the sensitivity of the system towards the helicity of the double helix, some nucleobases at the end of the strand were replaced by achiral polyethylene glycol moieties. If seven of the last twelve base pairs were replaced by polyethylene glycol units, no more stereo selectivity was observed. Replacing the polyethylene glycol units by the original nucleobases one by one, showed an almost linear increase of the stereoselectivity up to the initial value of $k_{S,app}/k_{R,app} = 5.0$, for the fully complete oligonucleotide. The chirality of the DNA strand was thus transferred to the products of the reactions.

Thioether formation is not the only reaction that can be catalyzed by this kind of templates. In 2003 the system was expanded to other reactions, like amine acylation, Wittig reactions, 1,3-dipolar cycloadditions and reductive aminations as shown in Figure 2.12 [23]. However, not all of these new reactions could be performed with the E and H architecture designed in 2001 [20]. Two new architectures were therefore designed, a so-called Ω-architecture and a T-architecture (Figure 2.13) [23].

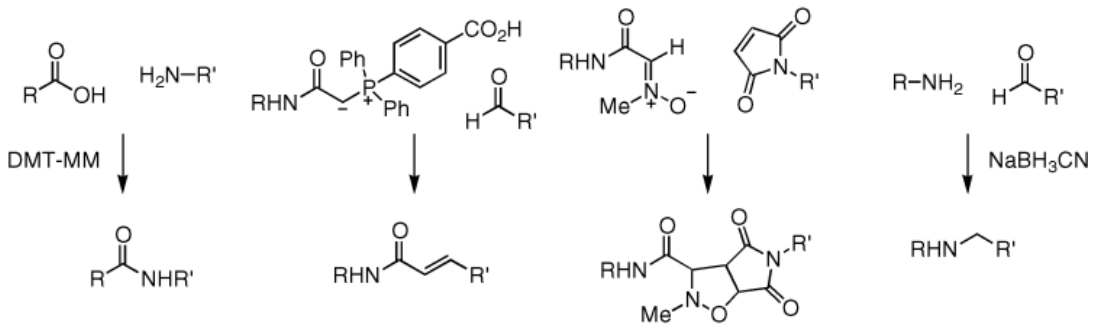


Figure 2.12. Various DNA templated reactions as shown by Liu.

The Ω-architecture resembles the End-of-helix architecture, having the reactive groups at the end side of the helix. But instead of a fully complementary duplex, the template contains a loop near the end of the double helix. This loop allows a more flexible end of the double helix, allowing reagents to come into close proximity of each other more easily.
The T-architecture has more similarities with the H-architecture, bearing the reactive moieties in the middle of the strand. But instead of a hairpin, the duplex consists of three individual strands; one template strand, and two strands bearing reactive groups. An additional reactant can be incorporated in the middle of the template strand, to have three reactive groups in close proximity to each other in the final complex.

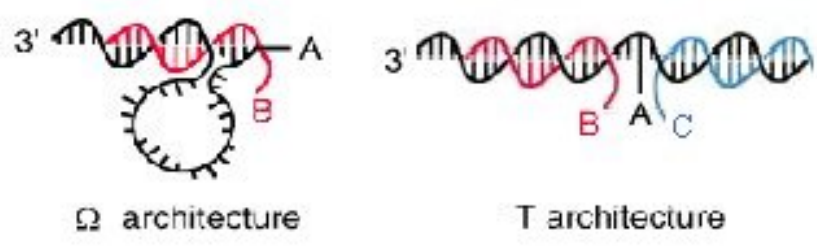


Figure 2.13. Ω and T DNA template.

The reactions shown in Figure 2.12 were tested with the E-, H-, and Ω-architecture. Comparison showed that reactions with the Ω-template had in general a higher yield than the E- and H-templates.
An interesting experiment was the one-pot mixing experiment as shown with similar compounds in 2002 by the same authors [22]. They mixed various reagents, which would normally form a multitude of products if mixed randomly. But the various DNA templates directed the reactions sequence-specifically to yield only the few desired products.

The T-architecture allowed the scope of DNA templated reactions to be extended beyond one-step reactions between two components. In 2003 the group of David Liu published two examples of two-step reactions involving three reagents overall [23]. The first case is the reaction shown in Figure 2.14. An initial S_N2 reaction between the phosphane and α-iodoamide yielded the corresponding phosphorane. This phosphorane subsequently participated in a DNA-templated Wittig reaction to generate the cinnanamide **22** depicted in Figure 2.14 in a 52% yield.

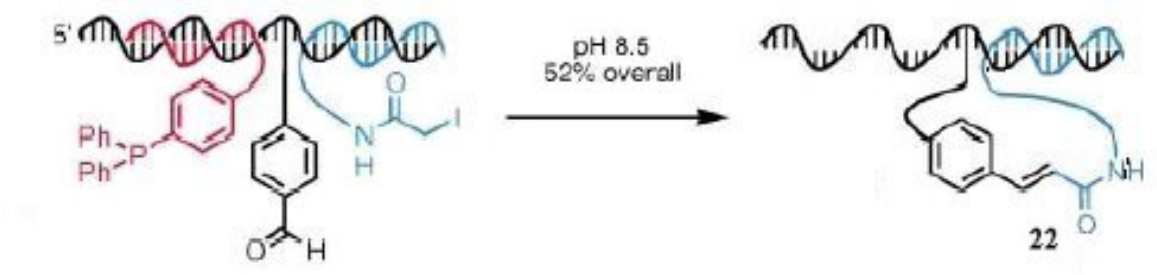


Figure 2.14. DNA templated Wittig reaction.

The second two-step reaction using the T-template was the one between an amine, a propargylglycine reagent and a phenylazide as depicted in Figure 2.15 [23]. Addition of the condensing agent DMT-MM (4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride) [26] induced amide formation. In the same pot this reaction was followed by the addition of copper(II) sulfate and sodium ascorbate to induce a Sharpless-modified Huisgen 1,3-dipolar cycloaddition [27]. 32% of the disubstituted triazoylalanine adduct **23** shown in Figure 2.15 was obtained.

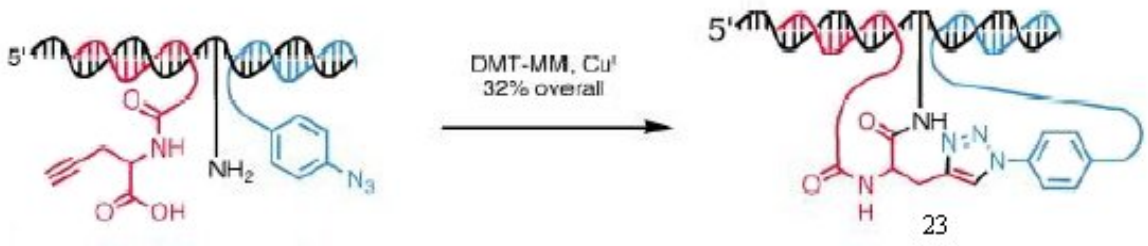


Figure 2.15.DNA templated three component reaction.

2.2.4 Ribozymes

All reactions shown in the previous paragraph take place on DNA strands binding to the template strand. Separation of template and products does not take place. This limits the yield of the reaction to one molecule per template. Recovery of the template may be possible, but the DNA strands bound to the products are most certainly lost. It is therefore much more attractive to synthesize a (DNA) oligonucleotide capable of multiple turnover.

Since this subject is beyond the scope of this thesis, I want to highlight only one example of this catalysis, a ribozyme catalyzed Diels Alder reaction as proposed by Jäschke *et al.* [28-32]. For more information about this subject, I refer to the excellent reviews of Jäschke *et al.* [30], Bartel and Unrau, [33] and Famulok and Jenne [34].

To perform DNA templating, Jäschke *et al.* optimized a ribozyme consisting of 49 nucleobases *via* directed evolution to perform Diels-Alder reactions. They started with a

sequence of 20 given nucleotides, which was able to bind anthracene derivatives, and 29 random nucleotides. They added maleimide reagents modified with a biotine unit and proceeded to the Diels alder reaction. After the reaction the ribozyme-anthracene compounds that had reacted with the biotin modified maleimide reagent were selected by fixing them to a streptavidine layer. These compounds were then multiplied by PCR and another cycle of reaction and selection with this enriched pool was started.

This directed evolution led to a ribozyme able to accelerate the Diels-Alder reaction between anthracene derivatives and the biotine modified maleimide derivative up to 20,000 fold [31]. The optimized ribozyme was then tested to various anthracene and maleimide reagents and reaction rate enhancements of up to 32,000 were obtained.

Since the ribozyme is a chiral moiety, it can be expected that the Diels Alder reaction takes place enantioselectively. Investigation whether this hypothesis was correct led to the conclusion that with $R_1 = (C_2H_4O)_6$-H and $R_2 = (CH_2)_5COOCH_3$ a selectivity of >95% *ee* was obtained, as shown by HPLC (Figure 2.16). Both enantiomers could be synthesized by changing the chirality of the ribozyme. The substrates are symmetrical, without the template the reaction therefore proceeds to a racemic mixture.

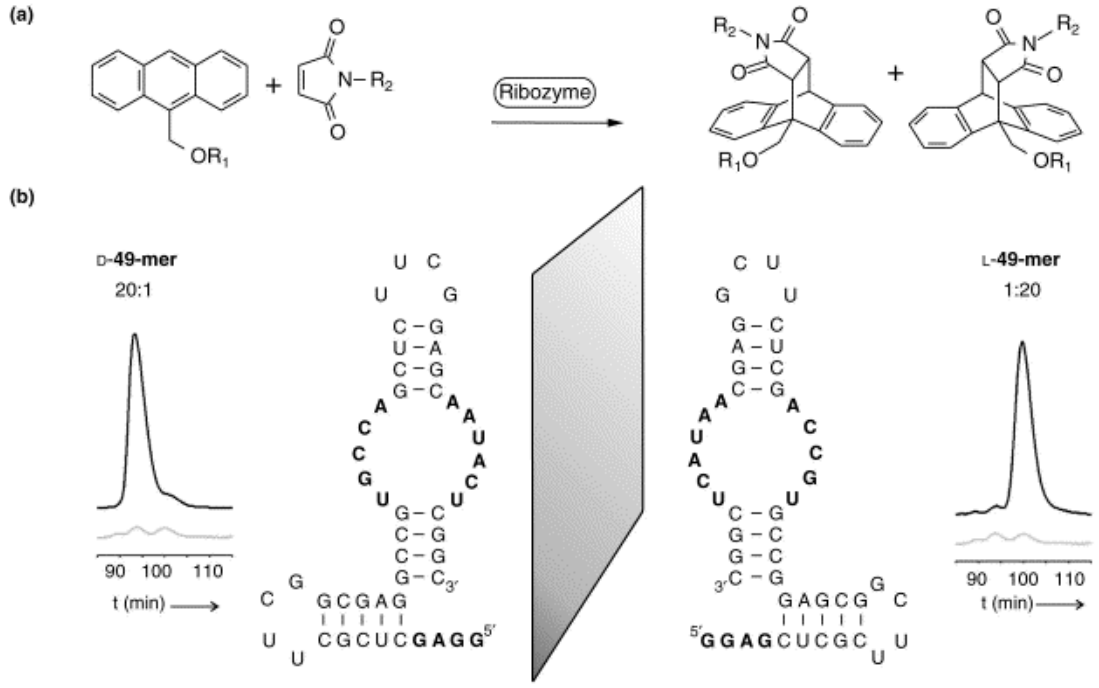


Figure 2.16. Ribozyme catalyzed enantioselective Diels-Alder reaction.

Other examples of this kind of reactions involve glycosidic bond formations, alkylations, acylations and amide bond formations [30, 33, 34].

2.3 Dynamic combinatorial chemistry and molecular imprinting

2.3.1 Dynamic combinatorial chemistry

Template directed synthesis has allowed the development of dynamic combinatorial chemistry (DCC). This principle is based on combinatorial chemistry, based on *reversible* linkages between the initial building blocks. These building blocks can, in principle, react reversibly in many ways to form multiple products. The population of products is not static as in traditional combinatorial chemistry. The proportions of the various compounds can vary in response to a change of the conditions, for example the addition of a template, which is also called target in this case.

The principle of dynamic combinatorial libraries can be made clear with the help of Figure 2.17 [35]. An imaginary reaction mixture containing compounds M_2-M_6 is shown. Under normal, equilibrium conditions, M_4 is the predominant compound. Upon addition of a template that binds product M_3 selectively, the mixture re-equilibrates in favor of compound M_3. The reactions may take place in solution as shown in Figure 2.17, but they also may or may not take place when the reagents are bound to the target. The final equilibrium is the same.

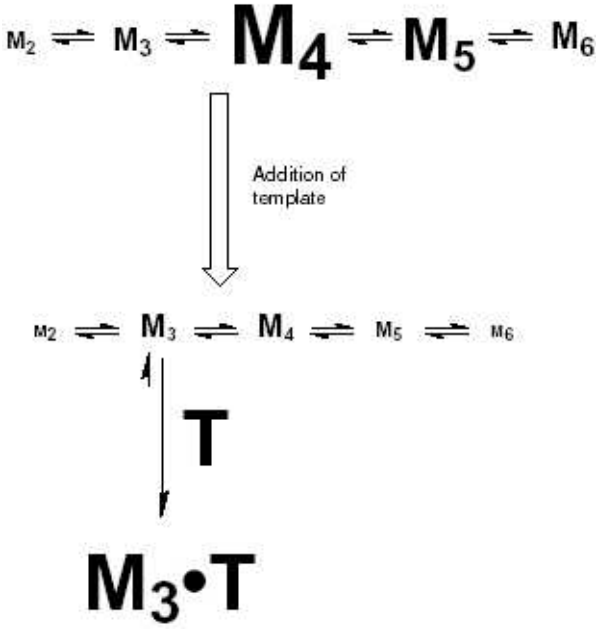


Figure 2.17. Principle of dynamic combinatorial chemistry.

Reversibility of the reaction is an essential feature of dynamic combinatorial libraries. The various products of the reaction have to be in dynamic equilibrium, but also interaction with the template has to be reversible to ensure that the formation of the best-binding compound is

enhanced. If reversibility is assured, DCC is a powerful methodology to rapidly screen many compounds for their (biological) activity.
It is desirable that the equilibrating constituents of the library are of similar free energy. If one of the constituents is highly favored, a biased library is created in which the interaction of a minor constituent with the template may not be strong enough to overturn the equilibrium situation. The template-product complex has to be very tight, especially with large libraries [36]. Mathematical models show that to overturn large libraries and form a sizable portion (2-3%) of the best binding compound, this compound must bind a factor of 1000 or more better than the other possible products [37].
In the ideal situation, the reactants form only one compound that binds best to the given template. The quantity of side products should diminish dramatically. The target serves as a 'thermodynamic' template, stabilizing one possible product and enhancing its formation.

DCC presents a huge advantage over traditional combinatorial chemistry. Traditional combinatorial chemistry initially involved the synthesis of chemical compounds as ensembles (libraries) and the screening of those libraries for compounds with desirable properties, such as binding to a given target [38]. After formation of the library, a target compound is added to screen the library for compounds that bind to this target. Addition of a target does not enhance the formation of the product that binds best to the target as in DCC, which makes screening very difficult. Identifying one compound out of a complex mixture is usually too complicated to be of practical use. It has been replaced by high throughput screening of all compounds individually. Screening of a dynamic combinatorial library (DCL) is *a priori* much easier, since the desired compound is also the most abundant. A DCL can shift towards the compound binding strongest to a given target, by rearrangement of its components. Whereas a traditional combinatorial library is a set of products, a DCL can be regarded as a set of basic components [39].

DCC requires in fact exactly the opposite way of thinking compared to the use of templates in traditional organic synthesis (Figure 2.18). For synthetic purposes one wants to enhance a specific reaction. For that, one can design a template, which is able to recognize reagents and/or products.
In DCC one is looking for strong ligands of a given target molecule. This molecule will be used as a template to enhance a specific reaction out of many possibilities and a ligand is

identified as the product of this reaction. Beforehand it is not known which product binds best and which reaction is enhanced. The system will select the best choice itself.

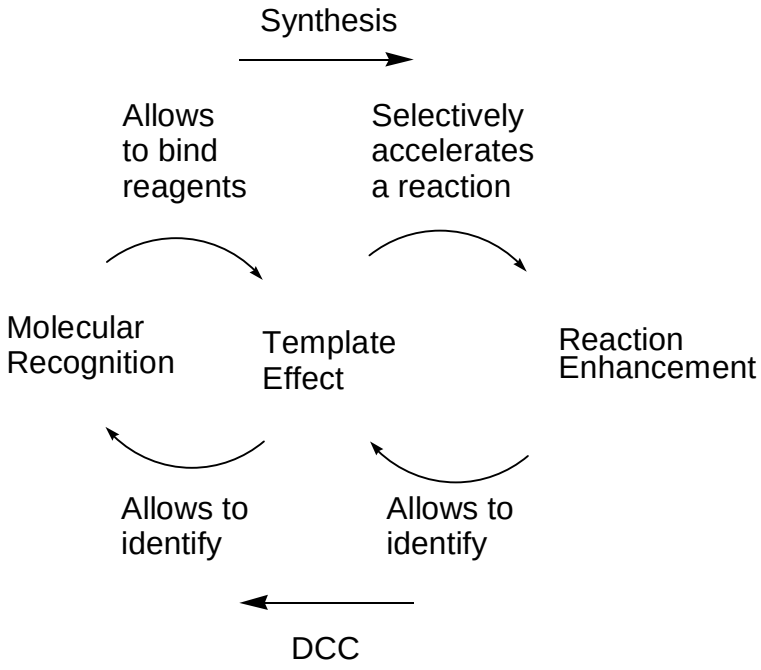


Figure 2.18. DCC versus templates for synthetic purposes.

Dynamic combinatorial systems can be divided into two categories, casting or molding, depending on whether a receptor or a substrate acts as a template [35, 40] (Figure 2.19). Casting consists of the assembly of a substrate, induced by a receptor. Molding is the opposite, a receptor is created that binds a given substrate.

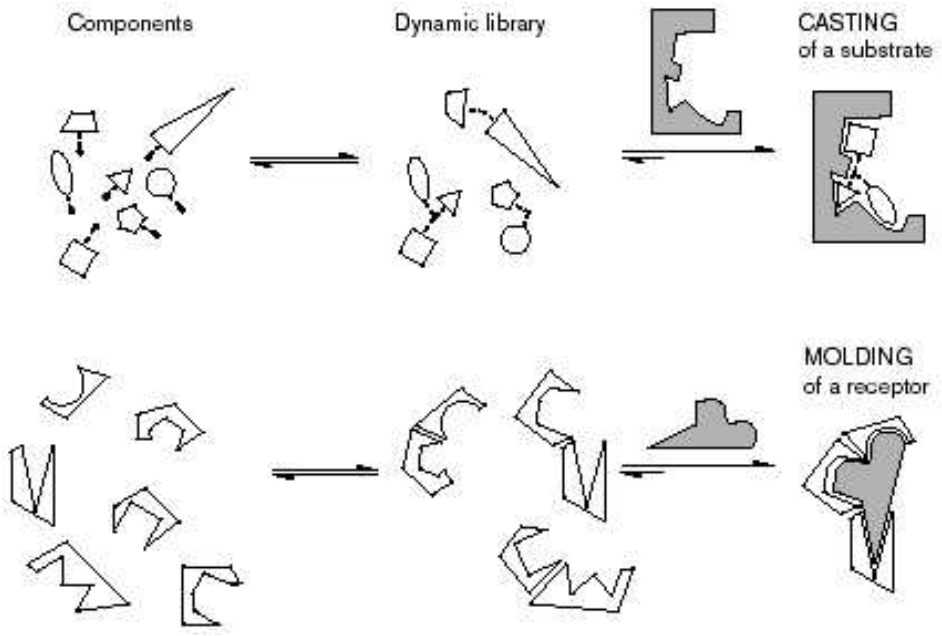


Figure 2.19. Casting and molding.

After casting or molding of the desired receptor or substrate, one may wish to lock the self-assembled structures, by performing a chemical reaction that irreversibly links together the components. The formed structure can then be analyzed and conserved, and used for further recognition events if desired.

The system by Hasenknopf *et al.* [9] as described in paragraph 2.2.1, can be regarded as an example of dynamic combinatorial chemistry. In principle all circular complexes can be formed in the reaction mixture, if no anionic template is present. Addition of one species of anion may lead to the conversion of the whole library to one species that is the best host for the anion. This was tested by the authors by synthesizing a hexanuclear helicate using $FeSO_4$ and subsequently exchanging the sulfate ions for chloride ions. After equilibrating under the new conditions, this resulted in the complete conversion of the initial hexanuclear helicates into pentanuclear helicates.

Another example of dynamic combinatorial chemistry was shown by Huc *et al.* [40]. It involves casting of an inhibitor for carbonic anhydrase (CA). Four amines and three aldehydes were mixed (Figure 2.20) to yield a library of twelve possible imines. Addition of carbonic anhydrase favored the formation of the imine that fitted best in the active site of the enzyme. Reducing the equilibrated mixture of imines with $NaBH_3CN$ locked its composition to facilitate isolation and analysis. HPLC analysis shows that the presence of one compound (**24**) is enhanced by a factor of 20 relative to competing products.
This is an example of casting of a substrate. In drug design the same principle may be used. A protein is targeted using a dynamic combinatorial library and the product that binds best to the protein is the one that should be further considered.

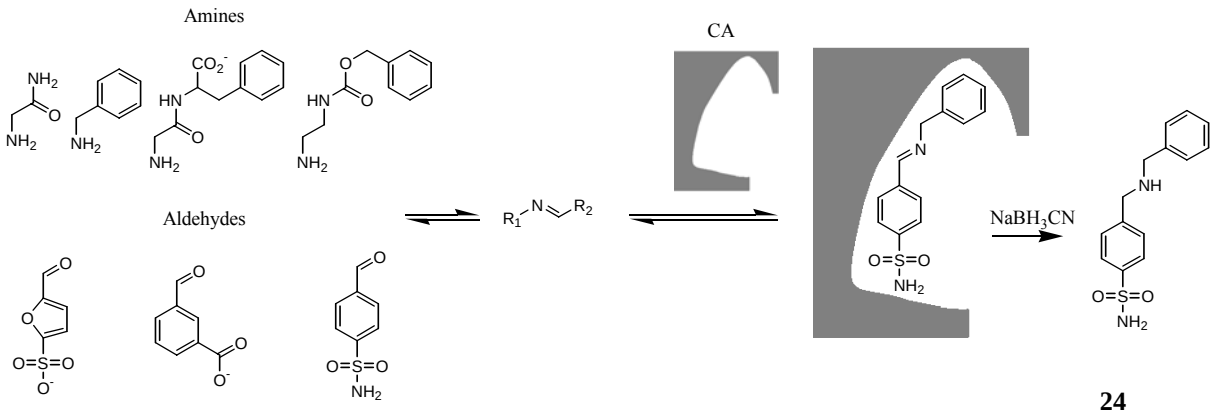


Figure 2.20. Selection of an inhibitor for carbonic anhydrase by DCC.

Examples of receptor molding are also common. One example was given by Eliseev *et al.* and involves isomerization of unsaturated bonds [41, 42]. The compounds shown in Figure 2.21, are present as their *cis-cis, cis-trans* and *trans-trans* isomers and can equilibrate under the influence of light. Arginine has a strong affinity for the *cis-cis*-isomer and arginine was loaded on a column to selectively retain the *cis-cis* isomer.

A starting mixture with the composition *cis-cis / cis-trans / trans-trans* of 3/28/69 was submitted to 30 cycles of irradiation and selection. After this treatment, 54% of the starting material was found on the column, with a product distribution *cis-cis / cis-trans / trans-trans* of 85/13/2. 11% of the starting material was found in solution, with a product distribution *cis-cis / cis-trans / trans-trans* of 49/29/23.

This last product distribution is very close to the product distribution in solution obtained by irradiating the mixture for 8 hours (*cis-cis / cis-trans / trans-trans* 52/31/17) or by performing 30 irradiation-selection cycles without arginine on the selection column (*cis-cis / cis-trans / trans-trans* 55/31/14). But in these cases there is no or only little material on the column, and no enhancement of the *cis-cis* isomer in that fraction is observed.

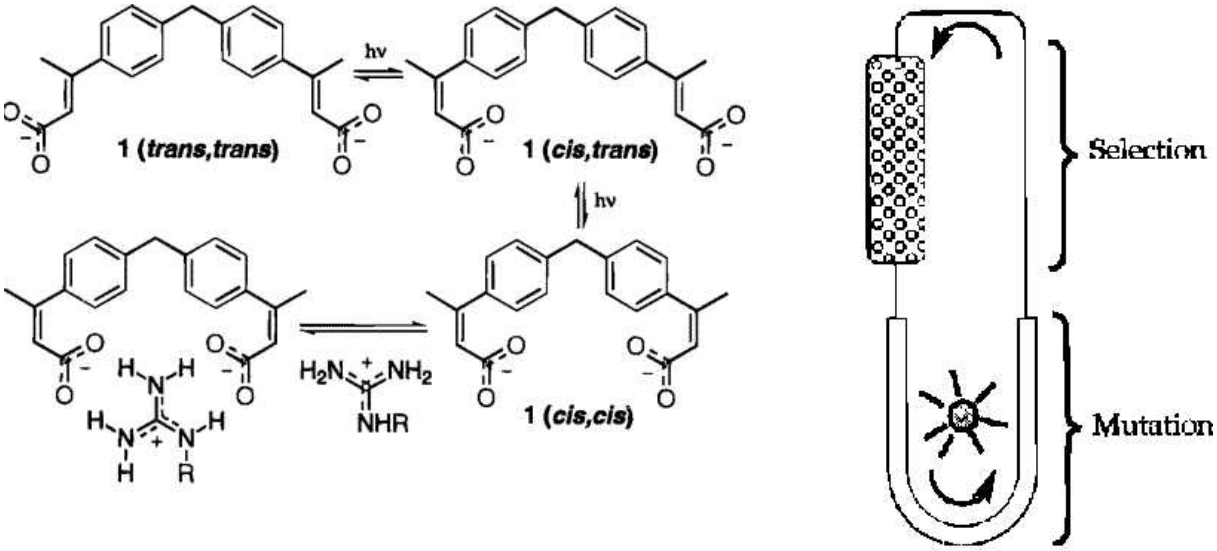


Figure 2.21. Cis-trans equilibrium of dicarboxylates.

Another example of molding of a receptor was published by Hioki *et al.* [43]. They noticed that compounds as shown in Figure 2.22 are good receptors for various tripeptide sequences bound to polystyrene (PS), like Ac-(D)Pro-(L)Val-(D)Val-PS. A disulfide linker was used to connect the two receptor units in a reversible way.

Addition of excess phenylsulfide yielded an initial reference mixture of 35% dissymmetric disulfide and 65% of the symmetrical receptor. Addition of a tripeptide guest on a resin

changed the composition of the mixture to 5% dissymmetric product and 95% of the two symmetrical ones. Of the formed receptor, 90% could be found on the resin, only 5% was present in solution. The diphenylsulfide could only be found in solution.

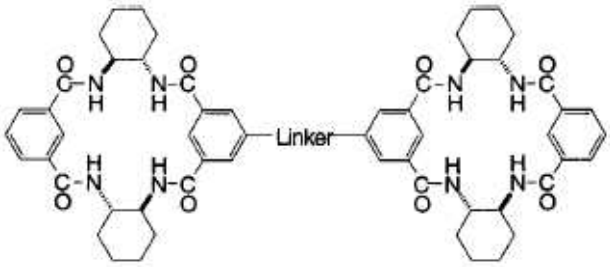


Figure 2.22. Receptor molding

2.3.2 Irreversible reactions

In dynamic combinatorial chemistry a reaction is favored because of the stabilization of one product. The target serves as a 'thermodynamic' template. This principle can be expanded to irreversible reactions. In that case, if a given reaction pathway is enhanced by the target, it means that the reaction transition states are bound. In practice, it has been shown that the product of this reaction is a good binder as well (see section 2.2). The presence of a product then suggests that that product is a good binder for the given template. The template serves here as a 'kinetic' template.

An example of this principle was published by the group of Sharpless [44]. They used acetylcholinesterase as a template for the Huisgen 1,3-dipolar cycloaddition of azides and acetylenes to give 1,2,3-triazoles (Figure 2.23).

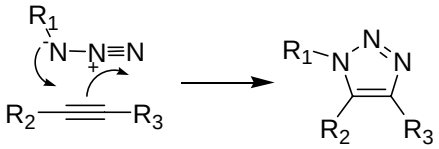


Figure 2.23. 1,2,3-triazole formation by Huisgen 1,3-dipolar cycloaddition.

They designed 8 azide and 8 acetylene compounds that were bearing tacrine or phenanthridinium groups, which are known to be site-specific inhibiting groups for acetylcholinesterase (Figure 2.24a). Each of the possible binary mixtures was incubated in the presence of the enzyme at room temperature for 90 minutes. The rate of the non-templated reaction is negligible under these conditions. Examination of all the reaction mixtures showed that in only one case a 1,2,3-triazole was produced. Since the reaction can take place both *syn* and *anti*, they found two lead compounds, ***syn*-25** and ***anti*-25** (Figure 2.24b). These had to be tested for their inhibition properties on acetylcholinesterase. It turned out that ***syn*-25** was the

best binder for acetylcholinesterase, having dissociation constants in the order of 10^{-13}M, two orders of magnitude less than its *anti*-isomer. The *syn*-isomer was also identified as the one that was preferentially synthesized by the enzyme.

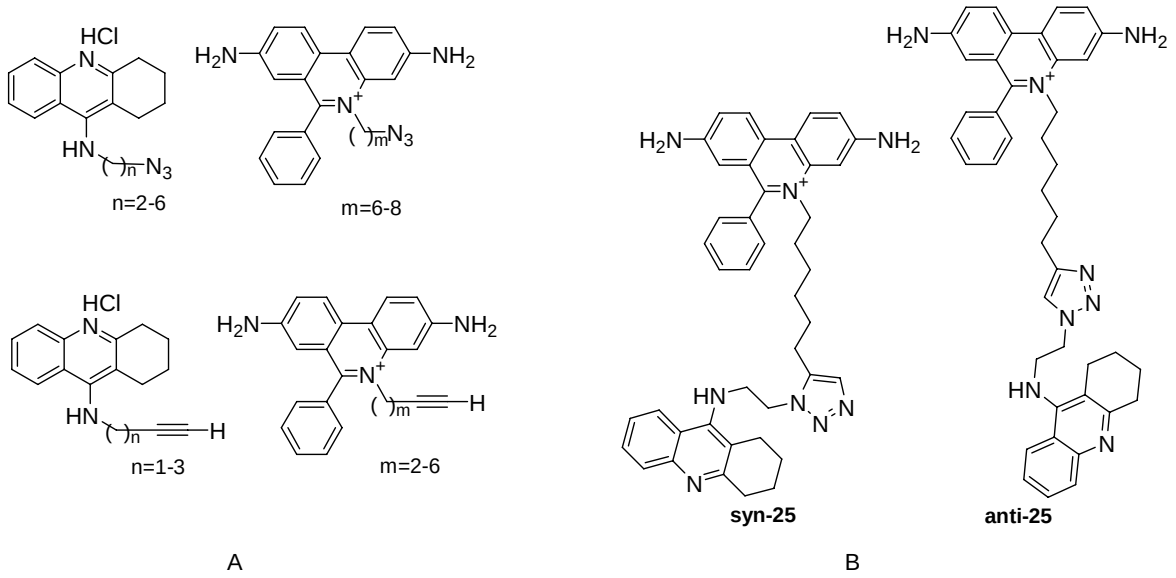


Figure 2.24. a: 16 building blocks for triazole formation. b: two best inhibitors for acetylcholinesterase.

2.3.3 Molecular imprinting

Molding of a receptor is nowadays often done using building blocks that are able to polymerize. This is called molecular imprinting (Figure 2.25) [45, 46]. The technique consists of dissolving a target molecule along with functional, cross-linking monomers in a solvent. Polymerization and cross-linking of the monomers create cavities with exactly the size of the target molecules that are trapped. Washing out the target molecules then gives a molecularly imprinted polymer, which is able to selectively recognize the desired target molecules.

To give just one example, Parmpi and coworkers used this technique to selectively recognize sugars [47]. Poly (allylamine hydrochloride) was mixed with D-glucose 6-phosphate monobarium salt, which serves as a template. The phosphate group can bind non-covalently to an amine group of the polymer. This mixture was then crosslinked, using epichlorohydrin and the glucose phosphate template was washed away with an aqueous, alkaline solution to give the imprinted polymer. When the imprinted polymer is now brought in a solution containing both glucose and fructose, it is capable to selectively bind glucose over fructose with a factor of three to four. This is not a very large factor, but it is more than necessary to achieve a chromatographic separation of glucose and fructose using the polymer as a stationary phase.

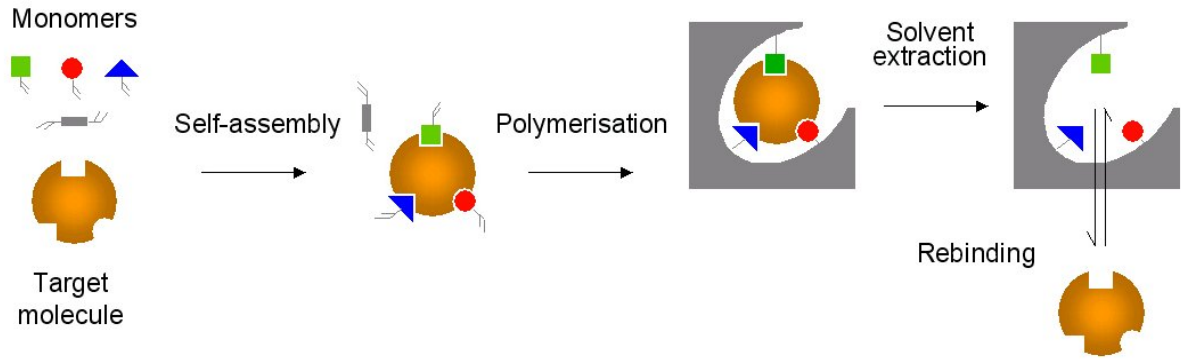


Figure 2.25. Principle of molecular imprinting.

Molecular imprinting is not limited to recognizing the molecule it was imprinted with. An imprinted polymer can also enhance the formation of that molecule from its precursors. The imprinted polymer then acts as a template. When various reagents are added to the imprinted polymer, the reagents that can form the product will preferentially be selected. Their reaction is enhanced to give the desired product that was previously imprinted in the polymer.

An example of a reaction that was enhanced by molecular imprinting was published by Visnjevski *et al.* [48]. They imprinted a polymer with the adduct of a Diels-Alder reaction (Figure 2.26b) Then, they reacted hexachloropentadiene with maleic acid in the presence of the polymer (Figure 2.26a). They found a decrease of the activation energy of the Diels-Alder reaction of 60 kJ/mol in the presence of the imprinted polymer, compared to a non-imprinted control polymer.

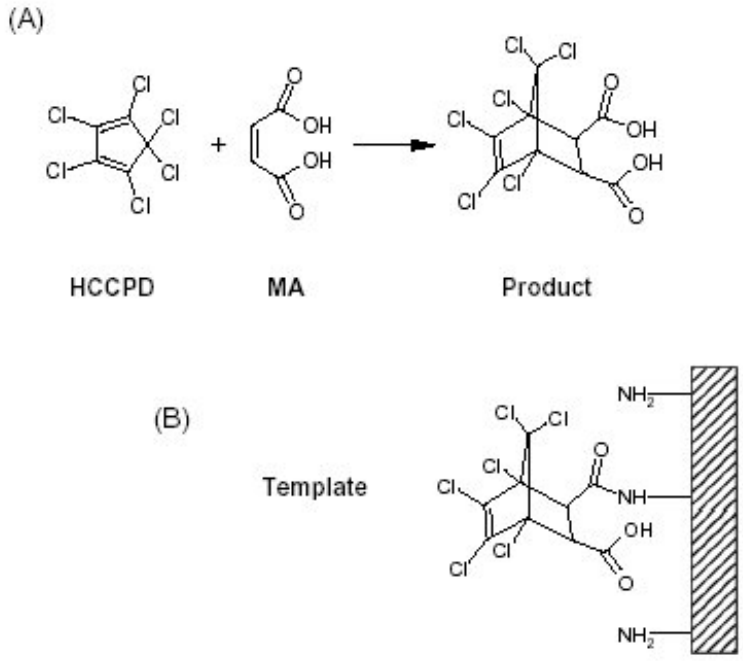


Figure 2.26. Templated Diels-Alder reaction with a molecularly imprinted polymer. a. Diels-Alder cycloaddition; b Template used for molecular imprinting.

2.4 Replication model systems

A special kind of templated reactions are replicating systems. These are systems that can transfer the information they contain to the molecules they template for. This phenomenon is of course associated with DNA. DNA holds our genetic information and transcribing DNA to form RNA molecules allows proteins to be synthesized.

Genetic copying of DNA is a special kind of information transfer. Copying of a DNA strand gives a complementary strand that, at its turn, can be replicated to give a copy of the original strand. This is called a cross-catalytic system since molecule A favors the formation of A', and A' favors the formation of A. Examples of this will be discussed in paragraph 2.4.3.

An easier system can be imagined, in which a molecule favors the formation of itself. This is called self-replicating. Auto-complementary DNA sequences are an example of this. Some examples of artificial self-replicating systems that have previously been developed will be discussed in paragraph 2.4.2.

2.4.1 Replicating systems

Catalysis of DNA polymerization is performed by DNA polymerases. The role of the copied DNA strand during its own duplication is to discriminate the incorporated base at each step, according to the corresponding nucleobase of the template. In the absence of the enzyme, no

catalysis takes place, but discrimination may be possible. An example of an experiment trying to replicate a DNA strand without polymerases was published in 1995 by the group of Leslie Orgel [49]. They showed the formation of amide bonds in PNA strands by DNA templating. PNA oligomers were synthesized using decameric DNA templates (Figure 2.27). They obtained mostly oligonucleotides with three, four or five nucleobases. Little product with the same length as the template was observed, but no products with a length exceeding the length of the template were formed at all. Relative to the non-templated case the reaction was significantly accelerated, since without the template only some dimer was observed.

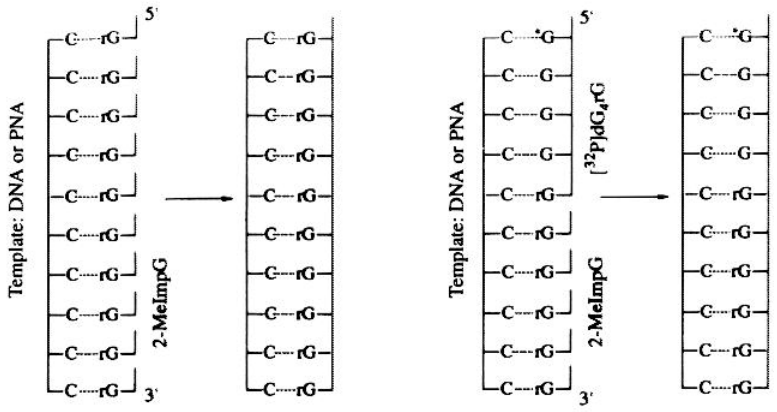


Figure 2.27. Peptide coupling using DNA and PNA templates.

Böhler *et al.* [49] found that the synthesis took place in the *3'-5' direction*, *parallel* to a DNA oligonucleotide. This is the inverse situation compared to the natural one. In nature DNA strands are synthesized in the *anti-parallel* fashion, in the *5'-3' direction*.
Catalysis by PNA was more effective, but only if a small primer sequence of three PNA nucleobases was already present. Synthesis of new PNA chains does not take place in this case. This behavior was explained by assuming that nucleobases should be pre-organized into an initial double helix segment, before templating becomes effective.

A different, but as successful method to polymerize nucleobase analogues from monomers was designed by the group of David Lynn in 2002 [50]. They used an all adenine DNA template, and added a thymine deoxyribose analogue with a 5'-amine group and a 3'-aldehyde (Figure 2.28) under reductive conditions ($NaBH_3CN$). The HPLC traces of the mixture (shown on the right of Figure 2.28), taken after various reaction times, show that polymerization yielded first the imine dimer, then tetramer and eventually the octamer, which has the same length as the template. This is in accordance with the classical step-growth polymerization claimed by the authors. No higher oligomers were detected.

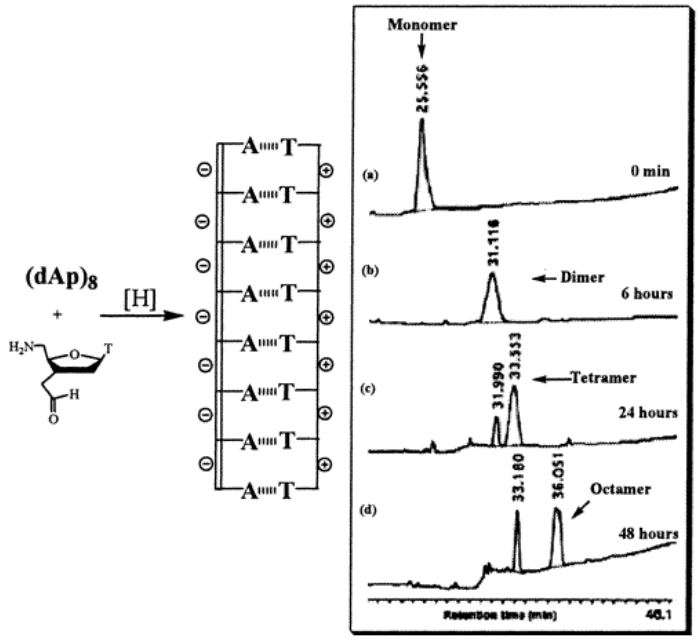


Figure 2.28. Peptide coupling using DNA and PNA templates.

Apart from the obvious capacity of DNA to transfer information linking DNA or DNA analogues, DNA is also capable of transferring information using other reactions.

Synthesis of covalent bonds between natural peptides (glycine-glycine / glycine-cystein) was performed in the group of Seitz [51, 52]. They synthesized short PNA strands (3 nucleobases) functionalized with glycine or cysteine units. Adding a template DNA strand with nucleobases complementary to the two oligonucleotides and a weak base as a catalyst, yielded a hexameric PNA strand with two glycine units in the middle. Yields were rather low (20%) but could be improved considerably (up to 80%) by using PNA as a template instead of DNA.

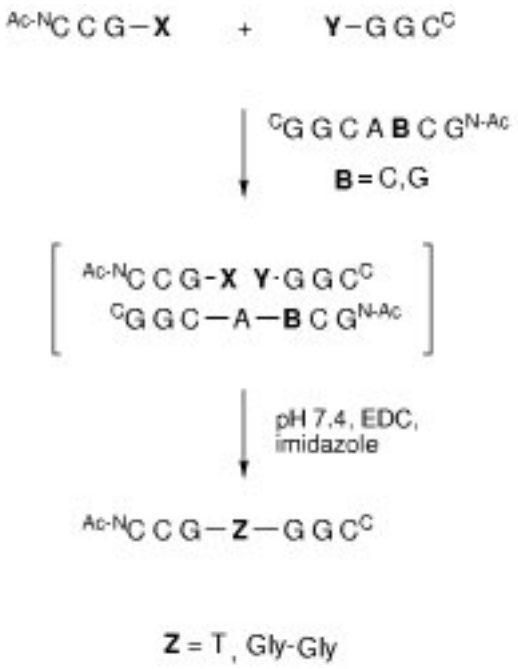


Figure 2.29. Glycine coupling.

Information about the length of an oligonucleotide is normally well transferred. However, transfer of the information about the nucleobase sequence is more difficult. I could only find a couple of examples of sequence specific information transfer of nucleobases, but only one showed information transfer with good fidelity. The group of David Lynn published copying of DNA dinucleotides to form a tetramer in 2002 [50].

A tetrameric DNA template coding $^{5'}$AAAT$^{3'}$ was mixed with two modified dimers, $^{3'}$TT$^{5'}$ and $^{3'}$AT$^{5'}$(Figure 2.30a). The dimers contained an amide linker and had an amine group at the 5' end and an aldehyde at the 3'end. The template and the two dimers were mixed 1/1/1. Coupling the dimers gave the imines, that were *in situ* reduced using $NaBH_3CN$ to give the tetramer $^{3'}$TTTA$^{5'}$ quantitatively.

No reaction took place when the $^{5'}$AAAT$^{3'}$ template was mixed with two equivalents of $^{3'}$TT$^{5'}$ in the absence of $^{3'}$AT$^{5'}$(Figure 2.30b). The single mismatch that would be present in the product $^{3'}$TTTT$^{5'}$ is obviously enough to prevent any reaction to occur. Also changing the template to $^{3'}$AATA$^{5'}$, showed no reaction (Figure 2.30c). In this case a fully complementary strand could be formed by parallel alignment of the template and the reactive strands, but this does not happen.

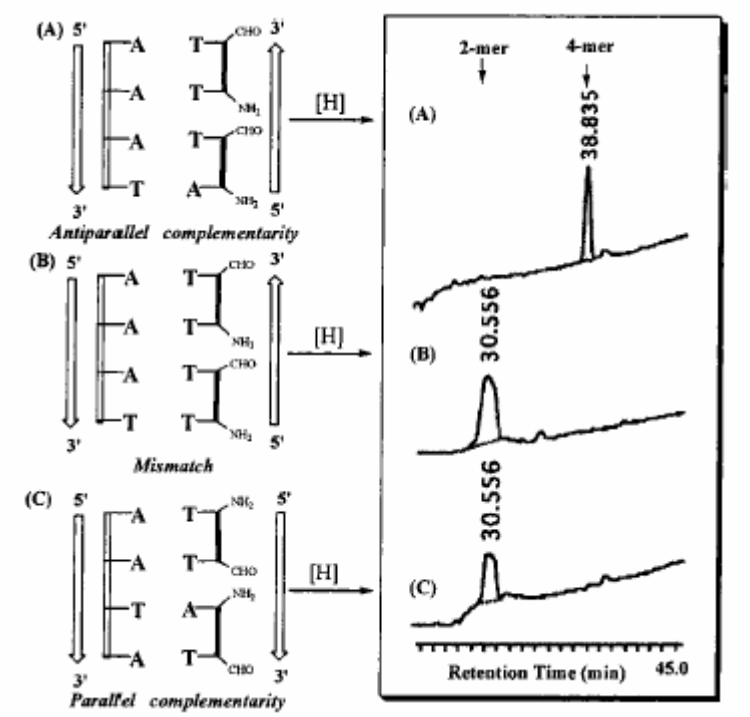


Figure 2.30. Selective information transfer in oligonucleotides. A) Successful anti-parallel copying. B) One mismatch prevents coupling. C) The strands do not align in the parallel fashion.

The previous example published by the group of Lynn dealt with the transfer of information contained in dinucleotides. The best attempt till now to copy DNA or RNA strands with mononucleotides, was published by Hey *et al.* in 2003 [5]. Starting from a primer they attempted PCR-type replications to give RNA duplexes.

They used nucleosides activated with 2-methyl-1H-imidazolides on the 5'-phosphate end. Transfer of information about the length of the oligonucleotide was successful. Copying of four cytosine units in a template strand gave the primer strand elongated with four guanines with a yield of 91% in 14 days (Figure 2.31).

In order to transfer the information contained in the sequence, they then incorporated two extra nucleotides in the template strand, a thymine and another cytosine. Addition of only the activated guanine nucleoside to a mixture containing the primer and the template strand, yielded 55% of the primer elongated with four nucleotides, 25% elongated with five nucleotides and 17% elongated with six nucleotides. Transcription was hindered by the thymine, but wobble pairing between the guanine and the thymine allowed the transcription to continue albeit the mismatch.

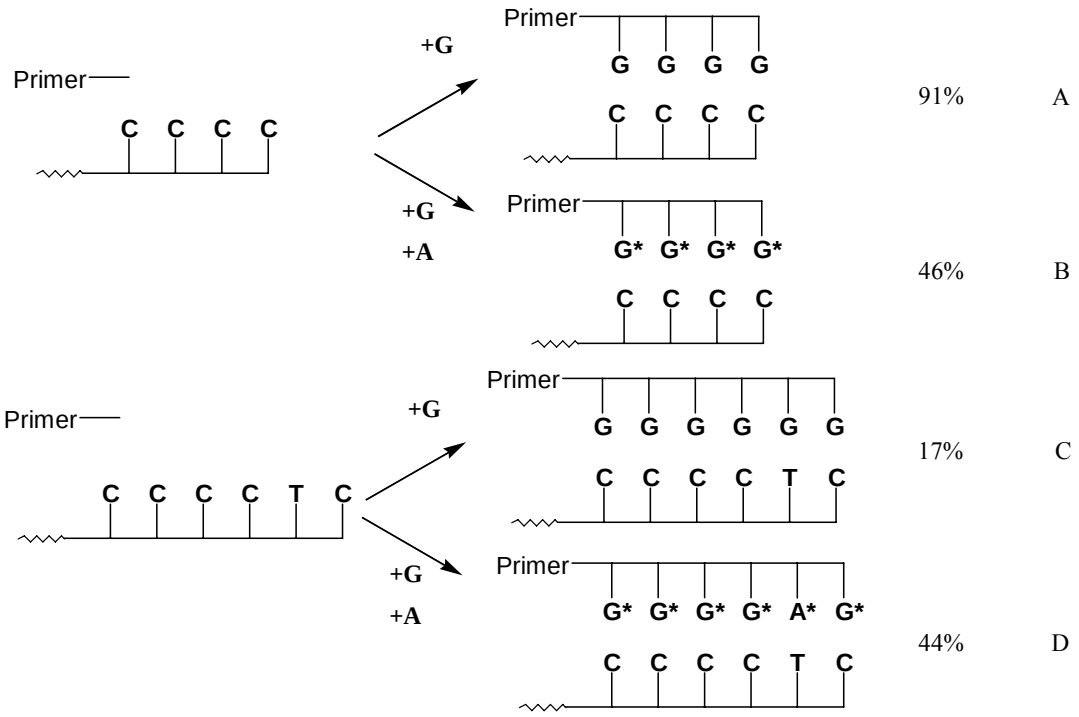


Figure 2.31. Experiments of Hey *et al.*. G* and A* represent the most likely incorporated base, but the incorporation of a mismatch is also possible. A. Information about the length of the oligonucleotide is correctly transferred. B. Transfer of information about the length of an oligonucleotide is hindered by addition of a mismatching nucleotide. C. A mismatch severely hinders information transfer. D. Information transfer with a both adenine and guanine present is better than with only guanine present.

Addition of both activated adenine and activated guanine nucleosides to the same template and primer yielded 19% of the primer elongated with four nucleotides, 23% elongated with five nucleotides and 44% elongated with six nucleotides under the same conditions. In the end much more oligonucleotides elongated with six nucleotides are thus obtained, but also much more primer is still present; 6% against 1.5% when only activated guanine is added. In the beginning of the process, when only cytosine is transcribed to guanine, the yields of the all-guanine experiments are much better. With the activated adenine also present in the reaction mixture, the reaction has difficulties starting, presumably due to product inhibition by wrongly incorporated units. When four units are incorporated the template prefers the adenine nucleoside, the reaction then proceeds faster compared to the situation where only activated guanine is present.

Inhibition equally takes place when the all-cytosine template strand that was used in the first experiment is mixed with activated guanine and activated adenine. Whereas with only activated guanine nucleosides a 91% yield was obtained after 14 days, only 46% of the primer is elongated with four nucleotides in case also adenine is present. 19% of the unmodified

primer was still present after 14 days against 2% in the experiment where only activated guanine was added.

Hey *et al.* do not identify the sequence of their extended primers in their studies, but only determine their length. Mismatches are thus not detected; the experiments only suggest that when the primer is inhibited from elongation, this is presumably due to a mismatch. Avoiding the incorporation of mismatches or *in situ* correction of them would presumably lead to a system with much higher fidelity.

2.4.2 Self replication

I mentioned before that copying of DNA strands yields it complementary strand and not the original. However, copying of auto-complementary DNA strands, does directly lead to the formation of a copy of itself and thus amounts to a duplication. This principle was used by the group of Von Kiedrowski [53-58].

They reported template-directed condensation of two trinucleotides on a hexameric template. The two starting trinucleotides have different activating and protecting groups (Figure 2.32). The first trinucleotide is protected on its 5'-end as its methyl ether (d($^{5'}$MeC-C-G$^{3'}$p)). The phosphate group at the 3'-end of the oligomer is activated *in situ* with 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide (EDC). The second oligomer is protected from activation with an *o*-chlorophenyl group on its 3'-terminus, whereas the 5'-end is not protected (d($^{5'}$C-G-G$^{3'}$p)). The template strand is protected with both a methyl on its 5'-terminus and an *o*-chlorophenyl group on its 3'-end (Figure 2.32).

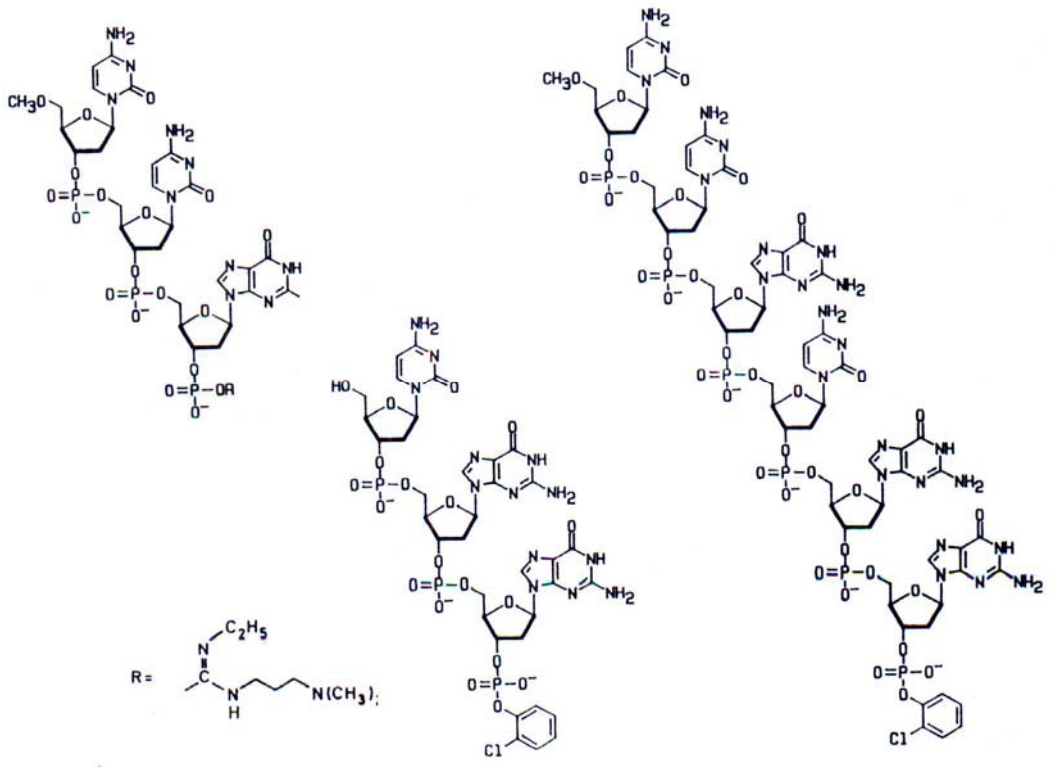


Figure 2.32. Autocatalytic template synthesis of DNA

Condensation between the two trimers in presence of the template at 0°C leads to two reaction products. Head to tail condensation yields the desired copy of the template, tail to tail condensation of the first trimer yields a 3'-3'-linked pyrophosphate (d($^{5'}$MeC-C-G$^{3'}$p-p$^{3'}$G-C-CMe$^{5'}$)) as a side product.

The group of Julius Rebek illustrated the self-complementarity of molecules with Figure 2.33 [59]. A molecule can be synthesized from two building blocks. In its turn, two molecules can self-assemble to form a complex. However, it is also possible that one molecule assembles the two building blocks to make a complex consisting of three species. The two building blocks are now perfectly aligned for reaction and the reaction kinetics will thus be enhanced and a complex of two molecules is formed. When the complex now falls apart, two template molecules are released. Each of them is able to template a new reaction to form new copies of themselves and the self-replicating cycle is complete.

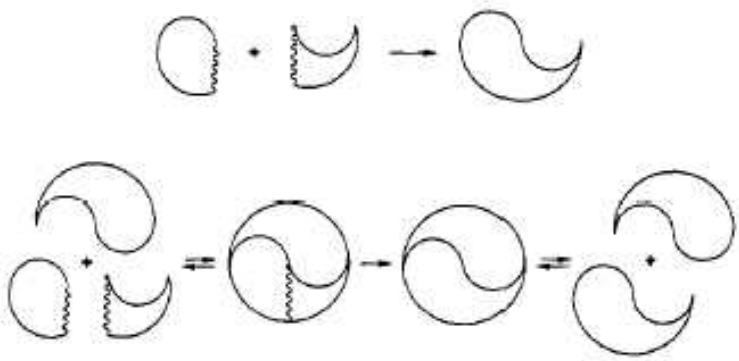


Figure 2.33. Self-complementary molecule.

The group of Rebek designed an artificial self-replicating system consisting of an amino adenosine unit and a pentafluorophenyl ester [59, 60]. The two molecules were prearranged by their own reaction product, *via* hydrogen bonding between an adenine and a thymine-like unit and π–π-stacking between the aromatic rings of the template and the reactants (Figure 2.34).

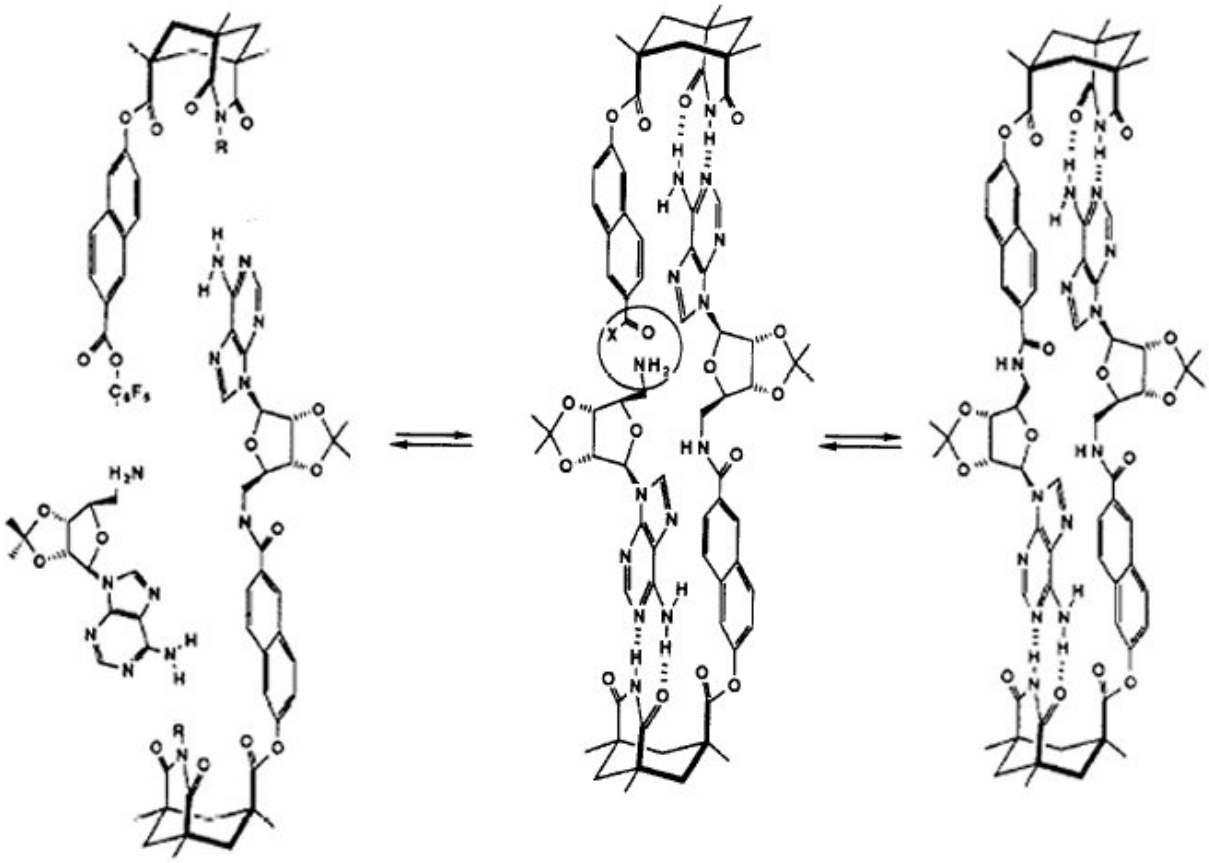


Figure 2.34. Self-replicating system.

Rate enhancements of 43% were observed by the addition of 20% of product (8mM, $CHCl_3$, 4eq. Et_3N), 50% of product increased the reaction rate up to 73%. Three reaction pathways leading to the reaction product were distinguished. The normal, uncatalyzed, bimolecular pathway, a preassociative, bimolecular pathway and the termolecular pathway as depicted in Figure 2.34.

Kinetic studies on the system revealed that by addition of 30% of catalyst, 67% of the amine and the ester are free in solution, 27% are associated with each other and only 1.5% are present as their termolecular complex. The relative reaction rates for background bimolecular pathway / preassociative bimolecular pathway / termolecular pathway are 1.0 / 4.3 / 4.6 respectively. The two latter mechanisms thus contribute much more to the reaction, albeit the fact that the reactants are predominantly present in their uncomplexed forms. It is important to note that substitution of the imide hydrogen for a methyl group at the ester side of the template led to a molecule that was not a catalyst for amide formation. Molecular recognition is required for replication.

Further work on this system by Menger *et al.* [61] and Reinhoudt *et al.*[62], revealed that the system is not as simple as presented by the group of Rebek. Menger *et al.* pointed out that the presence of amides accelerates acylations. Since the template contains an amide, they stated that this amide might catalyze the reaction and that the template effect might only play a minor role. Rebek *et al.* defended their work by proving that not all amides accelerated acylations [63].

Reinhoudt *et al.* solved the controversy by pointing out that the groups of Menger and Rebek did not work under the same reaction conditions. Most important difference in the reaction conditions was the difference in concentration, that was a factor of 10 higher in the experiments of Menger *et al.* [62]. Since hydrogen bonding is involved, the concentration of the reaction mixture has a major influence on the outcome of the reaction. Furthermore, Reinhoudt *et al.* did not identify three pathways, but five. The individual contributions of each of those five pathways to product formation are highly concentration dependent. Calculations led to their conclusion that the templated termolecular reaction is 6.8 times faster than the background reaction, proving the autocatalytic effect. To what extent this pathway contributes to the overall product formation depends on which conditions are used.

To make their results less ambiguous, the group of Rebek designed a reaction in which the preassociative bimolecular pathway was ruled out. They increased the spacer length between the two hydrogen bonding groups (Figure 2.35) [64] to keep the amine and ester groups from

reacting in a bimolecular complex. Addition of the 0.5 eq. of product to the mixture of the ester and amine, led to a relative enhancement of the reaction that was similar to the enhancement observed with the previously designed replicating system (54%).

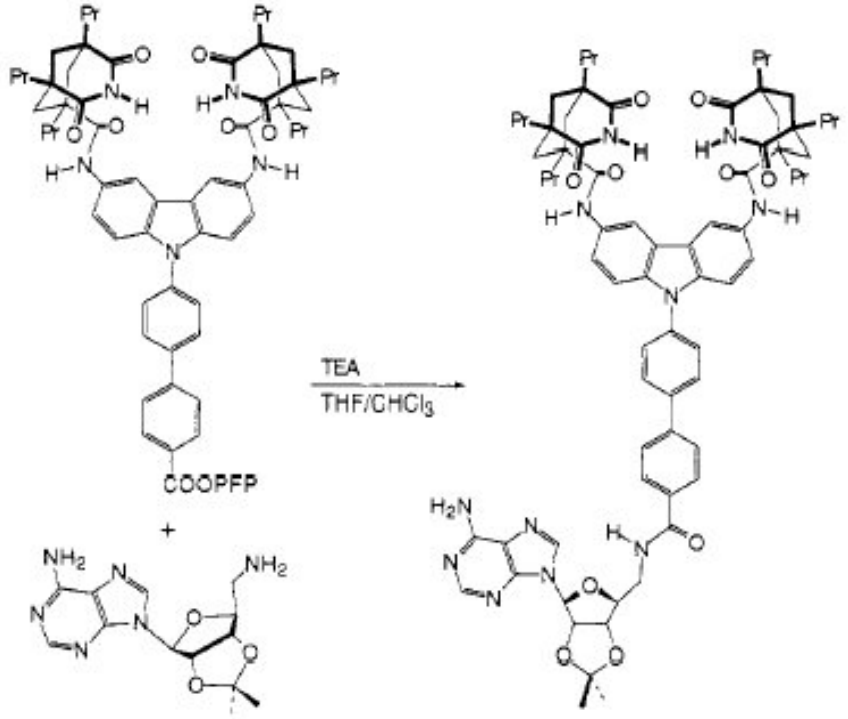


Figure 2.35. Self-replicating molecules, second generation

A smaller self-replicating system was published by Terfort and Von Kiedrowski in 1992 [58, 65]. They condensed an aldehyde (**27**) with an amine (**26**) to form an imine (**28**) in DMSO-*d6* at 50°C. Recognition of the units by the template was obtained by an amidinium-carboxylate salt bridge (Figure 2.36).

Calculation of the rate enhancement of the reaction between **26** and **27** in Figure 2.36 in the presence of an initial amount of **28** led to an acceleration factor of 370 M^{-1} for the complex **29** to give **30**.

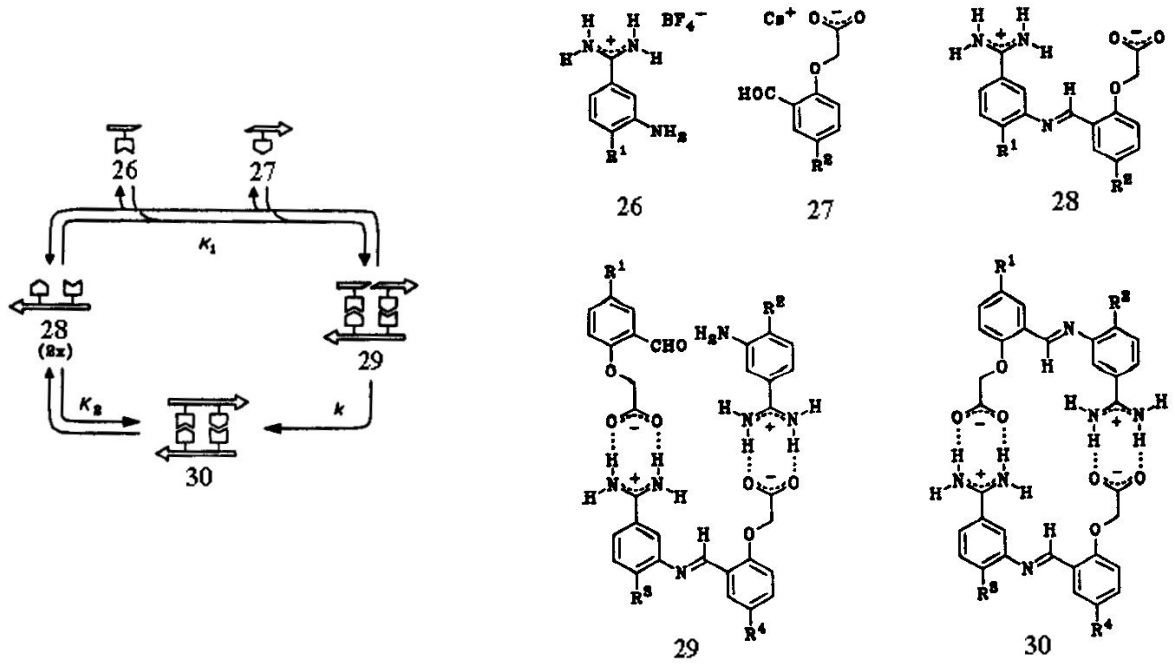


Figure 2.36. Salt bridge templating as shown by Von Kiedrowski. Simplified diagram (left) and chemistry (right).

Remarkable catalytic efficiencies were obtained by the self-replication of peptides, as described by the group of Chmielewski [66, 67]. Their helical peptides (Figure 2.37) were formed from two peptide fragments at pH 4.0 at a non-catalytic rate of $5*10^{-4}M^{-1}s^{-1}$. Their calculated catalytic rate constant was 51 $M^{-1.9}s^{-1}$, indicating a catalytic efficiency of 10^5. This is comparable to the efficiency observed for some enzymatic systems, such as glutathione transferases [68].

At pH 7.0 no rate enhancement was found upon addition of the template. The reason for this most certainly lies in the helical content of the sample. At pH 7.0, only 28% of the peptide is in a helical conformation. Lowering the pH to 4.0 increased this amount to 87%. Addition of the template to the two fragments at pH 4.0 increased the helical contents of the two fragments of 45% and 16%, indicating the template effect.

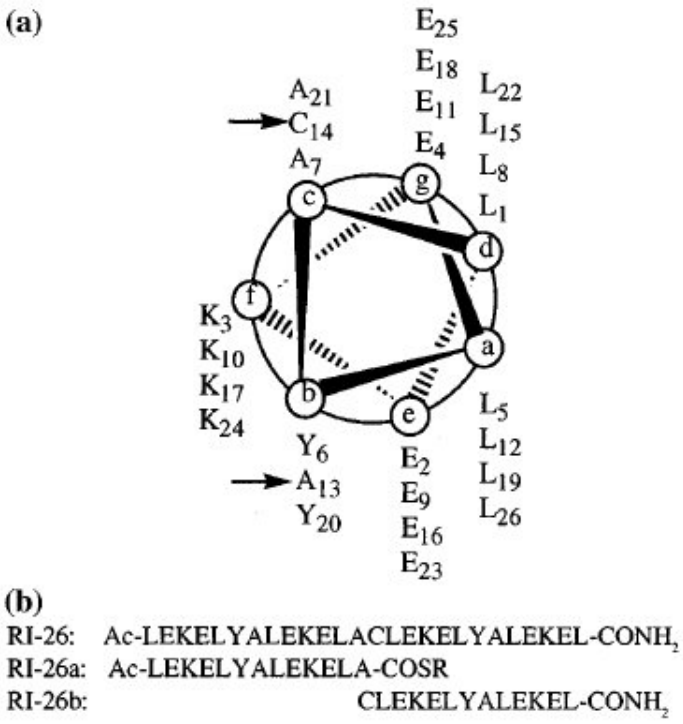


Figure 2.37. Helical wheel diagram (a) and sequence (b) of peptide RI-26 and its fragments. Arrows indicate the amino acids that ligate.

A very different artificial self-replicating system was published by Bachmann *et al.* in the beginning of the nineties [69, 70]. Instead of copying molecules, as in all previous examples, they focused on the replication of entire physical structures, namely reverse micelles. This can be compared with replication of living cells.

A reverse micelle is basically a water droplet in an organic solvent, stabilized by a layer of surfactant. The system designed by Bachmann *et al.* involved a bulk (organic) solvent of isooctane/1-octanol (9/1), containing octanoic acid octyl ester. The reversed micelles contained an aqueous solution of lithium hydroxide, which is able to catalyze the hydrolysis of esters. The used surfactant is octanoic acid sodium salt.

Hydrolysis of the octanoic acid octyl ester will predominantly take place at the interface, since the reactant is present in the bulk and the catalyst in the reverse micelles. The reaction products are 1-octanol, which is one of the bulk solvents, and octanoic lithium salt, which acts as a surfactant. If the micelles accumulate too much surfactant at their interfaces, they can diminish this amount of surfactant by splitting. New micelles are thus created, causing an increase in the concentration of micelles over time. Since the amount of water in the system remains constant, 'daughter micelles' are always smaller than the 'mother micelle'. This latter problem could in principle be overcome by continuous addition of water, but for simplicity the authors kept the amount of water in the reaction mixture constant.

In a control experiment suspending aqueous LiOH in the organic mixture without the surfactant, no reaction took place. This proves the 'self-reproduction' character of the micelles.

2.4.3 Cross catalytic systems

Replication of DNA strands in living cells does not occur via simple self-replication as described in the last paragraph. One of the two DNA strand templates the formation of a complementary strand. This complementary strand serves as a new template to produce a copy of the original strand. This process is called cross catalysis.
Li and Nicolaou worked on the subject of cross catalysis [71], achieving this *via* the formation of a DNA triple helix. Two pieces of a complementary third strand were aligned in the major groove of a DNA duplex. The phosphate end of one piece and the 5'-hydroxygroup of the other were thus close enough to react and form one strand to yield a full triplex. The equilibrium between triplex and duplex DNA was then driven towards the duplex by rising of the pH or the addition of extra equivalents of DNA fragments complementary to the newly formed strand. This newly formed strand could act as a template and catalyze the formation of its complementary strand, yielding a full copy of the original duplex.
Two factors limit the applicability of this method. First, one of the strands must be an all purine strand in order for triplex formation to occur. Second, only symmetric, or palindromic, sequences can be used. The Hoogsteen bonded strand aligns parallel with its complementary strand instead of the anti-parallel alignment adopted by the Watson-Crick bonded strand. 3' and 5' end would thus be inverted in case a dissymmetric sequence is used. It is possible to copy a non-symmetric sequence *via* the same system, but then a second cycle and the appropriate substrates will be needed to make a copy of the initial duplex in the correct sense.

Von Kiedrowski and coworkers continued their research on the formation of DNA hexamers from two trimers [56, 57]. The two trimers CCG (called A) and CGG (called B) can be ligated to form four different products; CCG CCG (AA), CGG CGG (BB), CCG CGG (AB), CGG CCG (BA) (Figure 2.38). In the original system, shown in Figure 2.32, only the formation of AB (CCG CGG) was considered. It was shown that the formation of AB was catalyzed by AB; auto-catalysis. Formation of BA is of the same principle; it is templated by another molecule of BA. Formation of AA is not catalyzed by AA, but by its complementary oligomer BB. This is also true *vice versa*, BB is templated by AA. A cross-catalytic cycle is thus created. Mixing of the two trimers A and B (with the condensing agent EDC) thus leads to one cross-catalytic cycle and two auto-catalytic cycles (Figure 2.38).

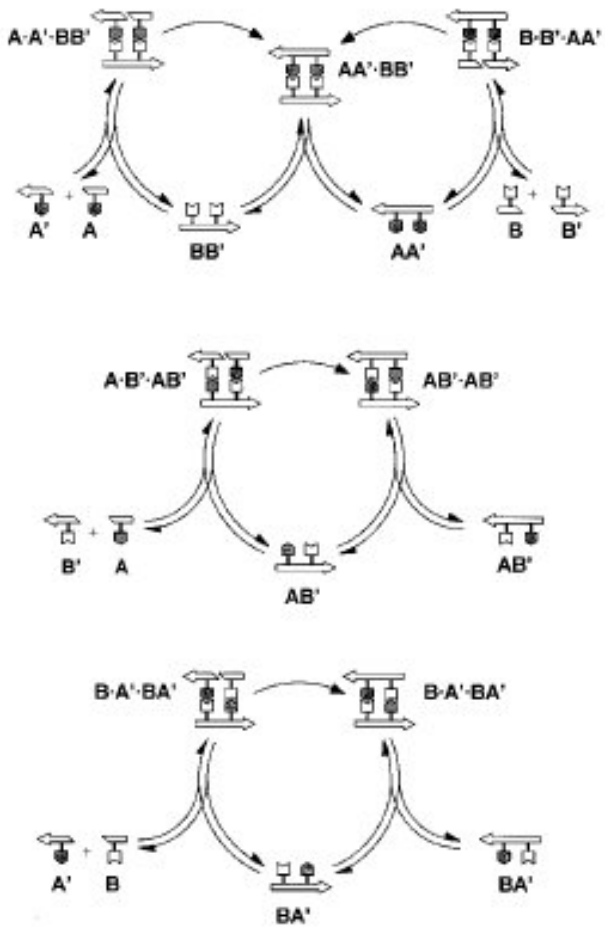


Figure 2.38. Cross-catalytic (top) and two autocatalytic (bottom) replication cycles.

Mixing of A and B without addition of one of the templates led to the formation of AA, BB, AB and BA at similar rates. Addition of 0.16 eq. of AB led to a 50% increase in the formation of AB. The formation of BA was hardly affected, but the formation of AA and BB was slightly reduced, presumably due to reduction of the availability of the monomer. Addition of BA showed similar results, enhancement of the formation of BA, no effect on AB, and slight reduction of AA and BB formation.

Addition of template AA (0.16eq.) led to a relative increase of BB of 100%. AB and BA formation decreased by around 20%, since less monomer is available. The amount of AA increased slightly, by around 20%. It is not known whether this is due to the increased availability of A, or due to templating by newly formed BB.

When both templates AA and BB are added, the cross catalysis effect becomes obvious. Both AA and BB are enhanced similarly to the case of the self-catalysis products AB and BA. Formation of the products AB and BA is decreased by around 30%.

A non-biological system capable of cross catalysis was published by R.J. Pieters *et al.* [72, 73]. A guest (**35**) and a host (**36**) were designed.

Both host and guest contained an amide bond, which could be formed from an amine (**31** and **33** respectively) and an ester (**32** and **34** respectively). Addition of the host to amine **31** and ester **32** accelerated this reaction twelve fold relative to the background reaction. The reciprocal template, guest **35**, proved to be capable of catalyzing the coupling of **33** and **34** five fold. These two reciprocally templated reactions form a replication cycle. A cross catalysis situation in one pot could unfortunately not be obtained. If compounds **31**, **32**, **33**, and **34** are mixed, **31** and **34** react preferentially to form a self-complementary product 500 times faster than the reaction between **33** and **34**. Similarly, **32** and **33** react 3500 times faster than **31** and **32**. Two self-replicating cycles will be formed instead of one cross-replication cycle.

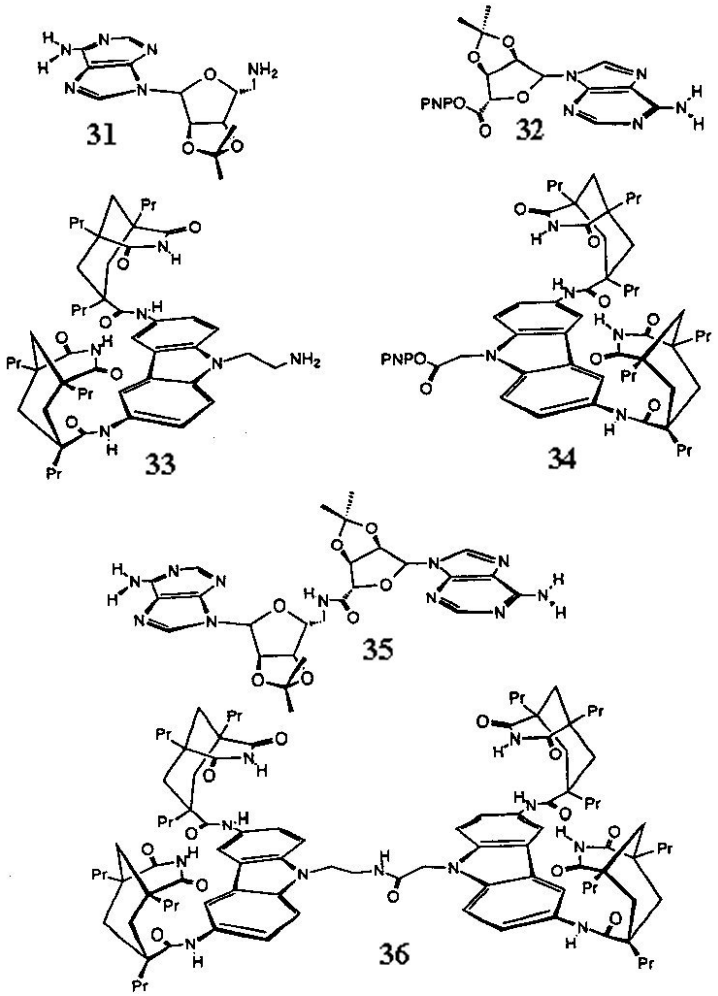
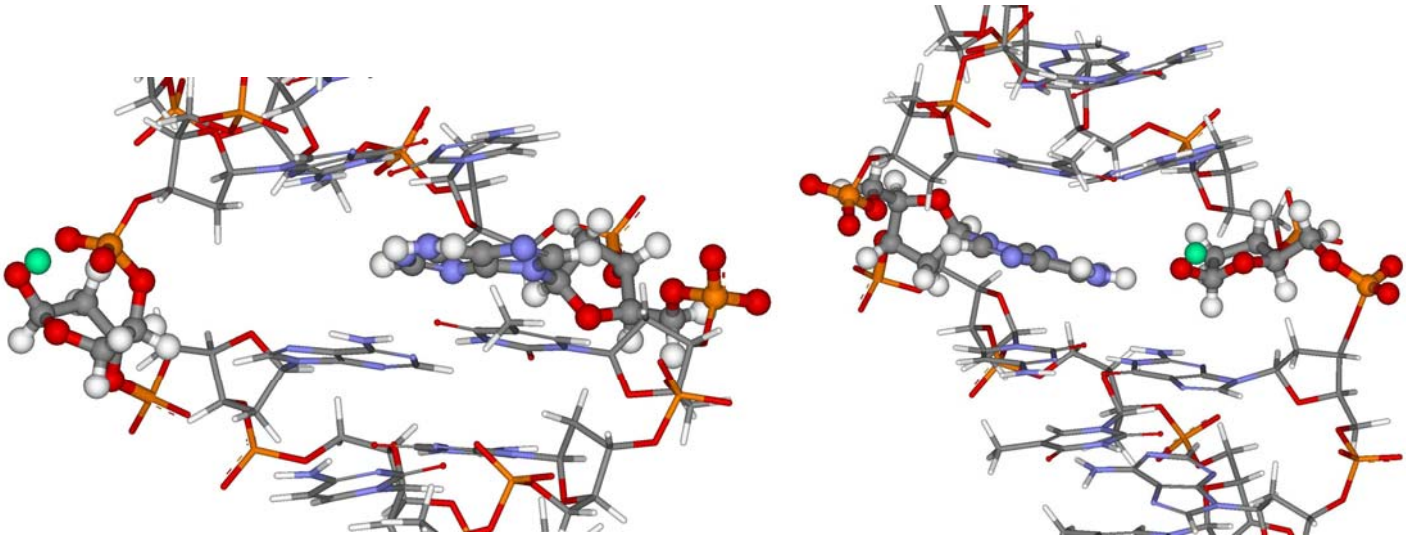


Figure 2.39. Cross catalysis of a non-biological system.

3 Design of a natural abasic site repair system

3.1 Introduction and goal

In this chapter we want to develop a repair model system for abasic sites within a nucleic acid duplex. An abasic site is a nucleotide in which the nucleobase is missing. The sugar is still present and bears a hydroxyl group on carbon $C^{1'}$, where the nucleobase is normally located. An abasic site is very destabilizing for the DNA duplex in which it is located because of the loss in hydrogen bonding and π–π-stacking. Abasic sites are naturally occurring moieties and arise for instance from the damage by radiation or from spontaneous hydrolysis [74]. NMR structures showed that the deoxyribose moiety of the abasic site can either point in or out of a double helix (Figure 3.1) [75].



a b

Figure 3.1. NMR structures of an abasic site, the deoxyribose group of the abasic site and the opposite base are represented with thicker bonds. a: abasic site is located outside the double helix. b: abasic site is located inside the double helix. The opposite adenine is inside the helix in both cases.

We want to explore the possibility of restoring the stability of a DNA duplex containing an abasic site by covalently reacting it with a repair unit (Figure 3.2). This unit should be involved in π–π stacking in the duplex. The group of Eric Kool developed so called universal bases that restore π–π stacking in a DNA duplex (Figure 3.3) and can ‘replace’ any of the four nucleobases. Hydrogen bonding with the opposite base is not very strong, but a duplex containing this base mimic is more stable than a duplex containing an abasic site, regardless of the opposite base [76, 77].

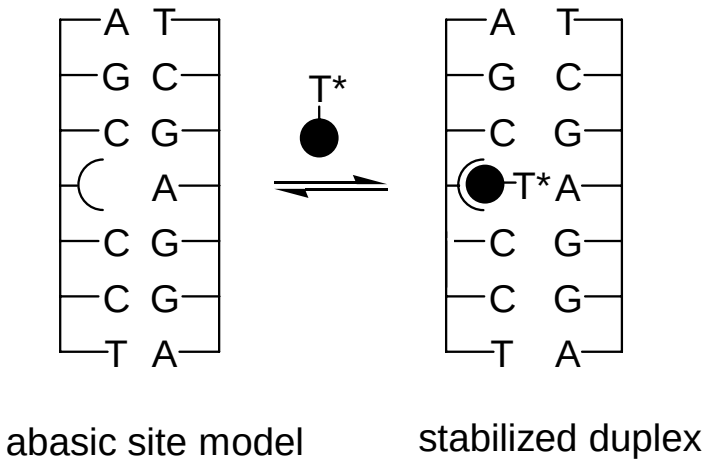


Figure 3.2. Principle of stabilization of DNA duplexes containing an abasic site. T*, any group complementary to A that stabilizes the DNA duplex.

If the repair unit resembles a nucleobase it may also be able to hydrogen bond with the base opposite to the abasic site. In that case we can envisage selection of a specific reagent in competition experiments. When mixing various repair units mimicking the four nucleobases, incorporation of one unit may be favored over the other. Which base is favored may depend on the base that is facing the abasic site. If for instance, like in Figure 3.2, an adenine is facing the abasic site, a thymine mimicking group may be incorporated selectively over repair units mimicking adenine, guanine or cytosine. The adenine serves here as a template. Changing the opposite base into a guanine for example, would probably favor selection of a cytosine mimic. This principle can also be regarded as information transfer from a template strand into the abasic site, other examples of templated reactions and information transfer were discussed in chapter 2.

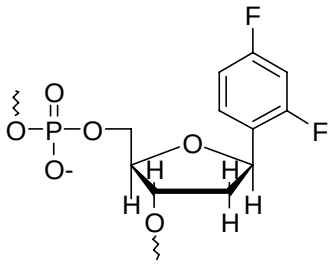


Figure 3.3. Universal base mimic as designed by Kool *et al.*

We would like the reactions to be reversible on the time scale of laboratory experiments (seconds to days). Error correction would then be possible in competition experiments. An incorporated repair unit that does not stabilize the duplex can be replaced by a unit giving more stabilization. From a pool of reagents the best stabilizing unit is then most present in the abasic site. If the reaction is irreversible, there can only be kinetic competition between the

reagents. The reagent that reacts fastest will be the most abundant in the final mixture of repaired duplexes.

Reactions that until now have been developed to recognize and stabilize abasic sites involve the use of intercalators that bind to DNA duplexes non-covalently. Compound **1** (Figure 3.4) for instance can be inserted non-selectively in a DNA duplex containing an abasic site [78]. One of the acridine moieties binds in the abasic site, the other intercalates at a distance of one base pair, in the 3' end direction relative to the abasic site. The effect of the nature of the opposite base upon intercalation was not studied by the authors.

Selective intercalation was obtained with compound **2**. The adenine like part of the molecule shows a higher affinity for an opposite thymine than for another nucleobase. Compound **2** temporarily stabilizes the DNA duplex, but can also lead to the cleavage of the DNA strand at the abasic site [78, 79].

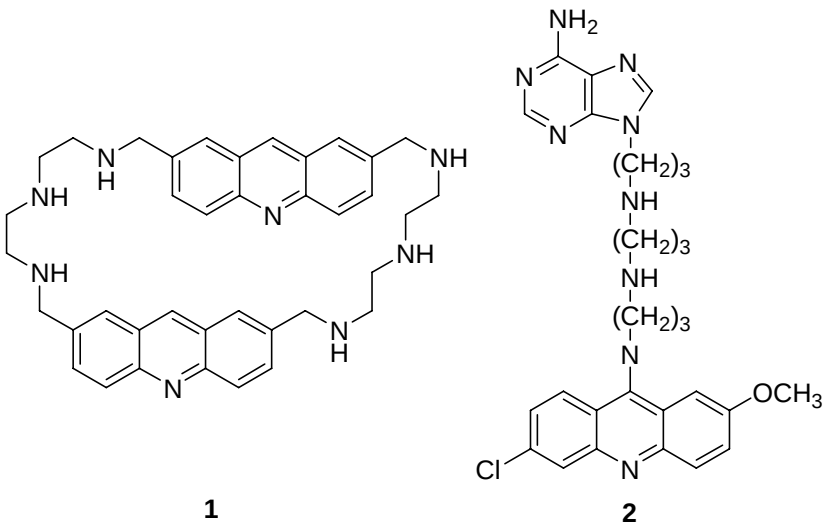


Figure 3.4. Example of an intercalating molecules.

Covalent repair of an abasic site has not yet been achieved. In the previous examples, intercalation only involves non-covalent interactions. We want to covalently link a repairing unit to the abasic site.

Covalent stabilization of an abasic site was obtained by Smirnov *et al*.[80]. They incorporated a pyrene unit opposite to an abasic site (Figure 3.5). The pyrene occupies the space that is normally filled by the base in the other strand. It restores π–π stacking in the DNA duplex and therefore enhances its stability. However, this does not amount to 'repairing' an abasic site.

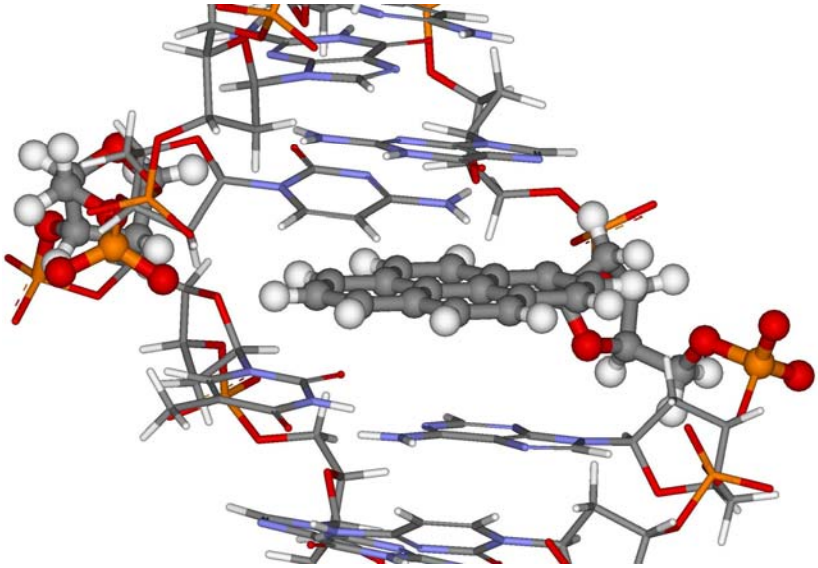


Figure 3.5. NMR structure of a pyrene unit incorporated opposite to an abasic site. The deoxyribose is in an extrahelical position. The deoxyribose moiety of the abasic site and the pyrene moiety are represented with thicker bonds.

3.2 Design of the system

3.2.1 Biological relevance

If not repaired, the occurrence of abasic sites has serious consequences for living cells. One obvious consequence is the loss of genetic information caused by the lack of a base pair. Not very likely under biological conditions but much more serious, is the occurrence of β-elimination, which can cause DNA strands to break (Figure 3.6). β-elimination is catalyzed by both acidic (<pH 3) and slightly basic (>pH 8) conditions [81, 82] and the degradation products are prone to further degradation.

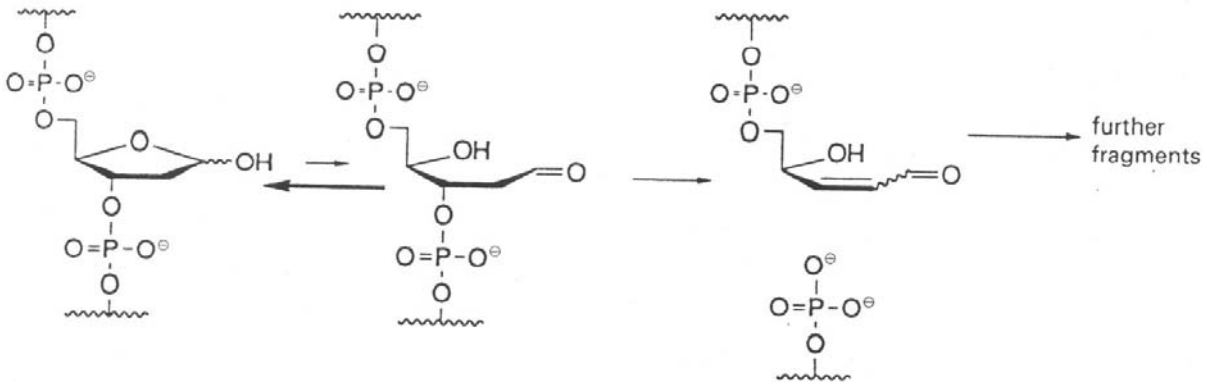


Figure 3.6. β-elimination of an abasic site.

Enzymatic repair of abasic sites involves the scission of the deoxyribose moiety (Figure 3.7 step c) [74, 83], followed by incorporation of a completely new nucleotide (Figure 3.7 step

d/f). Two covalent bonds have to be formed in the process and the supramolecular complex involves the alignment of five molecules; the two pieces of the strand that has to be repaired, the complementary strand, the new nucleotide and the enzyme.

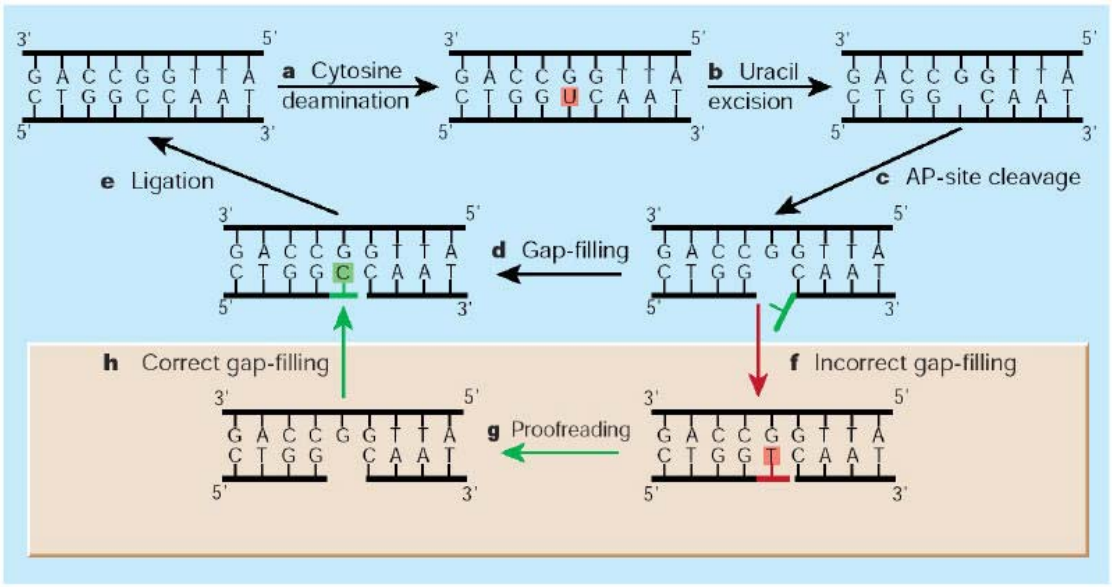


Figure 3.7. Natural point mutation reparation.

The model system for abasic site repair that we propose is very different from the natural situation. Without the use of enzymes, it will be difficult to align five molecules and form two covalent bonds. We will have to reduce the number of species involved and reduce the number of bonds that have to be formed. We will therefore not cut the backbone, but cut between backbone and base. Only one covalent bond has to be formed in this case and only three species have to be aligned; the strand containing the abasic site, the complementary strand and the repair unit.

As can be seen from Figure 3.7, the enzymatic repair system is capable of error correction. A wrongly incorporated base that generates a mismatch base pair is removed by another enzyme. We would like this error correction to be incorporated in our model system, the template strand should self-correct the errors it made.

3.2.2 Choice of the reaction

As mentioned above, to make error correction possible, we want our reactions to be reversible (section 3.1). Second we need a biologically compatible reaction. That means we want the reactions to proceed in water, at neutral pH. At last our selection method will involve DNA templating, which means that our reaction should not interfere or cross-react with the various functional groups of DNA.

Two reactions that suit these conditions are the formation of disulfide bonds followed by reversible disulfide exchange and reactions between aldehydes and amines to form imines. We thus have the choice of incorporating a thiol, an aldehyde or an amine in the DNA backbone and a complementary group on the repair units (Figure 3.8). This would generate a non-natural mimic of an abasic site with a pendant reactive group.

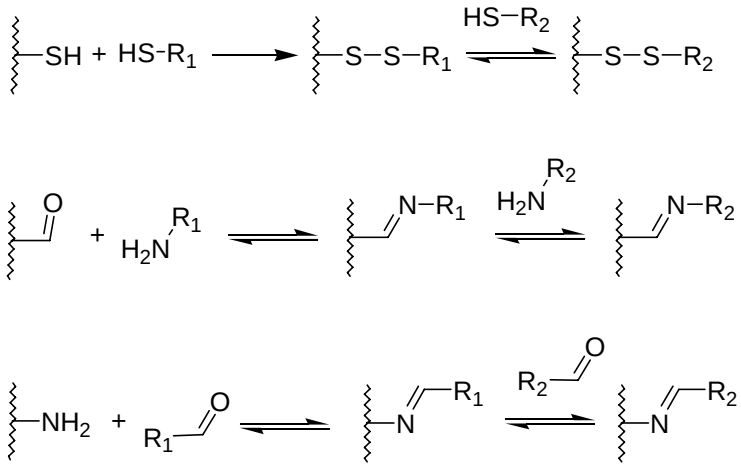


Figure 3.8. Design of a reversible abasic site repair model system by dithiol exchange or imine/aldehyde-amine equilibria.

However, a natural abasic site consists of a deoxyribose, which is in equilibrium between its closed α- and β- hemiacetal form and its open aldehyde form (Figure 3.9). Rather than to develop non-natural abasic sites, it seemed easier and more relevant to try to use this aldehyde as an anchor point for our reagents. To validate the possibility of using this aldehyde as an electrophile, we did a literature search on reactions at this position.

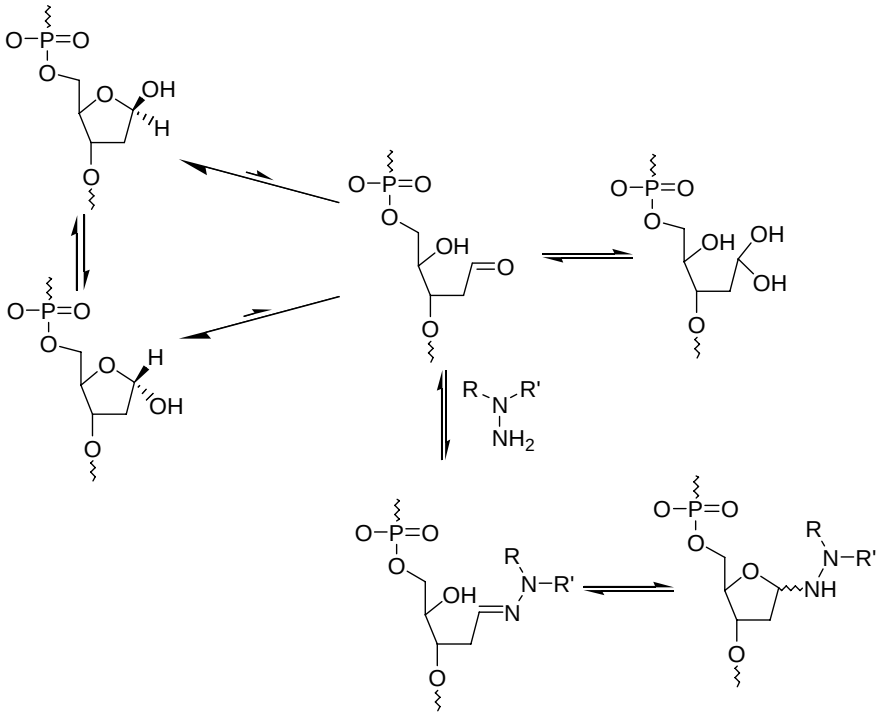


Figure 3.9. Abasic site in equilibrium between the α and β hemiacetal, aldehyde and aldehyde-hydrate, and its reactivity towards amines.

3.2.2.1 Reactions on abasic sites

The destabilizing effect the abasic site has on a DNA duplex may favor the reaction of other compounds with the abasic site, since the products will easily be more stable than the starting duplex. However, surprisingly little literature concerning reactions on abasic sites could be found. This can be explained by the occurrence of β-elimination, as well as by the relative inertness of the deoxyribose moiety towards various reagents at neutral pH. The hemiacetal that is present 99% of the time [84-86] is poorly reactive. The aldehyde form is more reactive, but it is hardly available to reaction.

The first reported reactions on the DNA abasic site were designed for detection. Talpaert-Borlé *et al.*[87] for instance reacted abasic sites with ^{14}C labelled methoxyamine and assessed the quantity of abasic sites in their mixtures by radioactivity measurements.

Kubo *et al.* [88, 89] designed a reagent to react with the aldehyde form of the abasic site that could be easily detected (Figure 3.10). The reactive side of the molecule consisted of a hydroxylamine residue, which reacts selectively with the aldehyde function of the deoxyribose. The detection side consists of a biotin moiety to fix the DNA with the abasic site on a target. An indicator enzyme was used to determine the number of biotin-tagged abasic sites.

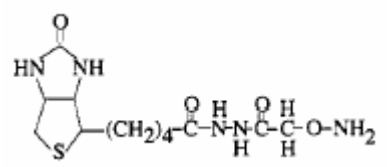


Figure 3.10. Aldehyde Reactive Probe.

Very little material (100 ng) is necessary for the method designed by Kubo and Ide, but the DNA must be adsorbed onto the surface of the wells in the micro titer plates. Some of the abasic sites are not accessible to the macromolecular reagents, causing quantification of the number of abasic sites impossible.

A better and easier detection method was designed by Boturyn *et al.* [90, 91]. Instead of a biotin, a fluorescent probe was attached to the hydroxylamine *via* a peptide or an ethyl ether linker. Fluorescence measurements make it possible to easily quantify the number of abasic sites present in an oligonucleotide sample.

Purifying an oligonucleotide sample by removing the DNA oligomers containing abasic sites was performed by Krotz *et al.* [84]. Reaction of the abasic site with a hydroxylamine derivative containing a hydrophobic tag, yielded DNA strands with reduced hydrophilicity. These strands could be separated from undamaged DNA using reversed-phase HPLC. Note that the group of Krotz *et al.* used phosphorothioate DNA, but that does not change the principle.

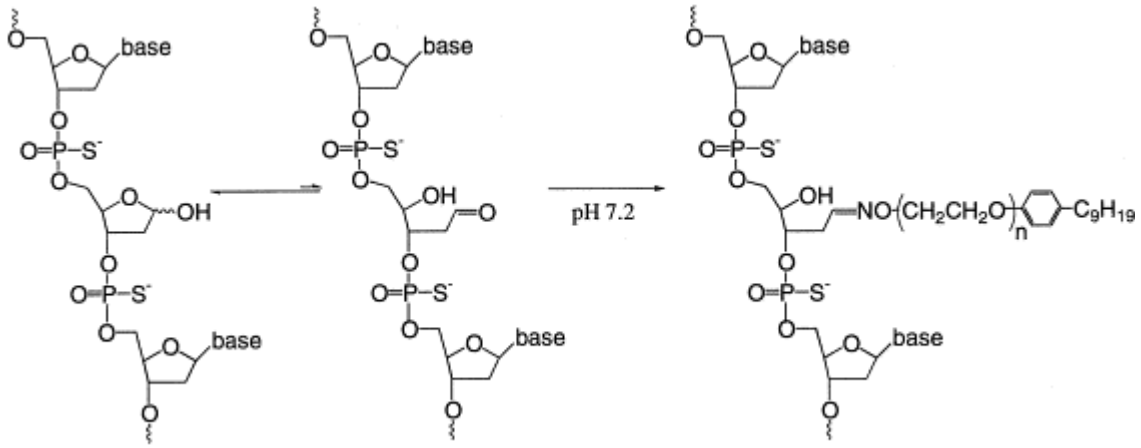


Figure 3.11. Hydrophobic group introduction.

Manoharan *et al.* [92] did not use hydroxylamines to react with the aldehyde form of the deoxyribose. They synthesized a DNA duplex with an abasic site on one strand and opposite to this abasic site a 2'-modified adenine, containing an amino group connected *via* an alkyl spacer (Figure 3.12).

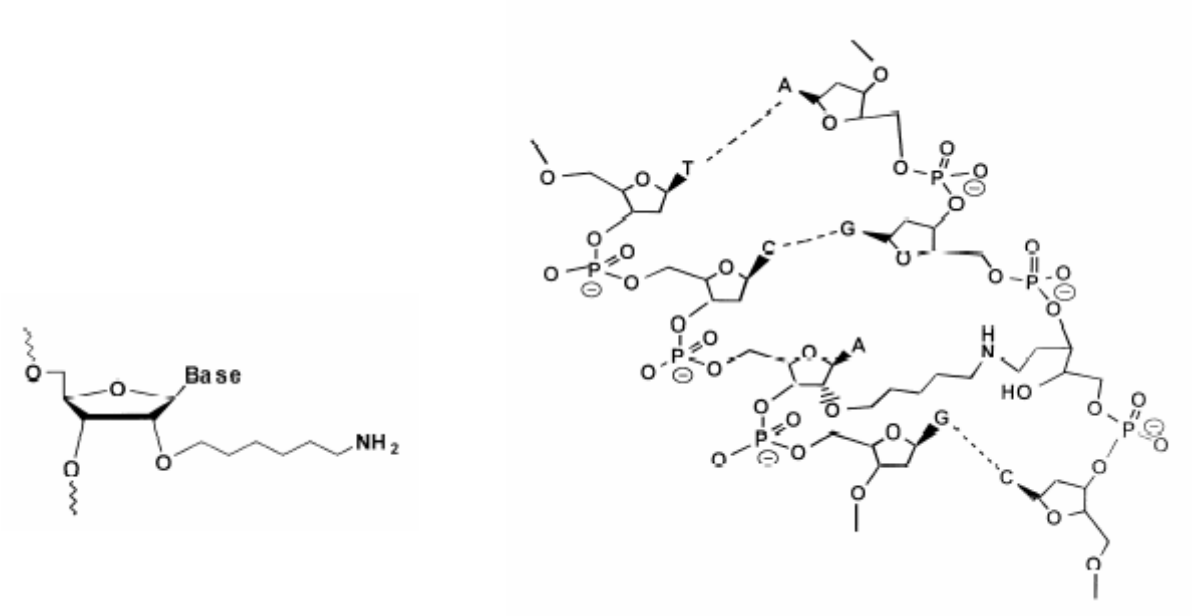


Figure 3.12. 2'-amine modified oligonucleotide and crosslinking with a complementary DNA strand

This amine can react reversibly with the aldehyde form of the deoxyribose to form an imine (water, pH 5), but the equilibrium is laying at the aldehyde/hemiacetal side. Under reductive amination conditions ($NaCNBH_3$) the imine is reduced preferentially over the aldehyde, pulling the equilibrium to the imine side and forming a secondary amine. The DNA duplex is now covalently crosslinked at the level of the abasic site. The crosslinked duplex, containing eleven base pairs including the crosslinked pair, maintains Watson-Crick base pairing up to 76°C and then changes conformation to form a random coil. The melting temperature (T_m) of the non-crosslinked duplex was 30°C and the natural duplex without the abasic site had a T_m of 48°C. This shows the huge increase of stability obtained by crosslinking.

The conclusion of the literature study is that reaction of the aldehyde with amine type compound is possible. Hydroxylamines react with the abasic site under physiological conditions to form oximes. Primary amines react as well, but the equilibrium between aldehyde/hemiacetal and imine lies on the aldehyde/hemiacetal side.
Reversibility of the reactions has not yet been reported though. The oximes that are obtained by the groups of Cole and Ido [84, 88] are bad leaving groups. The amines used by Manoharan *et al.* [92] do not react in sufficient quantity to be detected without reducing them.
We thus have to find a suitable amine type reagent that can react reversibly with the aldehyde and stabilize the DNA duplex in which the abasic site is incorporated.

3.2.2.2 Optimization of the amine component

Reaction of aliphatic amines and aliphatic aldehydes hardly take place in water at pH 7. Below pH 8, the amines are predominantly protonated and not nucleophilic enough to react.

The equilibrium is strongly shifted in favor of hydrolysis and only small quantities of imines are obtained. To circumvent this problem and pull the equilibrium to the product side, reduction of the imine with $NaBH_3CN$ is often used [92, 93], assuming that the proportion of secondary amines obtained after reduction reflects the proportion of imines.

On the other hand, hydrazines, hydrazides and hydroxylamines react very well with aliphatic or aromatic aldehydes at pH 7. However, hydrolysis of the corresponding hydrazones, acyl hydrazones and oximes is very slow. Only under acidic conditions (<pH 4) does hydrolysis take place rapidly [93].

Nguyen *et al.* speculated that hydrazine derivatives bearing electron withdrawing groups might be sufficiently nucleophilic to react with aldehydes and be sufficiently good leaving groups, to form easily hydrolyzable products (Figure 3.13). They tested various compounds and showed rapid formation of the hydrazone derivatives of trifluoro acetylhydrazine, acetylphenyl hydrazine and 3-aminorhodanine at pH 6-8. Hydrolysis rates were determined upon dilution of the hydrazone samples. At pH 6 hydrolysis took place in minutes, at pH 8 in around an hour.

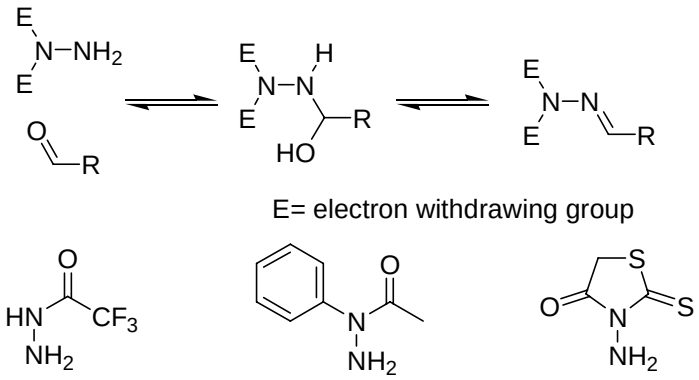


Figure 3.13. Imine chemistry with activated amines.

The deoxyribose moiety is not a simple aliphatic aldehyde, but the aldehyde is in equilibrium with its closed α and β hemiacetal forms. Under physiological conditions the equilibrium lies on the hemiacetal side, the α- or β-hemiacetal are present 99% of the time. The reaction of the open aldehyde form with the hydrazides can, in principle, shift the equilibrium away from the hemiacetal towards the hydrazone products. The hydrazone itself may cyclize to form α- or β-hemiaminal products. This hemiaminal-hydrazone equilibrium resembles the hemiacetal-aldehyde equilibrium and may be able to pull the reaction towards completion.

The hydrazide derivatives that we want to use to repair the abasic site are mimicking the natural nucleobases (Figure 3.14). Compounds **3-6** are mimicking thymine, adenine, cytosine and guanine respectively and all possess the structure $H_2N\text{-}NE_2$, as in the examples of Figure

3.13. Mixing these compounds with isobutyraldehyde [93] confirmed that they react reversibly with aliphatic aldehydes.

Acyl hydrazides **7** and **8** have only one electron withdrawing group and are thus stronger nucleophiles. They may react faster with the aldehyde form of the abasic site but may compromise the reversibility of the reaction. Compound **7** mimics a thymine, **8** is a universal base mimic. It is expected not to hydrogen bond with a complementary strand, but should restore π–π stacking [76, 77].

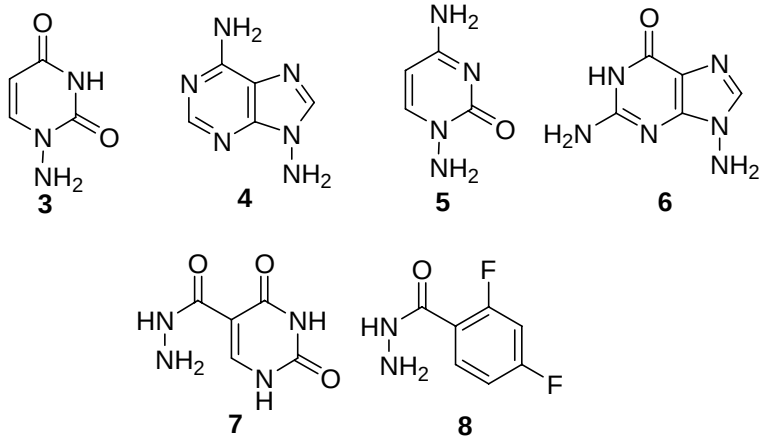


Figure 3.14. Nucleobase mimics.

3.3 Possible behavior of the system

3.3.1 Introduction

We would like to know more about the conformation of the abasic site. We know that it destabilizes the system, but not to what extent. The conformation of the abasic site determines its availability to reagents, and the possibility to stabilize the products. In Figure 3.1 I showed NMR-structures of two different conformations of the same abasic site, in the same DNA duplex. In one case the abasic site is positioned in the double helix, in the other case the abasic site is extrahelical. In the given sequence it thus seems that there is an equilibrium between the two conformations. The opposite adenine is in both cases intrahelical.

Both extrahelical and intrahelical conformations have advantages and inconveniences for the repair system that we want to design. An extrahelical abasic site (Figure 3.15b) should not be hindered for reaction, but the products should not bring about any stabilization of the duplex if they remain in an extrahelical conformation: in that conformation they cannot contribute to

π–π-stacking and hydrogen bonding. For stabilization, the products should flip inside the double helix. This seems not evident at first hand.

If the abasic site preferentially adopts an intrahelical conformation (Figure 3.15a), stabilization of the products may take place, but the repair units have to diffuse to the interior of the double helix. They may observe steric hindrance when entering the double helix, which might limit the reaction rates. The requirements for availability and product stabilization are thus in conflict. A prediction which of the situations will be best for DNA repair, cannot be made simply.

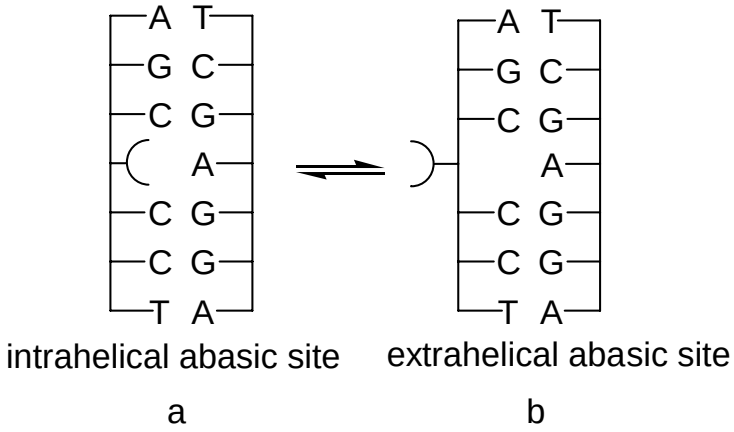


Figure 3.15. Schematic drawing of the conformations of the abasic site.

A literature study should give more information of the effect of an abasic site on a DNA double helix. We want to know whether an abasic site adopts an intra- or extrahelical conformation, and what the influence of the sequence of the oligonucleotide has on the conformation of the abasic site.

3.3.2 Conformation of the abasic site

The general structure of the backbone of a DNA double helix containing an abasic site is not very different from a normal, fully complete duplex as indicated by ^{31}P- and ^{1}H-NMR. It is still right-handed and in its B-form.

Locally, at the place of the abasic site, changes are more important. A kink of 10-30 degrees was observed in various DNA strands. [85, 94, 95]. Furthermore, according to Singh *et al.* [96], *both* the deoxyribose and the unpaired base opposite to it may be stacked inside the helix or expelled from it, or in equilibrium between these to extremes. Figure 3.15 thus does not tell the whole truth, but should be completed by Figure 3.16. The conformation of the abasic site and the opposite base depends on the nature of the base opposite to the abasic site, as well as to the neighboring basepairs [96].

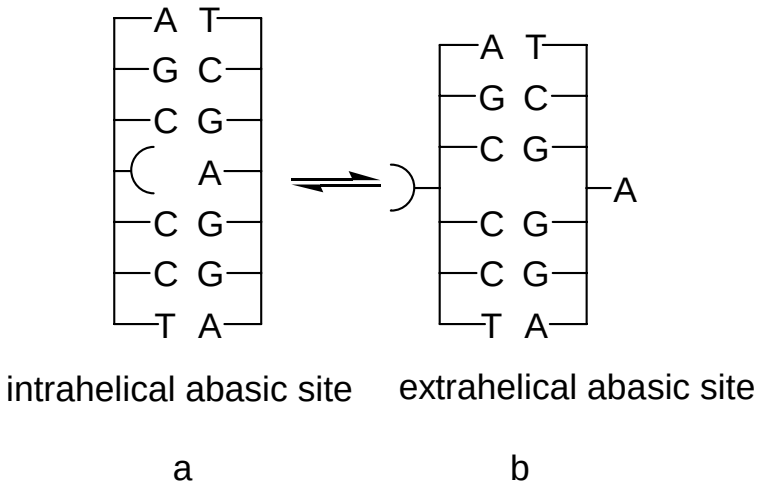


Figure 3.16. Schematic drawing of the conformation of the abasic site and the opposite base.

Various research groups have investigated the reasons why, in a particular duplex, the abasic site and the opposite base are intra- or extrahelical. Both π–π-stacking and hydrogen bonding play a role. The factor of π–π-stacking with neighboring bases may be evident. The missing base will reduce π–π-stacking and the remaining bases have two possibilities; The base pairs neighboring the abasic site may either stack with the unpaired base, or may be better stabilized by stacking with each other. In the first case the abasic site and the opposite base are positioned inside the helix (Figure 3.16a). In the second case the abasic site and the opposite base are expelled and the helix 'collapses' to allow π–π-stacking between the neighboring bases (Figure 3.16b).

The role of hydrogen bonding is less obvious. The group of Philip Bolton [97, 98] showed that hydrogen bonding can occur between the free deoxyribose and the opposite base. Molecular modeling and NOESY spectroscopy showed that the OH group on $C^{1'}$ of the deoxyribose and the N^3 of an opposite adenine could form a hydrogen bond *via* a complexed water molecule. Both the deoxyribose and the adenine remain stacked inside the helix. If an unpaired cytosine is incorporated in an otherwise identical duplex, the distance between the OH on the $C^{1'}$ of the deoxyribose and the cytosine is too large for hydrogen bonding to occur. Both the deoxyribose and the cytosine are then expelled from the double helix.

In most studies [94, 99, 100] a THF unit is used instead of a deoxyribose. The THF unit lacks a $C^{1'}$-OH, preventing hydrogen bonding. Nevertheless, this does not significantly influence the results compared to a natural abasic site, as shown by Singh *et al.* [96]. They repeated experiments performed by Cuniasse *et al.* [100] (*vide infra*) using a natural abasic site instead of a THF unit in an otherwise identical DNA oligomer. Singh *et al.* found approximately the same destabilizing effect for a natural abasic site as Cuniasse *et al.* for a THF unit. From these

experiments can be concluded that π–π-stacking is the main driving force that determines the conformation of the abasic site.

Multiple studies about the conformation of the abasic site have been reported [29, 76, 83, 85, 86, 91, 94-103], but prediction of whether an abasic site is in an intra- or extrahelical conformation in a given sequence remains difficult. The multiple studies are sometimes contradicting each other.

Various authors [85, 95, 101] indicated that in a DNA duplex containing an unpaired purine (a pyrimidine base is missing), this purine and the deoxyribose moiety remain predominantly stacked inside the double helix (situation as in Figure 3.16a). The purine bases have more π-electrons and thus contribute more to the π–π-stacking in the DNA duplex. Their conclusion was that the nature of the neighboring base-pairs is not of decisive importance.

However, as shown by Goljer *et al.* [98], this is not correct in all cases. An adenine incorporated opposite to an abasic site in the sequence d($^{5'}$CGC GAY ACG CC$^{3'}$)· d($^{3'}$GCG CTN TGC GG$^{5'}$) (with Y= natural abasic site, N is one of the four nucleobases), yields an equilibrium situation consisting of equal amounts of intrahelically and extrahelically positioned deoxyribose and opposite base.

Goljer *et al.* linked the conformation of the sugar and the opposing nucleobase to the fact that the hydroxyl group on C$^{1'}$ of the deoxyribose may be in its α or β conformation. Molecular modeling and NOESY-NMR experiments showed that the intrahelical fraction consisted only of β-deoxyribose, whereas the α-deoxyribose fraction was completely expelled from the helix. Hydrogen bonding *via* a molecule of water was proposed as an explanation. In the α-form, the hydroxyl group of the deoxyribose points downwards, preventing hydrogen bonding with the opposite base.

Cuniasse *et al.* [100] studied abasic sites with various complementary strands, each containing a different nucleobase opposite to the abasic site. They showed that in the sequence d($^{5'}$CGT GXG TGC$^{3'}$)·d($^{3'}$GCA CNC ACG$^{5'}$) (with X=THF unit, N is one of the four nucleobases) the conformation of the unpaired nucleobase (N) and the abasic site depend on the nature of the opposite nucleobase. The abasic site and the opposite base are fully stacked into the double helix in case an adenine or guanine is positioned opposite to the abasic site. Raising the temperature from 25°C to 37°C does not change the position of the adenine, but at 37°C the abasic site and an opposite guanine are extrahelical half of the time. In the case of a cytosine,

the abasic site and the cytosine are expelled from the helix, apparently driven by π–π–stacking between the two guanines neighboring the abasic site. In case of a thymine, the abasic site and the thymine are in equilibrium between a stacked and an extrahelical conformation. From these studies the authors concluded that the nature of the opposite base is the determining factor for the conformation of the abasic site.

In all these examples, the behavior of the abasic site is identical to the behavior of the opposite base. This is in contrast to the NMR structures of Hoehn *et al.* (Figure 3.1), who found a partially extrahelical abasic site, in combination with a fully intrahelical opposite base. But this is one of the few studies [75, 99] in which abasic site and opposite base are not in the same conformation.

The conclusion of Cuniasse *et al.* [100], that the conformation of the abasic site and the opposite base depends mostly on the nature of the opposite base, is not correct. In their studies they only varied the opposite base. The influence of the neighboring bases was ignored. Eight years later Gelfand *et al.* pointed out that the nature of the neighboring bases is more important than the nature of the opposite base. Their studies led to the conclusion that extrahelical conformations appear only in case the deoxyribose moiety is flanked by purines. Intrahelical conformations or an equilibrium situation arise in all types of sequence contexts and regardless of the base facing the abasic site [102].

However, one year after the publication of Gelfand *et al.,* an example of an extrahelically located lesion with only one flanking purine was published by Cline *et al.* [94]. NOESY spectroscopy of the sequence d($^{5'}$CGT CAG XTC GGA$^{3'}$)·d($^{5'}$GCA GTC CAG CCT$^{3'}$) (with X=THF unit) showed that the abasic site and the cytosine opposite it are completely expelled from the helix and relocated in the minor groove. This disproves the generality of the rule of Gelfand.

We found no other rules of thumb to predict the conformation of the abasic site, making it uncertain to predict the conformation of the abasic site without experimental data. The multiple studies that concern the intra- or extrahelical conformation of abasic sites in DNA did not give unambiguous rules to predict the conformation of an abasic site in a given DNA double helix. Two purines flanking the abasic site increase the chances of an extrahelical abasic site, as does a pyrimidine base opposite the abasic site. But to be sure that the abasic site is in an intra- or extrahelical conformation, we should use a DNA sequence that has already been characterized by others.

We will now proceed with molecular modeling to see if our proposed repair units fit into the cavity left by the missing base. We then continue with test reactions to get the ideal reaction conditions under which the reaction proceeds best. Reactions of deoxyribose and ribose will be performed to see whether we better choose a DNA or an RNA oligonucleotide. Compounds **3-8** will be used to test which compounds are best reversible with both ribose and deoxyribose.

3.3.3 Molecular modeling

Before starting any experiment, we would like to get an idea of whether the molecules that we have designed fit in a DNA double helix and might be able to repair an abasic site. Energy minimization of the reaction products of aminobases **3-6** and hydrazides **7** and **8** (Figure 3.14) with a natural abasic site were performed. These calculations were not intended as systematic searches of energy minima (no Monte-Carlo search or Molecular Dynamics were performed), but only as simple preliminary studies to evaluate whether energy minima exist where Watson-Crick type base pairing occur with the repair unit. We used an Amber* force field implemented in the MacroModel software. Amber* takes into account the hydrogen bonds in the DNA duplex very well.

As a starting point for molecular modeling we took a known decameric DNA duplex, $(d^{5'}(CCA\ GGC\ CTG\ G)^{3'})_2$ [104-106]. A crystal structure of the B-helix of this sequence was available. We modeled our modification in the middle of the strand, at the fifth base pair.

We were looking whether our artificial repair units (Figure 3.14) find a position favorable for hydrogen bonding with the opposite base and for π–π-stacking with its neighboring base pairs. The positions that were found may not be the optimal position for the repair unit, but when a good position is found, it is likely that the modified base can successfully repair a DNA duplex.

3.3.3.1 Aminouracil 3

As can be seen from Figure 3.17, the influence of an abasic site on the general structure seems not very large. The sugar moiety is positioned a bit more inside the helix than when a natural base is present. A kink in the helix, as observed by other authors [85, 94, 95], is not observed here.

Since the available hydrazide building blocks **7** and **8** both resemble a thymine unit, we choose to place an adenine unit opposite the abasic site. The duplex we use for modeling will thus be $d^{5'}(CCA\ GGX\ CTG\ G)^{3'\cdot 3'}(GGT\ CCA\ GAC\ C)^{5'}$ (with X = repaired abasic site). We

first minimized the strand containing the natural bases. We then cut the linker between base and backbone and reconnected base and backbone with our modified linker causing as little strain in the linker as possible. This allowed us to have a favorable position for the thymine mimic before starting the energy minimization.

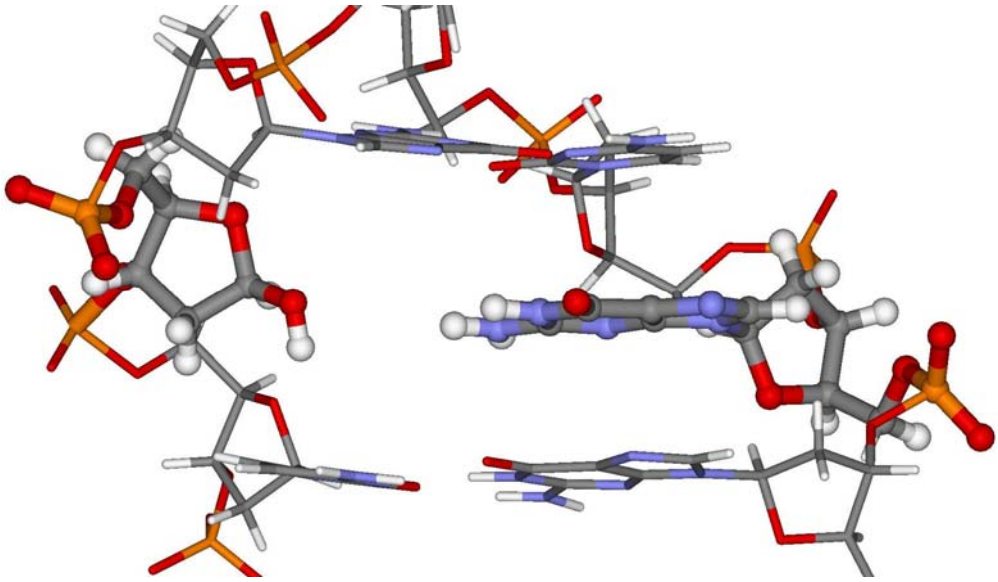


Figure 3.17. Energy minimization of a B DNA double helix containing an abasic site (β-epimer) with a guanine opposite the abasic site.

Upon energy minimization, A-T base pairing was overall well conserved as well as π–π–stacking with the neighboring base pairs. Only the sugar phosphate backbone underwent significant conformational changes. However, the difluoro benzene unit apparently generates much greater disturbance in the double helix.

Modeling of the product of aminouracil **3** in its hemi-aminal form showed a significant disturbance in the helix compared to a natural thymine (Figure 3.18). The extra atom between the base and the sugar seems to be difficultly accommodated by the backbone and the base is tilted compared to its normal position. The helix shows a distinct kink at the position of the repaired abasic site. Hydrogen bonds between repair unit **3** and the opposite adenine are still possible.

The hydrazone form of the same repair unit **3** shows much less disturbance of the duplex (Figure 3.19). The repair unit is almost at its natural position in the double helix, the sugar moiety is flexible enough to allow a backwards displacement to keep the uracil moiety at a favorable place for hydrogen bonding and π–π-stacking. The strain of one extra atom between the base and the backbone seems to be easily relaxed in the backbone. Hydrogen bonding is well accounted for by the force field, the system thus prefers to keep the base in place and move the backbone to keep that position.

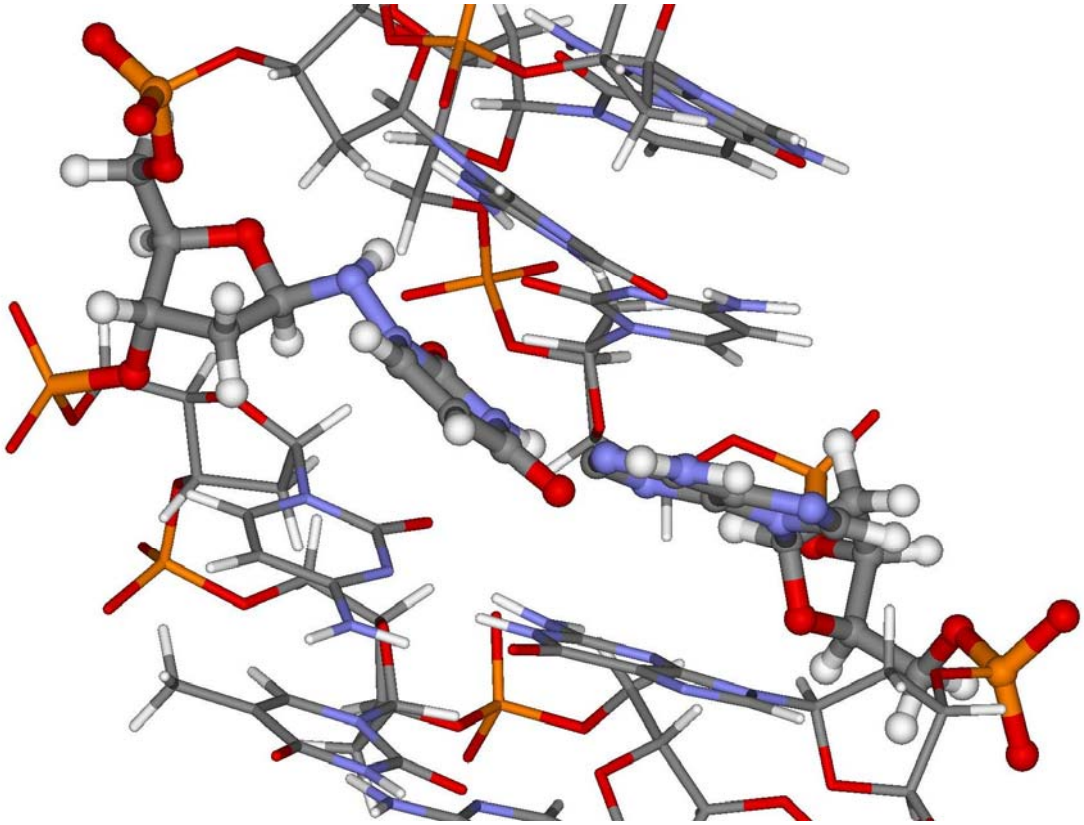


Figure 3.18. Energy minimization of aminouracil 3 at the abasic site position. The base is in its hemi-aminal form (β-epimer). Repair unit 3 and the opposite adenine are tilted at an angle of about 21 degrees. The O(=C^4) of 3 and the N(H_2) of the adenine are at a distance of 2.83 Å.

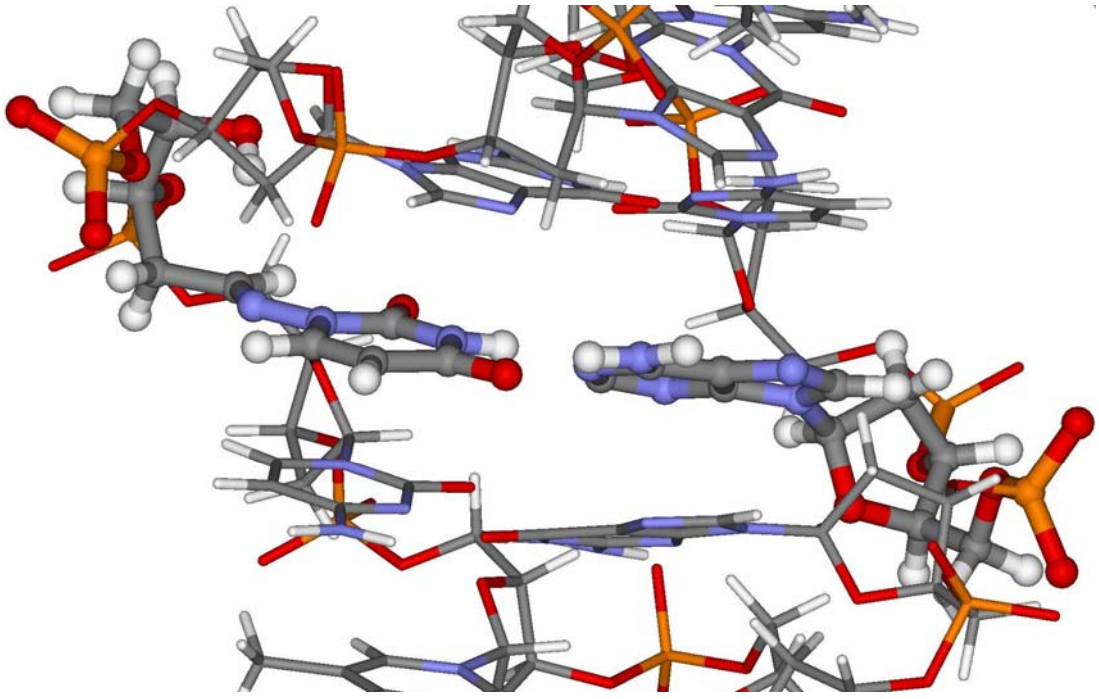


Figure 3.19. Energy minimization of aminouracil 3 at the abasic site position. The base is in its hydrazone form, *trans* configuration. Repair unit 3 and the opposite adenine are tilted at an angle of about 11 degrees. The O(=C^4) of 7 and the N(H_2) of the adenine are at a distance of 2.75 Å.

3.3.3.2 Uracil mimic 7

A stable duplex was also found for the hemi-aminal form of the reaction product of thymine mimic **7** and the abasic site. The deoxyribose unit is moved into the major groove to keep the base in a good position for hydrogen bonding. The amide bond of the base mimic is in a *trans*-configuration in Figure 3.20, but a *cis*-configuration is also well accommodated. The amide is tilted out of the plane of the base, presumably because of repulsions between the amide C=O and the uracil C^4=O. The trans conformation of the amide allows favorable hydrogen bonding between the C=O of the amide and the NH in the linker. The NH of the amide shows a hydrogen bond with one of the oxygen atoms of the 3'phosphate.

The sugar is in its α-form, whereas in a B-helix it is normally in its β configuration. To reach this epimer of the sugar, we first created the hemi-aminal without the proton on the $C^{1'}$ carbon. We then calculated an energy minimum and subsequently added a proton at the most favorable position, which turned out to be the α-form of the sugar. Finally, we calculated the energy minimum including the proton. The α-conformation of the repair unit in a duplex is thus more stable than the α conformation. The thymine mimic **7** is slightly tilted with respect to its complementary adenine, but this is also observed in natural base pairs.

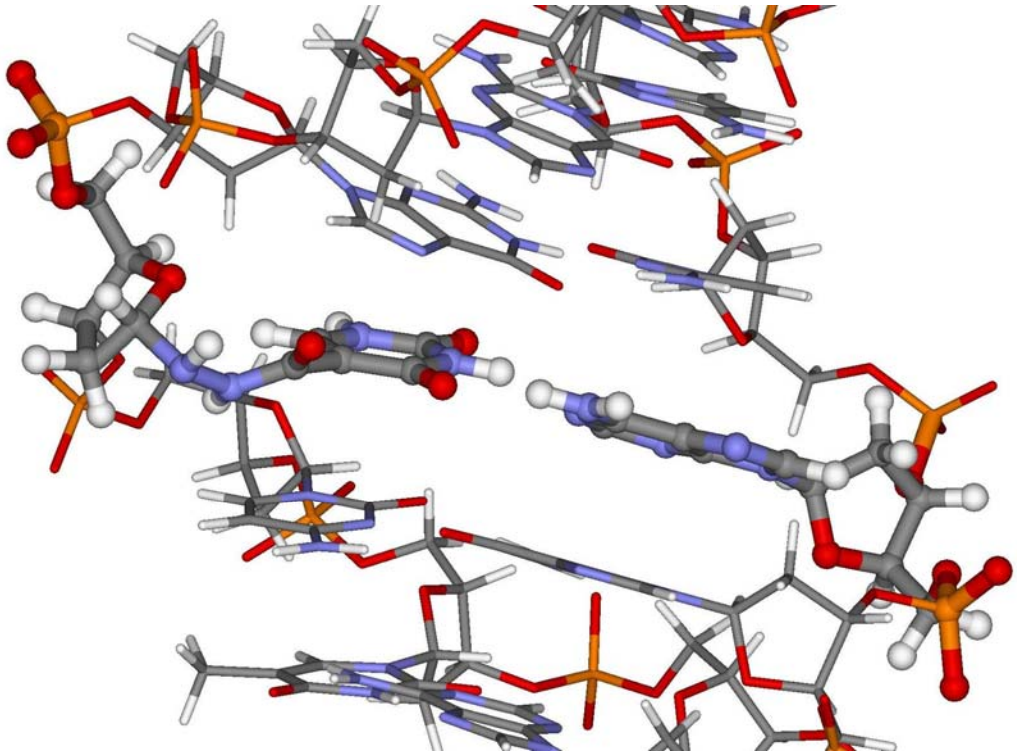


Figure 3.20. Energy minimization of thymine mimic 7 at the abasic site position in its hemi-aminal form (α-epimer). The linker is located in the major groove and the amide is in a *trans* configuration. Compound 7 and the opposite adenine are tilted at an angle of about 25 degrees. The O(=C^4) of 7 and the N(H_2) of the adenine are at a distance of 2.92 Å.

Besides the hemi-aminal form, the repaired abasic site may also exist as a hydrazone. The hydrazone has a more flexible linker and should be able to more easily find a stable conformation.

Since the hydrogen bonding pattern of uracil (or thymine) is symmetrical (CO-NH-CO), **7** may in principle be flipped around itself and still hydrogen bond correctly. This should provide two different ways of linking the base to the DNA backbone. If the base mimic is present as a hemi-aminal, the linker between backbone and base is not flexible enough to accommodate both possibilities of linking. Only the hemi-aminal structure shown in Figure 3.20 with the linker in the major groove was found to be stable. If the base mimic is present as a hydrazone, both the structures with linker in the major or the minor groove are stable.

In the first hydrazone we modeled (Figure 3.21) the linker was positioned in the major groove. In the second hydrazone (Figure 3.22) the linker was positioned in the minor groove. In both cases we were able to find a position in which both the amide bond and the hydrazone are *trans*. The *cis*-amide and *cis*-hydrazone were also modeled and a stable structure was found for both.

The structures shown in Figure 3.21 and Figure 3.22 differ from the position of the linker in the major or minor groove, but also from the conformation of the amide bond of the linker. In Figure 3.21 one can clearly observe that the amide bond is not coplanar with the uracil base. This is presumably caused by electronic repulsion between the C=O of the amide and the C^4=O of the uracil. The carbonyl pointing upwards apparently disturbs π–π-stacking in the double helix. The guanine that is positioned above the modified base in Figure 3.21 is at a larger distance compared to a natural base.

In Figure 3.22 one can observe that the amide bond is coplanar with the uracil base. There is no repulsion of the C=O of the amide and the C^4=O of the uracil. Instead, a hydrogen bond is observed between the NH of the amide and the $C^{4'}$=O of the uracil. This favors the coplanarity and increases π–π-stacking in the double helix.

Since the two structures are different conformations of the same compound, we can compare the formation energies of the two structures. The structure depicted in Figure 3.21 with the linker in the major groove has an energy of formation of –15 347.2 kJ/mol. Having the linker positioned in the minor groove results in a slightly more favorable energy of formation of -15 355.8 kJ/mol. The conformation with the linker in the minor groove is thus favored, presumably because of the favorable conformation of the amide bond.

In both cases we find a position that resembles a position adopted by natural bases. We can thus expect that a favorable conformation will be found in a practical experiment.

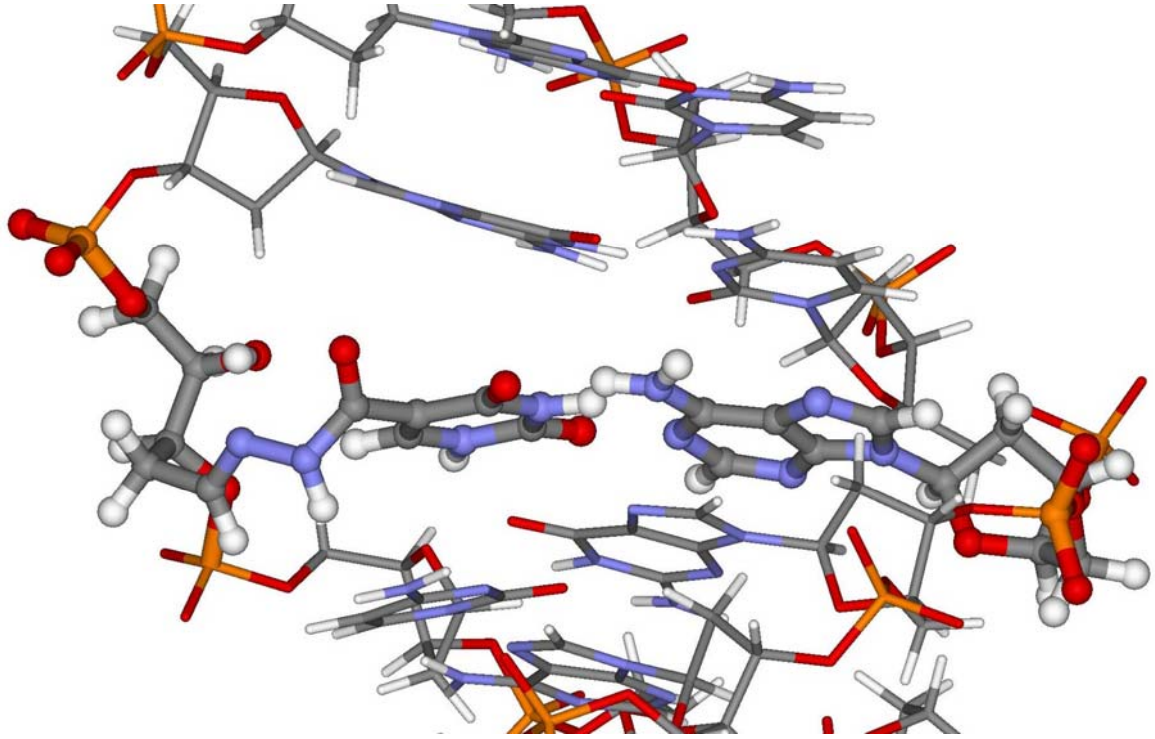


Figure 3.21. Energy minimization of thymine mimic 7 at the abasic site position in its hydrazone form. The linker is located in the major groove and the amide is in a *trans* configuration. Compound 7 and the opposite adenine are tilted at an angle of about 17 degrees. The O(=C^4) of 7 and the N(H_2) of the adenine are at a distance of 3.02 Å. An additional hydrogen bond is observed between the $^{4'}$OH and the nitrogen atom of the hydrazone.

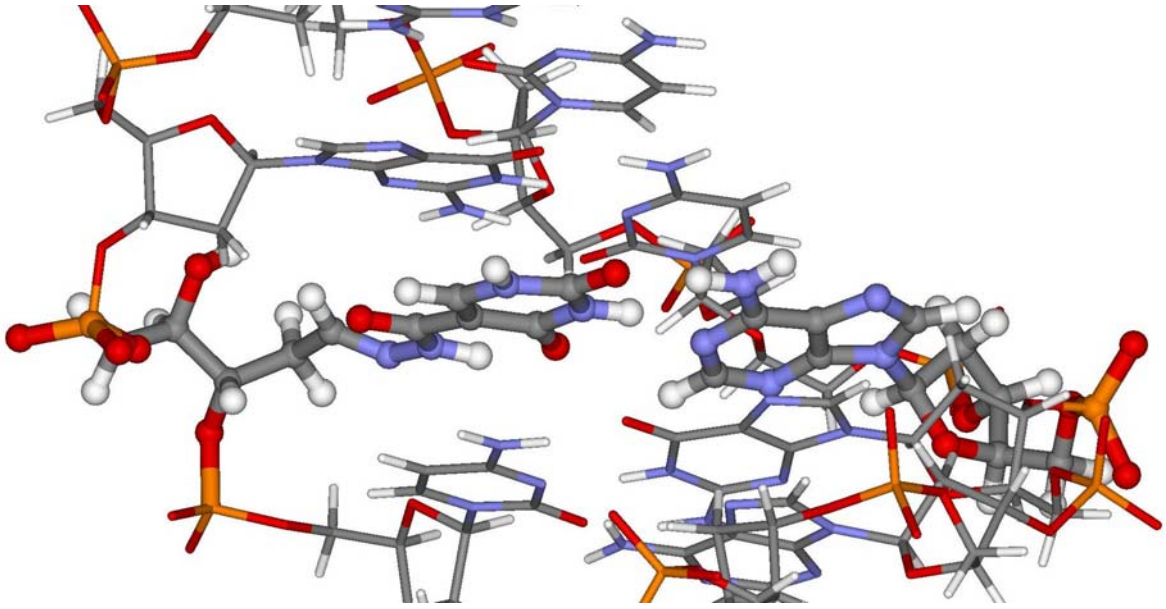


Figure 3.22. Energy minimization of thymine mimic 7 at the abasic site position in its hydrazone form. The linker is located in the minor groove with the amide coplanar to the base in its *trans* configuration. Compound 7 and the opposite adenine are tilted at an angle of about 11 degrees. The O(=C^4) of 7 and the N(H_2) of the adenine are at a distance of 2.91 Å. An additional hydrogen bond is observed between the NH of the amide and the O=C^4 of the uracil.

3.3.3.3 Universal base mimic 8

As a starting point for energy minimization of the difluorobenzene derivative **8** we took the minimized structures of **7**. We replaced the oxygen atoms in the uracil moiety by fluor atoms and the nitrogen atoms by carbon atoms and added or removed double bonds where necessary. For the two hydrazones and the hemi-aminal we saw the same thing happening, the base moved backwards in the direction of the backbone (Figure 3.23 - Figure 3.25). The force field does not recognize a hydrogen bond or any type of attraction between the C-H of **8** and the lone pair of the nitrogen of the opposite adenine. In the hemi-aminal form (Figure 3.23) and in the hydrazone with the linker in the major groove (Figure 3.24) there is a hydrogen bond present between the fluor atom and the NH_2 of the adenine. In the second hydrazone, when the linker is in the minor groove, (Figure 3.25) even this hydrogen bond is not observed. After starting energy minimization, the C-H is immediately flipped out of the plane, followed by the retreat of the base to the position it has in Figure 3.23 - Figure 3.25. This happened in all the three cases in the same way. The amide bonds in the final structures are of a similar conformation compared to the structures with uracil mimic **7**. In the *trans* configuration with the linker in the major groove, the amide is not coplanar with the phenyl ring, but pointing upwards towards the guanine above (Figure 3.22 and Figure 3.23). This may hinder π–π-stacking.

It must be mentioned here that AMBER* is known to enhance hydrogen bonding. Thymine may thus be more weakly bound than can be expected from these models. In the difluorobenzene derivative AMBER* does not recognize the N···H-C attraction, this compound may thus be bound stronger than expected from the minimized structures. We tried energy minimization using another force field less biased towards hydrogen bonding to test these hypotheses. However, the MM3 force field we tried did not recognize any hydrogen bond in the double helix, resulting in its complete dissociation. MM3 is thus not appropriate for modeling DNA double helices.

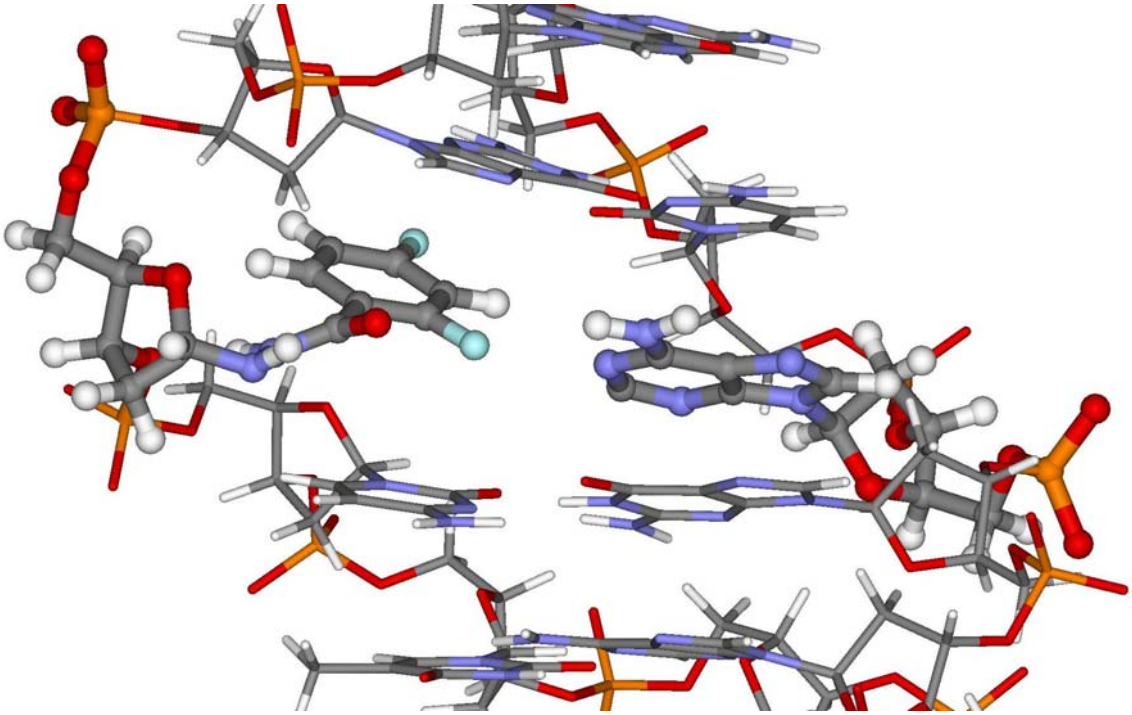


Figure 3.23. Energy minimization of universal base mimic 8 at the abasic site in its hemi-aminal form. The linker is located in the major groove and the amide is in its *trans* configuration. The F^2 of 8 and the $N(H_2)$ of the adenine are at a distance of 3.39 Å, which suggests a hydrogen bond is observed between F^2 and the NH_2 of the adenine.

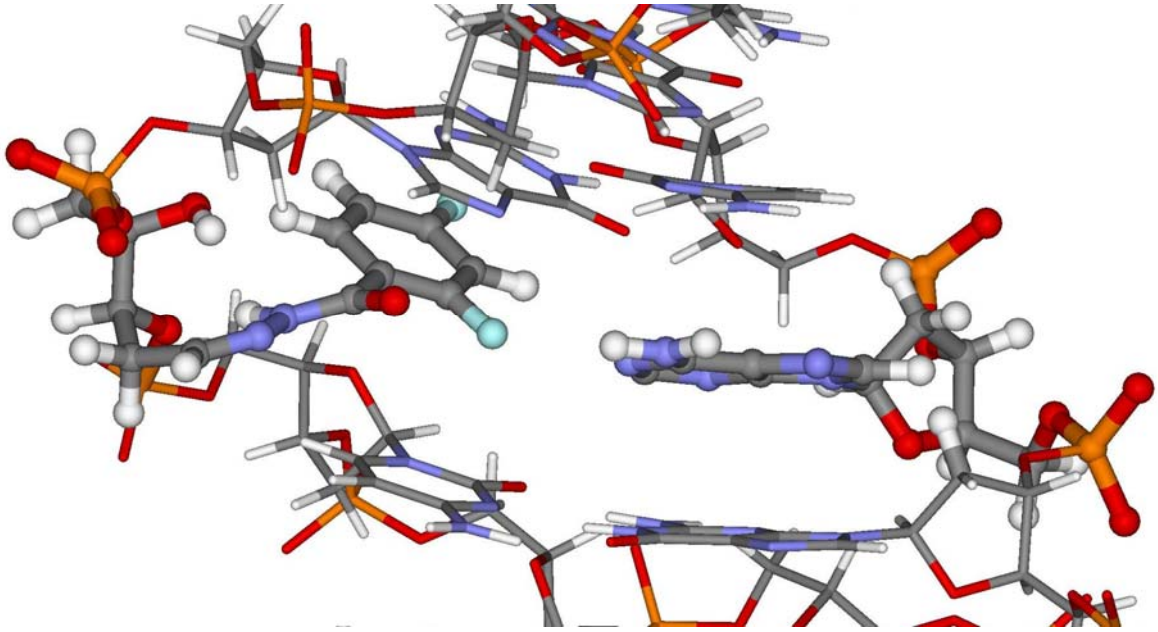


Figure 3.24. Energy minimization of universal base mimic 8 at the abasic site position in its hydrazone form. The linker is located in the major groove and the amide is in its *trans* configuration. The F^2 of 8 and the $N(H_2)$ of the adenine are at a distance of 3.34 Å, which is 0.3 Å larger compared to thymine mimic 7 (Figure 3.21). A hydrogen bond is possible between the F^2 and the NH_2 of the adenine and four additional hydrogen bonds are observed between $H(O^{4'})$ and N(=C), $O^{4'}$(H) and H(N) of the amide, and between the H(N) of the amide and $O^{3'}$.

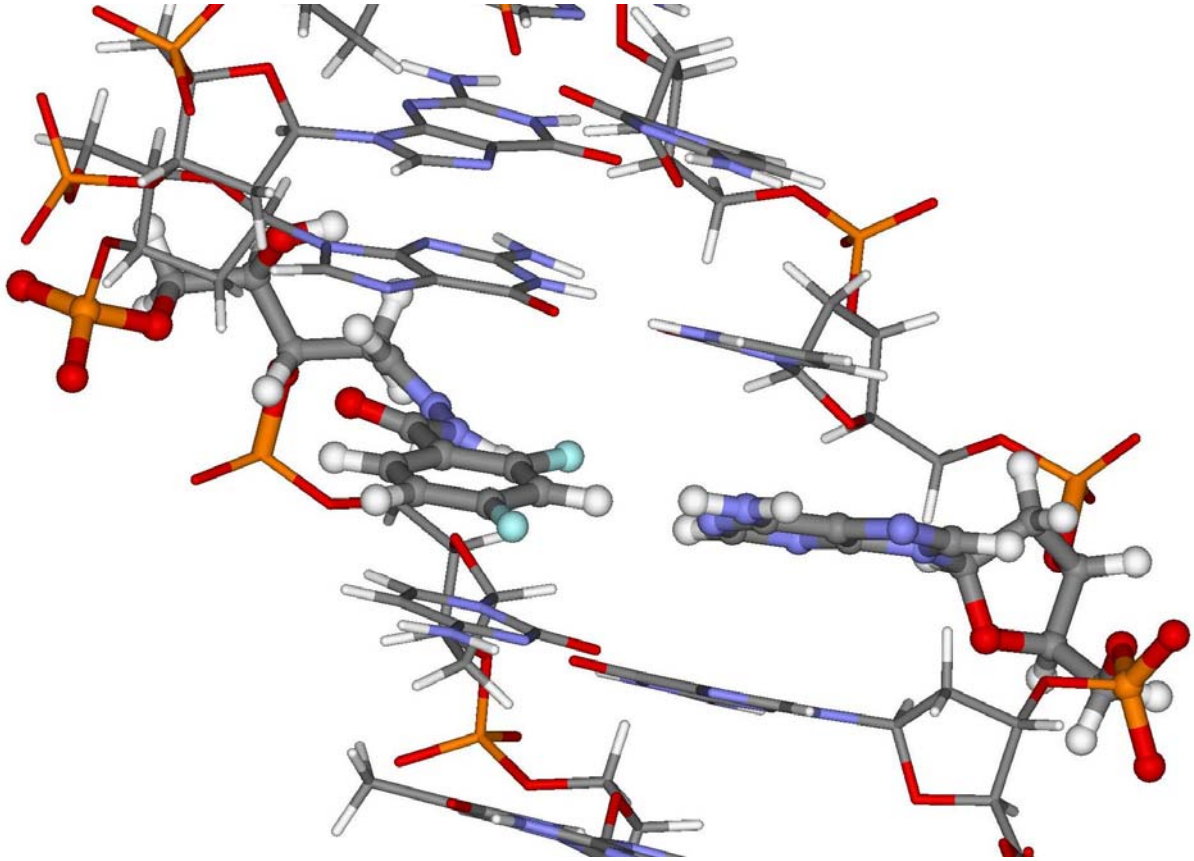


Figure 3.25. Energy minimization of universal base mimic 8 at the abasic site position in its hydrazone form. The linker is located in the minor groove with the amide bond in its *trans* configuration. The base mimic is moved towards the major groove. The F^4 of 8 and the $N(H_2)$ of the adenine are at a distance of 4.48 Å, which is about 50% more than observed in the similar thymine mimic hydrazone (Figure 3.22). No hydrogen bond is observed. An additional hydrogen bond is observed between the F^2 and the NH of the amide.

From these modeling experiments can be concluded that a stable conformation can be found for the thymine mimics **3** and **7**. The atoms inserted between the base and the backbone seem not to be problematic, a good position for hydrogen bonding and π–π-stacking can be found. Universal base mimic **8** moves backwards during the energy minimization, apparently because of unfavorable contacts between the aromatic protons and the lone pair of the N^1 of the opposite adenine.

To further test whether this was caused by the strain in the linker or because some interactions are not well described by the AMBER* force field, we tried modeling 2,4-difluorobenzene at exactly the same position as a natural thymine. This unit was designed by Kool and coworkers to be a good steric replacement for thymine [107, 108] and should fit well in a DNA double helix opposite an adenine. However, also when the unit designed by Kool was modeled at the place of a thymine, the difluorobenzene unit showed the same behavior. First the CH between

the fluor atoms is flipped out of the plane, then the difluorobenzene moiety moves in the direction of the backbone, away from the complementary adenine. The structures found for the difluorobenzene unit may thus be better than can be predicted from these models.
Aminobases **4-6** were not modeled, but we expect them to have a similar behavior as **3** opposite their respective complementary bases.

3.4 Synthesis of aminobases and hydrazides

The syntheses of the four aminobases **3-6** (Figure 3.14) were performed as shown in Figure 3.26 and Figure 3.27, according to literature procedures [109-117].
The strategy for synthesizing 9-aminoadenine (**4**) was to incorporate the hydrazine via nucleophilic substitution. Thus, 9-Aminoadenine was synthesized in five steps from commercial 5-amino-4,6-dichloropyrimidine (Figure 3.26). In a first step, hydrazine monohydrate was added to 5-amino-4,6-dichloropyrimidine give **9** in 71% yield by selective nucleophilic substitution of one chlorine atom. The yield is significantly higher than the yield reported by Montgomery *et al.* (40%) [117].
Selective condensation of the hydrazine in the presence of an aromatic amine with benzaldehyde gave compound **10** with a yield similar to the one obtained by Montgomery *et al*. (44%). The pure product was obtained after purification performing column chromatography. Refluxing **10** in diethoxymethyl acetate for 1h yielded **11** in high yield (90%), which was a bit better than the 75% obtained by Harnden *et al.* [109]. Treatment of **11** with ammonia gave **12**. The yield of 75% was a bit lower compared to Harnden *et al.*
Finally the protecting benzylidene group was removed by refluxing **12** in $CHCl_3$/MeOH 2/1 with methylhydrazine. The deprotection reflects the better nucleophilicity of methyl hydrazine versus aminoadenine. Precipitation and filtration of 9-aminoadenine, followed by washing with chloroform and diethyl ether gave the pure product. The yield of the reaction was only 43%, which is half the yield obtained by Harnden *et al*. Part of the product might have been lost due to too extensive washing.

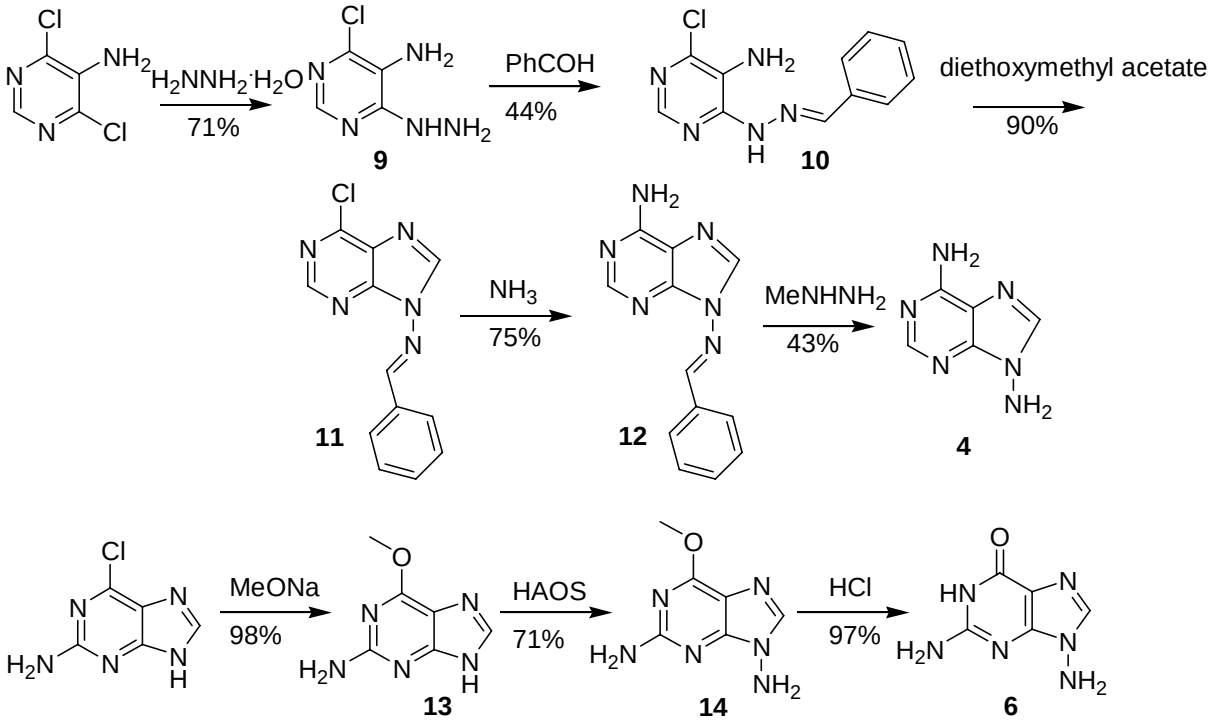


Figure 3.26. Aminopurine synthesis.

Introduction of the hydrazine in guanine was performed by selective electrophilic amination. The synthesis of 9-aminoguanine starts by treating commercial 2-Amino-6-chloropurine with sodium methoxide to give **13** in high yield [116]. This compound could be transformed in **14** using hydroxylamine-*O*-sulfonic acid (HAOS) under strongly basic conditions (NaOH, 4M). The nucleophilic substitution preferentially takes place at N^9, but some N^7 substitution equally takes place. Separation of the two compounds can be done by column chromatography using $CHCl_3$/MeOH/NH_4OH 95/4.5/0.5 as an eluent to give **14** in 71% yield. Finally, 9-aminoguanine **6** was obtained by deprotection of the 2-hydroxylgroup using a 1M HCl solution.

Synthesis of aminothymine has not been described in the literature, but aminouracil has. Basepairing and π–π stacking for the two units is similar, so it was decided to use aminouracil instead of aminothymine in DNA repair.

The syntheses of aminouracil and aminocytosine both start with the same compound, uracil (Figure 3.27). Uracil was disilylated using hexamethyldisilazane in the presence of ammonium sulfate [110]. Selective electrophilic amination of **15** by reacting it with *O*-mesitylenesulfonylhydroxylamine (MSH), followed by the addition of benzaldehyde yielded **16** in a 62% yield. The literature procedure [110] does not pass through the benzylidene intermediate, but proceeds to 1-aminouracil directly from **15**. However, separation of the two

products of that reaction, aminouracil **3** and uracil, is difficult. Passing through the benzylidene intermediate allows chromatography after this step and assures the purity of the aminouracil, obtained by addition of methyl hydrazine to **16** in a final step.

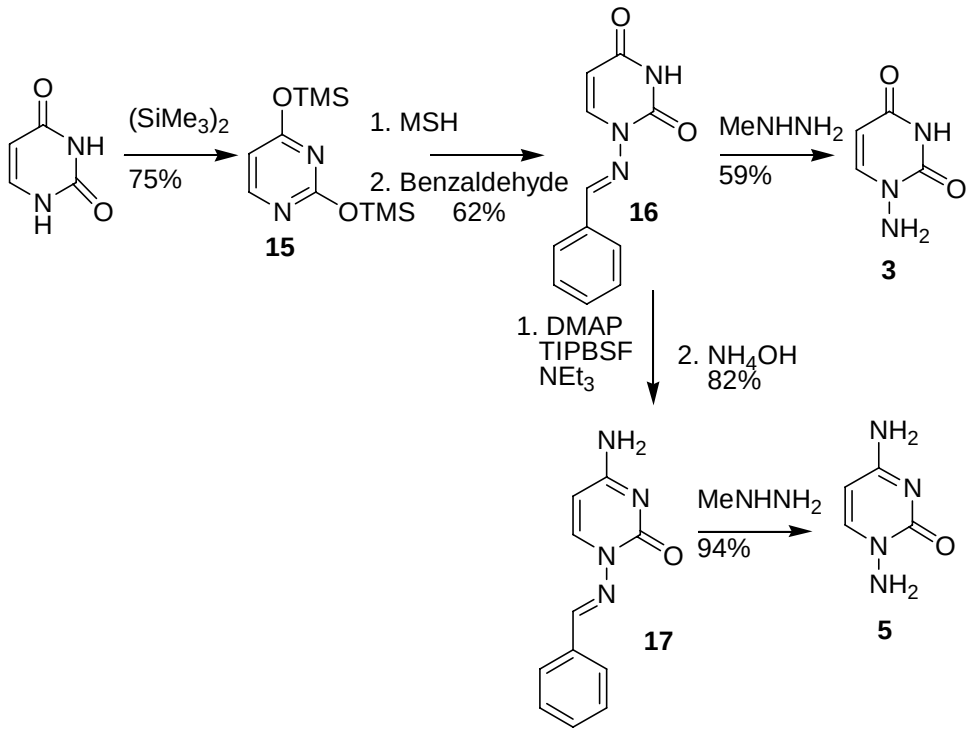


Figure 3.27. Aminopyrimidine synthesis.

The synthesis of aminocytosine starts from the benzylidene intermediate **16** used in the aminouracil synthesis (Figure 3.27). The 4-carbonyl group was selectively transformed in an amino group using 4-dimethylaminopyridine and triisopropylbenzenesulfonyl fluoride activation under basic conditions, and subsequent quenching of the reaction with ammonia to give **17** in 82% yield. This is high in comparison to the 25% obtained by Harnden *et al.*[109], but a bit lower than literature procedures performing similar transformations on other substrates [111-113]. Finally, 1-aminocytosine **5** was obtained from **17** by treatment with methyl hydrazine in 94% yield. This is significantly higher than the yields obtained e.g. during the deprotection of 1-aminouracil and 9-aminoadenine. Satisfying quantities (54 mg of 9-aminoadenine and 35 mg of 1-aminouracil) were obtained and the reactions were not optimized.

The synthesis of hydrazides **7** and **8** was started from their corresponding acids **18** and **20** respectively, which are both commercially available (Figure 3.28) [118-121]. Isoorotic acid (**18**) was esterified with decanol using DCC activation and DMAP as a catalyst. Compound **19** was obtained in a moderate yield (50%) by crystallization from MeOH/water 90/10.

Hydrazine hydrate was then added to cleave the ester and give hydrazide **7** in 44% yield. The universal base mimic **8** was prepared by a coworker, with similar yields compared to compound **7**.

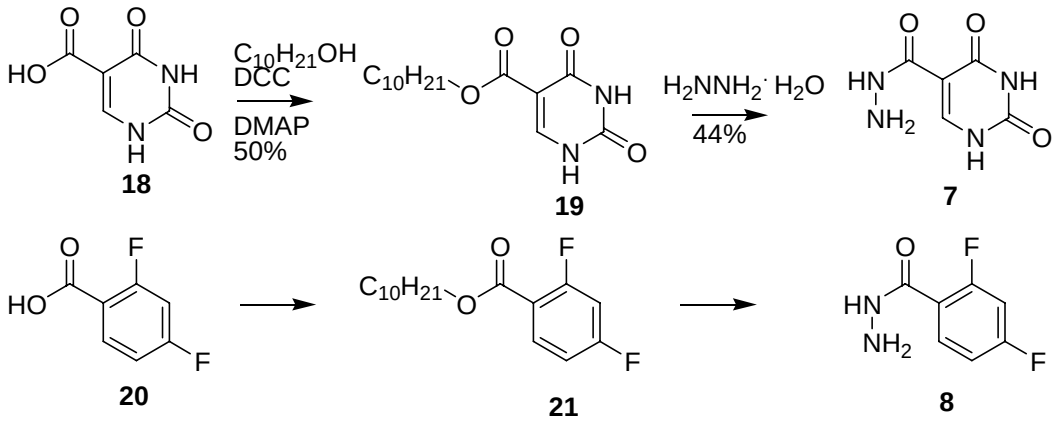


Figure 3.28. Synthesis of hydrazides 7 and 8.

3.5 Model reactions with ribose and deoxyribose

3.5.1 Test reactions with ribose

Before incorporating the abasic site in a DNA strand, we want to optimize the reaction conditions. We want to find conditions under which ribose or deoxyribose is reactive towards hydrazides and hydrolysis of the products can take place. Either ribose or deoxyribose can be used, they simulate RNA and DNA repair respectively. The 2' hydroxyl group of ribose will influence the reactivity of the aldehyde and the reactivity of ribose and deoxyribose will not be the same.

We know that the hemiacetal-aldehyde equilibrium lies on the hemiacetal side, the aldehyde is thus little available for reaction. However, after formation of the imine, the imine can form a hemiaminal. This imine-hemiaminal equilibrium looks similar to the aldehyde-hemiacetal equilibrium. We therefore speculated that the hemiaminal might be as stable as the hemiacetal. This might pull the reaction to completion (Figure 3.29).

Test reactions of 1-aminouracil with ribose and deoxyribose (H_2O, r.t., 12.5 mM) were performed by a coworker [122, 123] to see if this hypothesis was correct. Unfortunately, no reaction between the hydrazines and the sugars was observed. The hemiaminal form is thus not as favored as expected. We therefore started test experiments with hydrazides **7** and **8**, which are stronger nucleophiles and we expected them to react better with ribose. These test experiments are described in detail in the next paragraphs.

The goal of these test reactions is to learn which products are formed under which reaction conditions and obtain information about the kinetics of both the nucleophilic substitution and the hydrolysis of the products. It is likely that the products are in equilibrium with each other, but the time scale of this equilibrium cannot be predicted beforehand.

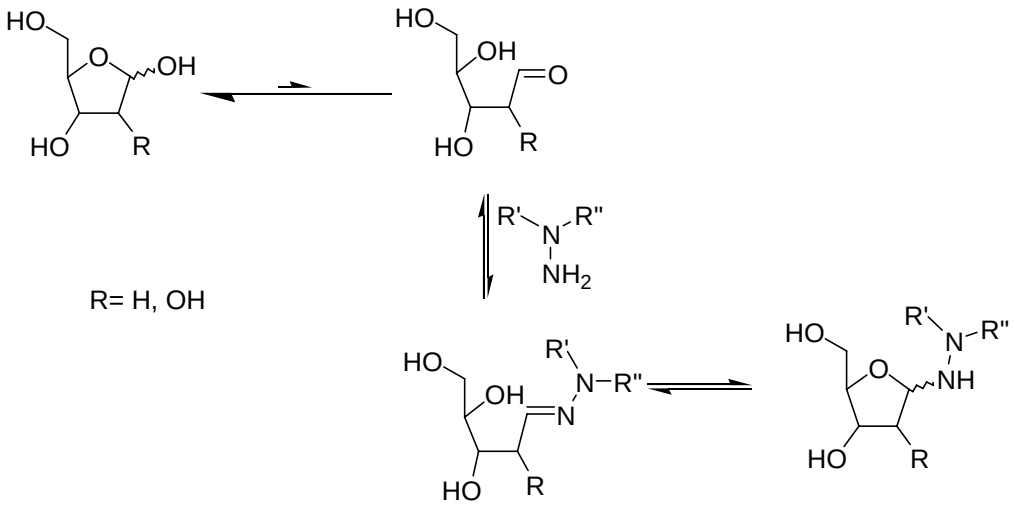


Figure 3.29. Equilibria between closed and open (deoxy)ribose, imine and hemiaminal.

3.5.1.1 Test reactions of thymine mimic 7 with ribose

The first test experiments involved mixing of ribose with the thymine mimic **7**. Some products that can be expected from this reaction are drawn in Figure 3.30. Apart from the obvious imine, cyclization can lead to the hemiaminal, and the formation of the imine hydrate is possible by addition of a water molecule. Furthermore, cyclization from the $^{5'}$hydroxy-group is possible to give the pyranose form of the sugar. Two stereoisomers of the hemiaminal, the imine hydrate and the pyranose form can be expected, as well as the *E*- and *Z*- product of the imine. It is possible that not all compounds can be observed; the proportion of some of the compounds may be very small.

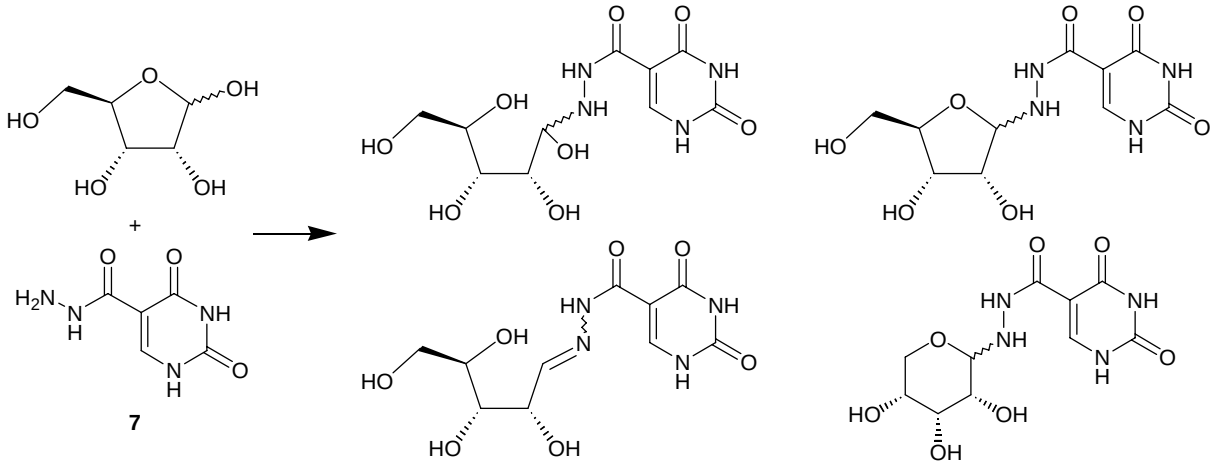


Figure 3.30. Expected products upon mixing of ribose with the thymine mimic. The amide bonds in the products are in reality in a *trans* conformation, but are depicted *cis* here to maintain the ribose and the uracil in conventional orientation.

The ^{1}H NMR spectrum of the thymine mimic in D_2O shows only one peak, at 8.26 ppm (Figure 3.31, spectrum b), corresponding to the H^6. The other protons are exchangeable and, since the experiments are run in D_2O, exchanged for deuterons. Because of residual H_2O in the D_2O, NMR spectra were recorded using a pre-saturation sequence. Upon mixing the thymine mimic with ribose (1.5 eq), nothing happened, but after seven days a new peak appears between 8.5 and 7.5 ppm.

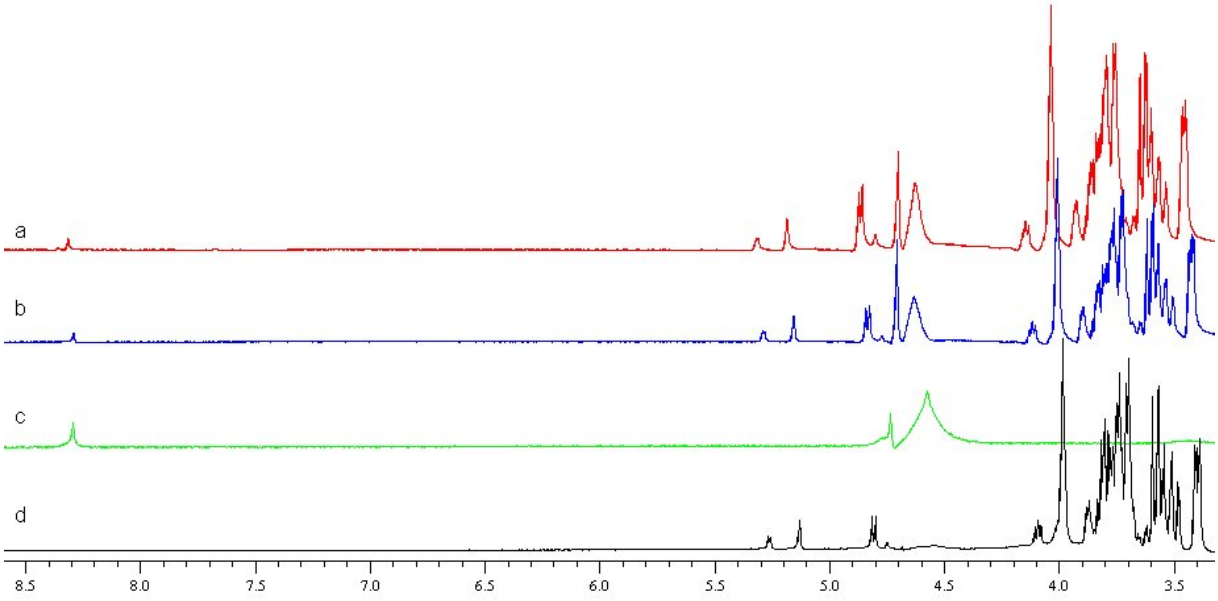


Figure 3.31. ^{1}H NMR spectra at r.t. in D_2O of a) ribose (1.5 mM) + thymine mimic (1.0 mM) after 14 days, b) ribose (1.5 mM) + thymine mimic (1.0 mM) after 7 days, c) thymine mimic (1.0 mM), d) ribose (1.5 mM).

Magnification of the region between 8.5 and 7.5 ppm shows the new peak at 8.30 ppm (Figure 3.32). The peak is very small though, only 15 % compared to the starting thymine mimic. Between 5.5 and 3.0 ppm, where the signals of the ribose protons are observed, no changes can be observed. If there are any new peaks appearing in this region they cannot be distinguished due to the large peaks of the ribose.

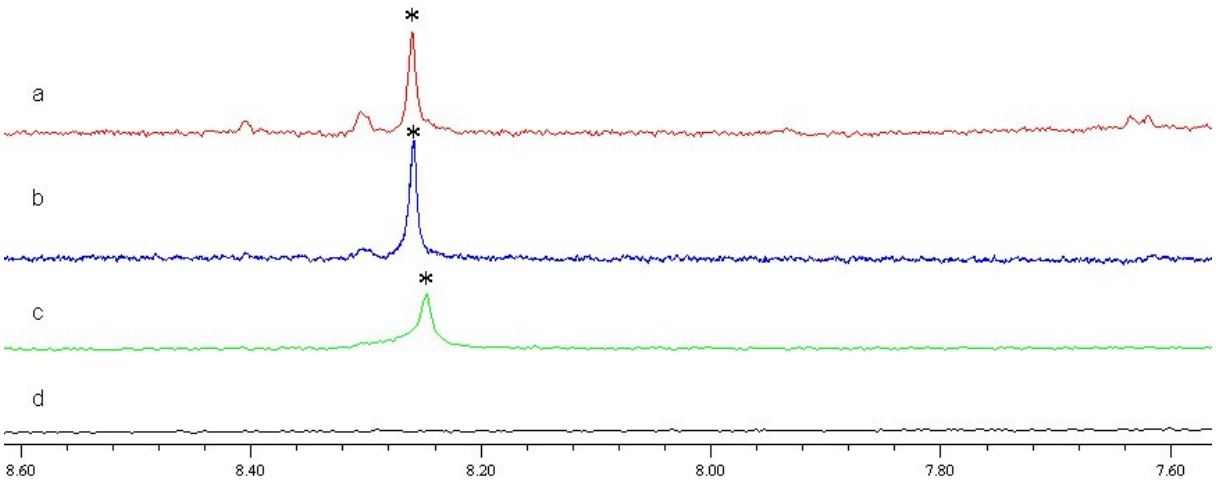


Figure 3.32. Magnification of the ^{1}H NMR spectra at r.t. in D_2O of a) ribose (1.5 mM) + thymine mimic (1.0 mM) after 14 days, b) ribose (1.5 mM) + thymine mimic (1.0 mM) after 7 days, c) thymine mimic (1.0 mM), d ribose (1.5 mM). * indicates signals of starting 7.

The reaction is not finished after seven days. After 14 days the peak at 8.30 has increased in intensity, it is now at about a third of the intensity of the starting material. Three more peaks now clearly arise, at 8.41, 7.64 and 7.62 ppm. With the peak at 8.30 ppm that makes a total of four products.

From these spectra, conclusions could of course not be drawn about which of the proposed structures the reaction products could correspond to. We therefore tried whether ^{13}C NMR could tell us more. If there are any hydrazones present in the reaction mixture, their signals are expected around 150 ppm, where no reactant peaks are observed. Unfortunately ^{13}C NMR spectra of the same mixtures did not yield any subsequent information. The expected peaks around 150 ppm for the hydrazones were not observed. This does not imply their absence though. Response of a tertiary ^{13}C is low and the concentration of the products is low as well. The formation of hemiaminals is another possible explanation, their signals are not expected around 150 ppm.

To increase the amount of product, conditions have to be changed. To favor the reaction, we tried decreasing the pH to 6. This slightly increased the reaction rate, but did not influence the final outcome of the reaction. Little product was formed (data not shown).

Increasing temperature was more successful. As can be seen from Figure 3.33, after 96h at 90°C three new peaks appeared at 8.34, 8.28 and 7.92 ppm, that may correspond to three products (all NMR spectra were recorded at room temperature). The starting material disappeared completely over time. Only the peak at 8.34 ppm is observed in both the spectrum at 25°C and 90°C. Heating the sample thus leads to other products compared to the experiment at r.t..

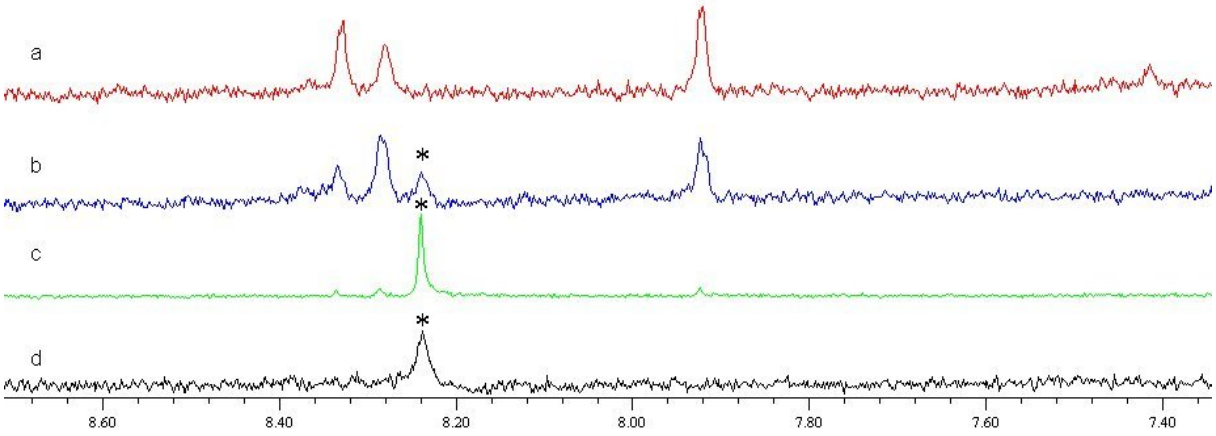


Figure 3.33. ^{1}H NMR spectra at r.t. in D_2O with a NaH_2PO_4/Na_2HPO_4 buffer at pH 6 of a mixture of ribose (1.5mM) and thymine mimic (1.0mM) a) after 96h at 90°C, b) after 24h at 90°C, c) after 3h at 90°C, d) just after mixing. * indicates signals of 7.

^{13}C NMR spectra unfortunately could not tell more about the nature of the products. A conclusion about which products were formed could not be drawn from these data. Nevertheless, we decided to perform similar tests with the universal base mimic **8**, hoping that it would work better and give a better view of the nature of the different products.

3.5.1.2 Universal base mimic with ribose

Upon mixing ribose with the fluorinated universal base mimic, four products similar to the thymine mimic analogues can be expected (Figure 3.34). Each of them may exist as a mixture of two isomers.

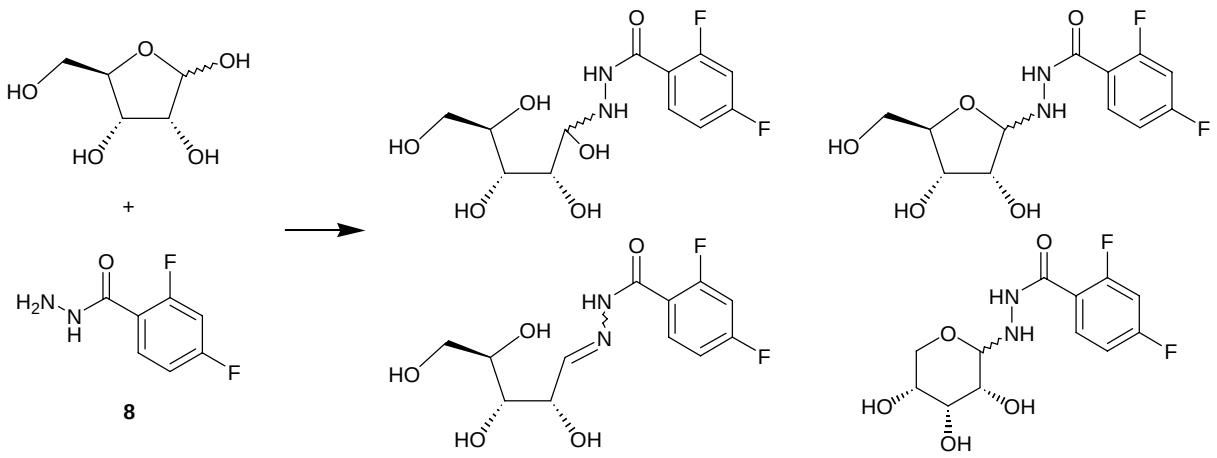


Figure 3.34. Possible products upon mixing ribose and the fluorinated universal base mimic. The amide bonds in the products are in reality in a *trans* conformation, but are depicted *cis* here to maintain the ribose and the uracil in conventional orientation.

Since the reaction between the thymine mimic and the ribose yielded some reaction products at pH 6 at 90°C, we let the universal base mimic and the ribose react at pH 6 at elevated temperature, 60°C and at a higher concentration. All NMR spectra were again recorded at r.t. As can be seen from Figure 3.35, the reaction proceeds over seven days, the spectrum after nine days (Figure 3.35a, purple) is almost identical to the one measured after seven days (Figure 3.35b, red). Changes at the ribose part of the NMR spectrum are hardly visible due to the multitude of peaks in that region, but new peaks clearly appear around 5.15 ppm and in the aromatic region. The multiplicity of the peaks prevents conclusion about the nature of the products or even the number of products. It can just be concluded that the starting material does not disappear completely in this case. ^{13}C NMR did not give any more information.

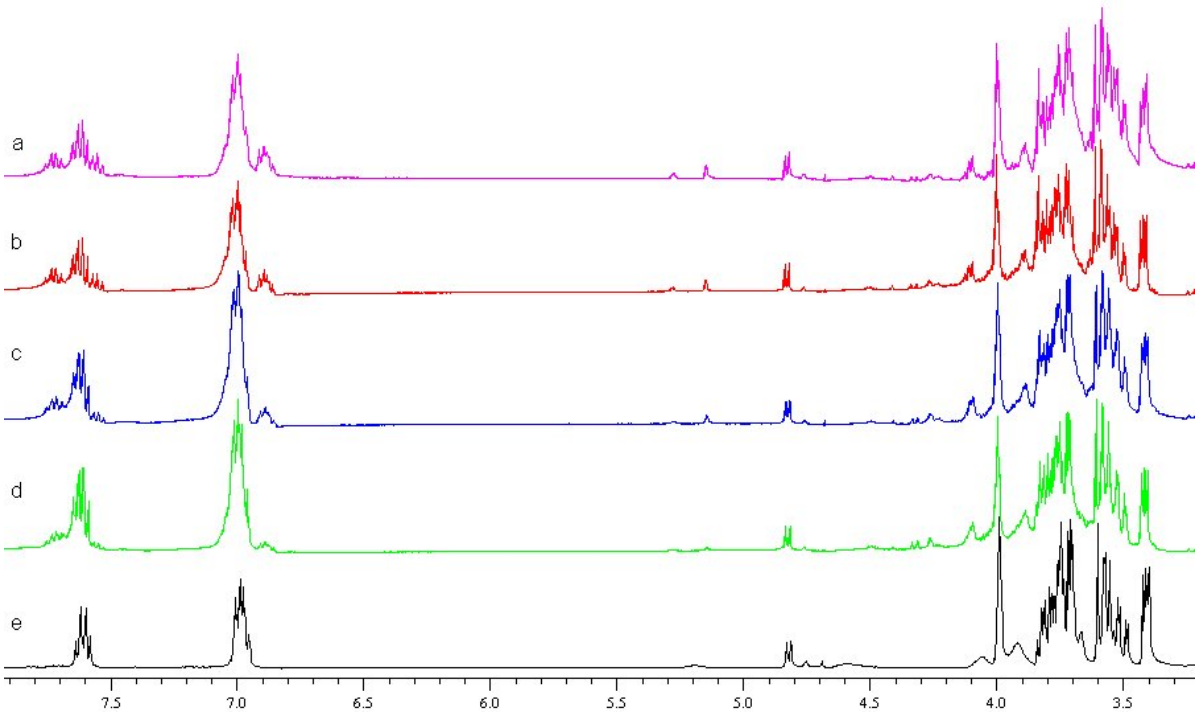


Figure 3.35. ^{1}H NMR spectra at 60°C in D_2O of a mixture of ribose (30mM) and universal base mimic (30mM) a) after 9 days, b) after 7 days, c) after 3 days, d) after 1 day, e) just after mixing.

The multiplicity of the peaks is caused by coupling of the protons with the fluorine atoms. These are also magnetically active. Fluorine-proton decoupling turned out to be impossible with the available NMR machine, but ^{19}F-NMR spectroscopy gave us an extra tool to analyze our products.

The ^{19}F NMR spectra of the reaction mixture after 96h at 60°C and of the starting material are depicted in Figure 3.36. It can be seen that various products are formed. Two singlets are observed in the spectrum, at -76 ppm and -123 ppm. In the region between -103 and -112 ppm six multiplets are observed. Two of them correspond to the starting material, the four other peaks correspond to two reaction products. The multiplicity arises from J-coupling with the other fluor atom and with the aromatic protons, but as mentioned before, proton-fluor decoupling experiments are impossible with our NMR spectrometer.

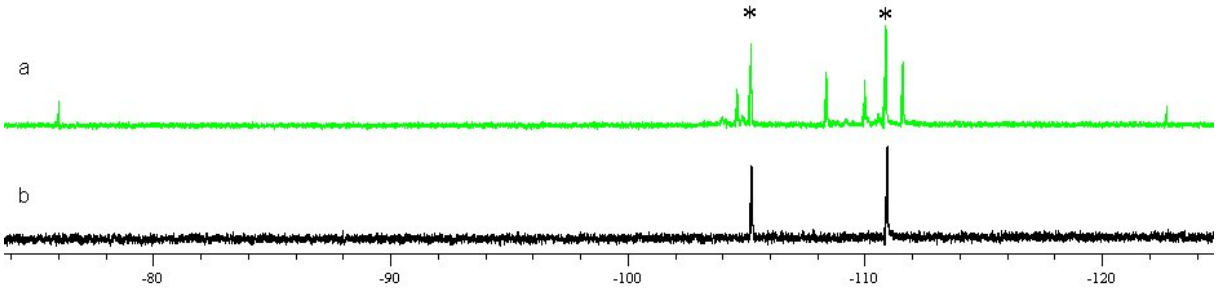


Figure 3.36. ^{19}F NMR spectra in D_2O of a) mixture of ribose (30 mM) and universal base (30 mM) after 4 days at 60°C, b) free universal base mimic. * indicates signals of 8.

Since all our fluor atoms should be coupled with protons, the singlets at -76 ppm and -123 ppm cannot belong to products that we predicted (Figure 3.34). The singlet at -76 ppm could not be attributed, but the singlet at -123 ppm corresponds to free F^-. This means that degradation takes place during the reaction. Nucleophilic aromatic substitution of a fluorine atom by water can give compounds like **22** and **23** as well as F^- (Figure 3.37). Of the four product peaks observed in Figure 3.36, it may thus be possible that one or two correspond to compounds **22** and **23** and are thus side reaction product signals instead of main reaction product peaks. Milder reaction conditions have to be chosen to avoid these species.

To check whether the observed reactions are really due to reaction between ribose and **8**, we ran two control experiments in which either ribose or **8** was missing. In that case none of the new signals of Figure 3.36 appeared. Not the multiplets, but no signal of the F^- was observed as well. That means that nucleophilic substitution by water is unlikely to be the reason for the degradation of the products, since in that case **8** should degrade also in the absence of ribose. Intramolecular substitution by the hydroxyl groups of the sugars or by the NH of the imine may be the cause for the degradation (Figure 3.37). Nucleophilic attack of the NH of the imine gives a five-membered ring (compound **25**). Compound **24** can be obtained after attack of the $^{2'}$OH on the *o*-fluor atom. The 8-membered ring that should be the product of the reaction is not very likely to form.

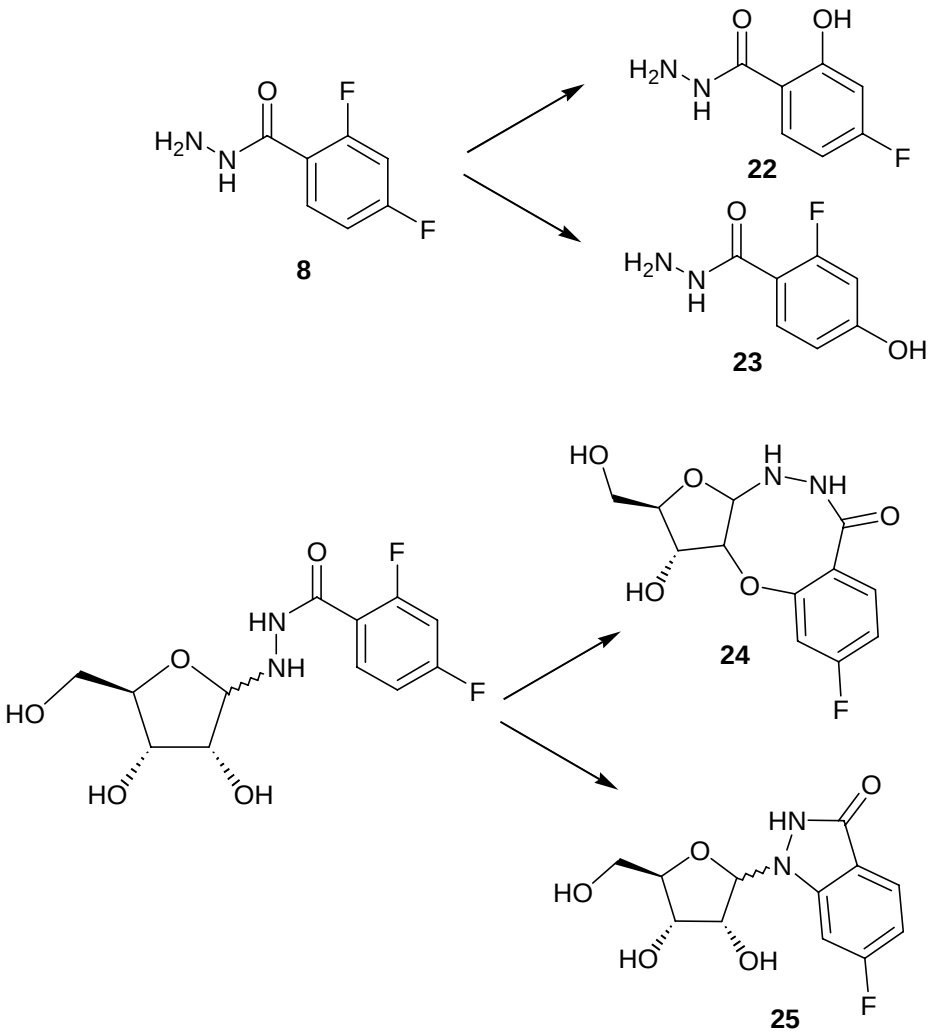


Figure 3.37. Nucleophilic fluorine substitution.

To limit degradation, we decreased the temperature. Experiments at 37°C showed significantly less F^- formation, but even at 37°C side reactions occur as shown by F^- formation. It was therefore decided to perform the experiments at room temperature. To accelerate the reaction and to obtain a better ratio of product over starting material in the ^{19}F spectrum, the amount of ribose was increased. The excess ribose is not visible in the ^{19}F NMR spectrum and the analysis is just as accurate.

Four new reaction conditions were tested at room temperature (Figure 3.38). The reaction was either carried out in pure D_2O, or the pD was decreased to 6 by using a Na_2DPO_4/NaD_2PO_4 buffer. Both at pD 6 and pD 7 two or twenty equivalents of ribose were added relative to the universal base mimic. As can be seen from Figure 3.38d (black), after 3 days at room temperature, no product was observed at pD 7 with only 2 equivalents of ribose. But F^- formation was avoided and the unexplained peak at -76 ppm was not observed. No degradation takes place under these conditions.

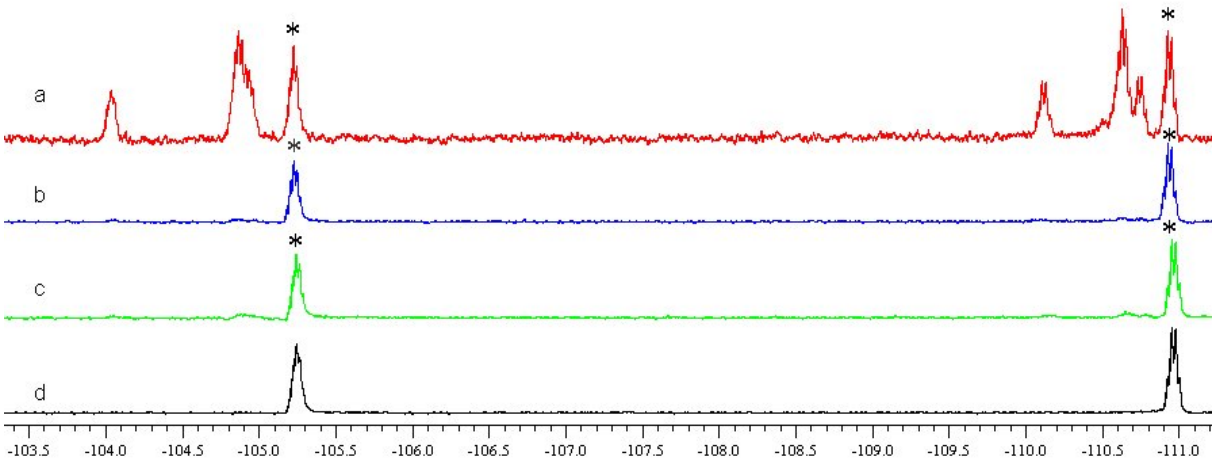


Figure 3.38. ^{19}F NMR spectra after 72h at r.t. of a mixture of ribose and universal base mimic a) universal base mimic 6 mM, ribose 120 mM, buffer pD 6, b) universal base mimic 6 mM, ribose 120 mM, D_2O, c) universal base mimic 6 mM, ribose 12 mM, buffer pD 6, d) universal base mimic 6 mM, ribose 12 mM, D_2O. * indicates signals of 8.

Decreasing pD (Figure 3.38c, green) or increasing the amount of ribose (Figure 3.38b, blue) resulted in formation of small amounts of products after three days. But measuring the sample after another three days revealed that both reaction mixtures were not yet at equilibrium and continued to react.

The reaction mixture at pD 6 with 20 equivalents of ribose clearly reacts much faster (Figure 3.38a, red). About 70% of the starting hydrazide is converted into the three products. The reaction evolves no further after 3 days. The distribution of products shown in Figure 3.38a is unaltered after 6 days. The sample in D_2O with 20 equivalents of ribose reaches approximately the same product distribution after 7 days.

Two domains can be distinguished in the ^{19}F NMR spectrum, one around -105 ppm and another around -110 ppm. The domain around -105 ppm presumably corresponds to the fluor atom *ortho* to the hydrazide, the domain around -110 ppm corresponds most certainly to the *para* positioned fluor atom. The hydrazide might deshield the ^{19}F atoms, therefore their signals will be observed downfield. The atom that is closest to the hydrazide is influenced most. Furthermore, changes at the configuration of the nitrogen atoms upon product formation provokes more important changes to the atoms resonating around -105 ppm than the ones around -110 ppm.

The region around -110 ppm clearly shows the formation of three different products, although two peaks show overlap. The overlap is more serious in the -105 region. To simplify the

spectra, proton decoupling experiments are the obvious way to go. But this turned out to be technically impossible (*vide supra*).

3.5.1.3 Summary

Reaction between ribose and aminobases **3-6** was not possible as shown by Régis Nguyen [122]. A probable explanation for this is shown in Figure 3.39. The equilibrium between hemiacetal and aldehyde lies on the hemiacetal side. On the contrary, the imine-hemiaminal probably lies on the imine side. The hemiaminal is not able to stabilize the imine and pull the reaction of the aldehyde with the amine towards the imine/hemiaminal side. We therefore started test experiments with hydrazides. Hydrolysis of the formed acyl hydrazones may be slower than the normal imines, therefore higher quantities of products may be obtained.

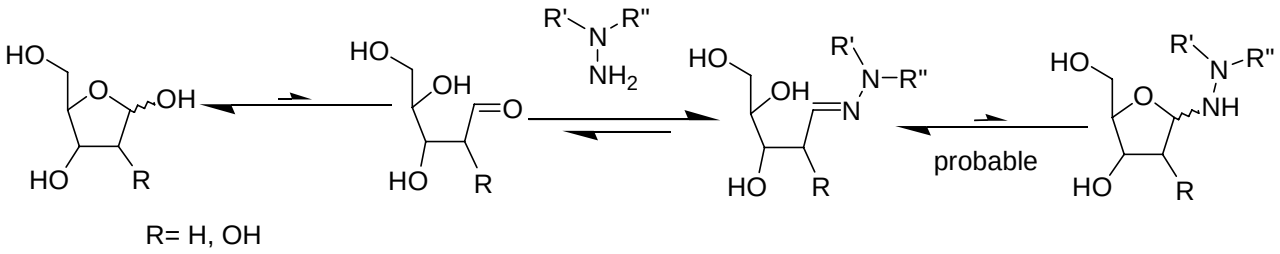


Figure 3.39. Hemiacetal, aldehyde, imine, hemiaminal equilibrium.

From the first experiments with ribose and hydrazides **7** and **8** can be concluded that hydrazide nucleophiles, contrary to the aminobases, react with the aldehyde form of ribose. However, the yield of the reaction is low (~10% when ribose and universal base mimic are mixed 1/1). Drawing any conclusions about the nature of the reaction products was not possible. Reacting ribose with the thymine mimic showed the formation of three products at elevated temperature and complete disappearance of the starting material. Performing the reaction with the universal base mimic at high temperature showed the formation of side products. The reaction at room temperature could be accelerated by decreasing the pD to 6 and the addition of 20 equivalents of ribose.

The reaction is slow, the fast equilibrium that we would have liked, is most certainly not reached. The reaction takes three days to reach its limit. This is too slow for the oligonucleotide experiments we envisage. When positioned in a DNA oligonucleotide, we would like the aldehyde function to react fast with the hydrazides. The reaction should be complete in less than about an hour. No data on the backward reaction were obtained.

Since our goal is to 'repair' DNA abasic sites, the addition of 20 equivalents of ribose is not desirable either, even though it helps the analysis of the products. In the oligonucleotide

experiments we want to 'repair' a maximum of abasic sites. Thus, the reaction with an abasic site in an oligonucleotide might be performed using a 20 fold excess of hydrazide. However, this may in turn lead to other products than in the test experiments. For instance, the formation of aminals is in that case more likely.

No obvious difference in the reactivity of **7** and **8** towards ribose was found. Analysis of the reaction products by ^{19}F NMR is very practical. We therefore decided to continue the test experiments with deoxyribose and the universal base mimic **8**.

3.5.2 Test reactions with deoxyribose

3.5.2.1 Forward reaction between deoxyribose and 8

We now would like to test the DNA model and therefore take deoxyribose to react with the hydrazides. With the encouraging results of the reactions between ribose and **8** in hand, we started test reactions between deoxyribose and **8**. The absence of the $^{2'}OH$ is expected to change the properties of the aldehyde, we want to find out whether reactions take place faster or slower. The products that can be possibly expected from the deoxyribose case (Figure 3.40) are of course similar to the ones expected with ribose (Figure 3.34).

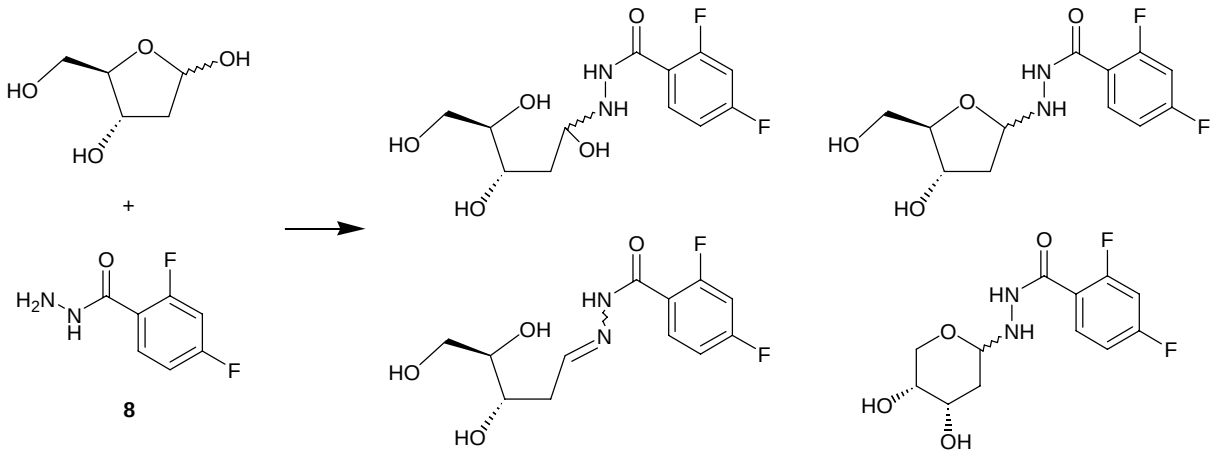


Figure 3.40. Possible products upon mixing deoxyribose and the fluorinated universal base mimic. The amide bonds in the products are in reality in a *trans* conformation, but are depicted *cis* here to maintain the ribose and the uracil in conventional orientation.

Reaction of compound **8** with deoxyribose turned out to be significantly faster than with ribose. When repeating the experiments described in Figure 3.38 replacing ribose by deoxyribose under otherwise identical conditions, reaction reached a steady state after 72h for all four cases (Figure 3.41). This in contrast with the ribose case, where only with 20

equivalents of ribose at pD 6 no more changes in the reaction mixture were observed after three days. Even in the experiment in D_2O with only 2 equivalents of deoxyribose reaction a steady state was reached (Figure 3.41d). With ribose, no reaction at all was observed under these conditions, even after 3 days.

In the test tube with 20 equivalents of deoxyribose at pD 6, the reaction was already at 90% of its final product distribution after only 3 hours of reaction time. After 20 hours the reaction did not proceed any further.

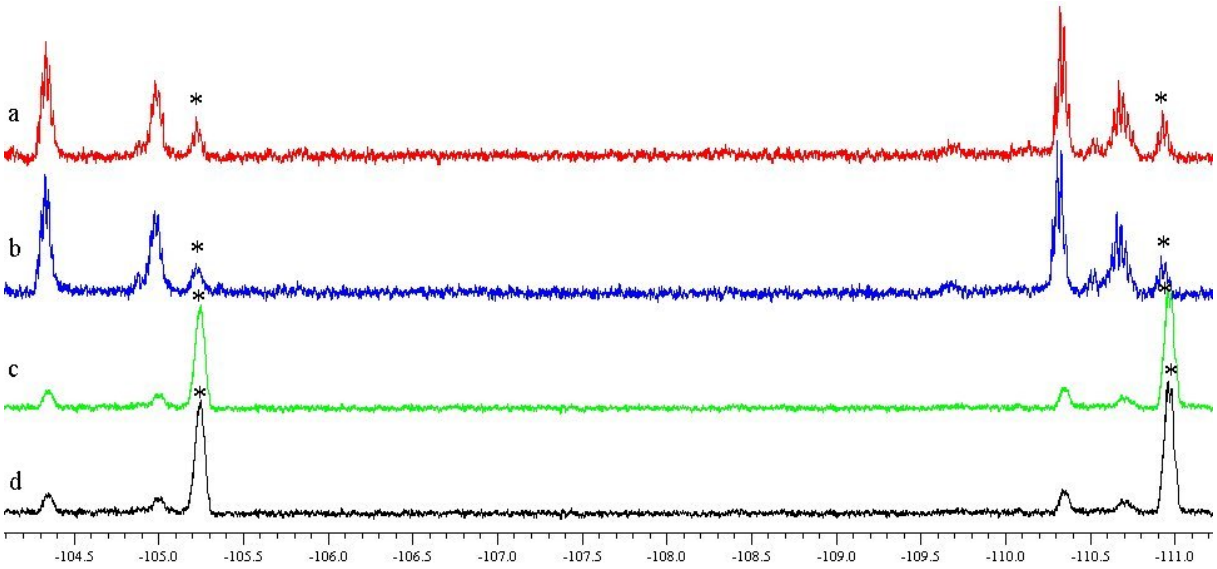


Figure 3.41. ^{19}F NMR spectra after 72h at r.t. of a mixture of deoxyribose and universal base mimic a) universal base mimic 6 mM, deoxyribose 120 mM buffer pD 6, b) universal base mimic 6 mM, deoxyribose 120 mM, D_2O, c) universal base mimic 6 mM, deoxyribose 12 mM, buffer pD 6, d) universal base mimic 6 mM, deoxyribose 12 mM, D_2O. * indicates signals of 8.

Apart from the reaction being faster, it also seemed to proceed differently. Reacting **8** with ribose gave three products in significant amounts, the ^{19}F NMR spectrum of the reaction between deoxyribose and **8** shows only two major products. A shoulder under one of the major product peaks can be observed, that indicates formation of small amounts of a third product. Some more minor products in the mixture are also observed in the ^{19}F NMR spectrum.

The ^{1}H NMR spectrum is again difficult to analyze. When using 20 equivalents of deoxyribose relative to **8** (Figure 3.42a), the deoxyribose protons obstruct a clear view of the changes in the region between 3.0 and 1.5 ppm. At 2 equivalents of deoxyribose, the product signals are hard to distinguish (Figure 3.42b). A clear downfield shift of the aromatic protons of the universal base mimic can be observed. The changes around 5.15 ppm as seen in the ribose case did not occur with the deoxyribose.

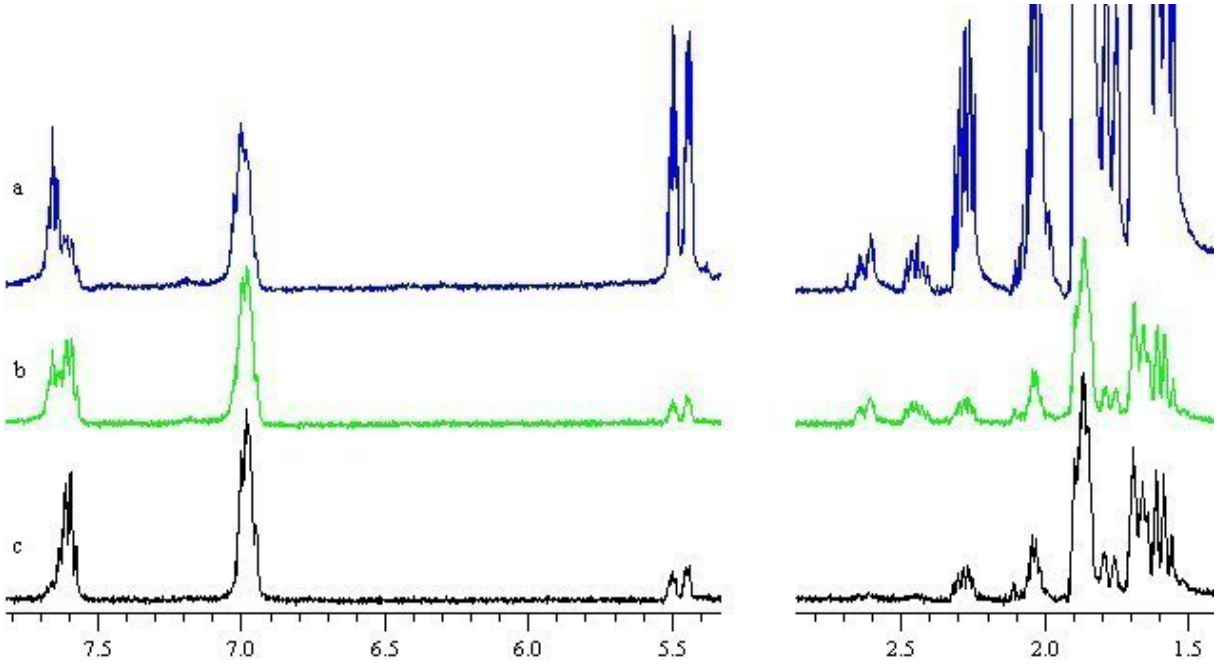


Figure 3.42. ^{1}H NMR spectra 72h at r.t. in D_2O with a NaH_2PO_4/Na_2HPO_4 buffer at pD 6 of a mixture of deoxyribose and universal base mimic a) universal base mimic 6 mM, deoxyribose 120 mM, after 72h, b) universal base mimic 6 mM, deoxyribose 12 mM, after 72h, c) universal base mimic 6 mM, deoxyribose 12 mM, just after mixing.

Two new multiplets can be observed around 2.5 ppm (Figure 3.42 a, b). Since these peaks are the only ones changing in the ribose part of the spectrum, they may correspond to protons close to the reactive center. The peak at 2.08 ppm corresponds to one of the two deoxyribose $C^{2'}$ protons, so it might be expected that the multiplets at 2.45 and 2.65 correspond to $C^{2'}$ protons of the new products. No changes are visible in the shift of the 1' proton (resonance at 5.5 ppm). This might suggest the formation of the imine. The imine signal is expected around 7-8 ppm and may be masked by the aromatic signals of the universal base mimic

From the NMR experiments, it cannot be concluded whether the products are the imine, hemiaminal or imine-hydrate. Crystallization experiments of the products in order to obtain an X-ray structure were unfortunately not successful. To discriminate between the three proposed structures we need to make use of additional experimental methods.

The molecular weight of the products is not the same for all compounds, $M_{w\ imine}$ = 288.2 g/mol, $M_{w\ hemiaminal}$ = 288.2 g/mol, $M_{w\ pyranose}$ = 288.2 g/mol, $M_{w\ imine\ hydrate}$ = 306.3g/mol. We therefore decided to analyze the reaction mixture (H_2O, r.t.) by electron spray mass spectrometry. The major peak in this experiment corresponded to a product with a mass of 311.3g/mol. This corresponds exactly to the mass of the hemiaminal or the imine plus a sodium ion. At 289 g/mol no peak corresponding to the product ionized by H^+ could be

observed. Another major peak was observed at 173.4 g/mol. This peak corresponds to compound **8**, equally ionized by Na^+. A peak corresponding to deoxyribose was not observed, neither was a peak observed at 329 or 307, that could indicate the formation of the imine hydrate. When the reaction was performed in D_2O, like in the NMR experiments, a peak at M_w 315 g/mol is observed. This confirms the formation of the deuterated imine, the hemiaminal, or the pyranose, in each molecules four exchangeable protons are present. From these mass spectrometry data the formation of the imine hydrate can be excluded. Discrimination between the imine, the hemiaminal and the pyranose could not be made, since they have an identical molecular mass.

To discover whether the products are the imines or the hemiaminal, we need a clean ^1H NMR spectrum. The reaction appears to be slow and we hoped that the products would not exchange during a separation experiment by for instance reversed phase HPLC and allow separation.

Analytical RP-HPLC chromatography of the mixture showed two product peaks with a retention time of 7 and 16 minutes (Figure 3.43). This is consistent with the ^{19}F NMR spectrum, where we also found two major products. Both peaks are rather broad compared to the peak of the starting universal base (4.5 min). This may indicate that the products, presumably the imine and the hemiaminal, inter-convert during the experiment. The equilibrium prevented obtaining a clean NMR spectrum.

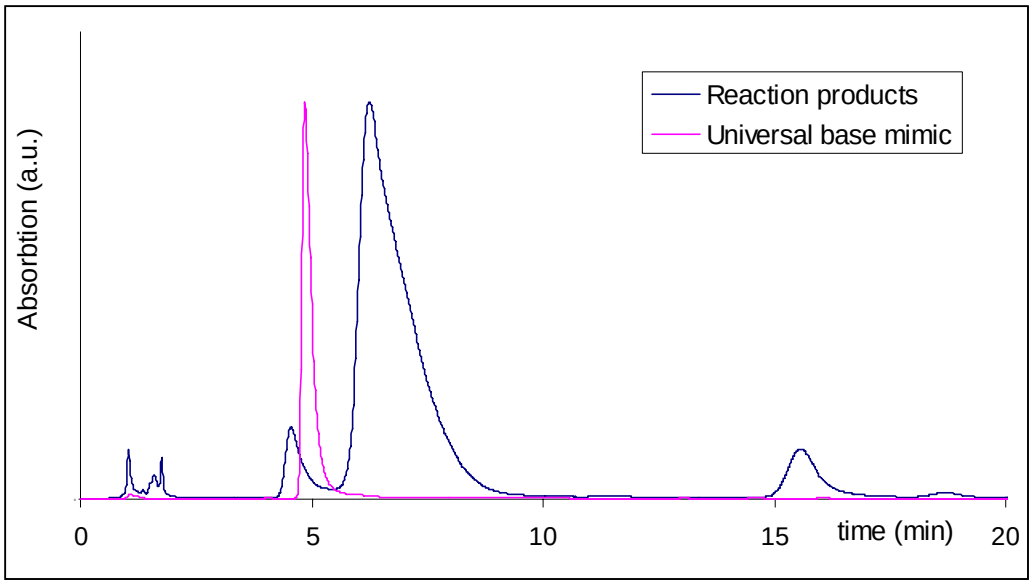


Figure 3.43. RP-HPLC spectrum of the reaction mixture recorded by UV at 254nm. Solvent system: H_2O/CH_3CN (H_2NaPO_4/HNa_2PO_4 buffer pH 6) 100/0 to 98/2 in 20 minutes.

3.5.2.2 Reversibility of the reaction

We established the reaction between deoxyribose and universal base mimic **8**. We found two major products and at pH 6 in D_2O the reaction proceeded to a steady state where around 10% of starting material was still present.

We now want to know whether the hydrolysis of the products also takes place, which would mean that there is an equilibrium situation. Also the kinetics of the backward reaction are of great interest. The RP-HPLC trace showed asymmetrical peaks, which might indicate reversibility of the reaction, or equilibrium between the different products. We used three different methods in the analysis of the products, hydrolysis by dilution, separation of the products by reversed phase column chromatography and NMR experiments at elevated temperature.

We also thought of performing exchange experiments between **7** and **8**, but these experiments are not possible. When deoxyribose and for instance **8** are mixed 1/1, only little product is formed, not more than around 15%. If in that situation compound **7** is added, it will not compete with **8** for the deoxyribose, but take the available deoxyribose. The changes in the fluor spectrum will not be visible. Mixing deoxyribose and **8** 1/20 will increase the yield relative to deoxyribose, but analysis with ^{19}F NMR will not be possible due to the large excess of **8**. Moreover, solubility problems prevent the addition of a large excess of **7**.

The simplest way to determine the kinetics of the backward reaction is diluting the mixture. Hydrolysis of the products should then lead to another steady state situation.

We mixed deoxyribose (120mM) and compound **8** (6mM) at pH 6 and let the mixture react for 24h to reach a steady state. We then diluted the sample a factor twice and observed that the reaction mixture reaches a new product distribution in about 24h. It remains unchanged after that. This means that there is an equilibrium with the starting materials, but the equilibrium is slow. Hydrolysis of the products takes place at similar rates as the forward reaction and reversible conditions are not practical with this reaction system, making error correction in the oligonucleotide experiments unlikely.

To investigate whether the assumption about an equilibrium between the different products is correct, we attempted a separation of the products that are formed. The RP-HPLC trace (Figure 3.43) showed that the two products have different retention times and it may thus be possible to separate them. RP-HPLC showed that the reaction products are retained longer than compound **8**, this compound may thus also be separated from the products to study the product-starting material equilibrium. Deoxyribose was not detectable during the RP-HPLC experiment, because it does not absorb in the long UV region. Deoxyribose is expected to pass very fast through the RP-HPLC column, because of its high affinity for water.

To run this experiment, the reaction products were prepared by mixing the universal base mimic **8** with 20 equivalents of deoxyribose. After incubation for 24h, the water was evaporated and the products were put on a reversed phase column and eluted with water.

No universal base mimic was collected, presumably because it completely reacted during the evaporation. ^{19}F NMR shows that the first fractions collected from the column contained several products (Figure 3.44b). Unexpectedly, the excess deoxyribose could not be separated from the products and can be seen on the ^{1}H NMR spectrum.

We placed the mixture obtained from the reversed phase column in an NMR tube and followed its evolution over time (Figure 3.44). In the -110 region, new signals can be observed in the spectrum taken directly after passing the sample through the ion exchange column (Figure 3.44d). This suggests the formation of new products during the separation with the reversed phase column. These products are not stable in water at room temperature, as can be seen from Figure 3.44a. After 24h the spectrum shows clearly that the mixture reverts to the products present before using the ion exchange column (Figure 3.44e) and compound **8** is released. Two products remain present in the reaction mixture, which are not formed when the products are not passed through the column.

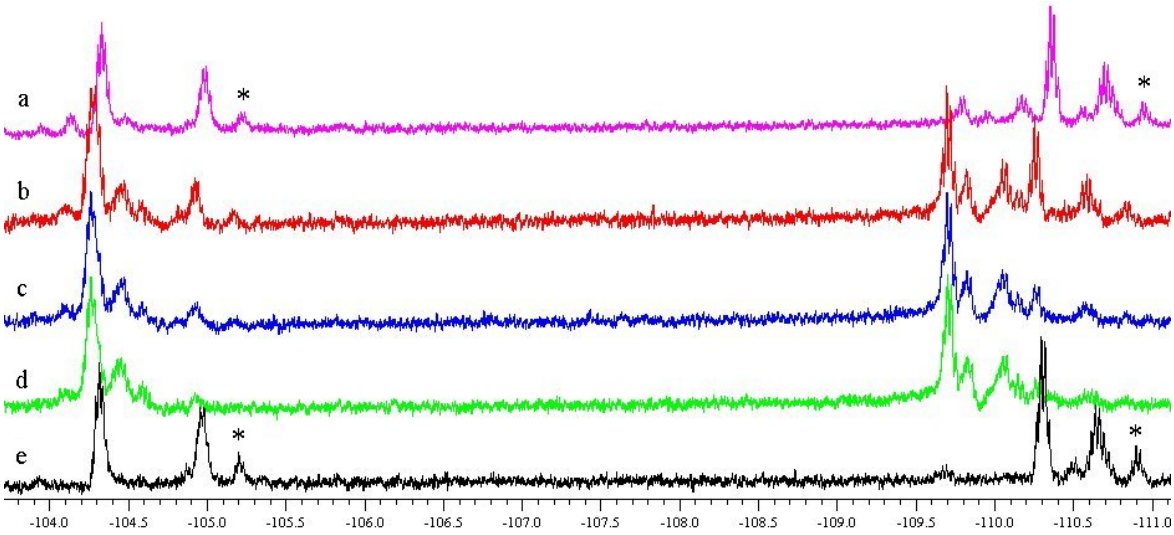


Figure 3.44. ^{19}F NMR spectra of a) First fractions reversed phase column, 24h after separation in D_2O, pH 6, concentrations unknown, b) first fractions reversed phase column, 4h after separation in D_2O, pH 6, concentrations unknown, c) first fractions reversed phase column, 1h after separation in D_2O, pH 6, concentrations unknown, d) first fractions reversed phase column, directly after separation in D_2O, pH 6, concentrations unknown, 4h, e) reference mixture of deoxyribose (120 mM) and universal base mimic (6mM) at r.t., pH 6. * indicates signals of compound 8.

From these experiments can be concluded that hydrolysis of the products takes about 24 hours to reach the same product distribution as the forward reaction. This confirms the conclusion of the previous dilution experiment. The backward reaction is thus approximately as fast as the forward reaction. Practically this means that hydrolysis will not take place during the oligonucleotide experiments that we envisage. The occurrence of other products when passing the mixture through the reversed phase column is noteworthy. When repairing an abasic site in a DNA oligonucleotide, the products may be able to rearrange to form the product that stabilizes a DNA oligonucleotide best.

To develop optimal conditions for our oligonucleotide experiments, we wanted to know more about the influence of temperature on the equilibrium. Our special interest goes to a possible displacement of the equilibrium in favor of the products or a possible increase in reaction kinetics. It was previously shown that increasing the temperature provokes the formation of side products during the reaction. However, these experiments submitted the reaction products to high temperatures for a substantial time. Due to this degradation we did not investigate in detail the consequences of temperature on the formation and equilibration of the products. It can be expected that degradation will be less important when the products are exposed to high temperatures for short times only.

We placed a mixture of the fluorinated universal base mimic **8** and deoxyribose (1 to 1) in an NMR tube. We let the mixture react at 4°C for one week to obtain a stable product distribution and then measured the ^{19}F NMR spectrum. The other spectra were obtained by incubating **8** and deoxyribose at r.t. for 24h, as in the other experiments. The ^{19}F NMR spectra of the sample were measured at various temperatures as shown in Figure 3.45. The equilibration time for the spectra is about 10 minutes.

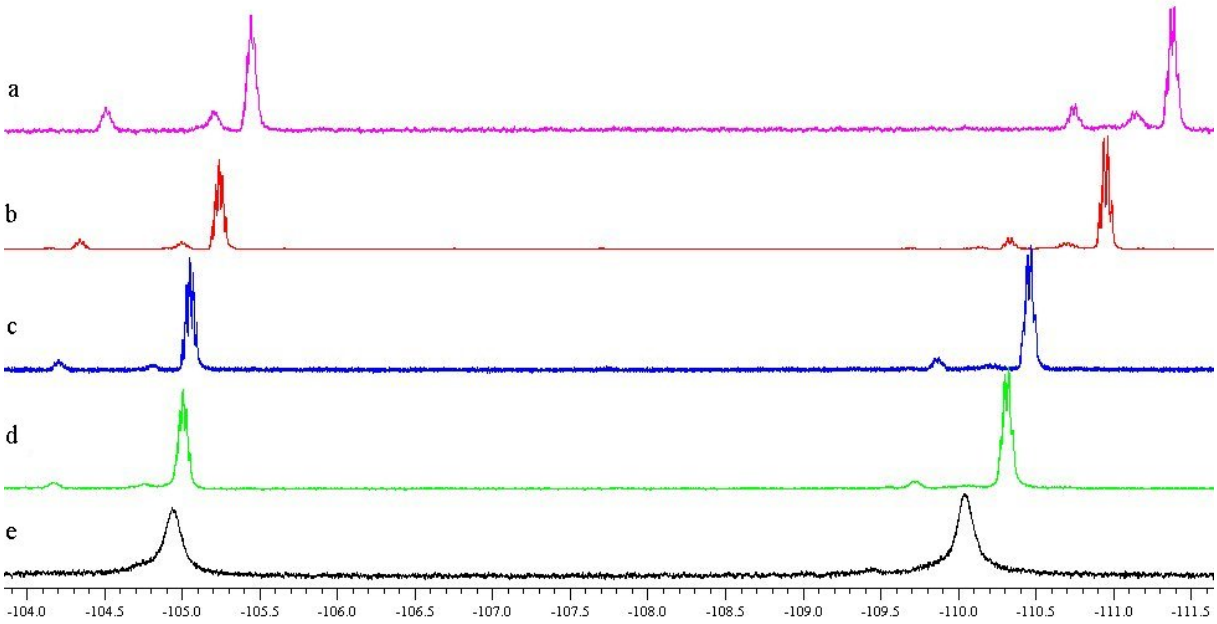


Figure 3.45. ^{19}F NMR spectra of a mixture of fluorinated universal base (6mM) deoxyribose (6mM) at pH 6 at a) 5°C, b) 25°C, c) 50°C, d) 60°C, e) 80°C.

When looking at the ^{19}F NMR spectra the displacement of the chemical shift upon temperature change is remarked immediately. Decreasing the temperature to 5°C leads to an upfield shift of around 0.5 ppm in the -110 ppm region. Increasing the temperature from r.t. to 50°C leads to a similar downfield shift. At -105 ppm the changes are less important, but here a similar downfield shift by increasing temperatures is observed. These kind of variations in chemical shifts are known in literature, although the variation we observe here is large compared to shifts found by other authors [124, 125]. Since the signal for F^- (not shown in Figure 3.45) also shifts with temperature, we think that the shifts are due to temperature and not due to the formation of other products.

Besides the change in chemical shift, a change in product distribution can be observed. Increasing temperature leads to less product. Integration of the signals showed that at 5°C approximately 24% of **8** is transformed in product, at 25°C this has decreased to 19% and at 50°C only around 12% of the initial **8** is present as one of the two products. Further must be

noted that at 80°C only one broad peak is obtained. This might point to fast exchange between the various products at that temperature.
Not shown in Figure 3.45 is the increasing amount of F^- ions that are formed with increasing temperature. Even short exposure to high temperatures thus leads to side reactions.

The same experiment was reproduced using 10 equivalents of deoxyribose to increase product formation relative to the universal base. A slightly better product to starting material ratio was obtained and the previous experiment is confirmed. The downfield displacement of the chemical shift with increasing temperature is reproduced, as well as the fast equilibrium conditions that seem to exist at 80°C.
Degradation products are observed, among which F^-, which forms in increasing quantities in the mixture above 50°C. Besides NMR spectroscopy the mixture was analyzed by mass spectrometry before and after heating. Albeit the new products that show up in the ^{19}F NMR spectrum, the mass spectrum does not change upon heating and cooling down. This means either that the formed products have the same mass as one of the compounds already present, or they are present in too small quantities to measure by electron spray mass spectrometry.

3.5.2.3 Summary

In conclusion we can say that the reaction between deoxyribose and universal base mimic **8** yields two major products and a few minor ones. In D_2O at pD 6, a good yield is obtained (approximately 85%). In around 20h a steady state is obtained that is easily analyzed by ^{19}F NMR spectroscopy. RP-HPLC trace analysis confirms the formation of two products. It could not be established whether these products were hemiaminals, imines, or another species. Mass spectrometry excluded the formation of the imine hydrate.
The backward reaction is approximately as fast as the forward reaction. It takes approximately 24h to reach a steady state.
When increasing the temperature, a fast equilibrium on the time scale of the NMR is obtained, but at this temperature degradation takes place.
Passing the mixture through a reversed phase column changes the product distribution. Signals of new products can be observed, that partially disappear over time. Some of the final products in this experiment were not observed by simply mixing deoxyribose with **8**.

3.5.3 Conclusion

The test reactions that we described in the last paragraphs were set up to find good conditions for dynamic repair of an abasic site in an oligonucleotide. We have to determine which reagents are capable of reacting with an abasic site, and whether a RNA or DNA would be easier to repair. Since a RNA abasic site is in fact a ribose unit and a DNA abasic site a deoxyribose, we used ribose and deoxyribose as cheap and available test components.
Reaction between ribose and the aminobases **3-6** did not yield any products, we therefore started test reactions with more reactive hydrazides **7** and **8**.

Reaction of thymine mimic **7** with ribose gave only low yields when mixing them 1 to 1 at r.t.. The yield could be increased by performing the reaction at elevated temperature. The kinetics of the reaction could be increased by either increasing the temperature or by decreasing the pH of the reaction mixture to 6.
Mixing universal base mimic **8** with ribose revealed that at elevated temperature degradation took place. Reaction at room temperature was possible, and analysis of the reaction mixture with ^{19}F NMR allowed the addition of 20 equivalents of ribose relative to **8**. This increased the yield relative to **8** significantly, without giving any additional signals in the ^{19}F NMR spectrum. Reaction of a mixture of 120 mM of ribose and 6 mM of compound **8** in D_2O with a NaD_2PO_4/Na_2DPO_4 buffer at pD 6 at r.t. took about 72h to reach a steady state situation. Three major products were observed when performing this reaction. No obvious difference between the reactivity of **7** and **8** towards ribose was observed.

Because of the easy analysis of the products with ^{19}F NMR spectroscopy, we decided to test the reactivity of **8** towards deoxyribose. This reaction proceeds much faster, within approximately 20h a steady state situation is attained. Two major products are observed in this case, and some minor products are also present. The nature of the products could not be determined. A molecular mass of 288 g/mol suggests the formation of either the acyl hydrazone or the hemiaminal as shown in Figure 3.46. Both products can be present as two stereoisomers.

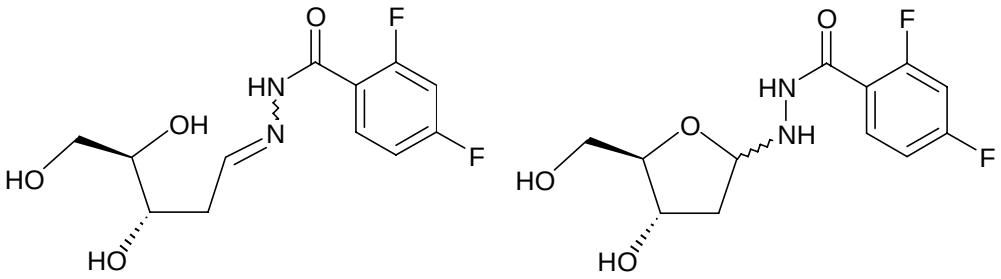


Figure 3.46. Acyl hydrazone and hemiaminal. The amide bonds in the products are in reality in a *trans* conformation but are depicted *cis* here to maintain the ribose and the uracil in conventional orientation.

The hydrolysis of the products takes approximately the same time as the formation of the products, approximately 24h. Equilibrium of the reaction is attained rapidly at elevated temperature, 80°C. However, the yield of the reaction decreases with increasing temperature. It is most certainly better to perform the reactions at low temperature to increase the yield, albeit the loss of reaction rate.

Changing the reaction conditions can lead to other products easily. In an attempt to separate the product mixture by a reversed phase column, a few new products appeared, which, in 24h, equilibrated to approximately the same product distribution as before passing the mixture through the column. That means that the products can rearrange to form the most stable product. In the oligonucleotide experiments this will be useful, since the compound that will form first may not be the one that stabilizes an abasic site in a DNA double helix best.

3.6 Design of an oligonucleotide sequence

In the last paragraph we presented reactions that tested the reactivity of the system that we want to use to repair abasic sites. A deoxyribose unit is more reactive than a ribose, we will thus focus our design on DNA sequences and not on RNA.

Besides being sufficiently reactive, the abasic site must also be available for reaction, and the products must be stabilized by hydrogen bonding and π–π-stacking, as discussed in paragraph 3.3. In that paragraph is shown that it will not be evident to have the abasic site available for reaction, and to have the products to be in an intrahelical conformation for stabilization. We do not know beforehand which of the two situations will be most beneficial for the repair process.

To be able to gain as much stability as possible in the repair process, we would like the abasic site to have a large destabilizing effect on the double helix. Therefore, we have to incorporate the abasic site in the middle of a strand. At the extremity of a strand, only the last base is

unpaired and destabilized, the rest of the oligonucleotide is hardly influenced. In the middle of the strand the stability of both neighboring base pairs will be negatively affected.

The first point that is important for the stability of the abasic site, is the nature of the opposite base [94, 99, 100, 102]. It is also important whether an A-T or a G-C base pair is broken. In the former only two hydrogen bonds are broken, in the latter three. Gelfand *et al.* [102] investigated this in the duplex d($^{5'}$CGC ATG X_1GT ACG C$^{3'}$).($^{5'}$GCG TAC X_2CA TGC G$^{3'}$) (with X =THF unit or one of the four nucleobases). In case an adenine or thymine is missing (two hydrogen bonds broken), the melting temperature[1] decreases by 12 to 16°C. In case a guanine or a cytosine is removed (three hydrogen bonds broken) a decrease of melting temperature in the range of 14 to 21°C is observed.

An unpaired purine results on average in a slightly smaller reduction in T_m (ΔT_m= -13°C) than an unpaired pyrimidine (ΔT_m= -15°C) [102]. This is probably due to reduced loss of π–π stacking in case of a missing pyrimidine relative to a missing purine. For our purpose an unpaired pyrimidine may give the best results based on this theory.

Neighboring base pairs also play a role in the stability of the double helix. A-T base pairs neighboring the abasic site have a lower destabilizing effect on the DNA duplex than a G-C base pair. Flanking cytosines have on average a slightly more destabilizing effect on the double helix than flanking guanines, but the difference is only 1°C, which is hardly significant. A missing guanine flanked by cytosines turned out to have the most destabilizing effect (ΔT_m= -21°C) on the double helix. A missing cytosine flanked by two guanines also has a considerable effect on duplex stability, with a ΔT_m of -19°C [102].

So to destabilize the DNA strand containing the abasic site as much as possible, we want to synthesize a DNA strand in which a G-C base pair is cut. A free cytosine opposite to the abasic site would be the best choice, and the neighboring base pairs also have to be G-C pairs. But we have to take into account that, at this stage, we only have synthesized thymine mimic **7** and universal base mimic **8** as repair units for the abasic site. An adenine opposite to the abasic site is thus the only base with which we can expect hydrogen bonding. Albeit the fact that this would less destabilize the double helix compared to a cytosine.

A long oligonucleotide will have a higher melting temperature than a small one. There are more hydrogen bonds between the strands, and the effect of π–π-stacking is more important.

[1] The melting temperature of a DNA duplex is defined as the temperature at which half of the two complementary strands is present as a duplex, and half of the sample as a single strand. It is widely used as an indicator for the stability of a DNA duplex.

The melting temperature of the strand is important, since we want to perform the oligonucleotide experiments between 0 and 15°C. The test reactions showed that the yield of the reaction is higher when decreasing temperature.

We chose to synthesize the DNA duplex d($^{5'}$CGC AGY GAC GC$^{3'}$)·($^{5'}$GCG TCA CTG CG$^{3'}$) (with Y equals a natural abasic site). This sequence has an adenine opposite to the abasic site to make base recognition between the base mimic and the opposite base possible. The two guanine units on both sides may cause the destabilizing effect of the abasic site to be large. Deprotection of the abasic site after synthesis involves irradiation with UV light. Two adjacent thymine groups can in principle photodimerize under these conditions, it was made sure that the sequence did not contain two such thymines. Synthesis of the fully complementary equivalent of this duplex has an estimated melting temperature of 35°C. Since the abasic site will decrease the melting temperature of the duplex by around 15-20°C, the duplex containing an abasic site will have a melting temperature of around 15-20°C. The binding studies should be performed at a temperature at which we are sure the DNA oligonucleotides are present as a duplex. A temperature around 5-10°C can thus be chosen. At this temperature the test reaction between deoxyribose and universal base mimic **8** showed the best yields. The reactivity of **7** is expected to be similar to **8**.

3.7 Synthesis

3.7.1 Introduction

The importance of abasic sites has led to many studies on them. The synthesis of abasic sites is not trivial, because of their instability. Nevertheless, the first method to incorporate abasic sites in DNA strands dates from 1972[126, 127].

One of the natural ways abasic sites occur, is hydrolysis under acidic conditions. This prompted Lindahl *et al.* in 1972 to treat a plasmid DNA in a buffer of NaCl and sodium citrate at pH 5 to create abasic sites in the plasmid [126, 127]. They proved experimentally that in this buffer at 70°C, an average of two abasic sites per plasmid DNA molecule per 8 minutes are created. The major drawback of this method is of course the random positioning of the abasic sites in the DNA duplex.

A more controlled way to incorporate abasic sites in predetermined places was developed by Stuart and Chambers [128]. They incorporated deoxyuridine residues in selected positions in a DNA oligomer and used the enzyme uracil-DNA-glycosylase to selectively cut out the uracil units after automated DNA synthesis.

The first synthetic way to prepare abasic sites without the use of enzymes was published by Groebke and Leumann [81]. They prepared a deoxyribose phosphoramidite monomer, bearing a TBDMS protecting group on the 1'-hydroxyl group. This method made the use of standard automated DNA synthesis possible and allowed incorporation of the abasic site at any desired place in a DNA-oligonucleotide. The drawback of this method is the synthesis of the phosphoramidite monomer. The selectivity of incorporation of the silyl group on the 1'-hydroxyl group was low, yielding only 30% of the desired product.

A year later, Péoc'h *et al.* published an optimized method, using an *o*-nitro benzyl group to protect the 1'-hydroxyl group of the deoxyribose [129]. This group can be removed after DNA-synthesis by irradiation with UV-light (Hg-lamp) and incorporation of the *o*-nitro benzyl group has a yield more than twice as high (75%) as the incorporation of the silyl group reported by Groebke and Leumann. We will thus used the method published by Péoc'h *et al*..

3.7.2 Synthesis of the phosphoramidite and the oligonucleotide

Synthesis was started from *o*-nitro benzyl 2-deoxy-D-ribofuranoside that the group of Bernard Rayner still had in stock and kindly provided us with. The $^{5'}$OH group of the ribose had to be protected with a dimethoxytrityl group and the $^{3'}$OH activated with a phosphoramidite group (Figure 3.47).

The first step was performed by selective nucleophilic substitution of the chlorine atom of dimethoxytrityl chloride by the primary hydroxyl group of the deoxyribose to give compound **26** in 34% yield. **26** is very sensitive to acidic conditions, which has to be taken into account during purification by column chromatography and explains the moderate yield.

The second step also involves a nucleophilic substitution. The chlorine atom of compound **27** is substituted for the $^{3'}$OH of the deoxyribose to give the desired compound **28** in 58% yield. This phosphine derivative is sensitive to humidity and acidic conditions and should be stored under an inert atmosphere. Purification by column chromatography in air is possible, as long as dry solvents with a slightly basic pH are used and the solvents are evaporated quickly after chromatography.

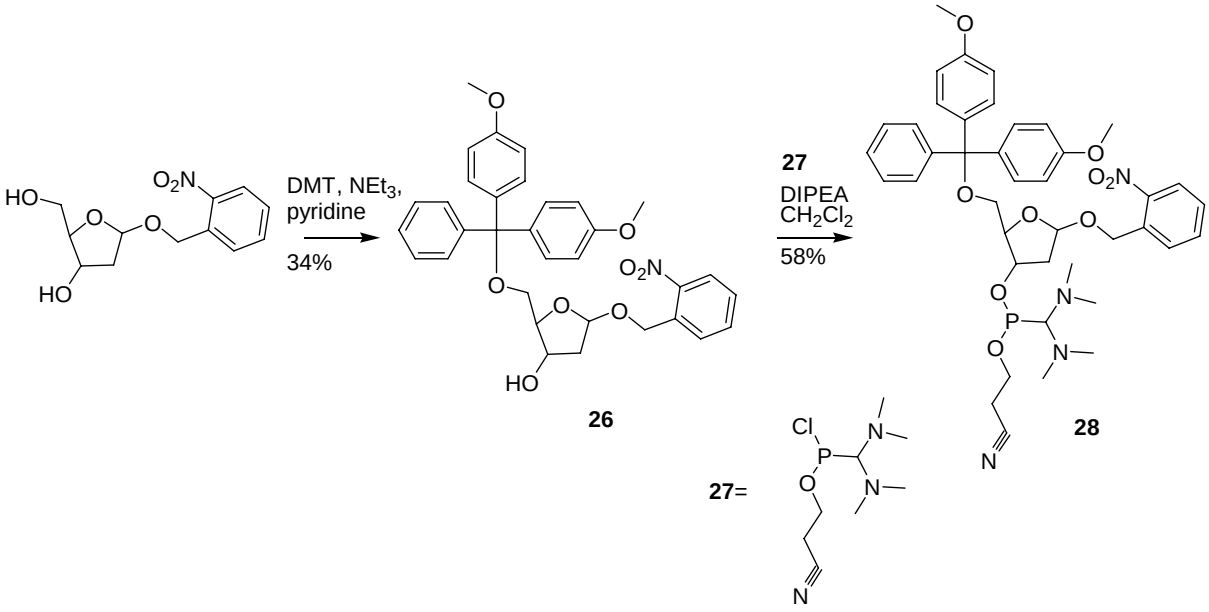


Figure 3.47. Phosphoramidite synthesis abasic site.

After synthesis of the monomer, we could proceed to DNA synthesis. The oligonucleotides were assembled ('DMT-on') on a DNA synthesizer, using a standard 1μmole cycle. The abasic site building block was incorporated manually with prolonged coupling times (15-20 minutes) and a 50-100 fold excess of **28**. The coupling efficiency was better than 98%, as judged by photometric trityl assay. The oligonucleotide was cleaved from the solid support using NH_4OH for 5 minutes. This solution was then heated at 60°C overnight to cleave the various protecting groups. The mixture was then washed with ethyl acetate and the ammonia was evaporated [130].

3.7.3 Purification of the oligonucleotide

After applying the standard procedure for the cleavage from the support and deprotection, the nitro benzyl protected oligomers were characterized by MALDITOF-MS (Figure 3.48) and purified by reversed-phase HPLC (Figure 3.49).

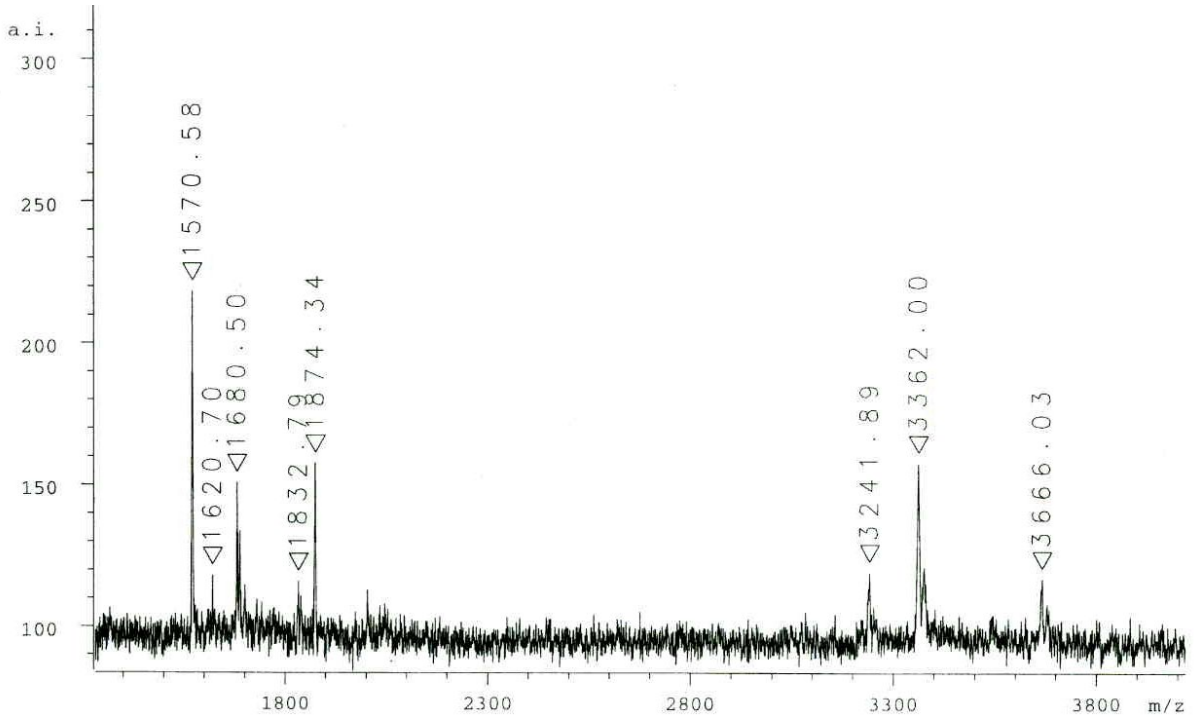


Figure 3.48. MALDITOF spectrum of the purified oligomer. The peak at 3666g/mol corresponds to the oligomer DMT on. The peak at 3362 corresponds to the oligomer DMT off. The peak at 3242 corresponds to the oligomer DMT off and nitro benzyl off.The peak at 1680 corresponds to the oligomer DMT off $^{2-}$.

During the MS analysis, the DMT group is partially or completely cut off, and the nitro benzyl group is partially cut off. The DMT group is sensitive to the slightly acidic conditions used for mass spectrometry. This is due to the 2,4,6-trihydroxyacetophenon matrix that is used in MALDITOF mass spectrometry. The nitro benzyl group is sensitive to the laser that is necessary for ionization. A mixture of nitro benzyl group on and off is observed.

RP-HPLC purification (Figure 3.49) was not an easy task. Even with the nitro benzyl group in place, the oligonucleotide turned out to be very sensitive to manipulation. Mass spectrometric analysis showed that the oligomer was cut at the position of the abasic site if serious precautions about sterility and pH were not taken.

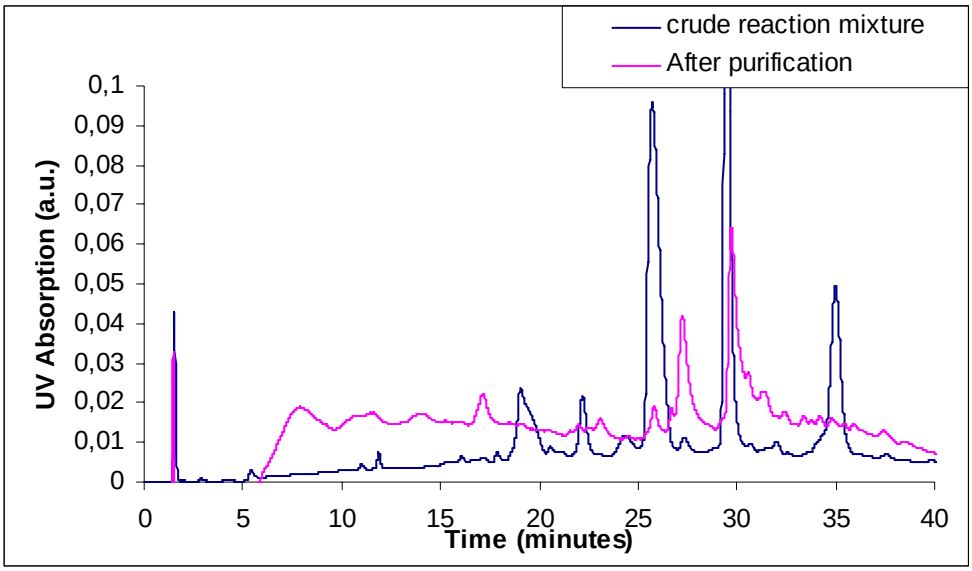


Figure 3.49. Typical RP-HPLC trace before and after purification using a water/acetonitrile at pH 6 as a solvent mixture. At t=0 H_2O / AcCN = 80/20 and the amount of acetonitrile increases with 1% per minute. The peak at 30 minutes corresponds to the DNA oligomer DMT on. The peak at 26 minutes corresponds to the DNA oligomer without the DMT group. The peaks that appear before 25 minutes correspond to smaller oligomers.

After purification by RP-HPLC, the DMT group was cut off by stirring the mixture in 80% acetic acid for 30 minutes. We then proceeded to the cleavage of the nitro benzyl group. Péoc'h *et al.* cleaved the nitro benzyl group by irradiation with a high pressure mercury lamp through a 2mm pyrex filter in diluted acetic acid (pH 4) for 10 minutes at room temperature [129]. We used the same lamp, but in absence of the described pyrex filter, we used pyrex Eppendorfs. These are expected to fulfill the same function as the pyrex filter, blocking the low-wavelength photons. We placed the Eppendorfs in a water bath to keep the temperature constant.

Unfortunately, irradiation of the oligonucleotide at room temperature in diluted acetic acid for 10 minutes with the mercury lamp led to severe degradation of the DNA strand. MALDI-TOF mass spectrometry, RP-HPLC and electrophoresis revealed that the original undecamer was cut in pieces (Figure 3.50). The mixtures thus consisted predominantly of pentamers. Some of the pentamers still contained a free deoxyribose moiety, others did not. Some tetramers and trimers were also observed. The presence of pentamers suggests that the strand is cut at the abasic sites. No undecamer was found in the mixture, neither protected, nor deprotected.

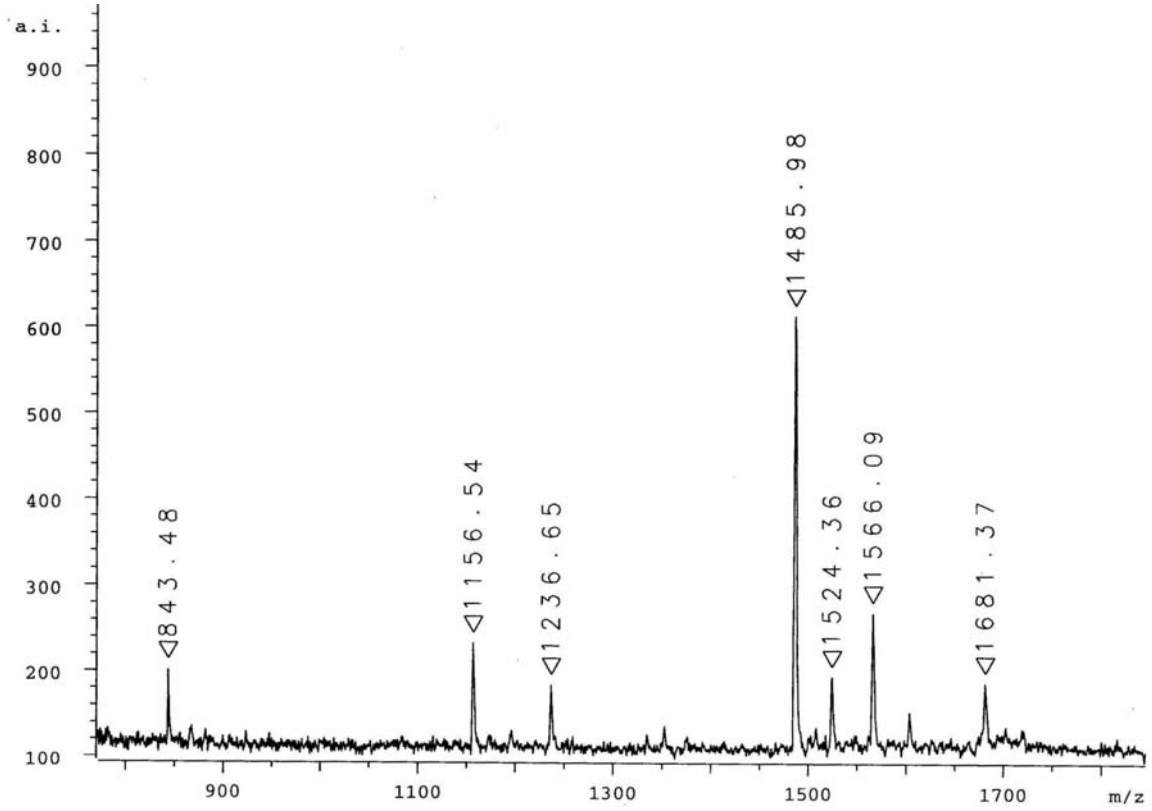


Figure 3.50. MALDITOF-MS spectrum after irradiation during 10 minutes. The peak at 1681 corresponds to the pentamer with a deoxyribose moiety attached. The peak at 1486 corresponds to the pentamer. No oligomers with higher masses were detected.

The deprotection conditions were obviously too harsh. But the fact that we did not find any protected starting material in the reaction mixture let us hope that we could use milder deprotection conditions. Since the deprotected abasic site is very labile to β-elimination under basic conditions, it was suspected that the diluted acetic acid solution (5%) might not be acidic enough to assure a low pH. The irradiation experiment was repeated in an 80% solution of acetic acid in water, using otherwise identical conditions. In this experiment there is no risk of the pH being above 7. However, in this case as well, only pentameric strands were found after irradiation, some still contained a free deoxyribose moiety.

Shorter irradiation times improved the outcome of the reaction slightly. Irradiation times of two or five minutes still showed a lot of decomposition, but some deprotected undecamer was found in the sample that was irradiated for 2 minutes. When irradiating less then 2 minutes, some protected undecamer was found. Two minutes is thus the minimum irradiation time to bring deprotection to completion. However, we observed by electrophoresis that when leaving the samples in solution at pH 4, the amount of undecamer decreased, in favor of pentameric species. The abasic site clearly makes the DNA oligonucleotide very fragile and β-elimination seems to occur at a lower pH than reported before by other authors. Conditions have to be

found under which the abasic site does not decompose over time. A non-constant concentration of oligonucleotides will severely hamper the analysis of the binding studies. A clean deprotection is also desirable, since purification by electrophoresis or RP-HPLC is not feasible with the labile abasic sites.

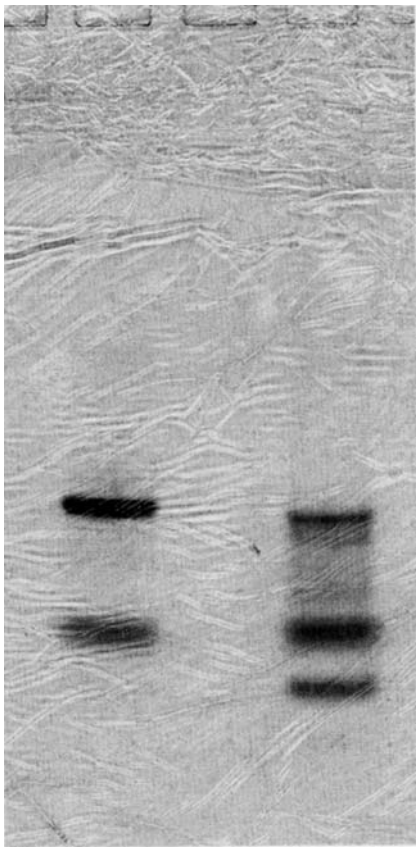


Figure 3.51. Electrophoresis diagram after irradiating the sample in 80% AcOH for 2 minutes at r.t.. Left; control undecamer. The upper band corresponds to the undecamer; the lower band corresponds to colorant, that is equivalent to an octamer. Right; deprotected oligonucleotide. The upper band corresponds to intact undecamer. The intense middle band corresponds to the colorant, equivalent to an octamer, the lower band corresponds to pentamer.

We thought that the pH might be the decisive factor for the DNA oligonucleotide stability and had a closer look at samples with a well defined pH. We irradiated samples at pH 2, 3, 4 and 5 for two minutes, to know more about the influence of the pH on the decomposition. The samples that were irradiated at pH 2 and 3 fully decomposed during the treatment, as expected. The samples at pH 4 and 5 suffered less, but they only contained 50% of intact, deprotected undecamer. The difference between the samples irradiated at pH 4 and pH 5 was only minor, as shown by electrophoresis. Decomposition of the sample after irradiation took place at both pH 4 and 5, and we could not find conditions under which no decomposition of the deprotected undecamer occurred. The problem of decomposition over time also remained.

We did a literature search to see if other groups equally had problems repeating the experiments of the group of Rayner [129]. It turned out that since the introduction of nitro benzyl method in 1991, some other methods have been developed, but incorporation of abasic sites *via* the nitro benzyl route seems to be exclusively used by the group of Rayner. One of the groups that developed other methods to incorporate abasic site mentioned that the irradiation step was inconvenient, because thymine dimerization can take place, when two adjacent thymines are present in the DNA oligonucleotide [131]. But this was the only reason that was found in the various articles that cite the method of the paper of the group of Rayner. No clue was found regarding the oligonucleotide sequence. We do not have adjacent thymines that can dimerize, that can thus not explain our problems. The oligonucleotide used by the group of Rayner in the original paper [129] contained two guanines adjacent to the abasic site. This strand was longer (30 bases) than ours, but shorter oligonucleotides (heptamers and nonamers) with other neighboring bases were also used by the same group [96, 129].

Shiskina *et al.* used oligonucleotides containing abasic sites similar to ours, but introduced the abasic sites protected as 1,2-diols [131]. The oligonucleotide they studied was a tridecamer containing an abasic site flanked by a cytosine and an adenine. This tridecamer is stable at 4°C. At room temperature the half-life time of the tridecamer is 30 days, at 55°C it is seven hours. It is not obvious why this oligonucleotide is stable, whilst we cannot find conditions where severe degradation does not take place.

3.8 Conclusion

In this chapter we designed a repair model system for natural abasic sites. Molecular modeling showed that the repair units that we developed could in principle be accommodated in a DNA double helix. We found favorable positions for hydrogen bonding with the opposite base and π–π-stacking with the neighboring bases in the helix.

The test reactions that we performed between ribose and deoxyribose and aminobases **3-6** showed that they did not react. However, the reaction of hydrazides **7** or **8** with ribose or deoxyribose yielded two or three products respectively. The reaction between **8** and deoxyribose takes about 24h to reach a steady state situation (r.t., water, pH = 6, 6 mM of **8**, 120 mM of deoxyribose). The hydrolysis of the products takes place in approximately the same time. The reactions are slower when decreasing temperature, but the equilibrium shifts to the product side at lower temperatures. The exact identity of the products could unfortunately not be determined.

Albeit the slow reactions, we wanted to test whether **7** and **8** are able to repair abasic sites in a DNA duplex. A natural abasic site nucleotide was therefore synthesized and incorporated in a DNA oligonucleotide. Purification of the oligonucleotide and deprotection of the nitro benzyl group were so troublesome that a pure oligonucleotide containing an abasic site was never obtained. The oligonucleotide containing the abasic site was not stable under conditions where others reported them to be stable.

3.9 Experimental section

General methods

All commercial reagents were used as supplied, without further purification. Petroleum ether (PE) refers to the fraction with distillation range 40-60°C. All reactions were performed under an atmosphere of nitrogen, unless stated otherwise.

Solvents

Water was purified over a MilliQ apparatus (Millipore). Pentane, hexane, cyclohexane, benzene, toluene, dichloromethane and chloroform were dried over calcium hydride and distilled under nitrogen atmosphere or dried over an alumina column. Diethylether and THF were dried over potassium hydride and distilled under nitrogen atmosphere.

Chromatography

All purifications by column chromatography were performed on silica gel ROCC 40-68μM with 0.2 bar of air pressure.

NMR spectroscopy

NMR spectra were recorded at 25°C on a Bruker Avance 400 MHz NB US NMR Spectrometer. ^{1}H NMR spectra were recorded at 400.1 MHz. ^{13}C NMR spectra were recorded at 100.6 MHz. ^{19}F spectra were recorded at 376.3 MHz. ^{31}P spectra were recorded at 161.9 MHz . Values for δ are in ppm relative to residual solvent signal. *J*-Values are in Hz.

Molecular modeling

All molecular modeling was perfomed on Silicon graphics workstations using MacroModel version 6.5. All simulations were run in the MacroModel version of the AMBER* force field at 300K with water as a solvent using a GB/SA solvent model. An energy minimized structure of a known structure was used as a starting point for the calculations and minimization was performed using a TNCG (Truncated Newton Conjugate Gradient) [132]. None of the calculations was intended as a systematic search of energy minima (no Monte-Carlo search or Molecular Dynamics were performed).

Infra red spectroscopy

Infrared spectra were measured on a Bruker IFS-55 spectrometer. Solids were measured in KBr pellets and liquids measured as a film between two KBr pellets.

Melting point determination

All melting points were measured in capillary tubes using an Electrothermal 9100 apparatus and are non-corrected.

Mass spectrometry

Mass analyses were performed on a MALDI mass spectrometer in the reflection mode (Bruker Reflex III) using R-cyano-4-hydroxycinnamic acid as a matrix.

HPLC

Reversed phase HPLC analysis was performed on a Waters 2690 instrument equipped with a Merck RP-Select B reversed phase C_{18} column (5mm x 15 cm, flow rate 1 $mLmin^{-1}$) (Alliance chromatographic system, Waters, Saint-Quentin-en-Yvelynes, France), with detection at 230 nm.

Oligonucleotides

DNA oligonucleotides were either purchased from Eurogentec (Brussels) or were synthesized on an Expedite 8908 synthesizer (Applied Biosystems) using a standard 1μmol cycle. Coupling efficiencies for natural bases were >98%. Non-natural bases were coupled manually with increased coupling times (15-20 minutes) and a large excess of nucleotide. Coupling efficiencies were high (>90%) for the nitro benzyl protected abasic site

Chemical procedures

All molecules synthesized in this part were described in literature. We thus refer to the original papers.

5-Amino-4-chloro-6-hydrazinopyrimidine (9) [117]

5-Amino-4,6-dichloropyrimidine (100 mg; 609.7 µmoles; 1 eq.) was added rapidly to 2.5 mL (2.58 g) of hydrazine monohydrate. The copious precipitate which formed was collected by filtration, washed thoroughly with n-propyl alcohol, then with ether, and dried *in vacuo* to yield 5-amino-4-chloro-6-hydrazinopyrimidine (70.9 %; 69 mg; 432.3 µmoles) as a green solid.

NMR: δ_H (DMSO); 8.018 (broad s, 1H, NH), 7.785 (s, 1H, CH), 5.016 (broad s, 2H, NNH_2), 4.399 (broad s, 2H, CNH_2). M.p. 182-184°C. υ_{max} (KBr)/cm^{-1}: 3412, 3550-3200, 1640, 1575, 1555, 1540.

5-Amino-4-benzylidenehydrazino-6-chloropyrimidine (10) [117]

A solution of 5-amino-4-chloro-6-hydrazinopyrimidine (194 mg; 1.215 mmoles; 1 eq.) and benzaldehyde (135.9 µL; 141.9 mg; 1.337 mmoles; 1.1 eq.) in ethanol was stirred at r.t. overnight. The solid that deposited was collected by filtration. The precipitate was purified by flash column chromatography using $CHCl_3$ as a solvent to yield 5-amino-4-benzylidenehydrazino-6-chloropyrimidine (43.5 %; 131 mg; 528.8 µmoles) as a white solid.

NMR: δ_H (DMSO); 11.164 (s, 1H, NH), 8.113 (s, 1H, PhCH), 7.877 (s, 1H, N_2CH), 7.649-7.629 (m, 2H, *m*-ArH), 7.464-7.396 (m, 3H, *o*-ArH + *p*-ArH), 5.693 (broad s, 2H, NH_2). M.p. 204-206°C. υ_{max} (KBr)/cm^{-1}: 3395, 3295, 3195, 3060, 1600, 1565, 1545, 1510, 1430.

9-Benzylideneamino-6-chloropurine (11) [109]

A solution of 5-amino-4-benzylidenehydrazino-6-chloropyrimidine (32 mg; 129.1 µmoles; 1 eq.) in diethoxymethyl acetate (5 mL) was heated at 120°C for 1h. The solution was allowed to cool down and after overnight storage at 4°C filtered to afford 9-benzylideneamino-6-chloropurine (90.1 %; 30 mg; 116.4 µmoles) as a white solid.

NMR: δ_H (DMSO); 9.729 (s, 1H, PhCH), 9.189 (s, 1H, $^8H/^2H$), 8.913 (s, 1H, $^2H/^8H$), 7.985-7.971 (m, 2H, *m*-ArH), 7.662-7.586 (m, 3H, *o*-ArH + *p*-ArH). M.p. 168-170°C. υ_{max} (KBr)/cm^{-1}: 1585, 1570, 1560, 1480, 1435.

9-Benzylideneaminoadenine (12) [109]

A suspension of 9-benzylideneamino-6-chloropurine (30 mg; 116.4 µmoles; 1 eq.) in a solution of NH_3 in ethanol was heated at 110°C for 4h in an autoclave. The solution was allowed to cool and the solid was filtered off. It was washed with ethanol followed by water to afford 9-benzylideneaminoadenine (75.7 %; 21 mg; 88.14 µmoles) as a white solid.

NMR: δ_H (DMSO); 9.871 (s, 1H, PhCH), 8.550 (s, 1H, $^8H/^2H$), 8.258 (s, 1H, $^2H/^8H$), 7.931-7.907 (m, 2H, *m*-ArH), 7.718-7.653 (m, 3H, *o*-ArH + *p*-ArH), 7.459 (broad s, 2H, NH_2). M.p. 239-240°C.

9-Aminoadenine (4) [109]

To a suspension of 9-benzylideneaminoadenine (44 mg; 184.6 µmoles; 1 eq.) in $CHCl_3$/MeOH 2/1 was added methylhydrazine (19.65 µL; 17.01 mg; 369.3 µmoles; 2 eq.) and the mixture was refluxed for 17h. The solution was cooled down to 4°C and the precipitate was filtered off. The precipitate was washed with chloroform and ether to afford 9-aminoadenine (43.2 %; 12 mg; 79.92 µmoles) as a white solid.

NMR: δ_H (DMSO); 8.154 (s, 1H, $^8H/^2H$), 8.000 (s, 1H, $^2H/^8H$), 7.211 (broad s, 2H, NNH_2), 6.013 (broad s, 2H, CNH_2). M.p. >300°C. υ_{max} (KBr)/cm^{-1}: 3290, 3100, 1675, 1630, 1600, 1580, 1485, 1415.

O-Methyl guanine (13) [116]

Sodium (94.14 µL; 84.73 mg; 3.685 mmoles; 2.5 eq.) was added to methanol (10 mL) and 2-amino-6-chloropurine (250 mg; 1.474 mmoles; 1 eq.) was added. This mixture was refluxed for 24h and cooled down to r.t.. The cooled reaction mixture was poured into water (20 mL) and 1 M HCl was added until pH = 6 to give a precipitate. The resulting solution was decolorized with charcoal and filtered. The residue was washed with MeOH and the solvent was evaporated *in vacuo*. The residue was crystallized from water/MeOH 80/20 to yield 2-amino-6-methoxypurine (97.7 %; 238 mg; 1.441 mmoles) as a white solid.

NMR: δ_H (DMSO); 7.813 (s, 1H, CH), 6.227 (broad s, 2H, NH_2), 3.942 (s, 3H, CH_3). M.p. >300°C.

9-Amino-O-methylguanine (14) [115]

2-Amino-6-methoxypurine (400 mg; 2.421 mmoles; 1 eq.) was dissolved in 15 mL of 4M NaOH. Hydroxylamine-O-sulfonic acid (2.738 g; 24.21 mmoles; 10 eq.) was gradually added

to the solution and the mixture was stirred for 48h at r.t.. The reaction mixture was neutralized with 1M HCl and the products were purified by flash column chromatography using $CHCl_3$/MeOH/NH_4OH 95/4.5/0.5 as an eluent to yield 9-amino-*O*-methylguanine (71 %; 310 mg; 1.72 mmoles) as a white solid.

NMR: δ_H (DMSO); 7.715 (s, 1H, CH), 6.394 (broad s, 2H, CNH_2), 5.801 (broad 2, 2H, NNH_2), 3.948 (s, 3H, CH_3).

9-Aminoguanine (9) [115]

9-Amino-*O*-methylguanine (50 mg; 277.5 µmoles; 1 eq.) was heated in 1M HCl at 95°C for three hours and the solvent was evaporated *in vacuo*. The product was purified by flash column chromatography using $CHCl_3$/MeOH/NH_4OH 80/17/3 as an eluent. 9-Aminoguanine (97.5 %; 45 mg; 270.8 µmoles) was recrystallized from 1M HCl as a white solid.

NMR: δ_H (DMSO); 10.533 (broad s, 1H, NH), 7.540 (s, 1H, CH), 6.441 (broad s, 2H, CNH_2), 5.743 (broad s, 2H, NNH_2). M.p. >300°C.

2,4-Bis(trimethylsilyloxy)-pyrimidine (15) [110]

A mixture of uracil (2 g; 17.84 mmoles; 1 eq.) and ammonium sulfate (34.29 µL; 60.66 mg; 459.1 µmoles; 0.0257 eq.) in 1,1,1,3,3,3-hexamethyldisilazane (7.494 mL; 5.732 g; 35.52 mmoles; 1.99 eq.) was heated under reflux overnight. Distillation of the residue after removal of the solvent yielded 2,4-bis(trimethylsilyloxy)-pyrimidine (75.3 %; 4.04 g; 13.43 mmoles) as a colorless oil. The oil is very sensitive to humidity and was therefore used immediately.

NMR: δ_H ($CDCl_3$); 8.192 (d, $^3J = 5.6$ *Hz*, 1H, OCCH), 6.297 (d, $^3J = 5.6$ *Hz*, 1H, NCH) 0.374 (s, 18H, $SiCH_3$).

1-(Benzylideneamino) uracil (16) [110]

To a solution of 2,4-bis(trimethylsilyloxy)-pyrimidine (502.7 mg; 1.672 mmoles; 3 eq.) in dichloromethane was added *o*-mesitylenesulfonylhydroxylamine (120 mg; 557.4 µmoles; 1 eq.) at 0°C. The suspension was allowed to gradually warm to r.t. and then stirred for 3 h. The solvent was evaporated *in vacuo* and the residue was partitioned between water and $CHCl_3$. The aqueous layer was adjusted to pH = 6 by the addition of ion exchange resin (Amberlite IRA 98, OH form). The mixture was filtered and the solvent was evaporated *in vacuo*. A solution of benzaldehyde (226.6 µL; 236.6 mg; 2.229 mmoles; 4 eq.) in 25 mL of ethanol was added and this mixture was stirred overnight. The solvent was evaporated *in vacuo* and the

residue was triturated in water. The suspension was filtered and the filtrate was washed with water. The water was evaporated. The residue was purified by flash chromatography using $CHCl_3$/MeOH 95/5 as an eluent to yield 1-(benzylideneamino) uracil (62.1 %; 74.6 mg; 346.6 μmoles) as a white solid.

NMR: δ_H ($CDCl_3$); 9.077 (s, 1H, N=CH), 8.019 (d, ^3J = *8.2 Hz*, OCCH), 7.848-7.820 (m, 2H, *m*-ArH), 7.572-7.507 (m, 3H, *p*-ArH + *o*-ArH), 5.796 (d, ^3J = *8.2 Hz*, 1H, NCH).

1-Aminouracil (3)[114]

To a suspension of 1-(benzylideneamino) uracil (100 mg; 464.6 μmoles; 1 eq.) in chloroform-methanol (2/1) was added methyl hydrazine (74.16 μL; 64.22 mg; 1.393 mmoles; 3 eq.) and the mixture was stirred for 6h. The solid was filtered off and washed with chloroform. Further solid was obtained by concentration of the filtrate to afford 1-aminouracil (59.2 %; 35 mg; 275.3 μmoles) as a white solid.

NMR: δ_H (DMSO); 11.310 (broad s, 1H, NH), 7.611 (d, ^3J = *8.1 Hz*, OCCH), 5.473 (broad s, 2H, NH_2), 5.397 (d, ^3J = *8.1 Hz*, NCH). M.p. 247-250°C. υ_{max} (KBr)/cm^{-1}:3292 3201 2986 2802 1686.

1-(Benzylideneamino) cytosine (17) [109, 111-113]

Triethylamine (273.6 μL; 199.1 mg; 1.968 mmoles; 2 eq.) was added to a mixture of 1-(benzylideneamino) uracil (200 mg; 984.2 μmoles; 1 eq.), 4-dimethylaminopyridine (240.4 mg; 1.968 mmoles; 2 eq.) and 1,3,5-triisopropylbenzenesulfonyl fluoride (563.7 mg; 1.968 mmoles; 2 eq.) in acetonitrile (10 mL). After the mixture was stirred for 22 h at r.t. under a nitrogen atmosphere, concentrated NH_4OH (6 mL) was added to the mixture and the whole was stirred for a further three hours at r.t.. The solvent was removed *in vacuo,* the residue was suspended in $CHCl_3$ and filtered over a glass filter. The residue was purified by flash chromatography using EtOAc/MeOH 90/10 as an eluent to yield 1-(benzylideneamino) cytosine (81.8 %; 163 mg; 806 μmoles) as a white solid.

NMR: δ_H ($CDCl_3$); 9.457 (s, 1H, N=CH), 7.824-7.786 (m, 3H, m-ArH + H_2NCCH), 7.520-7.479 (m, 3H, *p*-ArH + *o*-ArH), 5.790 (d, ^3J = *8.2 Hz*, 1H, NCH). M.p. 218-221°C υ_{max} (KBr)/cm^{-1}: 3360 3185 3090 1680 1840 1805 1595 1520 1485.

1-Aminocytosine (5) [109]

To a suspension of 1-(benzylideneamino) cytosine (163 mg; 806 µmoles; 1 eq.) in chloroform/methanol 2/1 was added methyl hydrazine (128.6 µL; 111.4 mg; 2.418 mmoles; 3 eq.) and the mixture was stirred for 6h. The solid was filtered off and washed with chloroform. Further solid was obtained by concentration of the filtrate to afford 1-amino cytosine (94.4 %; 96 mg; 761.1 µmoles) as a white solid.

NMR: δ_H (DMSO); 7.558 (d, ^3J = *7.7 Hz*, 1H, H_2NCCH), 6.938 (broad s, 2H, CNH_2), 5.547 (d, ^3J = *7.7 Hz*, 1H, NCH), 5,399 (broad s, 2H, NNH_2). M.p. 252-256°C υ_{max} (KBr)/cm^{-1}: 3365 3320 1660 1620 1525 1480 1385.

n-Decyl isoorotate (19) [118-121]

2,4-Dihydroxypyrimidine-5-carboxylic acid hydrate (1.3 g; 7.466 mmoles; 1 eq.) was added to a solution of decyl alcohol (1.568 mL; 1.299 g; 8.213 mmoles; 1.1 eq.), 4-dimethylaminopyridine (9.121 mg; 74.66 µmoles; 0.01 eq.) and N,N'-dicyclohexylcarbodiimide (1.544 mL; 1.925 g; 9.333 mmoles; 1.25 eq.) in dry $CHCl_3$. The mixture was stirred at r.t. overnight. The solvent was evaporated *in vacuo* and the residue was crystallized from MeOH/water 90/10 and filtered to yield isoorotic acid decyl ester (50.1 %; 1.11 g; 3.745 mmoles) as a white solid.

NMR: δ_H (DMSO); 11.565 (broad s, 1H, 2-NH), 11.288 (broad s, 1H, 4-NH), 8.089 (s, 1H, CH), 4.099 (t, ^3J = *6.6 Hz*, 2H, OCH_2), 1.336-1.249 (m, 16H, 8 CH_2). M.p. 190-194°C. υ_{max} (KBr)/cm^{-1}: 3096, 1730, 1704, 1666, 1629, 1493, 1399, 1261.

Uracil-5-carboxyhydrazide (7) [118-121]

A suspension of isoorotic acid decyl ester (137 mg; 462.2 µmoles; 1 eq.) in 2 mL of hydrazine hydrate was heated under reflux for 15 minutes and the solution was cooled to r.t.. Methanol (1mL) was added and on chilling, uracil-5-carboxyhydrazide (44.5 %; 35 mg; 205.7 µmoles) precipitated. The product was recrystallized from water.

NMR: δ_H (DMSO); 11.579 (broad s, 1H, 2-NH), 11.291 (broad s, 1H, 4-NH), 9.439 (broad s, 1H, H_2NNH), 8.095 (s, 1H, CH), 4.511 (broad s, 2H, NH_2) M.p. >300°C. υ_{max} (KBr)/cm^{-1}: 3236, 3077, 3030, 1770, 1730, 1669, 1642, 1587, 1504, 1445.

5-(*p*-Anisyldiphenyl)methoxy-*o*-nitrobenzyl 2-deoxy-*D*-ribofuranoside (26) [129]

o-Nitrobenzyl 2-deoxy-*D*-ribofuranoside (200 mg; 740 µmoles; 1 eq.) was dissolved in anhydrous pyridine (5 mL) and triethylamine (0.1 mL). The solution was cooled down to -10°C and a solution of 4,4'-dimethoxytrityl chloride (263.2 mg; 777 µmoles; 1.05 eq.) in 5 mL of pyridine was added dropwise. The reaction was stirred for 3h at -10°C and subsequently 1h at r.t..

The reaction was quenched by the addition of 3 mL of methanol. The solution was evaporated *in vacuo* to a yellow syrup and co-evaporated 3 times with toluene to remove traces of pyridine. The residue was dissolved in ethyl acetate and washed with saturated aqueous $NaHCO_3$ solution, water and brine. The organic phase was dried on Na_2SO_4, filtered and the solvent was evaporated *in vacuo*. The residue was purified by flash column chromatography using $CHCl_3$/MeOH/Et_3N 97/2/1 as an eluent to yield 5-(*p*-anisyldiphenyl)methoxy-*o*-nitrobenzyl 2-deoxy-*D*-ribofuranoside (33.9 %; 144 mg; 251.4 µmoles) as a light yellow viscous oil.

NMR: δ_H ($CHCl_3$); 8.019 (d, ^3J = *8.0 Hz*, 1H, O_2NCCH), 7.569-7.510 (m, 2H, O_2N-ArH), 7.424-7.376 (m, 3H, 1 O_2N-ArH + 2 ArH), 7.295 (d, ^3J = *8.8 Hz*, 4H, H_3COCCH), 7.239-7.139 (m, 2H, ArH), 6.821 (m, 1H, ArH), 6.759 (d, ^3J = *8.8 Hz*, 4H, $H_3COCCHCH$), 5.303 (dd, ^3J = *5.3 Hz*, ^3J = *2.2 Hz*, 1H, $H^{1'}$), 5.020 (d, ^2J = *15.1 Hz*, 1H, $PhCH_2$), 4.843 (d, ^2J = *15.1 Hz*, 1H, $PhCH_2$), 4.512-4.467 (m, 1H, $H^{3'}$), 4.061-4.021 (m, 1H, $H^{4'}$), 3.747 (s, 6H, 2 OCH_3), 3.281-3.244 (m, 1H, $H^{5'}$), 3.202-3.163 (m, 1H, $H^{5'}$), 2.345-2.297 (m, 1H, $H^{2'}$), 2.184-2.121 (m, 1H, $H^{2'}$) δ_C ($CHCl_3$); 144.7, 135.9, 134.7, 133.4, 129.9, 129.1, 128.7, 128.2, 128.1, 127.7, 126.7, 124.5, 113.0, 103.9, 86.0, 85.0, 72.7, 66.1, 64.8, 55.1, 41.2.

5-(*p*-Anisyldiphenyl)methoxy-3-*O*-(2-cyanoethoxy(diisopropylamino)-phosphino)-*o*-nitrobenzyl 2-deoxy-*D*-ribofuranoside (27) [129]

5-(*p*-anisyldiphenyl)methoxy-*o*-nitrobenzyl-2-deoxy-*D*-ribofuranoside (144 mg; 251.4 µmoles; 1 eq.) was dissolved in anhydrous dichloromethane (2 mL) and diisopropylethylamine (175.2 µL; 130 mg; 1.005 mmoles; 4 eq.) and 2-cyanoethyl N,N-diisopropylchloro- phosphoramidite (84.14 µL; 89.27 mg; 377.2 µmoles; 1.5 eq.) in anhydrous dichloromethane (2 mL) were added. The reaction was stirred at r.t. for 1h and quenched with methanol. The mixture was extracted with ethyl acetate twice (10mL). The solvent was evaporated *in vacuo* and the residue was purified by flash chromatography using PE/EtOAc/Et3N 79/20/1 as an eluent to yield 5-(*p*-anisyldiphenyl)methoxy-3-*O*-(2-

cyanoethoxy(diisopropylamino)-phosphino)-*o*-nitrobenzyl 2-deoxy-*D*-ribofuranoside (57.9 %; 113 mg; 145.8 μmoles) as a light yellow viscous oil.

NMR: δ_H ($CHCl_3$); 8.030 (d, ^3J = *8.0 Hz*, 1H, O_2NCCH), 7.613-7.508 (m, 2H, ArH), 7.444-7.377 (m, 3H, ArH), 7.333-7.301 (m, 4H, H_3COCCH), 7.236-7.145 (m, 3H, ArH), 6.787-6.735 (m, 4H, $H_3COCCHCH$), 5.375-5.329 (m, 1H, $H^{1'}$), 5.082 (d, ^2J = *15.1 Hz*, 1H, $PhCH_2$), 4.899 (d, ^2J =*15.1 Hz*, 1H, $PhCH_2$), 4.684-4.526 (m, 1H, $H^{3'}$), 4.206-4.116 (m, 1H, $H^{4'}$), 3.749 (s, 6H, H_3CO), 3.635-3.505 (m, 2H, $POCH_2$), 3.331-3.135 (m, 2H, $NCCH_2$), 2.442-2.165 (m, 2H, $H^{2'}$), 1.204-1.138 (m, 14H, NCH + $NCHCH_3$). δ_C ($CHCl_3$); 133.8, 130.5, 129.2, 129.1, 128.6, 128.1, 127.0, 125.0, 113.3, 104.5, 86.2, 85.1, 74.1, 66.7, 64.7, 58.7, 55.6, 43.5, 41.2, 25.0, 20.6. δ_P ($CHCl_3$); 149.5, 149.2.

4 Design of an artificial abasic site repair system

4.1 Introduction and goal

In chapter 3 I discussed our efforts to design a model system of natural abasic site repair. In these experiments we allowed the aldehyde function of the abasic site to react with hydrazides to form acyl hydrazones. The kinetics of these reactions turned out to be slow and practically irreversible on the time scale of our experiments. In a context where several hydrazides would compete to react with an abasic site, the outcome of the reaction is determined by the first hydrazide that reacts. If this hydrazide does not stabilize the duplex (for example because it does not match with the corresponding base), it cannot be replaced rapidly by a better match. We would now like the system to be able to self-correct such possible errors or mismatches. For that we need a fast, reversible reaction that allows the release of a wrongly incorporated repair unit and the selection of a unit that stabilizes the duplex best.

For example, we would like to incorporate an abasic site in a DNA strand (Figure 4.1 left) and mix it with four possible repair units, mimicking the four natural bases. If a guanine repair unit is incorporated opposite to an adenine (Figure 4.1 middle), this guanine will be deselected and replaced by a thymine during the experiment (Figure 4.1 right). If the reactions are not reversible, this cannot happen. The wrongly incorporated guanine remains in the DNA duplex until the end of the experiment.

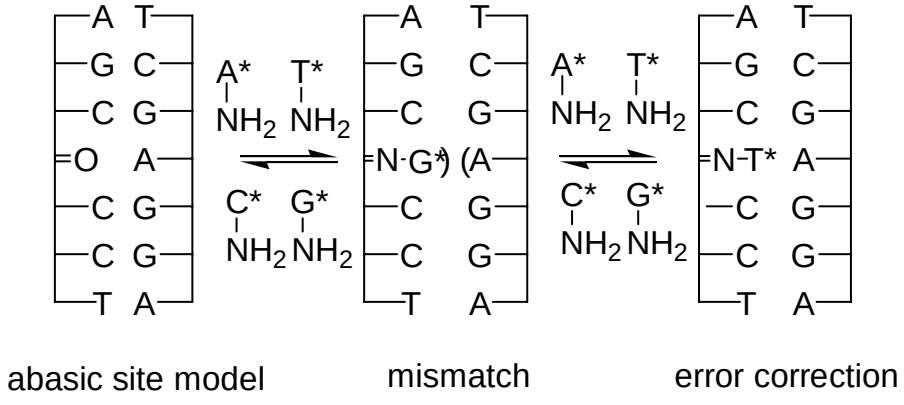


Figure 4.1. Principle of dynamic repair of DNA abasic sites. A*, T*, C*, G*: any group complementary to T, A, G, C respectively.

This chapter will focus on the design and synthesis of abasic site mimics. Contrary to chapter 3, the natural abasic site will not be considered a given object. We intend to incorporate an aldehydic function in the DNA backbone as an anchor point for the reagents and to optimize it for our needs.

The aldehyde has to fulfill two conditions. First, to optimize aromatic stacking and hydrogen bonding to the complementary base, the condensation products of the aldehyde should be located in a position as close to the natural situation as possible. Second, to make the aldehyde more reactive and the condensation products more hydrolyzable than with a natural abasic site, we want an aldehyde that is not able to form a cyclic hemiacetal (see chapter 3). Instead of being available only 1% of the time, the aldehyde should ideally be present 100% of the time. This will increase its apparent reactivity.

4.2 Design of the amines and the aldehydes

4.2.1 Choice of the reaction

As in chapter 3, dynamic repair model systems of abasic sites will consist of the reaction between aldehydes and hydrazides. Since we have to newly design both the amine and the aldehyde, we can choose between placing the amine in the backbone and the aldehyde on the base, or the reverse, placing the amine on the base-side and incorporate the aldehyde function in the backbone.

We choose for the latter solution, incorporation of the aldehyde in the backbone and incorporation of the amine on the base side. No examples of incorporation of an aldehyde function into a DNA strand had been published when this work was started. However, synthesis of hydrazides mimicking the four natural bases is known (Figure 4.2) [109-117, 133].

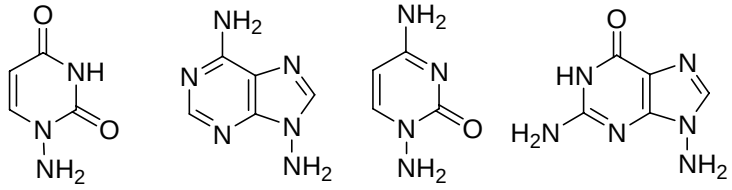


Figure 4.2. Four aminobases that can react reversibly with aliphatic aldehydes, 1-aminouracil, 9-aminoadenine, 1-aminocytosine and 9-aminoguanine.

Test reactions pointed out that the reaction between the aminobases and aliphatic aldehydes can be made reversible under physiological conditions [93]. Aliphatic amines do not react with aldehydes, since in water at pH 7 they are predominantly present as their protonated form. They are thus not nucleophilic. On the other hand, strong nucleophiles like oximes or hydrazides react very well, but the products do not hydrolyze. Hydrazides with electron withdrawing groups like in the case of the aminobases depicted in Figure 4.2 however, are sufficiently nucleophilic to be able to react with aliphatic aldehydes and the resulting imines

can still hydrolyze to give back the starting material. The four aminobases turned out to have all a similar reactivity, as proven with the reactions that were performed with isobutyraldehyde [123].

4.2.2 Aldehyde design

Figure 4.3 shows two possibilities to position an aldehyde function in a DNA backbone. The ring of the ribose has been cut at two different positions, to form two different aldehydes (Figure 4.3 middle right). If the aldehydes react with one of the aminobases (Figure 4.3, middle left), the aminobase can in principle be located in the same position as in a natural base (Figure 4.3, left). Since no free hydroxyl group is present in the molecule, the aldehyde cannot form a cyclic hemiacetal.

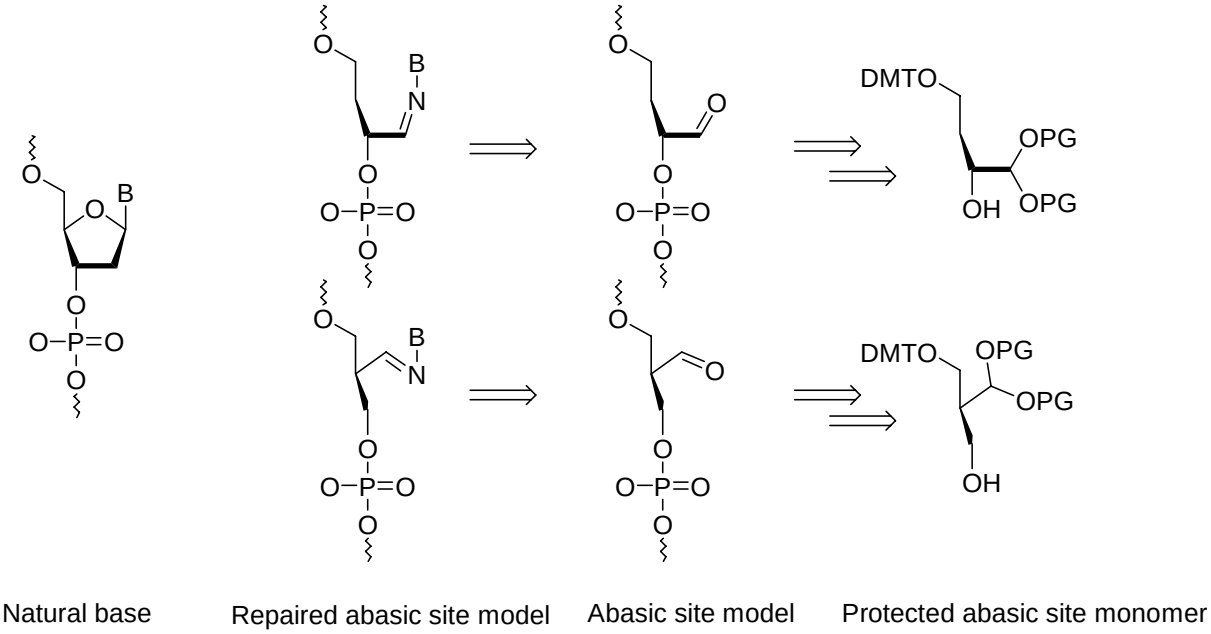
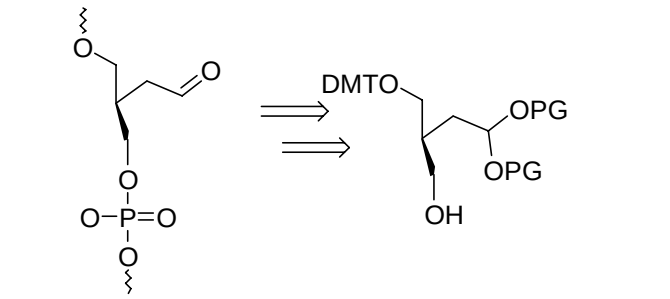


Figure 4.3 Abasic site design through ribose ring opening. PG = protecting group.

However, the loss of the sugar ring in the designed artificial abasic site increases the degrees of freedom of the system. Thus the product may find a better position for π–π stacking and hydrogen bonding compared to a natural, constrained nucleobase, creating a more stable duplex. But more likely the cost in entropy upon product formation may be higher, resulting in less stable binding.

A third compound was considered, containing an extra atom linking the base and the backbone (Figure 4.4). This extra carbon atom brings additional flexibility in the spacer between base and backbone. It may allow the molecule to find a good stacking and hydrogen bonding position in the DNA double helix, despite the fact that the spacer now contains more atoms than in a deoxyribose.



Abasic site model Protected abasic site monomer

Figure 4.4. Alternative homologous abasic site design. PG = protecting group.

Epple *et al.* had a closer look at the effect of the flexibility of backbones consisting of various modified nucleotides on the stability of double helices [134]. They concluded that binding to a complementary DNA strand is best if either the backbone or the linker between the base and the backbone is flexible, and the other is rigid. If both backbone and linker are rigid, a good position for π–π stacking is not likely to be found, yielding less stable duplexes. If both backbone and linker are flexible, the entropy cost upon binding is too high to bind strongly to a complementary strand. If either backbone or linker is rigid and the other flexible, a good position for π–π-stacking and hydrogen bonding can be adopted for a significantly lower cost in entropy.

Their research was focused on oligonucleotides consisting of the modified monomers exclusively. In our case, only one modified monomer should be incorporated in a DNA oligomer. The entropy cost of this one unit may thus be lower than expected from the calculations of Epple *et al.*

The artificial abasic sites shown in Figure 4.3 and Figure 4.4 were developed to have a better reactivity than the natural abasic site which is predominantly present as its hemiacetal form. A look at the structure of the aldehyde monomers shows that it is possible to form cyclic structures with the phosphate group (Figure 4.5). This may mask the aldehyde function part of the time, making it less available for reacting with the hydrazides.

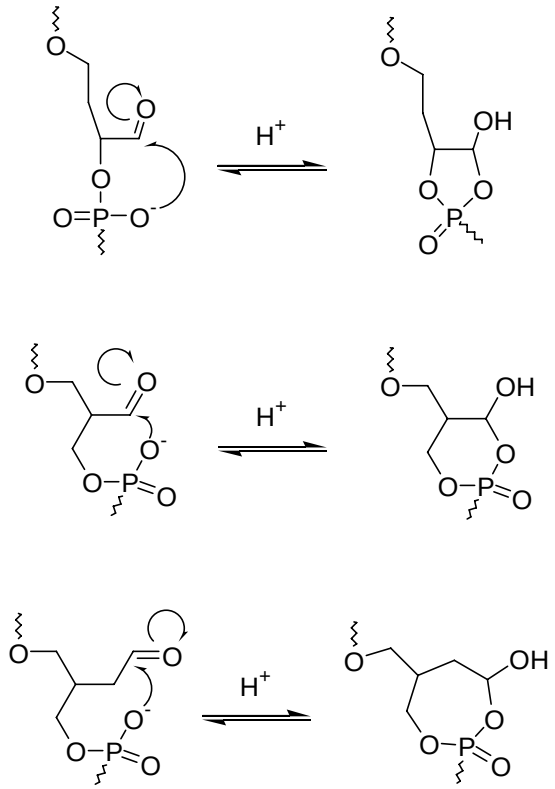


Figure 4.5. Possible cyclization reactions in artificial abasic sites

From Figure 4.5 it can be seen that the three designed compounds would form a 5, 6 and 7 membered rings with the phosphate respectively. The seven membered ring is unlikely, but the occurrence of five and six membered rings can be envisaged more likely.

However, in literature these cyclic hemiacetal structures are not known. Only cyclic phosphate acetals have been described [135]. It can be assumed that the aldehydes of Figure 4.3 and Figure 4.4 should be available for reaction with hydrazides.

4.2.3 Molecular Modeling

Similar to the energy minimization experiments performed with the natural abasic site (paragraph 3.3.3), we wanted to validate the design of an artificial abasic site before starting any experiment. We performed energy minimization of the products of the reactions between the three artificial abasic sites **1-3** and 1-aminouracil and 1-aminocytosine. The calculations were again not intended as systematic searches of energy minima (no Monte-Carlo search or Molecular Dynamics were performed), but only as simple preliminary studies to evaluate whether energy minima exist where Watson-Crick type base pairing occur with the repair unit. We used an AMBER* force field implemented in the MacroModel software.

We were looking whether our artificial repair units would find a position favorable for hydrogen bonding with the opposite base and π–π-stacking with its neighboring base pairs.

The positions that were found may not be the optimal position for the repair unit, but when a good position is found, it is likely that the modified base can successfully 'repair' a DNA duplex.

A selection of energy minimizations of the reaction products of abasic sites **A**, **B** and **C** is presented in sections 4.2.3.1, 4.2.3.2 and 4.2.3.3 respectively.

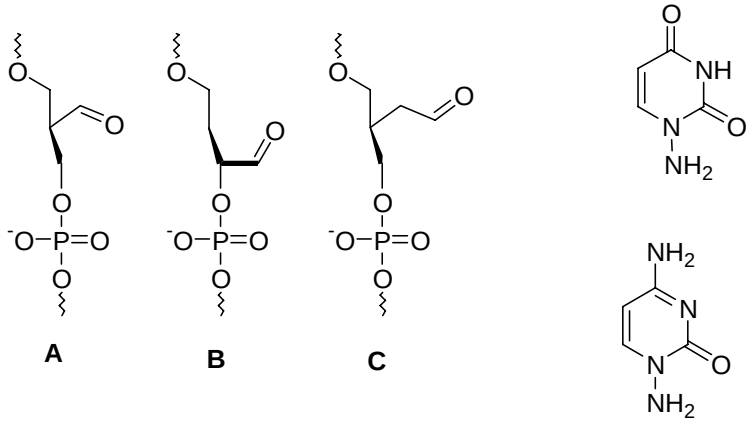


Figure 4.6. The three artificial abasic sites and 1-aminouracil and 1-aminocytosine that were used for energy minimizations.

4.2.3.1 Modeling of artificial abasic site A

Similar to the energy minimizations performed for the natural abasic site as described in section 3.3.3, we took the known decameric DNA duplex ($d^{5'}$(CCA GGC CTG G)$^{3'}$)$_2$ [104-106]. A crystal structure of the B-helix of this sequence was available. We modeled our modification in the middle of the strand, at the fifth base pair. The first product that we modeled is the reaction product of the reaction between abasic site **A** and aminocytosine. This product has a linker that is attached to the backbone at the carbon atom that corresponds to the 4'carbon in a deoxyribose moiety.

To obtain a correct position for hydrogen bonding and π–π-stacking, we started from the natural, energy minimized strand and then cut the 2'carbon atom out of the deoxyribose ring. We replaced the oxygen atom of the deoxyribose ring by a carbon atom, replaced the 1'carbon atom by a nitrogen atom and added a double bond to give the desired hydrazone.

The hydrazone linker that is formed has two possible configurations, *cis* (Figure 4.7) and *trans* (Figure 4.8). We modeled both and looked what the consequences for the base mimic were. In a first energy minimization we cut the 4'hydrogen atom to give the system a maximum of freedom. After the first energy minimization, the hydrogen was added back and the structure was minimized again.

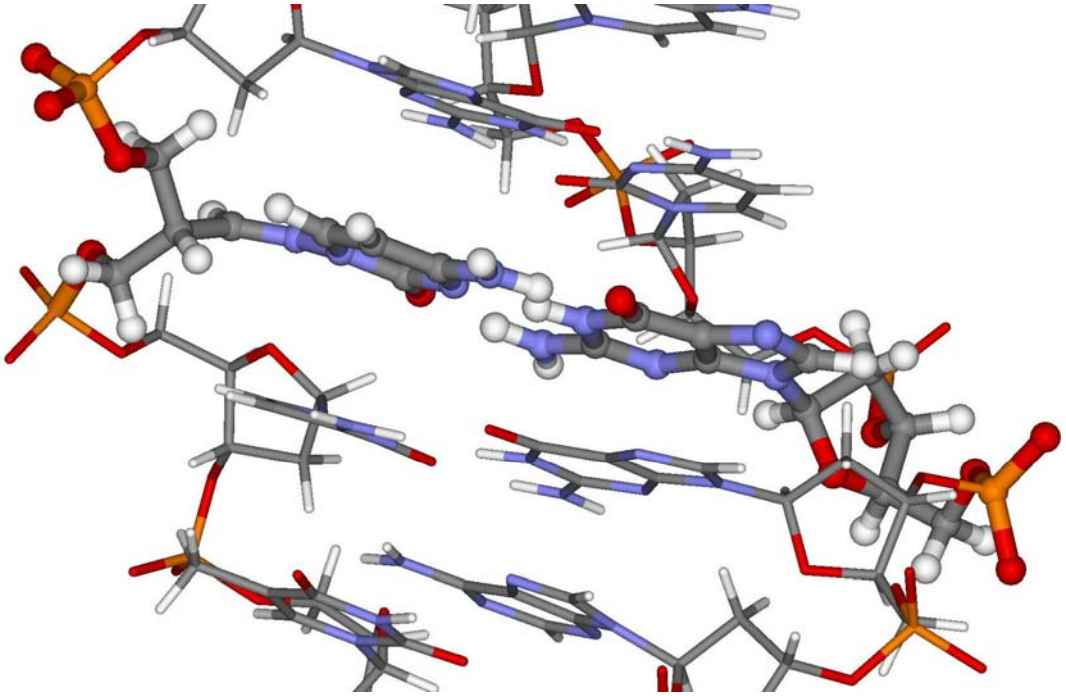


Figure 4.7. Energy minimization of 1-aminocytosine at the abasic site position in its hydrazone form. The hydrazone is in its *cis* configuration and the $C^{4'}$ has an R configuration. The O(=C) of the guanine and the $N(H_2)$ of the 1-aminocytosine are at a distance of 2.75 Å. The 1-aminocytosine and the opposite guanine are tilted at an angle of about 10 degrees.

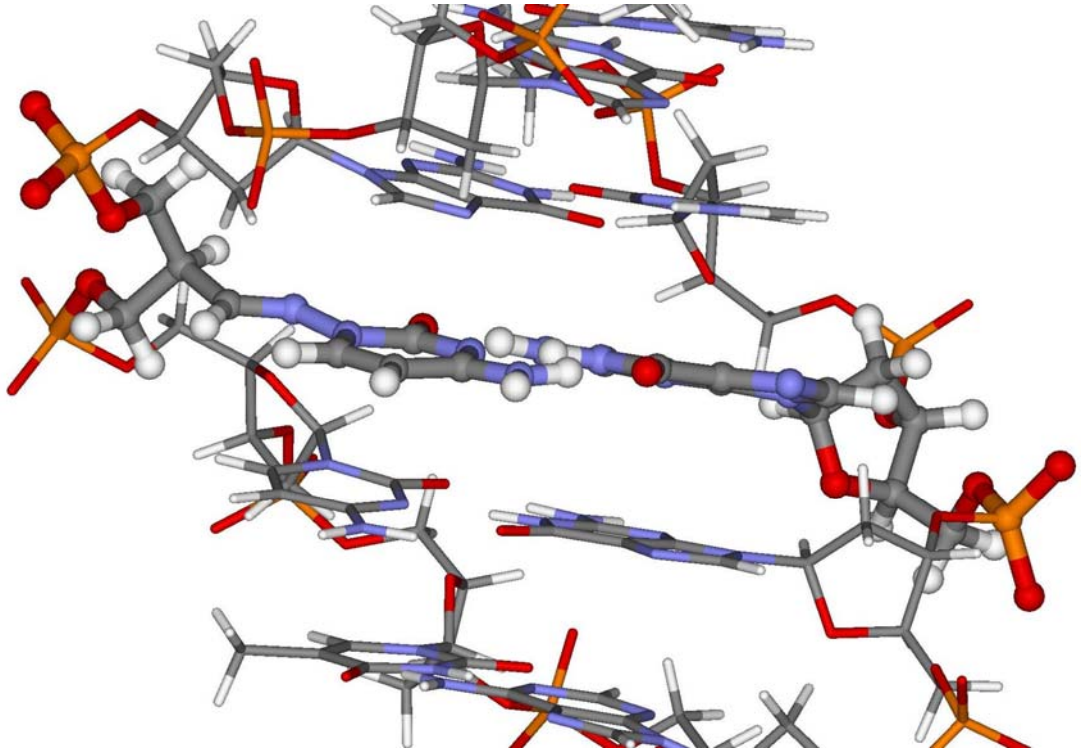


Figure 4.8. Energy minimization of 1-aminocytosine at the abasic site position in its hydrazone form. The hydrazone is in its *trans* configuration and the $C^{4'}$ has an S configuration. The O(=C) of the guanine and the $N(H_2)$ of the 1-aminocytosine are at a distance of 2.75 Å. The 1-aminocytosine and the opposite guanine are tilted at an angle of about 12 degrees.

It turned out that the more favorable stereochemistry of the 4'carbon depended on whether the hydrazone was either *cis* or *trans*. A *cis*-hydrazone results in an *R*-conformation of the 4'carbon, a *trans*-hydrazone results in an *S*-conformation of the 4'carbon. Both structures find a position that is favorable for Watson-Crick hydrogen bonding. The 4'carbon can in principle racemize *via* a keto-enol equilibrium. The abasic site may be present as both the R- and the S-conformation and after reaction the more favorable conformation may be predominant.
The additional flexibility in the backbone allowed the repair unit to find a favorable position easily, without major conformational changes in the rest of the DNA duplex. The conformation of the linker seems not to have much influence on the position of the base. The angle between the base mimic and the opposite base and the distance between the bases is approximately the same for the two models.

After modeling of the cytosine-guanine base pair, we changed the base pair to a uracil-adenine base pair. We minimized two structures, the trans hydrazone with *S*-chirality at its 4'carbon and its corresponding hydrazone hydrate. The hydrazone hydrate can be formed by simple addition of water to the hydrazone and may occur in the repaired abasic site. It is an additional conformation that increases the chances that a good position will be found. In Figure 4.9 and Figure 4.10 it can be seen that for both of the structures a good position for Watson-Crick hydrogen bonding can be found. Moreover, the position of the bases in the two structures is almost the same. The difference in the angle between the uracil mimic and the adenine is only one degree and the distance between the bases in the two minimized structures differs by a non-significant 0.01 Å.
The hydrazone hydrate gives rise to a second asymmetric carbon atom and both the R- and the S-hydrate can in principle be formed. The compound with *S*-chirality is depicted here, but the almost symmetrical backbone will presumably also accommodate the structure with the R-chirality.

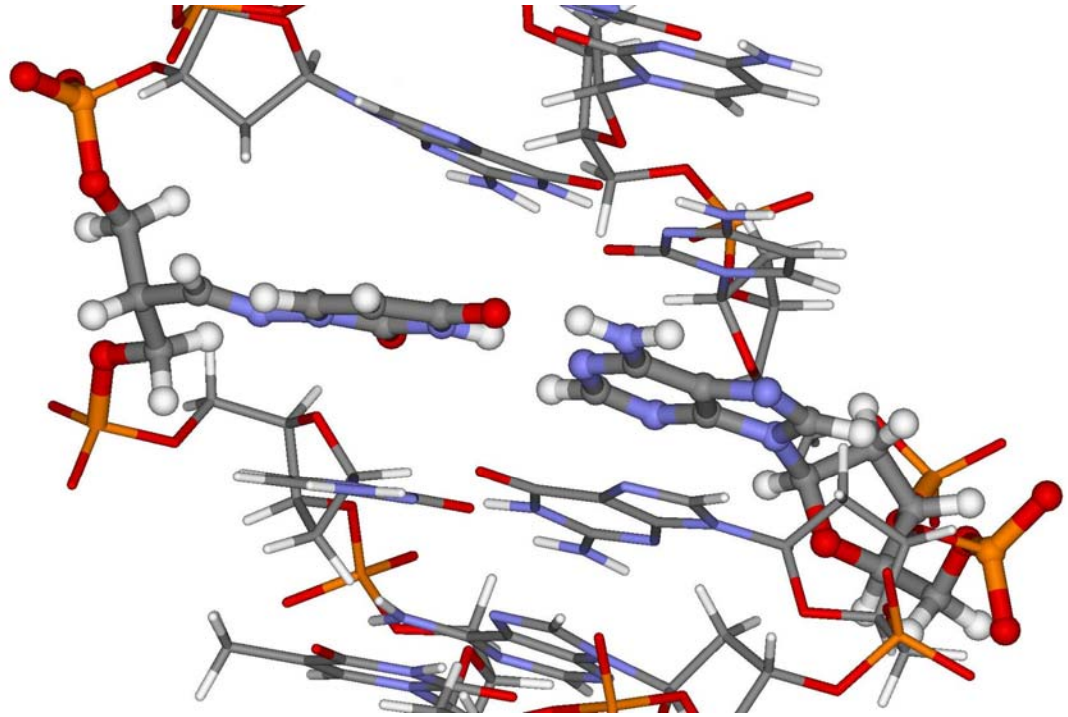


Figure 4.9. Energy minimization of 1-aminouracil at the abasic site position in its hydrazone form. The hydrazone is in its *trans* configuration and the $C^{4'}$ has an R configuration. The $N(H_2)$ of the adenine and the $O(=C^4)$ of the 1-aminouracil are at a distance of 2.73 Å. The 1-aminouracil and the opposite adenine are tilted at an angle of about 13 degrees.

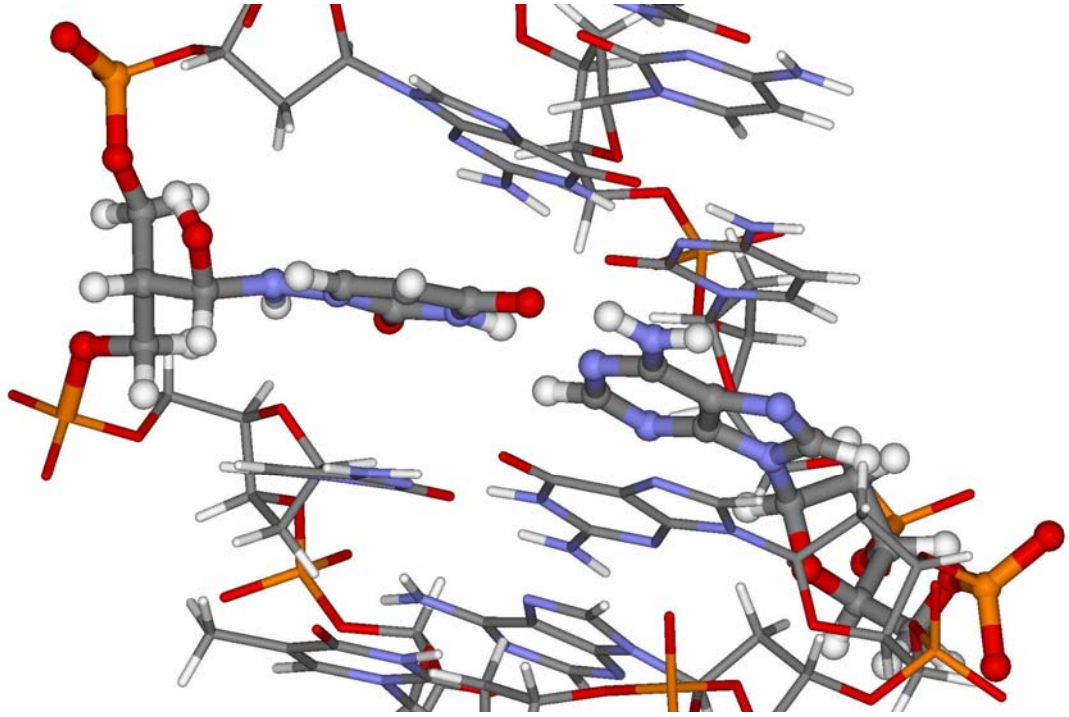


Figure 4.10. Energy minimization of 1-aminouracil at the abasic site position in its hydrazone hydrate form. Both $C^{4'}$and $C^{1'}$ have an S configuration. The $N(H_2)$ of the adenine and the $O(=C^4)$ of the 1-aminouracil are at a distance of 2.72 Å. The 1-aminouracil and the opposite adenine are tilted at an angle of about 14 degrees.

4.2.3.2 Modeling of artificial abasic site B

The second reaction product that we modeled has a linker that is attached to the backbone at the carbon atom that corresponds to the $^{3'}$carbon atom in a deoxyribose moiety. We modeled a cytosine-guanine and an adenine-uracil base pair, and for both of these base pairs a hydrazone and a hydrazone hydrate form.

To obtain a correct position for hydrogen bonding and π–π-stacking, we started again from the natural, energy-minimized strand and then cut the oxygen atom out of the deoxyribose ring. We replaced the 1'carbon atom by a nitrogen atom and added a double bond to give the desired hydrazone.

With the cytosine mimic in its hydrazone form, the distance between the cytosine mimic and the facing guanine is almost the same as in the case where the linker was attached to the backbone at the 4'carbon. However, the angle between the bases is significantly reduced to an almost coplanar conformation (4 degrees, Figure 4.11). In the hydrazone hydrate, the angle between the bases is 12 degrees (Figure 4.12), which is similar to the situation with the linker attached to the 4'carbon. The distance between the two bases is almost the same compared to the situation with the linker attached to the 4'carbon atom.

The uracil mimic linked to the backbone at the 3'carbon facing an adenine, showed similar behavior compared to the one linked to the 4'carbon atom. We tested both a hydrazone (Figure 4.13) and a hydrazone hydrate (Figure 4.14) structure. The 3'carbon can racemize *via* a keto-enol equilibrium, similar to artificial abasic site **A**. The configuration that is able to stabilize a DNA double helix best will be predominant when performing this reaction.

The hydrazone we modeled had a *trans*-configuration, with the 3'carbon in an *R*-configuration. An *S*-configuration on the 3'carbon is normally favored by a *trans*-hydrazone, but for this configuration a good position for Watson-Crick hydrogen bonding was found as well. The base is tilted at approximately the same angle and the distance between the bases is equivalent for both hydrazone and hydrazone hydrate, compared to their respective structures linked to the 4'carbon atom.

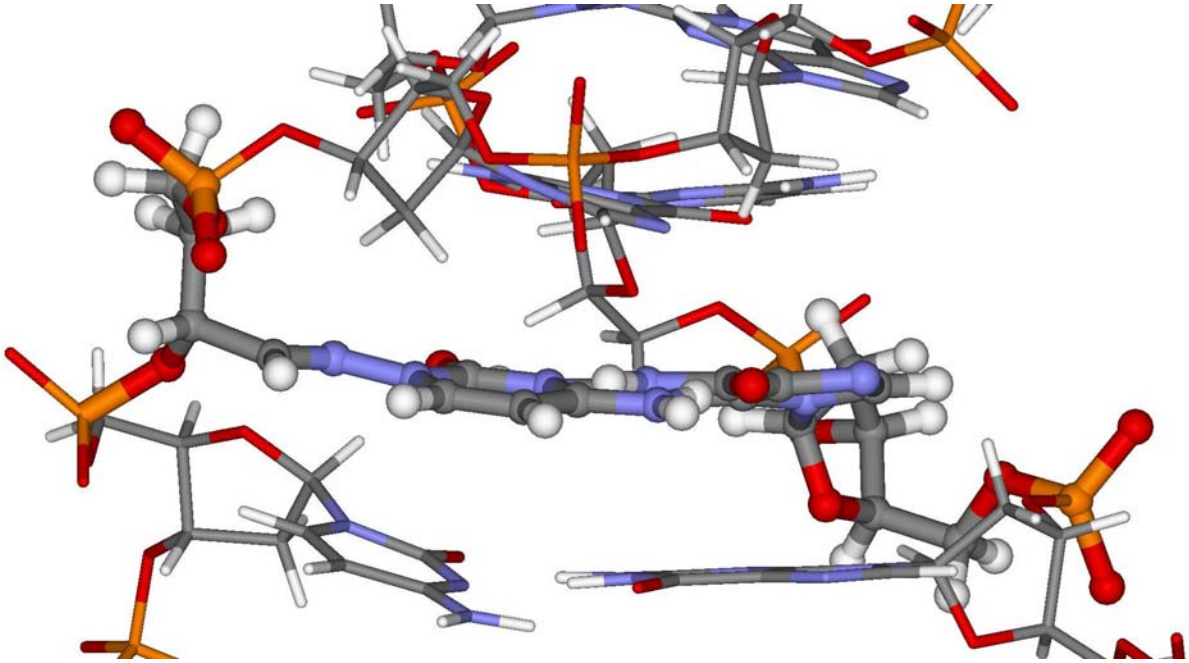


Figure 4.11. Energy minimization of 1-aminocytosine at the abasic site position in its hydrazone form. The hydrazone is in its *trans* configuration and the $C^{3'}$ has an R configuration. The O(=C) of the guanine and the $N(H_2)$ of the 1-aminocytosine are at a distance of 2.75 Å. The 1-aminocytosine and the opposite guanine are tilted at an angle of about 4 degrees.

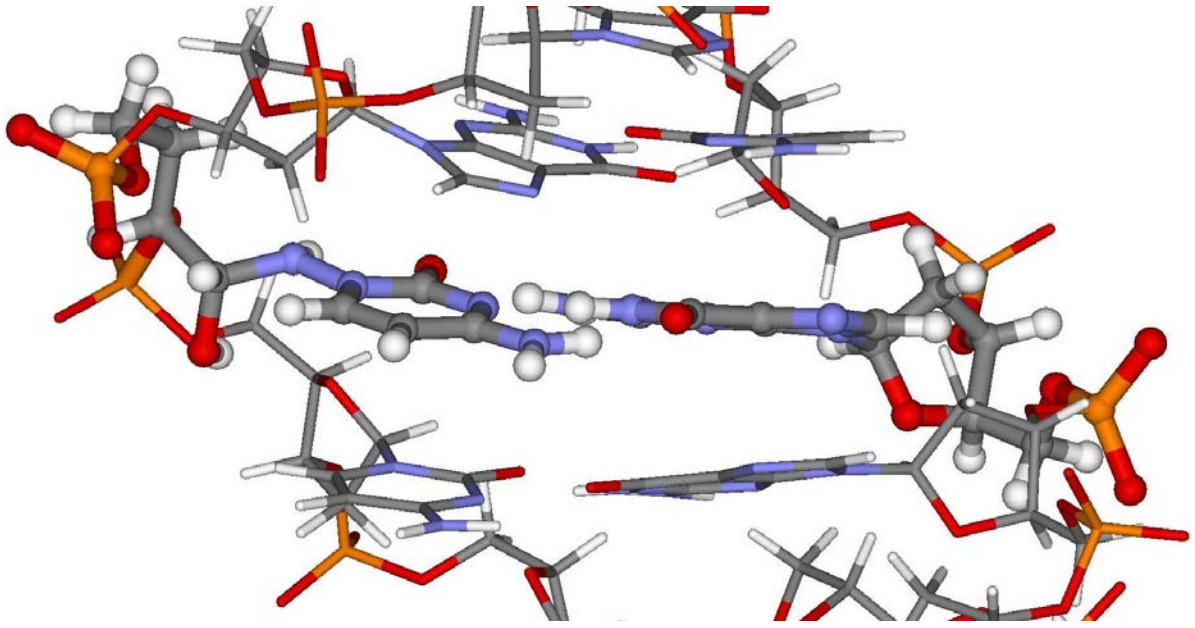


Figure 4.12. Energy minimization of 1-aminocytosine at the abasic site position in its hydrazone hydrate form. Both $C^{2'}$ and $C^{3'}$ have an R configuration. The O(=C) of the guanine and the $N(H_2)$ of the 1-aminocytosine are at a distance of 2.75 Å. The 1-aminocytosine and the opposite guanine are tilted at an angle of about 12 degrees.

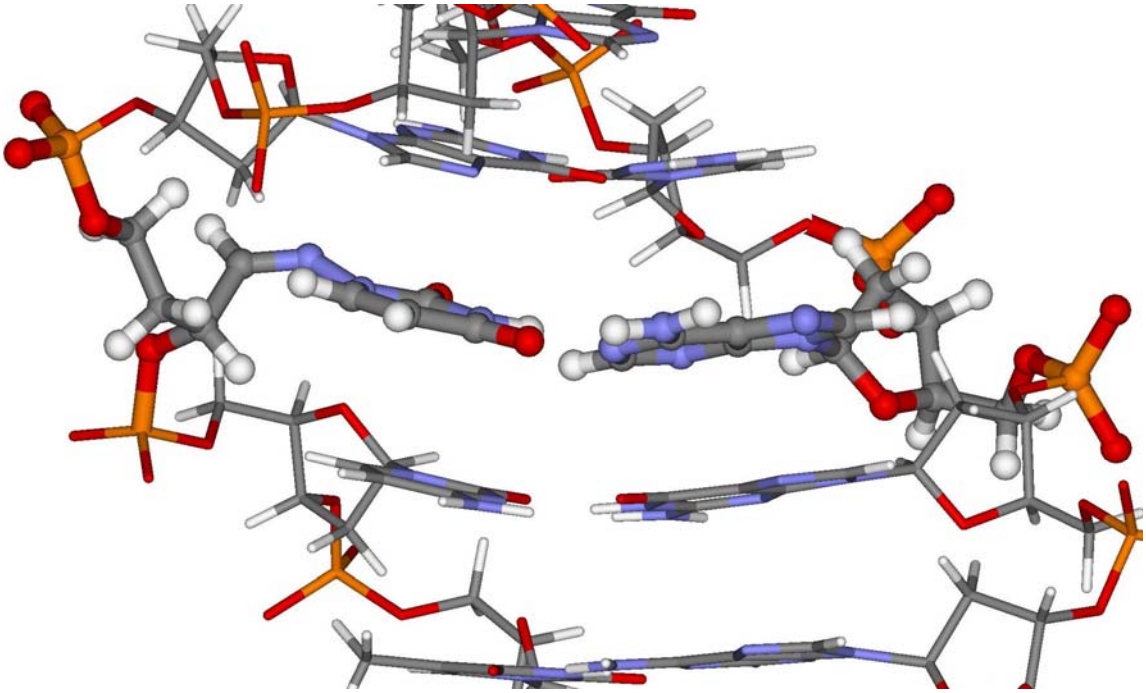


Figure 4.13. Energy minimization of 1-aminouracil at the abasic site position in its hydrazone form. The hydrazone is in its *cis* configuration and the $C^{3'}$ has an R configuration. The $N(H_2)$ of the adenine and the $O(=C^4)$ of the 1-aminouracil are at a distance of 2.82 Å. The 1-aminouracil and the opposite adenine are tilted at an angle of about 15 degrees.

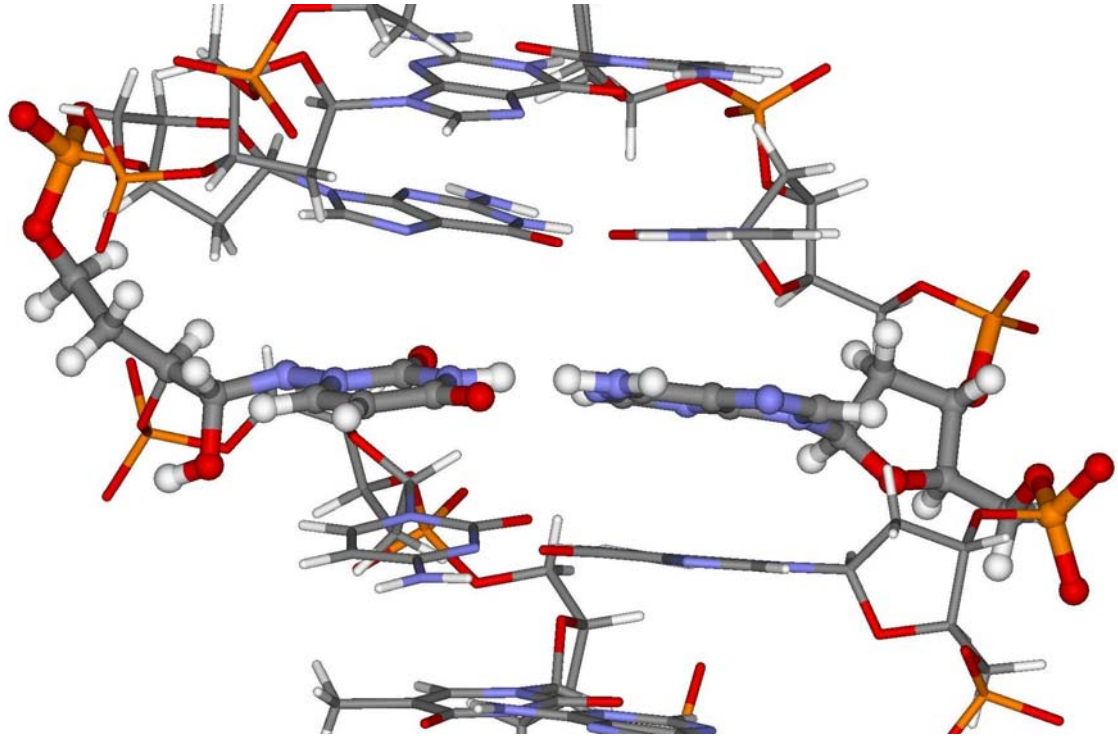


Figure 4.14. Energy minimization of 1-aminouracil at the abasic site position in its hydrazone hydrate form. Both $C^{2'}$ and $C^{3'}$ have an R configuration. The $N(H_2)$ of the adenine and the $O(=C^4)$ of the 1-aminouracil are at a distance of 2.76 Å. The 1-aminouracil and the opposite adenine are tilted at an angle of about 8 degrees.

4.2.3.3 Modeling of artificial abasic site C

Abasic site mimic **C** contains one more carbon compared to the linker in compound **A**. This adds flexibility to the linker, but the extra atom has to be accommodated correctly in the double helix. The extra atom prevents racemization of the 3'carbon *via* the keto-enol equilibrium, as is the case with the two other artificial abasic sites. The hydrazone can still form both the *cis*- and the *trans*-compound to compensate for the lack of flexibility at the 3'carbon. We only minimized structures with a uracil-adenine base pair for this case, with the 3'carbon atom in an *R*-configuration and the *trans*-hydrazone.

To obtain a correct position for hydrogen bonding and π–π-stacking, we started from the energy minimized structure of the reaction product of abasic site **A** with uracil mimic **4**. We cut between the carbon atom of the linker and the backbone. We added a hydrogen atom on both of the two carbon atoms and let the force field perform a few energy minimization iterations. Repulsion between the newly added hydrogen atoms made the backbone displace backwards and gave some space between the base and the backbone. We then removed the two hydrogen atoms again, inserted the desired extra carbon in the newly created space and connected base mimic and backbone to give the desired structure. This structure was then minimized to give the desired energy minimized structures.

A correct position for hydrogen bonding with the opposite adenine and π–π-stacking with the facing adenine was found. The uracil mimic and the facing adenine were almost coplanar in both the hydrazone and the hydrazone hydrate form. The distance between the two bases is almost similar to the distance found in the previous examples, 2.83 and 2.87 Å.

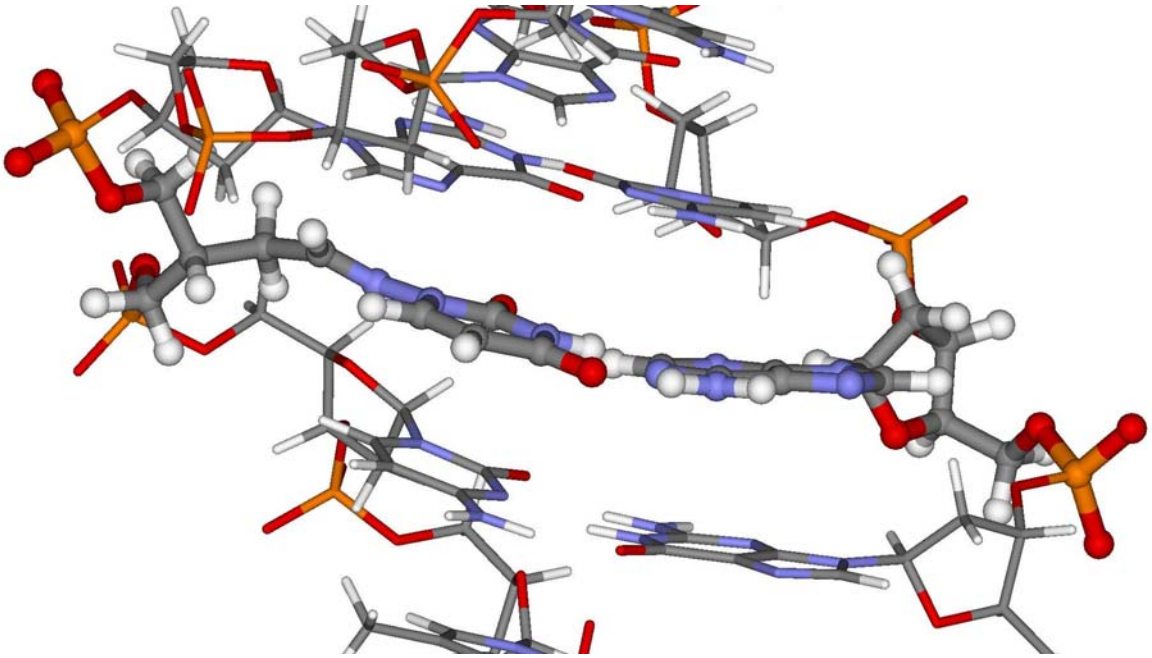


Figure 4.15. Energy minimization of 1-aminouracil at the abasic site position in its hydrazone form. The hydrazone is in its *trans* configuration and the $C^{4'}$ has an R configuration. The $N(H_2)$ of the adenine and the $O(=C^4)$ of the 1-aminouracil are at a distance of 2.87 Å. The 1-aminouracil and the opposite adenine are tilted at an angle of about 4 degrees.

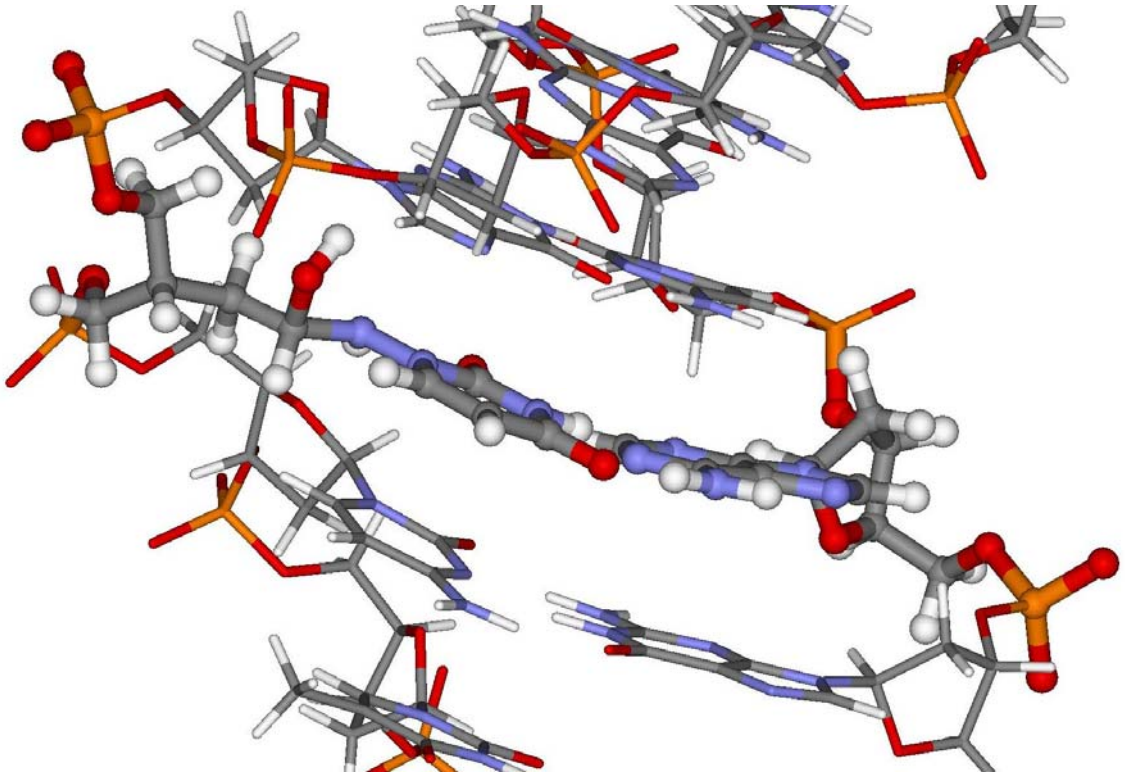


Figure 4.16. Energy minimization of 1-aminouracil at the abasic site position in its hydrazone hydrate form. "$C^{1'}$" has an S configuration and $C^{4'}$ has an R configuration. The $N(H_2)$ of the adenine and the $O(=C^4)$ of the 1-aminouracil are at a distance of 2.83 Å. The 1-aminouracil and the opposite adenine are tilted at an angle of about 5 degrees.

The energy minimizations shown in the paragraphs 4.2.3.1 through 4.2.3.3 show that a good position for hydrogen bonding and π–π-stacking can be found for all the modeled structures. Compared to the modeling results of section 3.3.3, the angle between the base mimic and the opposite base is low (4-15 degrees), on average about 10 degrees less compared to the uracil mimic in the natural abasic sites. Also the distance between the two bases is approximately 0.2 Å less on average compared to the results of section 3.3.3. It seems that the duplex comprising our artificial abasic site can better accommodate the repair units compared to the duplex comprising the natural abasic site. This is presumably caused by the fact that no additional atoms have to be accommodated in the major or the minor groove.

All reaction products of the three abasic site models with the amino bases seem able to find a position favorable for hydrogen bonding and π–π-stacking. The distance between the modified base and the opposite base is approximately the same in all cases. The extra atom that is present in artificial abasic site model **C** is easily accommodated in the backbone. A prediction of which of the three compounds would perform best in an experiment could thus not be made beforehand. Their synthesis was therefore started simultaneously and the compound that would turn out to be easiest to prepare would be used first for testing.

4.3 Aldehydes protected as 1,3-dioxolanes

4.3.1 Retro-synthesis

Synthesizing a DNA oligomer with a free aldehyde function is not possible *via* automated DNA synthesis. A suitable protecting group of the aldehyde is necessary. The protecting group should be easily cleavable after DNA synthesis, without damaging the rest of the oligomer.

Di-acetals are the most common protecting groups for aldehydes. Dimethyl, diethyl and dibenzyl acetals are the most common. Cyclic acetals like 1,2- and 1,3-dioxolanes are also popular. Dithioacetals are another variety of the same group. Imines and enamines are also possible, but less common [136]. The drawback of using acetals in our project is that they normally require strongly acidic conditions to be cleaved and give the aldehyde. A DNA chain is too sensitive to sustain this treatment.

However, the first deprotection of acetals under mildly basic conditions was published by Marko *et al.* [137, 138] in Angewandte Chemie in 1999. They reported the cleavage of a range of acetals and ketals using catalytic amounts of Cerium(IV) ammonium nitrate in

acetonitrile/aqueous borate-HCl buffer at pH 8 obtaining high yields (86-99%) of aldehydes. Other functional groups like secondary and tertiary alcohols, ketones, enones and triisopropylsilyl ethers were reported to be stable under these conditions. The obtained aldehydes were not oxidized to the corresponding carboxylic acids.

These results prompted us to synthesize the dioxolane derivatives of the designed aldehydes (**1-3** Figure 4.17) and to incorporate them into DNA oligomers. In the mean time the deprotection method was tested with simple dioxolanes.

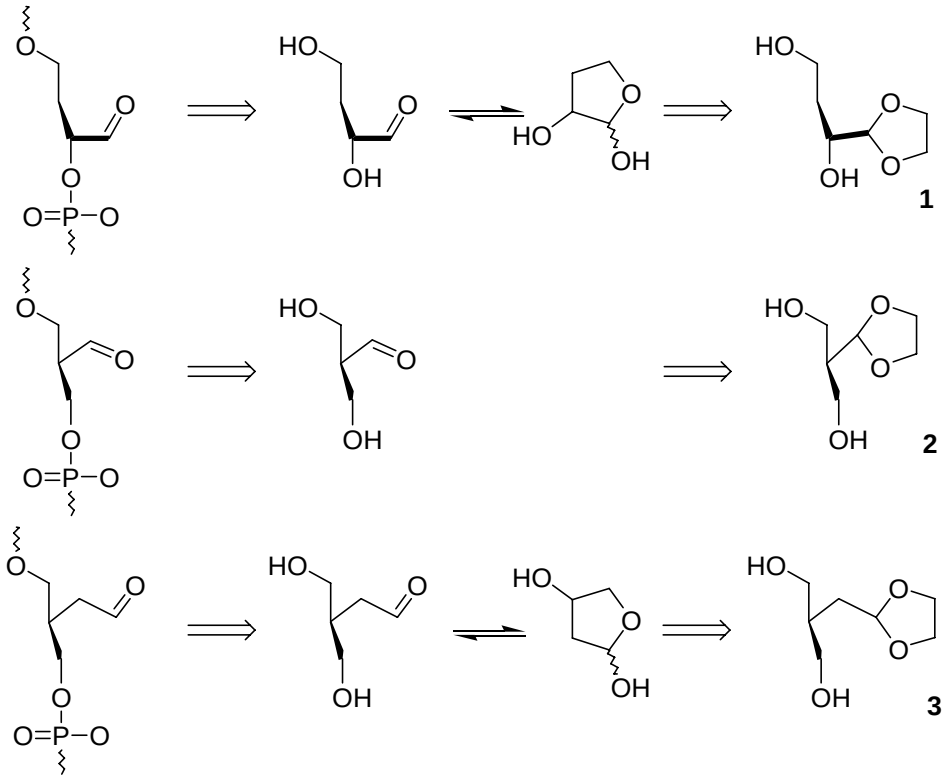


Figure 4.17. Dioxolane design

The free aldehyde forms of compounds **1** and **3** may in principle be able to cyclize as hemiacetals. The hemiacetals can be avoided if the dioxolane is introduced early in synthesis. If this is not possible, addition of a large excess of ethylene glycol may shift the equilibrium to the dioxolane.

It must be noted that compound **1** contains a chiral center, whereas compounds **2** and **3** contain a prochiral center. The synthesis of compound **1** may thus be envisaged as racemic or enantioselective. Compounds **2** and **3** will become chiral after protection of one side of the molecule with a standard protecting groups for DNA synthesis, such as monomethyltrityl (MMT) or dimethyltrityl (DMT) group. Introduction of the DMT group will be non-selective for both compounds, since the two hydroxyl groups are equivalent. Both enantiomers of the

monoprotected product should be obtained in equal quantities. In addition some diprotected compound can be expected.

After deprotection of the aldehyde, compounds **1** and **2** should in principle racemize via a keto-enol equilibrium. Asymmetric synthesis is thus not necessary for these two compounds. When incorporated in the DNA duplex, fast racemization is favorable for the goal we have. The goal is to stabilize a DNA duplex with base-mimicking hydrazides. If the keto-enol equilibrium is reversible, the reaction products will rearrange to the product that stabilizes a double helix best. The other stereoisomer of that product will be less abundant in the final reaction mixture.

The extra atom of compound **3** prevents racemization. The product distribution of this compound does not change after synthesis. If we incorporate the racemic mixture resulting from a racemic synthesis, only half the molecules are in the best position. To prevent this we can envisage to resolve and separate the enantiomers before oligonucleotide synthesis.

4.3.2 Synthesis

Synthesis of compound **3** seemed the most easy, when ignoring the chirality issue. In two steps the desired dioxolane can be synthesized from commercial products (Figure 4.18).

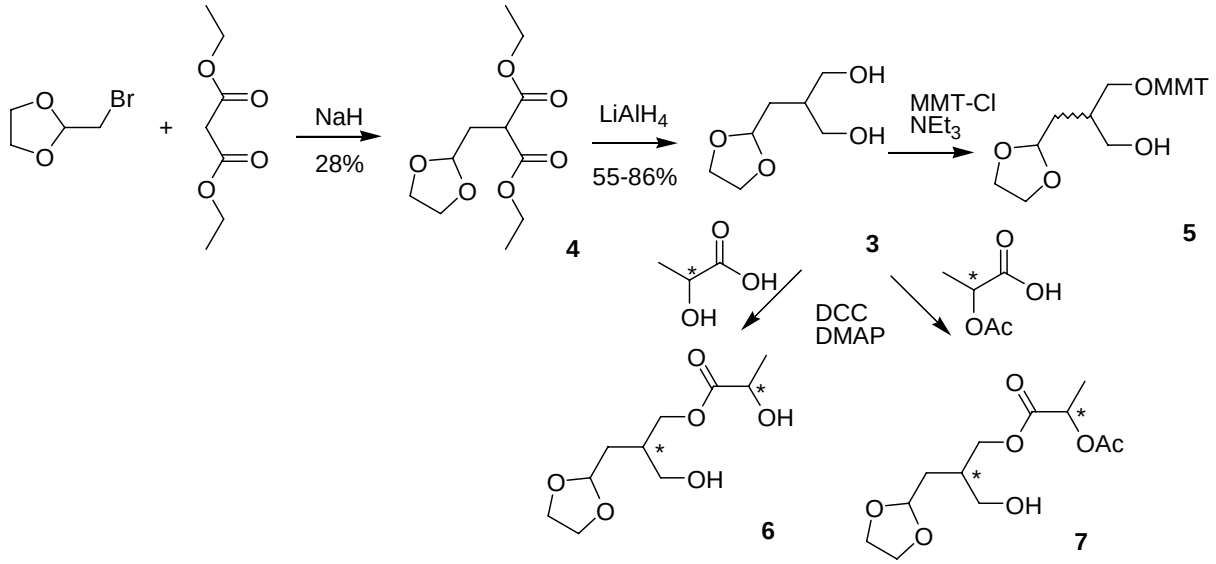


Figure 4.18. Synthesis compound 3.

The 1,3-diol function of compound **3** can be obtained by reduction of the corresponding malonate (**4**). This malonate is a compound that is known in literature and has been published by Cereda *et al.* [138] and Dang *et al.* [139]. It can be obtained in one step by alkylation of dimethyl malonate with 2-bromomethyl-1,3-dioxolane. The reaction proceeded with sodium hydride in a mixture of dry DMF and dry benzene at 80°C. The products were extracted with

diethyl ether and washed with water. Chromatography was used to purify the products, in contrast with both literature procedures, where distillation is used. This may cause our slightly lower yield (28%), compared to the yields obtained by both Dang *et al.* (32%) and Cereda *et al.* (49%).

The next step caused more difficulties. Lithium aluminum hydride reduction in diethyl ether is normally a good working procedure for reducing malonates, but the desired compound **3** was obtained in an 86% yield only once. Generally the yield was about 55% and the product contained some impurities. The product turned out to be unstable at room temperature, even under a nitrogen atmosphere. The decomposition products could not be identified, but with the two hydroxyl groups present, various reactions might be possible.

To avoid decomposition, the product was set through to be protected by monomethyl trityl (MMT) immediately. MMT was chosen for its better stability towards humidity compared to DMT. Product **5** seems to be a stable compound, but due to the instability of **3**, low yields (5%) were observed and the products could not be purified.

Resolution of the prochiral OH of compound **3** by esterification was attempted, using chiral L-lactic acid and S-2-acetoxypropionic acid. In each case the two diastereomeric monoesters could be separated from the diester and the diol **3**, but not from each other. To dissymmetrize compound **3** another chiral acid must be used that creates a larger difference between the two diastereoisomers. The low yields also hinder the resolution of the diastereoisomers.

In the synthesis of compound **1**, we had the choice of either incorporating the dioxolane early or late. Introducing the dioxolane early would avoid problems with the hemiacetal form of the monomer. When introducing it late the stable hemiacetal form could be used as an accessible intermediate from which the dioxolane could be obtained.

A literature study revealed that hemiacetal **9** had already been synthesized by others [141]. So choosing to incorporate the dioxolane in the last step was obvious. The synthesis started with TIPS protection of commercial 2-hydroxy-γ-butyrolactone with a yield of 70% to give compound **8** (Figure 4.19). The cyclic ester was then reduced using diisobutyl aluminum hydride in THF to yield the cyclic hemiacetal **9** [141] (73%). Reaction of **9** with ethylene glycol gave problems. The hemiacetal turned out to be very stable and thus difficult to open with ethylene glycol. Dioxolane formation with 50 equivalents of ethylene glycol with catalytic amount (~10%) of HCl in refluxing toluene yielded traces of the desired product, but multiple non-identified side products were equally detected. Increasing or decreasing the

amount of catalyst or the reflux time did not improve the situation. The small amount of product could not be separated from the side products.

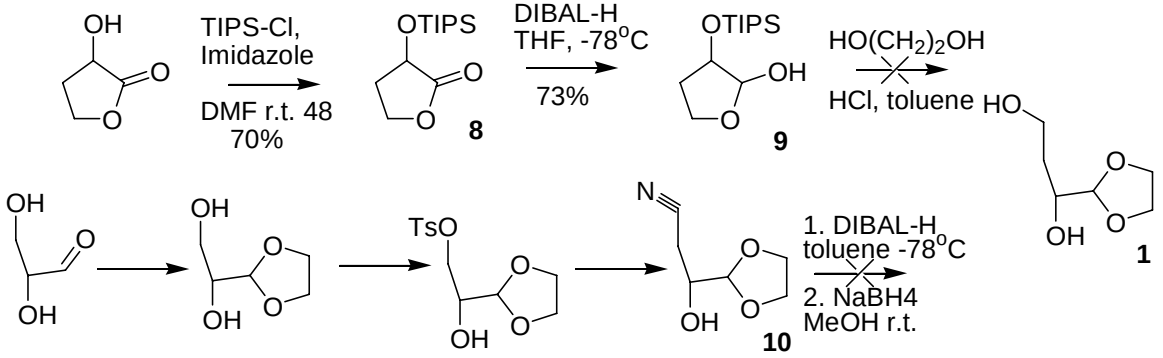


Figure 4.19. Synthesis compound 1.

Incorporating the dioxolane early was not more successful. The nitrile **10** was obtained by a coworker starting from glyceraldehyde. Reaction with ethylene glycol yields the 1,3-dioxolane. The primary hydroxyl group can be selectively substituted by a nitrile *via* a tosylate intermediate to give compound **10**. A two step reduction of this compounds *via* the aldehyde should then yield target **1**. Traces of the target compound were again found, but the compound could never be separated from the side products. Two step reduction of nitriles is commonly performed and should, in principle, be amenable to completion [142]. However, this route was stopped because a suitable deprotection method for deprotection could not be found, as will be discussed in paragraph 4.3.3.

The synthesis of compound **2** should not be constrained by formation of hemiacetal intermediates. The 1,3-dioxolane may thus be introduced late in synthesis when desired. Deconnecting the 1,3-dioxolane yields a known compound (Figure 4.20). Compound **13** was synthesized in 1973 by Vik *et al.* [143]. Their high yields could be reproduced. The synthesis starts by condensing diethoxymethyl acetate with diethyl malonate, catalyzed by a Lewis acid at 140°C [144]. This reaction yielded two products, the expected diethyl (diethoxymethyl) malonate (57%) (**12**) and diethyl (ethoxymethylene) malonate (35%) (**11**). Fuson *et al.* reported the formation of compound **12** only. However, depending on the time the mixture was refluxed and amount of catalyst added, different product distributions of **11** and **12** were found. Since **11** can be converted into **12** by reaction with sodium ethanolate [144] with a yield of 48%, formation of more of **11** and less of **12** is less problematic.

$LiAlH_4$ reduction of the diethyl (diethoxymethyl) malonate yielded 1,1-diethoxy-2,2-*bis*(hydroxymethyl) ethane **13** [143] in 95% yield. The ethyl protected hydroxyl groups of this compound could subsequently be transformed into a dioxolane by transetherification with excess ethylene glycol (68%). Incorporation of the DMT group was successful, but with low yields (25%) of **2a**. This can be explained by the non-selective incorporation of the DMT group on the two primary hydroxyl groups. There is no preference for mono- or disubstitution which resulted in a 1 / 2 / 1 mixture of non-substituted / mono-substituted /di-substituted products. The second reason for the low yield is the sensitivity to acidity of the DMT group, which causes loss of the compound during column chromatography. Incorporation of the phosphoramidite group also proceeded in low yields (33%), because of the sensitivity to acid and humidity of compound **2b**.

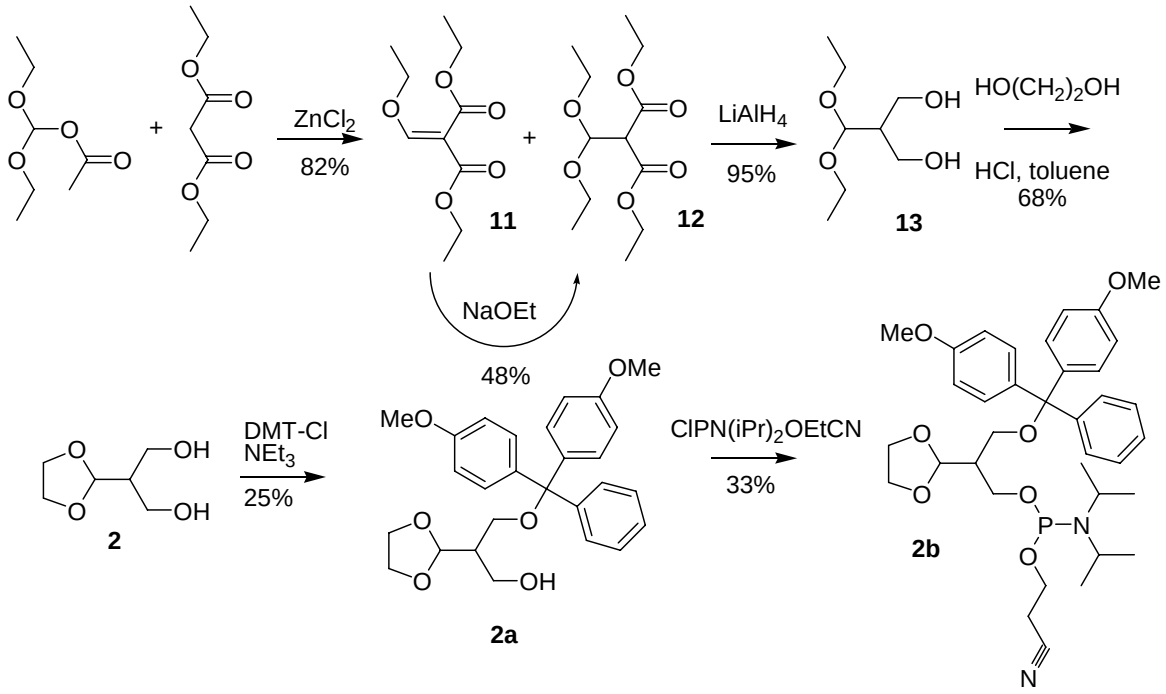


Figure 4.20. Synthesis artificial abasic site monomer 2.

Whilst performing the synthesis of the three products above, we also investigated the synthesis of a simpler abasic site model, which would not be incorporated within an oligonucleotide, but at the end. This is not the best position from a stability point of view when we will be 'repairing' the abasic site. A terminal abasic site destabilizes a DNA duplex less than an abasic site in the middle. Recovery of this lost stability will thus give a smaller energy difference for the repaired terminal abasic site. This results in less product formation if the reaction is under thermodynamic equilibrium. However, availability of the abasic site may be much better, making the reactions faster and synthesis of the monomer is much easier.

Figure 4.21 shows that synthesis of the abasic site building block including the activating group takes only two steps. Starting from commercial 2-bromomethyl-1,3-dioxolane the first step leads to the desired 2-hydroxymethyl-1,3-dioxolane (44%) [145]. When a nucleotide has to be incorporated in the middle of a DNA strand, the 5' hydroxyl group has to be protected for oligonucleotide synthesis. This compound does not have a "5' end" equivalent, which saves a protecting step. For incorporation in an oligonucleotide, we have to activate the nucleotide with a phosphoramidite group. Reaction of **14** with 2-cyanoethyl *N*,*N*-diisopropylchloro phosphoramidite gives compound **15** in moderate yield (53%).

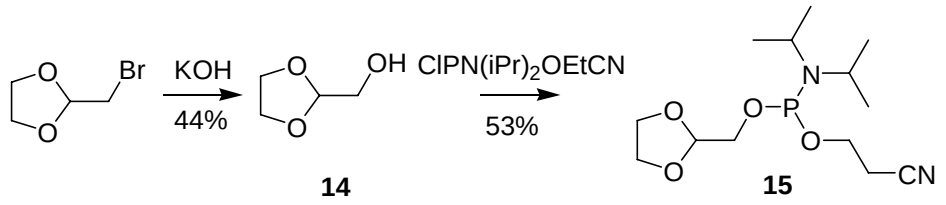


Figure 4.21. Synthesis of a terminal artificial abasic site monomer.

This building block was incorporated in a DNA strand *via* automated DNA synthesis to give the sequence $^{5'}$ **14**CG ATG TGT ACG T$^{3'}$. The modified nucleotide was coupled manually, using a 100 fold excess and a coupling time of 15 minutes reaction time. The yield of the oligomer could not be determined with the standard trityl viewer. Mass spectrometry indicated though that the coupling was far from 100%, a lot of dodecamer was found in the crude reaction product and much less tridecamer. Purification of the oligonucleotide sample has never been attempted.

4.3.3 Deprotection tests

Whilst performing the synthesis of compounds **1-3** and **15**, some test reactions were performed to optimize conditions under which the dioxolane could be deprotected once incorporated in the DNA oligonucleotide. The conditions used by Marko *et al.* [138] were used as a starting point. They used 4% of CAN (6mM) in a solution of acetonitrile/Borate-HCl buffer 1/1. The buffer should assure a high pH (8).

Five dioxolanes were tested, **16**, **18**, **20**, **22** and **2** (Figure 4.22). **16** and **18** can easily be synthesized from anisaldehyde and phenyl acetaldehyde respectively, by refluxing them in toluene using excess ethylene glycol and a catalytic amount of HCl. **20** and **22** were obtained from **14** by esterification with acetic acid and benzoic acid, respectively, using 4-dimethylaminopyridine (DMAP) and *N*,*N'*-dicyclohexylcarbodiimide. **14** would be an even

easier substrate to synthesize, but the product, 2-hydroxyethanal, may be volatile and hard to detect.

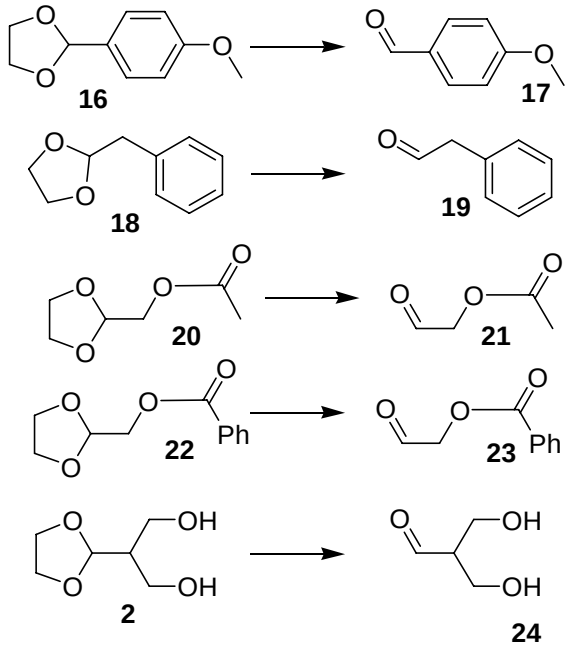


Figure 4.22. Test reactions for deprotection conditions.

Deprotection of **16** was successful, using 10% of CAN (6mM), instead of the 4% reported by Marko *et al*. After the reaction, the products were extracted with dichloromethane, which was then washed with water. Anisaldehyde was obtained as the only reaction product, no starting material was found in the reaction mixture. This confirms the efficiency of the reaction conditions published by Marko *et al.*

Deprotection of **18** was less successful. When using 10% of catalyst (4mM) a triplet at 9.76 ppm was observed in the ^{1}H NMR spectrum of the reaction mixture. The yield was only 26% however. A second, non-identified, side product was also formed (4%) and 70% of the starting material was still present in the reaction mixture.

A model closer to the desired abasic site mimic is **20**. Deprotection with 10% of cerium (IV) ammonium nitrate (2 mM) was attempted as with substrates **16** and **18**, but no aldehyde could be detected. The ^{1}H NMR spectrum showed some starting material, but it seemed that some degradation had equally occurred.

Increasing the amount of catalyst to 50% (concentration now 5 mM) improved the situation slightly. A small peak at 9.25 ppm was observed in the ^{1}H NMR spectrum of the washed dichloromethane phase, which might indicate aldehyde formation. The starting material had disappeared, but multiple side products could not be identified. To make sure the product was

not present in the aqueous phase of the extraction, a ^{1}H NMR spectrum of this phase was equally taken, but no signal of an aldehyde was observed.

To prevent this degradation and have a product that would be easier to detect, compound **22** was synthesized. Deprotection was again attempted, starting with 10% of catalyst (concentration about 1mM). Degradation or evaporation of the products was indeed prevented, but the reaction still hardly proceeded. The starting material was almost exclusively recovered, around 2-3% of aldehyde was detected. Increasing the amount of catalyst was not successful.

To increase the driving force of the reaction, it was attempted to trap the aldehyde with phenyl hydrazine. But also in this case the starting materials were recovered without significant amounts of imine present in the reaction mixture.

A last aldehyde tested was compound **2**. Deprotection of this abasic site mimic with 10% of catalyst (3 mM) yielded two peaks in the aldehyde region. The starting dioxolane had not disappeared, but was still present four times as much as the product. Increasing the catalyst concentration to 8 mM (25%) yielded only one product that had a peak in the ^{1}H NMR spectrum in the aldehydic region (Figure 4.23), no traces of any residual starting material were observed.

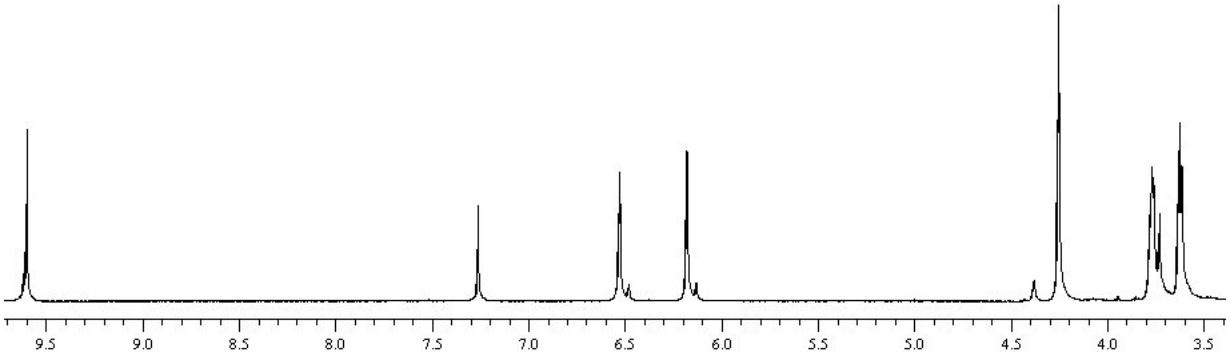


Figure 4.23. ^{1}H NMR spectrum obtained after deprotection of 2 with CAN.

However, the aldehyde peak is a singlet, where a doublet is expected. The rest of both the ^{1}H and ^{13}C NMR spectrum cannot correspond to the expected product either. Apart from the singlet at 9.60 ppm, two singlets are observed at 6.52 and 6.18, which is the region for protons at a C=C double bond. Equilibrium between the aldehyde and its enol form might explain the singlets, but remains that the aldehyde should be present as a doublet. Besides, no signal for the =CH proton of the hydrate could be observed. Equilibrium with its hydrate form would also be possible theoretically, but that would not explain the singlets either. Degradation thus took place, and the products could not be identified. The desired aldehyde could not be obtained with the deprotection method of Marko *et al.*.

The origins of these mixed results with the different aldehydes are not easily understandable when performing the experiments. It seemed that concentration of the catalyst and the substrate were important, but the effects are huge compared to common reaction kinetics. What was happening exactly could not be found out.

These negative results prompted us to look in literature if any other researchers had used the method of Marko *et al.* for aldehyde deprotection and if they found any limitations. One publication was found, published beginning 2002, three years after the publication of Marko *et al.* The group of Parrilli had tried to take advantage of this method for deprotection of ketals in carbohydrates [146]. They failed to reproduce the results obtained by Marko *et al*. They reported that under the conditions described by Marko *et al.* both ketals and glycosidic bonds were cleaved very easily due to the strongly acidic (pH 0.2) conditions. Deprotection of the ketal in a mixture of acetonitrile and borate/HCl buffer (1/1) at pH 8 did not occur.
Collaboration of the two groups pointed out that the buffer as used by Marko *et al.* was a dilution of the commercial buffer. It turned out that the dilute buffer could not assure a high pH.
In the paper of Manzo *et al.* it was concluded that cyclic ketal/acetal cleavage occurs in a protic acid environment, rather than under neutral or basic conditions as claimed by Marko *et al.*. Ketal cleavage by CAN acting as a Lewis acid at pH 4.4 was proven. Steric hindrance limited the reaction of certain ketals.
The problems we had in our test reaction are most certainly also related to the pH. When increasing the concentration of catalyst, the reactions worked better. In that case the pH of the reaction mixture is obviously low enough for the reaction to work.
These conditions make cleavage of acetals in nucleic acids impossible by using CAN. Steric hindrance may very well play a role in nucleic acids and more importantly, CAN is a very good cleaving agent of DNA in acidic conditions [147-149].

These reports pointed to the fact that dioxolane was not a suitable protection group for the artificial abasic site. However, to avoid restarting chemistry all over again, it was first tried to find other methods to deprotect 1,3-dioxolanes under milder conditions. A method to deprotect 1,3-dioxolanes was published in 1995 by Nishiguchi *et al.* [150]. Treatment with nitrogen dioxide in the presence of silica gel catalyses the oxidation of the dioxolane. But it was reported that only ketals can be cleaved, acetals remain unchanged.

Trimethylsilyl iodide in dichloromethane as proposed by Jung *et al.* [151] was not applicable because of the formation of methyl iodide during the reaction. This will certainly harm the DNA strand.

Slow hydrolysis of 1,3-dioxolanes in the presence of acid sensitive groups was reported using $LiBF_4$ in wet acetonitrile [152-154]. However, it was reported that during the reaction the pH decreased from an initial value of pH 5 to pH 2.5 after twelve hours, making the method unsuitable.

Three transacetalization methods in wet acetone at room temperature were tested to deprotect the abasic site monomer shown in Figure 4.20. Three different catalysts were tested, palladium dichloride [155], Amberlyst-15 resin [156] and pyridinium-p-toluene sulfate [157]. As a substrate was chosen compound **2** (Figure 4.24).

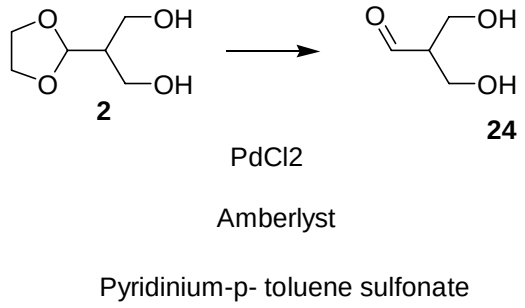


Figure 4.24. New dioxolane deprotection conditions.

With palladium chloride the starting material was almost completely recovered. In the ^1H NMR spectrum a small peak could be distinguished at 9.70 ppm, but similar to deprotection with CAN, this is a singlet and cannot be the desired product.

Deprotection with Amberlyst-15 gave the same aldehydic compound as when using CAN (Figure 4.23). This was not the only compound though, a large fraction of the products decomposes. Pyridinium tosylate leaves the aldehyde unchanged.

Deprotection of dimethylacetals was proposed by Colvin *et al.* [158]. Cleavage of dimethylacetals is reported to be easier than cleavage of dioxolanes. Since the diethylacetal **13** is an intermediate product in the synthesis of the abasic site monomer, this might be a suitable method. Colvin et al. did not use the transacetalization method, but since diethylacetals are reported easier to deprotect, the three methods I mentioned above were also tested on this new substrate (Figure 4.25).

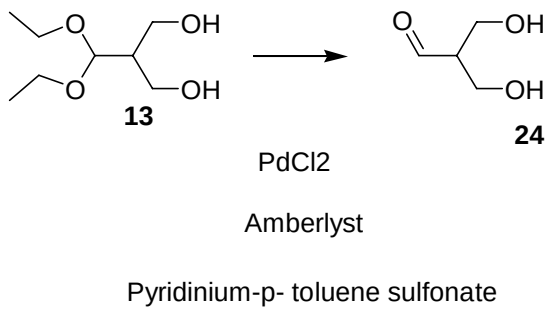


Figure 4.25. Deprotection diethyl acetal

Test results with the diethyl acetal changed the situation only slightly. The starting material almost completely disappears when treated with $PdCl_2$ in acetone after 48h. The unidentified aldehyde-like peak at 9.70 ppm was again observed and in addition severe side reactions took place.

Deprotection of the diethyl acetal with Amberlyst-15 gave a similar result as the 1,3-dioxolane. The undesired aldehydic compound was observed, together with side product formation. Side product formation was more important when cleaving **13** compared to cleavage of **2**.

Contrary to the 1,3-dioxolane case, pyridinium tosylate is active towards the diethyl acetal. It gives now the same products as the Amberlyst-15, unidentified aldehydic-type compounds that cannot be the products we desire. Apart from that, more side products are also observed.

In conclusion we can say that deprotection of acetals and dioxolanes is not possible in oligonucleotides. When cleaving the dioxolanes or acetals, the oligonucleotides will most certainly degrade as well. Different chemistry has to be developed to incorporate the desired aldehyde in the DNA strand.

4.4 Aldehydes masked as 1,2-diols

In the mean time a Russian group published a method to incorporate aldehydes in DNA [159, 160]. In a nine step synthesis Zatsepin *et al.* introduced an allyl substituent at the 2' position of a RNA nucleotide after a method described by Sproat *et al.* [161]. Oxidation of this allyl group with osmium tetroxide lead to the formation of a 1,2-diol. This 1,2-diol was then protected with benzoyl groups for oligonucleotide synthesis. After oligonucleotide synthesis the 1,2-diol was unmasked during the standard treatment with ammonia, and subsequently oxidized using sodium periodate to yield the aldehyde (Figure 4.26).

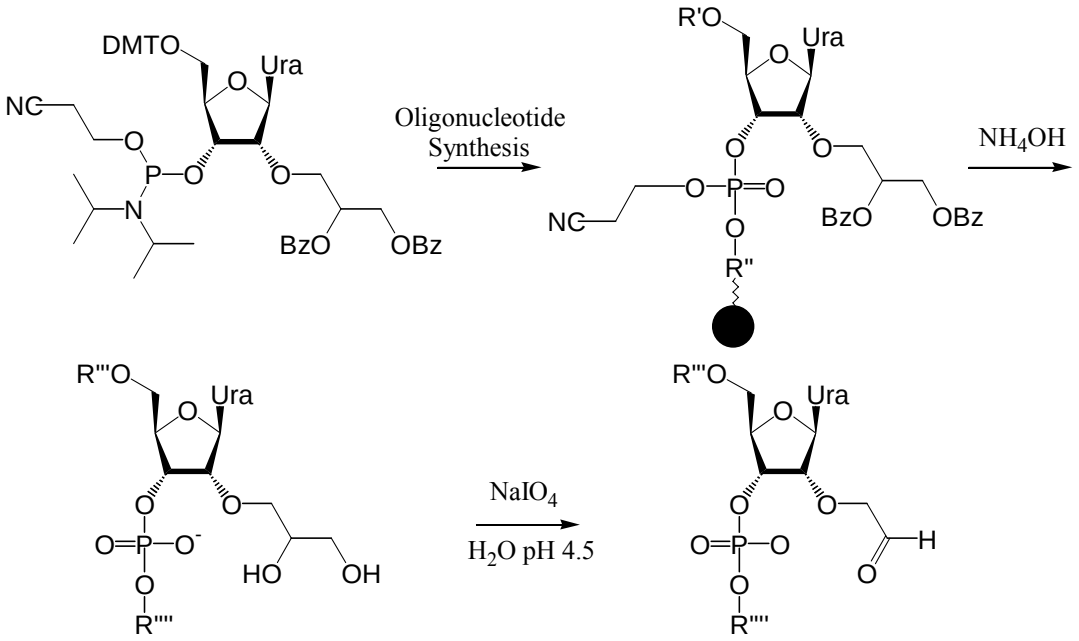


Figure 4.26. Incorporation of an aldehyde function at the 2'position in an oligonucleotide. R', R'' = protected oligonucleotide chain; R''', R'''' = unprotected oligonucleotide chain.

Based on the work of Zatsepin *et al.*, we designed a new version of abasic site monomers. Figure 4.27 shows the retro-synthetic analysis of the target monomers (**25-27**) that will lead to the three different abasic sites.

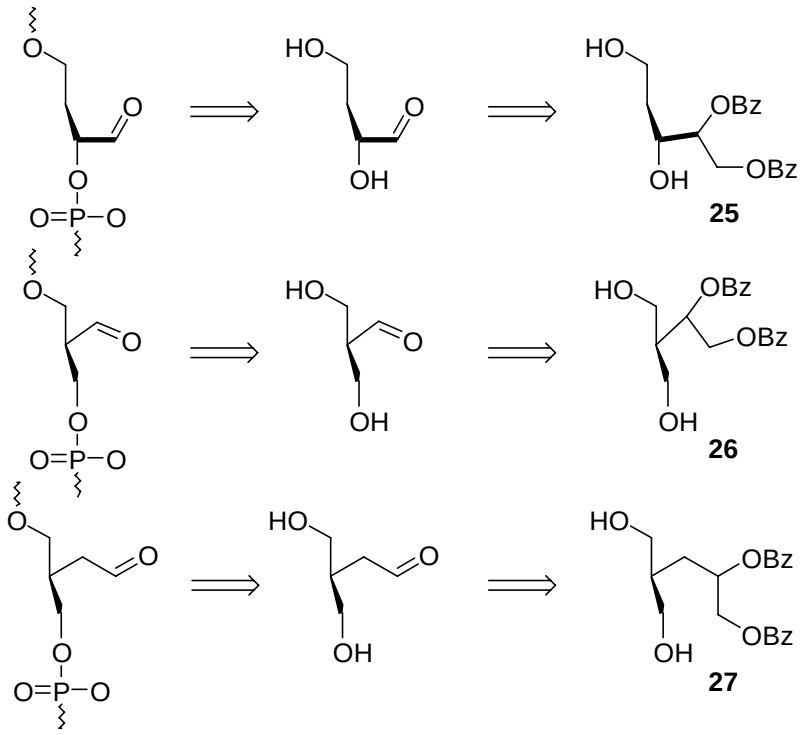


Figure 4.27. Retro-synthetic analysis to incorporate aldehydes *via* 1,2-diols.

Of the compounds **25-27**, **27** seemed the one that was most easy accessible. A synthetic pathway for compound **27** is shown in Figure 4.28. Oxidation of commercial dimethyl allylmalonate worked well, but instead of di-esterification of diol **28**, cyclization took place between the secondary alcohol and the malonate. Formation of the five membered ring is favored over di-esterification. Reducing the malonate to the alcohol before esterification of the 1,2-diol is not a good option, since discrimination between the three primary alcohols will not be feasible. A route that can be envisaged in the future is introducing the protecting groups of the 1,2 diol first and then do a nucleophylic substitution on the malonate. Selective reduction of the malonate is possible by using Li(*s*-but)$_3$BH.

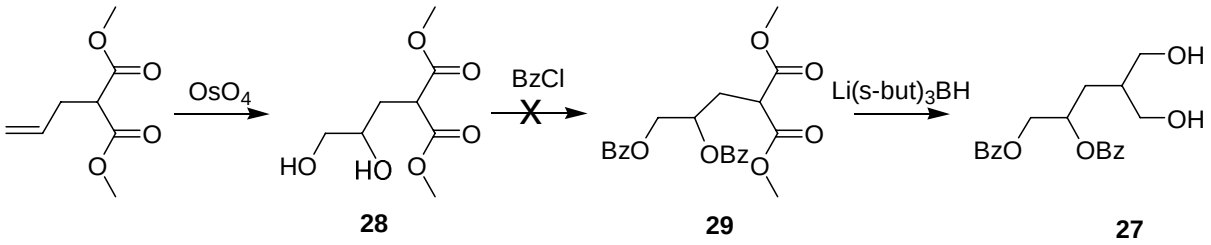


Figure 4.28. Synthesis benzoyl protected artificial abasic site monomer.

Incorporation at the terminus of a DNA strand, similar to the dioxolane protected aldehydes, was also considered. The hydroxide at the "5' end" equivalent of the oligonucleotide is again superfluous, reducing the desired molecules to compounds **30** and **31 (**Figure 4.29).

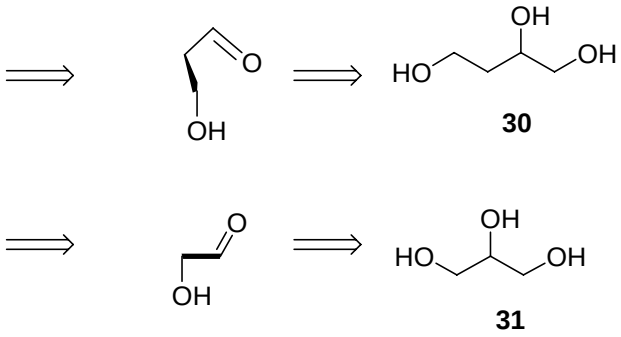


Figure 4.29. Monomers for abasic site incorporation at the end of an oligonucleotide

Compound **31** looks very easy to synthesize. Only the protective groups must be placed in the right position, which gives compound **34** as a target molecule (Figure 4.30). OsO_4 oxidation of TIPS protected allyl alcohol yields the 1,2-diol (99%). Esterification with benzoyl chloride followed by removal of the silyl group yields the abasic site monomer in a 50% overall yield. Incorporation of the phosphoramidite was performed with a yield of 31%, which is low compared to standard nucleotides. Due to lack of time the monomer has not yet been incorporated in a DNA strand.

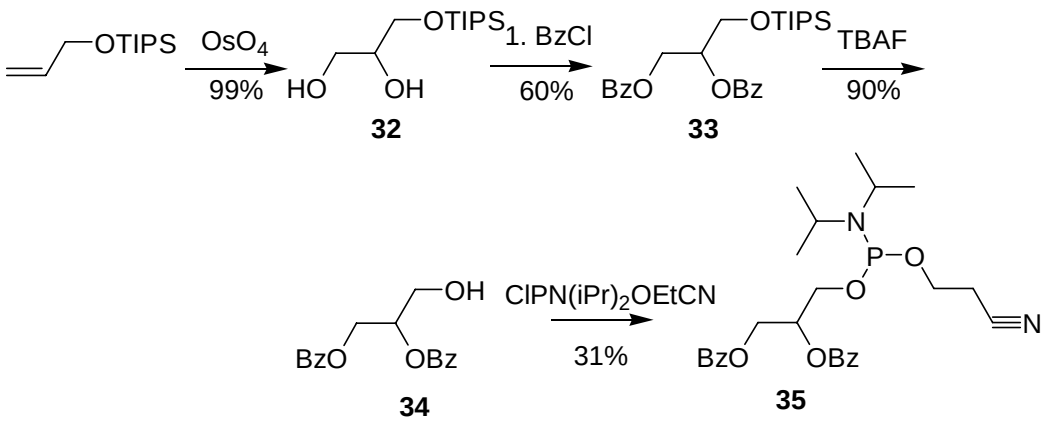


Figure 4.30. Terminal end artificial base mimic

4.5 Conclusion

In this chapter we designed an artificial abasic site replicating system. We wanted to take advantage of the reversible reactions between aliphatic aldehydes and the aminobases. The abasic sites that we designed are not able to cyclize and form hemiacetals as a natural abasic site and may therefore have a higher apparent reactivity. Molecular modeling showed that the products of the reaction between the aminobases and the aldehydes were easily accommodated in a natural DNA duplex. A good position for hydrogen bonding and π–π-stacking was found for the products of the reaction between the aminobases and each of the three abasic sites that we developed.

Deprotection of the 1,3-dioxolanes that we developed and synthesized to protect the aldehydes during oligonucleotide synthesis failed. Test reactions showed that the deprotection of 1,3-dioxolanes does not work under the conditions that were described. Incorporation of the aldehydes masked as 1,2-diols is hopefully more successful, but the multi-step synthesis of the monomers was not yet achieved. Synthesis of a monomer that can be incorporated at the end of an oligonucleotide was completed.

4.6 Experimental section

General methods

All commercial reagents were used as supplied, without further purification. Petroleum ether (PE) refers to the fraction with distillation range 40-60°C. All reactions were performed under an atmosphere of nitrogen, unless stated otherwise.

Solvents

Water was purified over a MilliQ apparatus (Millipore). Pentane, hexane, cyclohexane, benzene, toluene, dichloromethane and chloroform were dried over calcium hydride and distilled under nitrogen atmosphere or dried over an alumina column. Diethylether and THF were dried over potassium hydride and distilled under nitrogen atmosphere.

Chromatography

All purifications by column chromatography were performed on silica gel ROCC 40-68μM with 0.2 bar of air pressure.

NMR spectroscopy

NMR spectra were recorded at 25°C on a Bruker Avance 400 MHz NB US NMR Spectrometer. ^{1}H NMR spectra were recorded at 400 MHz. ^{13}C NMR spectra were recorded at 100.56 MHz. ^{19}F spectra were recorded at 376.3 MHz. ^{31}P spectra were recorded at 161.9 MHz . Values for δ are in ppm relative to residual solvent signal. *J*-Values are in Hz.

GC-MS

GC-MS were performed on a Finigan Trace MS run under Electron Impact conditions.

Molecular modeling

All molecular modeling was perfomed on Silicon graphics workstations using MacroModel version 6.5. All simulations were run in the MacroModel version of the AMBER* force field at 300K with water as a solvent using a GB/SA solvent model. An energy minimized structure of a known structure was used as a starting point for the calculations and minimization was performed using a TNCG (Truncated Newton Conjugate Gradient) [132]. None of the calculations was intended as a systematic search of energy minima (no Monte-Carlo search or Molecular Dynamics were performed).

Infra red spectroscopy

Infrared spectra were measured on a Bruker IFS-55 spectrometer. Solids were measured in KBr pellets and liquids measured as a film between two KBr pellets.

Melting point determination

All melting points were measured in capillary tubes using an Electrothermal 9100 apparatus and are non-corrected.

Chemical procedures

Some molecules synthesized in this part were described in literature. In that case we thus refer to the original papers.

Dimethyl-2-(2-ethylenedioxyethyl) malonate (4) [139, 140]

Dimethyl malonate (3.46 mL; 4 g; 30.27 mmoles; 1.01 eq.) was added dropwise to a stirring suspension of sodium hydride (1.6 g; 40 mmoles; 1.33 eq.) in dry dimethyl formamide (DMF, 20 mL) and dry benzene (5 mL) at 0°C. A heavy precipitate formed at this stage, mechanical stirring is needed.

After the evolution of hydrogen had ceased, 2-bromomethyl-1,3-dioxolane (3.099 mL; 5 g; 29.93 mmoles; 1 eq.) was added dropwise with stirring. The reaction mixture was then heated at 80°C for 20 h with stirring. The reaction mixture was poured into ice-water (20 mL), and the organic phase was separated. The water-DMF phase was extracted with diethyl ether (5 * 30 mL). The combined organic phase was washed with water twice, with saturated brine once, and dried over $MgSO_4$. The solvent was removed by evaporation *in vacuo*.

The resulting oil was purified by flash column chromatography using PE/Et_2O 50/50 as an eluent to yield dimethyl-2-(2-ethylenedioxyethyl) malonate (27.6 %; 1.807 g; 8.281 mmoles) as a colorless oil.

NMR: δ_H ($CDCl_3$); 5.001, (t, ^3J *= 3.7 Hz*, O_2CH), 3.963-3.903 (m, 2H, OCH_2), 3.883-3.823 (m, 2H, OCH_2), 3.740 (s, 6H, OCH_3), 3.626 (t, ^3J *= 7.1 Hz*, C_3CH), 2.361 (dd, ^3J *= 7.1 Hz,* ^3J *= 3.8 Hz*, 2H, CH_2). δ_C: 169.6, 101.8, 65.1, 52.6, 46.3, 32.4.

4-Hydroxy-3-(hydroxymethyl)butanal-1,3-dioxolane (3)

To a slurry of lithium aluminum hydride (284.5 µL; 260.9 mg; 6.874 mmoles; 3 eq.) in 100 mL of diethyl ether, a solution of dimethyl-2-(2-ethylenedioxyethyl) malonate (500 mg; 2.291 mmoles; 1 eq.) was added at a rate sufficient to keep the ether refluxing. After the first vigorous reaction had subsided the reaction mixture was heated under reflux overnight.

1mL of water was added dropwise to destroy the excess of $LiAlH_4$. The reaction mixture was filtered over Celite and the salts were washed repeatedly with ethyl acetate. The solvents in

the filtrates were removed and the residue was purified by flash column chromatography using C_6H_{12}/EtOAc 80/20 as an eluent to yield 4-hydroxy-3-(hydroxymethyl)butanal-1,3-dioxolane (86.1 %; 320 mg; 1.973 mmoles) as a colorless oil. The obtained product is unstable and decomposes over time. The 86% yield is exceptional, generally the yields were around 50%, because of decomposition of the product during workup.

NMR: δ_H ($CDCl_3$); 4.981 (t, ^3J = *4.4 Hz*, 1H, O_2CH), 4.024-3.963 (m, 2H, OCH_2), 3.921-3.859 (m, 2H, OCH_2), 3.785-3.700 (m, 4H, $HOCH_2$), 2.379 (broad s, 2H, OH), 2.014-1.957 (m, 1H, C_3CH), 1.802 (dd, ^3J = *4.6 Hz*, ^3J = *6.7 Hz*, 2H, CH_2).

4-((*p*-*A*nisyldiphenyl)methoxy)-3-(hydroxymethyl)butanal-1,3-dioxolane (5)

4-Hydroxy-3-(hydroxymethyl)butanal-1,3-dioxolane (280 mg; 1.726 mmoles; 1 eq.) was dissolved in anhydrous pyridine (5 mL) and triethylamine (0.1 mL). The solution was cooled down to -10°C and a solution of 2-(4-methoxyphenyl)-4,6-diphenylpyrylium tetrafluoroborate (662.2 mg; 1.553 mmoles; 0.9 eq.) in 10 mL of pyridine was added dropwise. The reaction was stirred overnight at -10°C and checked by NMR.

The reaction was quenched by the addition of 3 mL of methanol. The solution was evaporated to a yellow syrup and coevaporated 3 times with toluene to remove traces of pyridine. The residue was dissolved in ethyl acetate and washed with saturated aqueous $NaHCO_3$ solution, water and brine. The organic phase was dried on Na_2SO_4, filtered and evaporated *in vacuo*. The residue was purified by flash column chromatography using C_6H_{12}/EtOAc/NEt_3 88/10/2 as an eluent to yield 4-((*p*-anisyldiphenyl)methoxy)-3-(hydroxymethyl)butanal-1,3-dioxolane (4.26 %; 32 mg; 73.64 μmoles) as a yellow oil. The product could not be purified and isolated because of the low yields.

3-Triisopropylsilanyloxy-dihydro-furan-2-one (8) [141]

A solution of 2-hydroxy-γ-butyrolactone (819 mg; 8.022 mmoles; 1 eq.) and imidazole (1.638 g; 24.06 mmoles; 3 eq.) in DMF (25 mL) was prepared. Triisopropylsilyl chloride (2.3 mL; 2.072 g; 10.42 mmoles; 1.3 eq.) was added and the mixture was stirred for 48 h.

100 mL of hexane was added and the organic phase was washed with water 6 times. The solvent was evaporated and the residue was purified by flash column chromatography using C_6H_{12}/EtOAc 95/5 as an eluent to yield 3-triisopropylsilanyloxy-dihydro-furan-2-one (70.3 %; 1.18 g; 4.299 mmoles).

NMR: δ_H ($CDCl_3$); 4.507 (dd, *J = 8.0 Hz, J = 8.0 Hz*, 1H, TIPSO-CH) 4.416-4.365 (m, 1H, OCH_2) 4.224-4.163 (m, 1H, OCH_2) 2.528-2.453 (m, 1H, OCH_2CH_2) 2.306-2.211 (m, 1H, OCH_2CH_2) 1.204-0.986 (m, 21H, CH_3 + SiCH).

4-Hydroxy-2-triisopropylsilanyloxy-butyraldehyde (9) [141]

3-Triisopropylsilanyloxy-dihydro-furan-2-one (1.00 g; 3.643 mmoles; 1 eq.) was dissolved in 20 mL of toluene, and 3.7 mL of a 1M diisobutylaluminum hydride (3.7 mmoles, 1.02 eq.) solution in hexane was added at -78°C. The mixture was stirred for 3 h at -78°C, and then $NH_4Cl_{(aq)}$ was added. The mixture was allowed to warm up to r.t. and diethyl ether and $NaOH_{(aq)}$ (pH =10) were added.

The organic phase was washed with diluted NaOH (pH =10) three times, water and brine. The solution was dried on $MgSO_4$ and the solvent evaporated *in vacuo*. The product was purified by flash column chromatography using C_6H_{12}/EtOAc 90/10 as an eluent to yield 4-hydroxy-2-triisopropylsilanyloxy-butyraldehyde (72.2 %; 728 mg; 2.633 mmoles) as a colorless oil. The product consists of a mixture of its two epimers that can be separated but equilibrate over time.

NMR: First epimer: δ_H ($CDCl_3$); 5.254 (d, ^3J *= 3 Hz*, 1H, $HOCH_2$), 4.306 (dd, ^3J *= 1.4 Hz,* ^3J *= 4.9 Hz*, 2H, TIPSOCH), 4.118-4.080 (m, 2H, OCH_2), 2.275-2.187 (m, 1H, OCH_2CH_2), 1.877-1.814 (m, 1H, OCH_2CH_2), 1.092-1.041 (m, 21H, CH_3 and SiCH).

Second epimer. δ_H ($CDCl_3$); 5.254 (d, 1H, $HOCH_2$), 4.397-4.360 (m, 1H, TIPSOCH), 4.056-4.014 (m, 1H, OCH_2), 3.839-3.787 (m, 1H, OCH_2), 2.150-2.081 (m, 1H, OCH_2CH_2), 1.956-1.885 (m, 1H, OCH_2CH_2), 1.157-1.023 (m, 21H, CH_3 and SiCH).

(Diethoxymethyl) ethylmalonate (12) and (ethoxymethylene) ethyl malonate (11) [144]

A mixture of diethoxymethyl acetate (2.014 mL; 2 g; 12.33 mmoles; 1 eq.), diethyl malonate (1.872 mL; 1.975 g; 12.33 mmoles; 1 eq.) and zinc chloride (1.732 µL; 5.041 mg; 36.99 µmoles; 0.003 eq.) was heated at 120°C overnight.

The crude reaction mixture was purified by flash column chromatography using C_6H_{12}/EtOAc 95/5 as an eluent to yield (diethoxymethyl) ethylmalonate (46.8 %; 1.25 g; 5.78 mmoles) and (Ethoxymethylene) ethyl malonate (34.9 %; 1.13 g; 4.307 mmoles).

NMR: (Diethoxymethyl) ethylmalonate δ_H ($CDCl_3$); 7.590 (s, 1H, =CH), 4.301-4.152 (m, 6H, 3 CH_2), 1.390-1.249 (m, 9H, 3 CH_3).

(Ethoxymethylene) ethyl malonate δ_H ($CDCl_3$); 5.110 (d, ^3J = *8.5 Hz*, 1H, O_2CH), 4.230-4.177 (m, 4H, $COOCH_2$), 3.753-3.695 (m, 3H, OCH_2 + C_3CH), 3.650-3.574(m, 2H, OCH_2), 1.270 (t, ^3J = *7.1 Hz,* 6H, malonate CH_3), 1.179 (t, ^3J = *7.1 Hz*, 6H, OCH_2CH_3).

(Diethoxymethyl) ethylmalonate (12) [144]

(Ethoxymethylene) ethyl malonate (2.6 g; 13.96 mmoles; 1 eq.) was added to a solution of sodium (428.1 µL; 385.2 mg; 16.75 mmoles; 1.2 eq.) in 20 mL of anhydrous ethanol. The temperature was maintained at 40°C during the addition of ethyl alcohol, and for one hour afterward. The reaction mixture was neutralized with acetic acid and the solvent was evaporated *in vacuo*.

The residue was redissolved in EtOAc and filtered over a glass filter. The solvent was evaporated *in vacuo* and the residue was purified by flash column chromatography using C_6H_{12}/EtOAc/Et_3N 79/20/1 as an eluent. (Diethoxymethyl) ethylmalonate (41.7 %; 1.53 g; 5.832 mmoles) was obtained as a colorless oil, and (ethoxymethylene) ethyl malonate (22.8 %; 690 mg; 3.19 mmoles) was recovered.

3-Hydroxy-2-hydroxymethyl-propanal diethyl acetal (13) [143]

To a slurry of lithium aluminum hydride (22.58 µL; 20.7 mg; 545.5 µmoles; 2 eq.) in 100 mL of diethyl ether was added a solution of (diethoxymethyl) ethylmalonate (65 mg; 272.7 µmoles; 1 eq.) at a rate sufficient to keep the ether refluxing. After the first vigorous reaction had subsided the reaction mixture was heated under reflux overnight. 1mL of water was added dropwise to destroy the excess of $LiAlH_4$. The reaction mixture was filtered over Celite and the salts were washed repeatedly with ether and ethanol. The solvents in the filtrate were evaporated *in vacuo*. The residue was purified by flash column chromatography using EtOAc/PE 60/40 as an eluent to yield 3-hydroxy-2-hydroxymethyl-propanal diethyl acetal (96.6 %; 47 mg; 263.7 µmoles) as a colorless oil.

NMR: δ_H ($CDCl_3$); 4.63 (d, ^3J = *5.7 Hz*, 1H, O_2CH), 3.811 (d, ^3J = *5.5 Hz*, 2H, $HOCH_2$), 3.798 (d, ^3J = *5.5 Hz*, 2H, $HOCH_2$), 3.777 (q, ^3J = *7.0 Hz*, 1H, CH_3CH_2), 3.754 (q, ^3J = *7.0 Hz*, 1H, CH_3CH_2), 3.580 (q, ^3J = *7.0 Hz*, 1H, CH_3CH_2), 3.557 (q, ^3J = *7.0 Hz*, 1H, CH_3CH_2), 2.349 (broad t, ^3J = *5.8 Hz*, 2H, OH), 2.091-2.025 (m, 1H, C_3CH), 1.238 (t, ^3J = *7.0 Hz*, 6H, CH_3); δ_C ($CDCl_3$), 101.0, 62.7, 61.4, 41.2, 15.2.

3-Hydroxy-2-hydroxymethylpropanal-1,3-dioxolane (2)

3-Hydroxy-2-hydroxymethyl-propanal diethyl acetal (141 mg; 791.1 µmoles; 1 eq.) was placed in a 250 mL roundbottomed flask, which was fitted with a Dean Stark receiver. It was dissolved in 150 mL of toluene and 2.66 mL (2.961 g) of ethylene glycol and 35 microliters of HCl were added. The mixture was heated to 140°C for 20h. The solution was cooled down to r.t. and the solvent was evaporated *in vacuo*. The residue was purified by flash column chromatography using $CHCl_3$/MeOH 98/2 as an eluent to yield 3-hydroxy-2-hydroxymethylpropanal-1,3-dioxolane (68.2 %; 80 mg; 539.9 µmoles) as a colorless oil.

NMR: δ_H ($CDCl_3$); 4.996 (d, ^3J = *4.3 Hz*, 1H, O_2CH), 4.042-4.007 (m, 2H, OCH_2), 3.911-3.876 (m, 2H, OCH_2), 3.865-3.819 (m, 2H, CH_2OH), 2.357 (broad t, ^3J = *5.5 Hz*, 2H, OH), 2.135-2.071 (m, 1H, C_3CH); δ_C ($CDCl_3$); 103.9, 72.6, 68.8, 43.6.

GC-MS (EI) characteristic fragment ions: 75 ($HOCH_2CH_2OCH_2^+$, $HOCH_2CH{=}CHO^+H_2$, $HOCH{=}O^+CH_2CH_3$, $HO^+{=}CH_2CH_2CH_2OH$), 105 ($HOC^+HCH(CH_2OH)_2$), 120 ($HOCH_2CH_2O^+HCH{=}CHCH_2OH$, $HOCH_2CH_2OCH_2CH{=}CHO^+H_2$, $H_2C{=}O^+CH_2CH(CH_2OH)_2$).

3-(Di-(*p*-anisyl)phenyl)methoxy-2-hydroxymethylpropanal-1,3-dioxolane (2a)

3-Hydroxy-2-hydroxymethylpropanal-1,3-dioxolane (42 mg; 283.4 µmoles; 1 eq.) was dissolved in anhydrous pyridine (5 mL) and triethylamine (0.1 mL). The solution was cooled down to -10°C and a solution of *p*-anisylchlorodiphenylmethane (83.16 mg; 269.3 µmoles; 0.95 eq.) in 5 mL of pyridine was added dropwise. The reaction was stirred for 3h at -10°C.

The reaction was quenched by the addition of 3 mL of methanol. The solution was evaporated to a yellow syrup and co-evaporated 3 times with toluene to remove traces of pyridine. The residue was dissolved in ethyl acetate and washed with saturated aqueous $NaHCO_3$ solution, water and brine. The organic phase was dried on Na_2SO_4, filtered and evaporated *in vacuo*. The residue was purified by flash column chromatography using C_6H_6/EtOAc/Et_3N 89/10/1 as an eluent to yield 3-(*p*-anisyldiphenyl)methoxy-2-hydroxymethylpropanal-1,3-dioxolane (25.1 %; 30 mg; 71.34 µmoles) as a light yellow oil.

NMR: δ_H ($CDCl_3$); 7.432 (d, ^3J = *8.7 Hz*, 4H, H_3COCCH), 7.324-7.201 (m, 5H, H_{arom}), 6.836 (d, ^3J = *8.7 Hz*, 4H, $H_3COCCHCH$), 4.930 (d, ^3J = *5.0 Hz*, 1H, O_2CH), 3.959-3.761 (m, 6H, 3 OCH_2), 3.698 (s, 6H, CH_3), 3.343 (m, 2H, CH_2OH), 2.634 (broad dd, ^3J = *6.6 Hz* ^3J = *6.0 Hz*, 1H, OH), 2.174-2.103 (m, 1H, C_3CH), δ_C ($CDCl_3$); 158.5, 144.3, 135.4, 130.3, 128.4, 127.8, 126.8, 113.1, 104.4, 86.6, 64.9, 64.7, 61.6, 55,2, 44,9, 29,7. GC-MS (EI) characteristic

fragment ions: 75 ($HOCH_2CH_2OCH_2^+$, $HOCH_2CH{=}CHO^+H_2$, $HOCH{=}O^+CH_2CH_3$, $HO^+{=}CH_2CH_2CH_2OH$), 105 ($HOC^+HCH(CH_2OH)_2$), 274 ($C^+(C_6H_5)_2C_6H_4OCH_3$) .

3-(*p*-*A*nisyldiphenyl)methoxy-2-*O*-((2-cyanoethoxy(diisopropylamino))-phosphino)-hydroxymethyl-propanal-1,3-dioxolane (2b)

3-(*p*-Anisyldiphenyl)methoxy-2-hydroxymethylpropanal-1,3-dioxolane (59 mg; 151 µmoles; 1 eq.) was dissolved in anhydrous dichloromethane (5 mL) and diisopropylethylamine (105.2 µL; 78.11 mg; 604.3 µmoles; 4 eq.) and 2-cyanoethyl N,N-diisopropylchloro-phosphoramidite (50.55 µL; 53.64 mg; 226.6 µmoles; 1.5 eq.) were added. The reaction was stirred at r.t. for 1h and quenched with methanol. The solvent was evaporated *in vacuo* and the residue was purified by flash column chromatography using PE/Et_2O/Et_3N 49/49/2 as an eluent to yield 3-(*p*-*A*nisyldiphenyl)methoxy-2-*O*-((2-cyanoethoxy(diisopropylamino))-phosphino)-hydroxymethyl-propanal-1,3-dioxolane (39.6 %; 39 mg; 59.93 µmoles) as a light yellow oil.

NMR: δ_H ($CHCl_3$); 7.480 (d, ^3J = *7.9 Hz*, 2H, *o*-ArH), 7.361 (d, ^3J = *8.9 Hz*, 4H, H_3COCCH), 7.274-7.158 (m, 3H, *m*-ArH + *p*-ArH), 6.800 (d, ^3J = *8.9 Hz*, 4H, $H_3COCHCH$), 5.090 (d, ^3J = *4.9 Hz*, 1H, O_2CH), 3.970-3.927 (m, 2H, OCH_2CH_2O), 3.867-3.820 (m, 2H, OCH_2CH_2O), 3.763 (s. 6H, CH_3), 3.729-3.705 (m, 1H, O_2CHCH), 3.643-3.903 (m, 2H, $POCH_2$), 3.219-3.192 (m, 4H, 2 OCH_2), 2.612-2.512 (m, 2H, $NCCH_2$), 1.304-1.163 (m, 14H, CCH_3 + H_3CCH). δ_C ($CHCl_3$); 158.3, 145.2, 136.4, 130.1, 128.2, 127.7, 126.6, 113.0, 103.2, 103.1, 85.8, 70.8, 68.1, 64.9, 63.2, 55.2, 43.7, 30.3, 24.6. δ_P ($CHCl_3$); 149.2.

2-Hydroxymethyl-1,3-dioxolane (14) [145]

2-Bromomethyl-1,3-dioxolane (123.9 µL; 200 mg; 1.197 mmoles; 1 eq.) was dissolved in 10 mL of aqueous KOH (2M) and refluxed for 2h. The mixture was extracted with Et_2O (5 times) and the combined organic layers were washed with water and brine. The solvent was evaporated and the residue was purified by flash column chromatography using Et_2O/PE/Et_3N 49/49/2 as an eluent to yield 2-hydroxymethyl-1,3-dioxolane (44.1 %; 55 mg; 528.3 µmoles).

NMR: δ_H ($CDCl_3$); 5.009 (dd, ^3J = *3.1 Hz*, ^3J = *3.1 Hz*, 1H, O_2CH), 4.047-4.015 (m, 2H, OCH_2), 3.952-3.916 (m, 2H, OCH_2), 3.697 (d, ^3J = *3.1 Hz*, 1H, $HOCH_2$), 3.680 (d, ^3J = *3.1 Hz*, 1H, $HOCH_2$). δ_C ($CDCl_3$); 103.0, 65.4, 63.0.

2-*O*-((2-Cyanoethoxy(diisopropylamino))-phosphino)-methyl-1,3-dioxolane (15)

2-Hydroxymethyl-1,3-dioxolane (55 mg; 528.3 µmoles; 1 eq.) was dissolved in anhydrous CH_2Cl_2 (2 mL) and diisopropylethylamine (368 µL; 273.1 mg; 2.113 mmoles; 4 eq.) and 2-cyanoethyl-*N,N*-diisopropylchloro-phosphoramidite (176.7 µL; 187.5 mg; 792.4 µmoles; 1.5 eq.) in anhydrous CH_2Cl_2 (2 mL) were added. The reaction was stirred at r.t. for 1h and quenched with methanol. The mixture was extracted with ethyl acetate (10mL) twice, the ethyl acetate evaporated *in vacuo* and the residue was purified by flash chromatography using PE/EtOAc/Et_3N 89/10/1 as an eluent to yield 2-*O*-((2-cyanoethoxy(diisopropylamino))-phosphino)-methyl-1,3-dioxolane (52.8 %; 85 mg; 279.3 µmoles) as a light yellow oil.

NMR: δ_H ($CDCl_3$); 5.085 (t, ^3J *= 3.9 Hz*, 1H, O_2CH), 4.016-3.975 (m, 2H, OCH_2CH_2O), 3.915-3.850 (m, 3H, OCH_2CH_2O + O_2CHCH_2), 3.652-3.580 (m, 3H, O_2CHCH_2 + $POCH_2$), 2.655 (t, ^3J*= 6.2 Hz*, 2H, $NCCH_2$), 1.202-1.177 (m, 14H, CH_3 + H_3CCH). δ_C ($CDCl_3$); 103.0 (d, $^3J_{C\text{-}P}$ *= 8.0 Hz*), 65.2, 64.1 (d, $^2J_{C\text{-}P}$ *= 15.8 Hz*), 58.7 (d, $^2J_{C\text{-}P}$ *= 18.5 Hz*), 43.1 (d, $^2J_{C\text{-}P}$ *= 12.5 Hz*), 24.7 (t, $^3J_{C\text{-}P}$ *= 8.3 Hz), 20.3* (d, $^3J_{C\text{-}P}$ *= 5.5 Hz).* δ_P ($CDCl_3$); 150.2.

2-Acetoxymethyl-1,3-dioxolane (20)

Acetic acid (151.2 µL; 158.6 mg; 2.641 mmoles; 1.1 eq.) was added dropwise to a solution of 2-hydroxymethyl-1,3-dioxolane (250 mg; 2.401 mmoles; 1 eq.), N,N'-dicyclohexylcarbodiimide (476.8 µL; 594.5 mg; 2.881 mmoles; 1.2 eq.) and 4-dimethylaminopyridine (2.933 mg; 24.01 µmoles; 0.01 eq.) in dry CH_2Cl_2. The mixture was stirred at r.t. overnight. The solvent was evaporated *in vacuo* and the residue suspended in cold EtOAc. This suspension was filtered over a glass filter and washed with cold EtOAc. The solvent was evaporated *in vacuo* and the residue was purified by flash column chromatography using PE/Et_2O as an eluent to yield the 2-acetoxymethyl-1,3-dioxolane (39.8 %; 140 mg; 957.9 µmoles) as a colorless oil.

NMR: δ_H ($CDCl_3$); 5.139 (t, ^3J *= 3.4 Hz*, 1H, O_2CH), 4.127 (d, ^3J *= 3.4 Hz*, 2H, OCH_2), 4.033-3.974 (m, 2H, OCH_2CH_2O), 3.965-2.906 (m, 2H, OCH_2CH_2O), 2.108 (s, 3H, CH_3). δ_C ($CDCl_3$); 170.5, 101.1, 65.2, 62.1, 20.7.

Benzoic acid-1,3-dioxolan-2-ylmethyl ester (22)

Benzoic acid (238.9 µL; 258 mg; 2.113 mmoles; 1.1 eq.) was added dropwise to a solution of 2-hydroxymethyl-1,3-dioxolane (200 mg; 1.921 mmoles; 1 eq.), 4-dimethylaminopyridine (2.347 mg; 19.21 µmoles; 0.01 eq.) and N,N'-dicyclohexyl carbodiimide (397.3 µL; 495.4

mg; 2.401 mmoles; 1.25 eq.) in dry CH_2Cl_2. The mixture was stirred at r.t. overnight. The solvent was evaporated *in vacuo* and the residue suspended in cold EtOAc. This suspension was filtered over Celite and washed with cold EtOAc. The solvent was evaporated *in vacuo* . The solvent was evaporated *in vacuo* and the residue was purified by flash column chromatography using chloroform as an eluent to yield benzoic acid-1,3-dioxolan-2-ylmethyl ester (26.2 %; 105 mg; 504.2 µmoles) as a light yellow oil.

NMR: δ_H ($CDCl_3$); 8.066 (d, ^3J = *7.9 Hz*, 2H, *o*-H_{arom}), 7.579-7.390 (m, 3H, H_{arom}), 5.285 (t, ^3J = *3.7 Hz*, 1H, O_2CH), 4.377 (d, ^3J = *3.7 Hz*, 2H, OCH_2), 4.064-4.005 (m, 2H, OCH_2CH_2), 3.988-3.930 (m, 2H, OCH_2CH_2).

_p_-Methoxyphenyl-1,3-dioxolane (16)

p-Anisaldehyde (178.2 µL; 200 mg; 1.468 mmoles; 1 eq.) was placed in a round bottomed flask fitted with a Dean Stark receiver and dissolved in toluene. 5 mL (5.565 g) of ethylene glycol was added and the mixture was refluxed for 14h.

The solvent was evaporated *in vacuo* to yield *p*-methoxyphenyl-1,3-dioxolane (87.2 %; 231 mg; 1.281 mmoles) as a colorless oil.

NMR: δ_H ($CDCl_3$); 7.407 (d, ^3J = *8.8 Hz*, 2H, *o*-H_{arom}), 6.903 (d, ^3J = *8.8 Hz*, 2H, *m*-H_{arom}), 5.758 (s, 1H, O_2CH), 4.148-4.059 (m, 2H, OCH_2CH_2O), 4.088-3.947 (m, 2H, OCH_2CH_2O), 3.811 (s, 3H, CH_3).

2-Benzyl-1,3-dioxolane (18)

Phenylacetaldehyde (194.7 µL; 200 mg; 1.664 mmoles; 1 eq.) was placed in a roundbottomed flask fitted with a Dean Stark receiver and dissolved in toluene. Ethylene glycol (928.2 µL; 1.033 g; 16.64 mmoles; 10 eq.) was added and the mixture was refluxed for 14h. The solvent was evaporated *in vacuo* to yield 2-benzyl-1,3-dioxolane (80.4 %; 220 mg; 1.339 mmoles) as a colorless oil.

NMR: δ_H ($CDCl_3$); 7.328-7.251 (m, 5H, H_{arom}), 5.088 (t, ^3J = *5.2 Hz*, 1H, O_2CH), 3.981-3.838 (m, 4H, OCH_2CH_2O), 2.991 (d, ^3J = *5.2 Hz*, 2H, CH_2).

1-Trimethylsilanyloxy-2,3-propanediol (32) [162]

Osmium tetroxide (97.57 mg; 383.8 µmoles; 0.05 eq.) was placed in a two-necked flask and dissolved in THF (15mL). 4-methylmorpholine-N-oxide monohydrate (1.101 mL; 1.245 g; 9.212 mmoles; 1.2 eq.) and allyloxytrimethylsilane (1.293 mL; 1000 mg; 7.676 mmoles; 1

eq.) were added and the solution was stirred at r.t. and monitored by TLC. After 3h at r.t. the reaction was quenched by the addition of 7 µL of a sodium thiosulfate solution.
The mixture was filtered over a glass filter and washed with THF. The solvent was evaporated *in vacuo* to yield 1-trimethylsilanyloxy-2,3-propanediol (99 %; 1.245 g; 7.675 mmoles) as a colorless oil.
NMR: δ_H ($CDCl_3$); 3.761–3.603 (m, 5H, CH_2OH, CHOH, CH_2OSi), 0.142 (s, 9h, $Si(CH_3)_3$). δ_C ($CDCl_3$); 71.4, 64.1, 64.0, -0.7.

1-Trimethylsilane-2,3-dibenzoylpropane (33)
1-Trimethylsilanyloxy-2,3-propanediol (1.245 g; 7.676 mmoles; 1 eq.) was placed in a round-bottomed flask and dissolved in acetonitrile (20mL). Benzoyl cyanide (2.73 mL; 3.019 g; 23.02 mmoles; 3 eq.) and triethylamine (106.6 µL; 77.67 mg; 767.6 µmoles; 0.1 eq.) were added at 0°C and the mixture was stirred for 1h at 0°C. The reaction was quenched with methanol (1 mL). The solvent was evaporated *in vacuo* and the crude reaction mixture was purified by flash column chromatography using PE/EtOAc/Et_3N 95/4/1 as an eluent yielding 1-trimethylsilane-2,3-dibenzoylpropane (60.2 %; 1.722 g; 4.622 mmoles) as a colorless oil.
NMR: δ_H ($CDCl_3$); 8.07–8.00 (m, 4H, ArH), 7.58–7.53 (m, 2H, ArH), 7.45–7.40 (m, 4H, ArH), 5.506 - 5.452 (m, 1H, CH), 4.701-4.593 (m, 2H, CH_2OCO), 3.926 (d, ^1J = 4.8 Hz, 2H, CH_2OSi), 0.117 (s, 9H, $Si(CH_3)_3$). δ_C ($CDCl_3$); 134.2, 130.1, 129.7, 128.4, 71.7, 64.3, 45.6, 1.0.

2,3-Dibenzoyl-1-propanol (34) [163]
1-Trimethylsilanyloxy-2,3-propanediol (100 mg; 268.4 µmoles; 1 eq.) was placed in a round-bottomed flask and dissolved in 25 mL of dry THF. 0.295 mL of a 1M solution of tetrabutylammonium fluoride (TBAF) in THF was added via a dropping funnel. The mixture was stirred at r.t. for 1h and monitored by TLC.
The solvent was evaporated and the mixture was dissolved in CH_2Cl_2 and washed with water and brine. The organic phase was dried over $MgSO_4$ and filtered and the solvent was evaporated. The residue was purified by flash column chromatography using PE/Et_2O 50/50 as an eluent to yield 2,3-dibenzoyl-1-propanol (89.9 %; 72.5 mg; 241.4 µmoles) as a colorless oil.

NMR: δ_H ($CDCl_3$); 8.071-8.023 (m, 4H, ArH), 7.601-7.564 (m, 2H, ArH), 7.465-7.428 (m, 4H, ArH), 4.569-4.481 (m, 5H, CH_2OH, CH, CH_2OCO), 2.690 (broad s, 1H, OH). δ_C ($CDCl_3$); 132.2 131.2, 130.5, 128.4, 72.3, 65.9, 49.7.

1-*O*-((2-Cyanoethoxy(diisopropylamino)-phosphino))-2,3-dibenzoylpropane (35)

2,3-Dibenzoylpropanol (25 mg; 83.24 µmoles; 1 eq.) was dissolved in anhydrous CH_2Cl_2 (2 mL) and 2-cyanoethyl-N,N-diisopropylamino-chloro-phosphoramidite (27.85 µL; 29.55 mg; 124.8 µmoles; 1.5 eq.) and diisopropylethylamine (58 µL; 43.03 mg; 332.9 µmoles; 4 eq.) in anhydrous CH_2Cl_2 (2 mL) were added. The reaction was stirred at r.t. for 1h and quenched with methanol. The mixture was extracted with EtOAc twice (10mL) and the residue was purified by flash chromatography using PE/EtOAc/Et_3N 96/3/1 as an eluent to yield a mixture of stereoisomers of 1-*O*-((2-Cyanoethoxy(diisopropylamino)-phosphino))-2,3-dibenzoylpropane (31.2 %; 13 mg; 26.02 µmoles) as a light yellow oil.

NMR: δ_H ($CDCl_3$); 8.08-8.01 (m, 4H, ArH), 7.58-7.53 (m, 2H, ArH), 7.47-7.41 (m, 4H, ArH), 4.721-4.463 (m, 5H, CH_2OCO, CHOCO, CH_2OP), 3.877-3.717 (m, 2H, CH_2CH_2CN), 3.659-3.553 (m, 2H, CH_2CN), 2.612-2.553 (m, 2H, CHN), 1.161 (t, ^3J = 4.0 Hz, 12H, CH_3). δ_C ($CDCl_3$); 166.2, 133.2-133.1 (m), 129.8-129.6, 128.5-128.3 (m), 117.5, 71.5 (d, J_{C-P} = 7 Hz), 69.6 (d, J_{C-P} = 16 Hz), 64.8 (d, J_{C-P} = 7 Hz), 64.8 (d, J_{C-P} = 6 Hz), 58.3 (d, J_{C-P} =19 Hz), 43.3 (d, J_{C-P} = 12 Hz), 43.2 (d, J_{C-P} = 12 Hz), 24.7-24.4 (m), 20.2 (d, J_{C-P} = 7 Hz). δ_P ($CDCl_3$); 151.0, 150.3, M_w+H^+ = 500.9 (calc. 500.21).

5 Perspectives

5.1 Short term research on natural abasic site repair systems

The development of the repair model system for natural abasic sites was stopped during deprotection of the oligonucleotide containing a natural abasic site. The problems encountered at this stage, degradation of the sample in solution, were preventing us from continuing. Since it was proven possible to obtain these abasic sites [129, 131], the deprotection should be pursued. It may be recommended that biologists, who have a better feeling for handling DNA samples, perform these experiments. First synthesizing the same oligonucleotide sequence as Rayner used can also be recommended. The conditions to obtain this oligonucleotide are well described. When synthesis of this oligonucleotide is achieved, the same conditions can be used to synthesize the oligonucleotide that will be useful for our purposes.

Since only the group of Rayner has used the nitro benzyl route to incorporate abasic sites in oligonucleotides [129], it might be worthwhile to investigate whether other methods are more convenient.

After deprotection, the abasic sites should be tested for their ability to be repaired by the developed hydrazides. We thought of performing surface plasmon resonance (SPR) experiments. These experiments require very little material (a few nanomoles), are fast (30 minutes) and give information about the association and dissociation kinetics of the oligonucleotide duplexes.

Because of the slow reactions of the hydrazides, we will have to incubate the abasic sites with the hydrazides before starting the binding experiment. The test reactions needed 24h to reach equilibrium and the reactions with the abasic site may be slower due to the decreased concentrations. A large excess (20-100 equivalents) of hydrazide should be used.

We performed some preliminary experiments to see whether this technique might give the results we expect. A typical SPR experiment consists of four steps. In the first step a streptavidine layer is fixed on a gold surface. In the second step a target strand is fixed on the streptavidine layer by using a DNA strand containing a 3'end biotin moiety. The biotin group has a high affinity for the streptavidine and is fixed sufficiently well to remain fixed on the surface during the rest of the experiments.

In the third step we add the strand of which we want to measure its affinity for the target strand. We add this strand during a certain time (typically 5 minutes), in a continuous flow of solvent. The amount of oligonucleotides that are attached to the surface can be quantified by

measuring the mass that is attached to the surface. The quantity of oligonucleotides that is accumulated on the surface depends on both the association and the dissociation constant. We measure the kinetics of the sum of the association and the dissociation constant.

In the fourth step of the SPR experiment we continue adding solvent, but without adding more of the strand we want to measure. The oligonucleotides will gradually dissociate from the surface and the mass on the surface will decrease exponentially. Mathematical fitting of the curve will give the dissociation constant. With the dissociation constant in hand the association constant can be calculated from the curve obtained in step three.

In the preliminary studies we performed, we used the samples of impure oligomer containing an abasic site that we obtained. Incubation of the strands with our repair units was not performed, since in the 24h incubation time that is needed for those experiments the oligonucleotide almost fully degraded.

We measured the undecamer $^{5'}$CGC AGX GAC GC$^{3'}$ (with X = natural abasic site) for its affinity towards its complementary strand, which was fixed on the gold surface. The estimated melting temperature of the fully complementary duplex is 35°C. Since the presence of an abasic site lowers the melting temperature by 15-20°C, an affinity of the strand will only be observed below 15-20°C. We performed our first experiments at 10°C and indeed found an affinity of the strand containing the abasic site for the complementary strand. At 20°C no mass was fixed on the surface, the estimated melting temperature is thus approximately correct. The temperature should be kept low for the binding experiments, which is advantageous for the yield of the reaction between the hydrazides and the abasic sites. The test reactions revealed that the yield augments with decreasing temperature. However, the kinetics of the reaction will be very slow.

Because of the slow equilibrium, we need to incubate the single strand with the hydrazides. Displacement of the equilibrium to the product side because of double helix stabilization will thus not be possible. The incubated strands will be tested for their affinity towards a complementary target strand that is fixed on the streptavidine layer. Because of the slow equilibrium, we will measure the kinetics of a mixture of repaired and non-repaired DNA strands simultaneously. This is very well possible, but the fraction of repaired strands must remain constant during the experiment. This means that no degradation of the oligonucleotide should take place during the experiment. Without degradation it may be possible to resolve the association and dissociation kinetics of both the unrepaired and the repaired DNA oligonucleotides using a double exponential fitting method.

To be able to measure only the kinetics of the repaired duplexes, it may be envisaged to perform the experiments at a temperature that is above the melting temperature of the DNA duplex containing the abasic site, but that is still below the melting temperature of the repaired duplex. In that case the non-repaired DNA strands will not bind to the complementary strand and only the kinetics of the repaired strands are observed. This will facilitate the analysis of the results.

If these reactions turn out to be successful, it may be attempted to incorporate two or more abasic sites in the same duplex. This amounts to designing a DNA replicating system. From a synthetic point of view, the easiest way to incorporate two abasic sites in a DNA duplex is designing a self-complementary duplex to give one abasic site in both of the duplex strands. However, this method will fail if we want to measure the kinetics of duplex formation with SPR, because SPR requires the fixation of one of the strands on a streptavidine surface with a biotin moiety. Incorporating a biotin on the DNA strand requires an extra step that will complicate synthesis and purification.

Traditional binding experiments to determine a melting curve of the duplex with one abasic site in both strands may be feasible. The DNA oligonucleotides should be incubated with a 20 fold excess of hydrazides to give good amount of repaired abasic sites. The melting curve may be difficult to analyze, since three species may be present in the mixture; the fully repaired duplex, the duplex having only one abasic site repaired, and the duplex still having two unrepaired abasic sites.

5.2 Short term research on artificial abasic site repair systems

The design and the synthesis of an aldehyde that can be incorporated in the middle of a strand were not successful. Incorporation of the aldehyde using a 1,3-dioxolane failed because the published deprotection method turned out not to work as described. We then designed 1,2-diols that can be cleaved by treating them with $NaIO_4$ to give an aldehyde. Synthesis of the 1,2-diols that were suitable targets turned out to be difficult and was not achieved. The abasic site that can be incorporated at the end of a DNA strand was prepared, but not yet incorporated. That is the first thing that should be done. Test reactions with that compound should reveal whether it is possible to repair abasic sites with aminobases in a reversible way.

If the method can be validated, another route of synthesizing artificial abasic sites that can be incorporated in the middle of a DNA strand should be envisaged. Maybe one should look at the work of Herdewijn *et al.* [164]. They were interested in extrahelical sugar moieties to

incorporate extrahelical functional groups. Their DNA duplexes, as presented in Figure 5.1, do not loose much of their base stacking properties relative to a natural DNA duplex. The additional sugar moiety can be incorporated by protection with benzoyl protecting groups. Functionalization of the sugar to yield two aldehyde functions was performed by oxidation with sodium periodate. This periodate method was also used by the group of Zatsepin *et. al.* [159, 160], as shown in chapter 4.

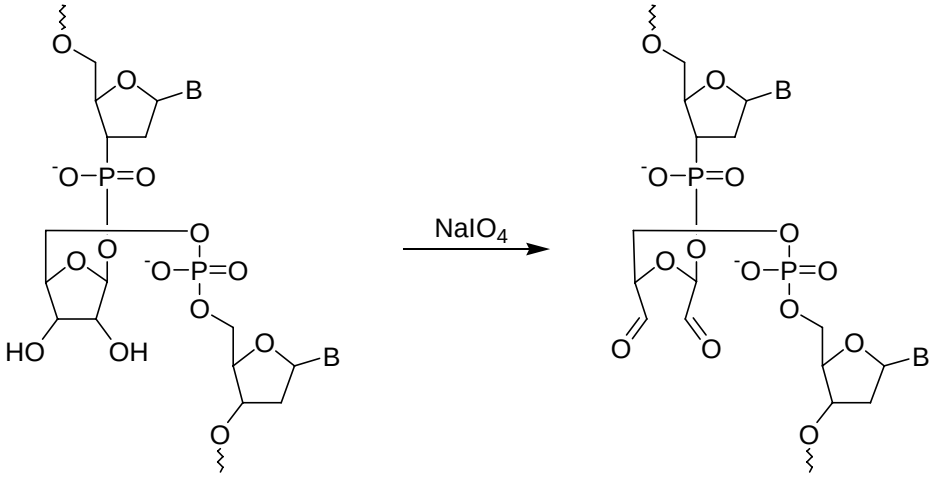


Figure 5.1. Incorporation of extrahelical, functionalizable sugar moieties

The structure as presented in Figure 5.1 is not useful for our purpose. The backbone in the case of Herdewijn *et. al.* is attached 5'-1', instead of the usual 3'-5'. This favors the extrahelical functionalization they desire, but our products need to be able to adopt an intrahelical position.

Attaching the sugar in the usual 3'-5' fashion and replacing the oxygen atom in the sugar ring by a carbon atom, gives three possible building blocks to yield three different di-aldehydes (Figure 5.2). Building block **1** corresponds to a ribose abasic site in which the oxygen atom of the sugar ring is replaced by a carbon. The aldehydes that are formed by treating this precursor with $NaIO_4$ are similar to the two aldehydes that were designed in chapter 4. Compound **3** is also a 1,2-diol and has the hydroxyl groups on other carbon atoms relative to compound **1.** One of the two aldehydes of compound **4** is in the same position as designed in chapter 4, the other is in a position that sofar has not been considered. Compound **5** is a 1,2,3-triol and during the treatment of **5** with $NaIO_4$, it will loose one carbon atom. The two aldehydes of **6** are in positions that were modeled in chapter 4.

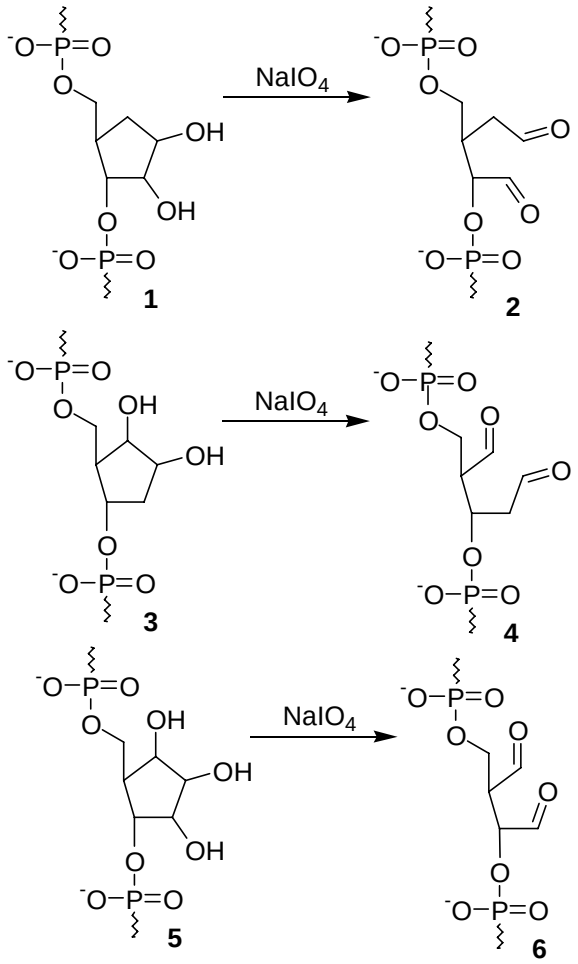


Figure 5.2. Revised abasic site model.

The advantage of the two aldehyde functions is that there in each case two possible positions for the aminobase to react. This increases the chance that a stable conformation is found for one of the reaction products.

The synthesis of the monomers of these three structures has been described. Compound **1** can be obtained with acetyl protecting groups on its 3'- and 5'-hydroxyl and a isopropylene group on its 1'- and 2'-hydroxyl group in eight steps starting from commercial 4.6-*O*-ethylidene-*D*-glucose [165]. Removal of the isopropylene group and protecting of the 5' hydroxyl group with a DMT group and activation of the 3' hydroxyl group as its phosphoramidite will cost three more steps. The monomer can thus be obtained in 11 steps.

The monomer of **3** without the protecting groups was prepared by Gathergood *et al*. in 5 steps [166]. Incorporation of the correct protecting groups may take another four steps to give a total of 9.

Synthesis of the skeleton of the last compound involves incorporation of the right protecting groups at the right place already. A four step synthesis starting from α-methyl-D-mannopyranoside will lead to the desired phosphoramidite [167].

Reaction with the four aminobases should be reversible. This can be expected, since it is known that the aminobases react reversible with aliphatic aldehydes in water. It should be tested though, because we think that to obtain a replicating system that shows high fidelity, we need a system that is able to self-correct its own errors.

5.3 Long term research on artificial abasic site repair systems

If we can make these systems work, we can extend them to more than one abasic site per strand. First two abasic sites, then tree abasic sites, four….. We might even envisage a strand that only consists of aldehydes and let it form a duplex with a complementary strand of the same length (Figure 5.3). If a fast equilibrium is present between the aminobases and the aldehydes, the strand that is perfectly complementary with its target strand will be the major product. If we use a temperature just below the melting temperature of the full duplex, this may be the only strand that binds successfully in an SPR experiment.

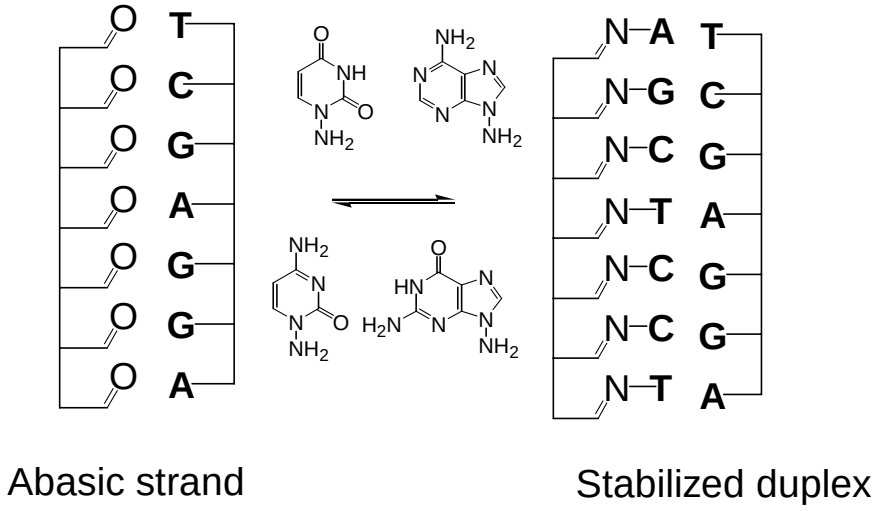


Figure 5.3. Empty strand that can be 'filled' with bases

This principle may be useful for the recognition of unknown DNA strands, with conditions under which normal DNA sequencing is impossible. One can for instance think of *in vivo* strand recognition. Addition of a blank, aldehydic strand and the four aminobases to an unknown strand, may give the complementary strand of the target strand. If one can then separate the modified strand from the rest of the reaction mixture, it can be read and the sequence of the unknown target strand is known. If the target strand is for instance a virus that

should be killed, a drug can be attached to a DNA oligomer that has the sequence complementary to the sequence of the target strand that has been developed. This drug will then selectively bind to the virus DNA and not kill any vital cells.

A second interest of this research is to get a better understanding of how replicating systems were developed by nature. The model system here gives a strand that resembles DNA, but will be much more flexible. For many years already it is debated how nature 'invented' the ability of DNA to store information and transfer the information to further generations [1, 2, 4, 33, 34, 168, 169]. Did other replicating systems exist before DNA or RNA, or did DNA and RNA evolve from other systems? Most people believe that DNA evolved from RNA, but that other systems existed before, making use of some complementary molecules. It is very improbable that a molecule like RNA appears *de novo* without any predecessors that would have had similar functions. But what these systems looked like and how they evolved to RNA and later to DNA is not known.

The hypothesis that DNA evolved from RNA is also not proven. Some people [2] think they may as well have developed together, both with their own functionality. Other researchers think that maybe other sugars may have functioned as replicating systems [170]. During evolution DNA and RNA may have been selected out of these various replicating systems as the most reliable replicating systems. The reason for this selection is unclear so far, but many research groups have synthesized DNA analogs like, PNA [171], LNA [172-175], bicyclo DNA [134], TNA [176] HNA [177] and various other analogues to investigate whether replication with these strands is possible. Whether any of these replicating systems might once have been the natural system, is unknown.

If other replicating systems than DNA, RNA or their analogues existed, they must have been very primitive. They must have developed under the conditions that were present on the early earth, which limits the amount of possible reactions. Only reactions in no other solvent than water at moderate pH and moderate temperature can be considered. Only a limited number of reagents was available.

A self-replicating clay was proposed by Cairns-Smith [3]. Cairns-Smith suggested that organic molecules were incorporated in the clay, and at one moment in history an organic species was self-replicating faster than the clay and thus became the predominant species.

The group of Leslie Orgel attempted the synthesis of RNA and DNA oligonucleotides under conditions that could have been present on earth millions of years ago [4]. They managed to synthesize nucleobases under 'prebiotic' conditions and even managed to synthesize small

quantities of nucleosides. However, phosphorylation attempts and polymerization of the nucleosides failed.

We do not pretend that our system was one of the intermediate systems that led to the development of DNA replicating systems. The aldehyde functions on the blank strand would have been much too vulnerable to attack by ammonia, CO or any amine that were present under ‘prebiotic’ conditions. We just hope that these systems may lead to a better insight about how replicating systems can be developed. And maybe, one day, someone may find out what the first replicating really looked like.

6 Glossary

A,T,G,C	Adenine, Thymine, Guanine, Cytosine
AP-site	Abasic site (apurinic or apyrimidinic)
CA	Carbonic Anhydrase
CAN	Cerium(IV)Ammonium Nitrate
CD	Circular Dichroism
d	Doublet
DCC	Dynamic Combinatorial Chemistry
DCC	N,N'-DiCyclohexylCarbodiimide
DCL	Dynamic Combinatorial Library
DIPEA	DiIsoPropylEthylAmine
DMAP	4-DiMethyAminoPyridine
DMF	DiMethyl Formamide
DMT	DiMethoxyTrityl
DMT-MM	4-(4,6-DiMethoxy-1,3,5-Triazin-2-yl)-4-MethylMorpholinium chloride
DNA	Deoxyribose Nucleic Acid
EDC	1-(3-dimethylaminopropyl)-3-Ethylcarbodiimide
Ee	Enantiomeric Excess
ELISA	Enzyme - Linked Immuno Sorbent Assay
HAOS	hydroxylamine-O-sulfonic acid
HPLC	High-Performance Liquid Chromatography
Hz	Hertz
J	Coupling constant
m	Multiplet
M.p.	Melting Point
MMT	MonoMethoxyTrityl
MSH	O-MesityleneSulfonylHydroxylamine
PE	Petroleum Ether
PNA	Peptide Nucleic Acid
PNP	*p*-NitroPhenyl
r.t.	Room Temperature
RNA	Ribose Nucleic Acid

RP-HPLC	Reversed Phase High-Performance Liquid Chromatography
s	Singlet
SPR	Surface Plasmon Resonance
t	Triplet
TBAF	TetraButylAmmonium Fluoride
TBDMS	Tert-Butyl DiMethyl Silyl
THF	TetraHydroFurane
TIPS	TriIsoPropylSilyl ether
TIPBSF	TriIsoPropylBenzeneSulfonyl Fluoride
TLC	Thin Layer Chromatography
Tm	Melting Temperature (of oligonucleotides)

7 Literature references

1. Strobel, S.A., *Repopulating the RNA world.* Nature, 2001. **411**: p. 1003-1006.
2. Orgel, L.E. and R. Lohrmann, *Prebiotic chemistry and nucleic acid replication.* Acc. Chem. Res., 1974. **7**: p. 368-377.
3. Cairns-Smith, A.G., *Genetic takeover and the mineral origins of life*. 1982: Cambridge university press.
4. Orgel, L.E., *Unnatural selection in chemical systems.* Acc. Chem. Res., 1995. **28**: p. 109-118.
5. Hey, M., C. Hartel, and M.W. Göbel, *Nonenzymatic oligomerization of ribonucleotides: Towards in vitro selection experiments.* Helv. Chim. Act., 2003. **86**: p. 844-854.
6. Anderson, S., H.L. Anderson, and J.K.M. Sanders, *Expanding roles for templates in synthesis.* Acc. Chem. Res., 1993. **26**: p. 469-475.
7. Thompson, M.C. and D.H. Busch, *Reactions of coordinated ligands. VI. Metal ion control in the synthesis of planar nickel(II) complexes of α-diketo-bis-mercaptoimines.* J. Am. Chem. Soc., 1964. **86**(213-217).
8. D'Acerno, C., et al., *Template effects and kinetic selection in the self-assembly of crown ether cyclobis(paraquat-p-phenylene) [2]catenanes - effect of the 1,4-dioxybenzene and 1,5-dioxynaphthalene units.* Chem. Eur. J., 2000. **6**(19): p. 3540-3546.
9. Hasenknopf, B., et al., *Self-assembly of tetra- and hexanuclear circular helicates.* J. Am. Chem. Soc., 1997. **119**: p. 10956-10962.
10. Kirby, A.J., *Enzyme mechanisms, models, and mimics.* Angew. Chem. Int. Ed., 1996. **35**: p. 707-724.
11. Stryer, L., *Biochemistry*. 4 ed. 1995, New York: W.H. Freeman and Company.
12. Breslow, R., M. Hammond, and M. Lauer, *Selective transamination and optical induction by a β-cyclodextrin-pyridoxamine artificial enzyme.* J. Am. Chem. Soc., 1980. **102**: p. 421-422.
13. Breslow, R. and A.W. Czarnik, *Transaminations by pyridoxamine selectively attached at C-3 in β-cyclodextrin.* J. Am. Chem. Soc., 1983. **105**: p. 1390-1391.
14. Kelly, T.R., C. Zhao, and G.J. Bridger, *A bisubstrate reaction template.* J. Am. Chem. Soc., 1989. **111**: p. 3744-3745.

15. Huc, I., R.J. Pieters, and J. Rebek, *Role of geometrical factors in template effects.* J. Am. Chem. Soc., 1994. **116**: p. 10296-10297.
16. Naylor, R. and P.T. Gilham, *Studies on some interactions and reactions of oligonucleotides in aqueous solution.* Biochemistry, 1966. **5**: p. 2723.
17. Xu, Y. and E.T. Kool, *Rapid and selective selenium-mediated autoligation of DNA strands.* J. Am. Chem. Soc., 2000. **122**: p. 9040-9041.
18. Fujimoto, K., et al., *Template-directed photoreversible ligation of deoxyoligonucleotides via 5-vinyldeoxyuridine.* J. Am. Chem. Soc., 2000. **122**: p. 5646-5647.
19. Liu, D.R. and P.G. Schultz, *Generating new molecular function: A lesson from nature.* Angew. Chem. Int. Ed., 1999. **38**(1): p. 36-54.
20. Gartner, Z.J. and D.R. Liu, *The generality of DNA-templated synthesis as a basis for evolving non-natural small molecules.* J. Am. Chem. Soc., 2001. **123**: p. 6961-6963.
21. Gartner, Z.J., M.W. Kanan, and D.R. Liu, *Multistep small-molecule synthesis programmed by DNA templates.* J. Am. Chem. Soc., 2002. **124**: p. 10304-10306.
22. Calderone, C.T., et al., *Directing otherwise incompatible reactions in a single solution by using DNA-templated organic synthesis.* Angew. Chem. Int. Ed., 2002. **41**(21): p. 4104-4108.
23. Gartner, Z.J., et al., *Two enabling architectures for DNA-templated organic synthesis.* Angew. Chem. Int. Ed., 2003. **42**(12): p. 1370-1375.
24. Li, X. and D.R. Liu, *Stereoselectivity in DNA-templated organic synthesis and its origins.* J. Am. Chem. Soc., 2003. **125**: p. 10188-10189.
25. Kanan, M.W., et al., *Reaction discovery enabled by DNA-templated synthesis and in vitro selection.* Nature, 2004. **431**: p. 545-549.
26. Kunishima, M., et al., *Formation of carboxamides by direct condensation of carboxylic acids and amines in alcohols using a new alcohol- and water-soluble condensing agent: DMT-MM.* Tetrahedron, 2001. **57**: p. 1551-1558.
27. Rostovtsev, V.V., et al., *A stepwise Huisgen cycloaddition process: Copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes.* Angew. Chem. Int. Ed., 2002. **41**: p. 2596-2599.
28. Jäschke, A. and B. Seelig, *Evolution of DNA and RNA as catalysts for chemical reactions.* Curr. Op. in Chem. Biol., 2000. **4**: p. 257-262.
29. Jäschke, A., *Artificial ribozymes and deoxyribozymes.* Curr. Opin. Chem. Biol., 2001. **11**(3): p. 321-326.

30. Jäschke, A., C. Frauendorf, and F. Hausch, *In vitro selected oligonucleotides as tools in organic chemistry.* Synlett, 1999. **6**: p. 825-833.
31. Seelig, B., et al., *Enantioselective ribozyme catalysis of a bimolecular cycloaddition reaction.* Angew. Chem. Int. Ed., 2000. **39**(24): p. 4576-4579.
32. Seelig, B. and A. Jäschke, *A small catalytic RNA motif with Diels-Alderase activity.* Chem. & Biol., 1999. **6**: p. 167-176.
33. Bartel, D.P. and P.J. Unrau, *Constructing an RNA world.* Trends Biochem. Sci., 1999. **35**(12): p. M9-M13.
34. Famulok, M. and A. Jenne, *Catalysis based on nucleic acid structures.* Top. Curr. chem., 1999. **202**: p. 101-131.
35. Cousins, G.R.L., S.-A. Poulsen, and J.K.M. Sanders, *Molecular evolution: dynamic combinatorial libraries, autocatalytic networks and the quest for molecular function.* Curr. Opin. Chem. Biol., 2000. **4**: p. 270-279.
36. Reymond, J.-L., *Outrunning the bear.* Angew. Chem. Int. Ed., 2004. **43**: p. 5577-5579.
37. Grote, Z., R. Scopelliti, and K. Severin, *Adaptive behavior of dynamic combinatorial libraries generated by assembly of different building blocks.* Angew. Chem. Int. Ed., 2003. **42**: p. 3821-3825.
38. Borman, S., *The many faces of combinatorial chemistry.* C&EN, 2003. **81**(43): p. 45-56.
39. Lehn, J.-M., *Dynamic combinatorial chemistry and virtual combinatorial libraries.* Chem. Eur. J., 1999. **5**(9): p. 2455-2463.
40. Huc, I. and J.-M. Lehn, *Virtual combinatorial libraries: dynamic generation of molecular and supramolecular diversity by self-assembly.* Proc. Natl. Acad. Sci. USA, 1997. **94**: p. 2106-2110.
41. Eliseev, A.V. and M.I. Nelen, *Use of molecular recognition to drive chemical evolution. 1. Controlling the composition of an equilibrating mixture of simple arginine receptors.* J. Am. Chem. Soc., 1997. **119**: p. 1147-1148.
42. Eliseev, A.V. and M.I. Nelen, *Use of molecular recognition to drive chemical evolution: mechanisms of an automated genetic algorithm implementation.* Chem. Eur. J., 1998. **4**(5): p. 825-834.
43. Hioki, H. and W. Clark Still, *Chemical Evolution: a model system that selects and amplifies a receptor for the tripeptide (D)Pro(L)Val(D)Val.* J. Org. Chem., 1998. **63**: p. 904-905.

44. Lewis, W.G., et al., *Click chemistry in situ: acetylcholinesterase as a reaction vessel for the selective assembly of a femtomolar inhibitor from an array of building blocks.* Angew. Chem. Int. Ed., 2002. **41**(6): p. 1053-1057.
45. Haupt, K. and K. Mosbach, *Molecularly imprinted polymers and their use in biomimetic sensors.* Chem. Rev., 2000. **100**(7): p. 2495-2504.
46. Mosbach, K. and O. Ramström, *The emerging technique of molecular imprinting and its future impact on biotechnology.* Biol/Technology, 1996. **14**(2): p. 163-170.
47. Parmpi, P. and P. Kofinas, *Biomimetic glucose recognition using molecularly imprinted polymer hydrogels.* Biomat., 2004. **25**: p. 1969-1973.
48. Visnjevski, A., E. Yilmaz, and O. Brüggemann, *Catalyzing a cycloaddition with molecularly imprinted polymers obtained via immobilized templates.* Appl. Cat. A., 2004. **260**(2): p. 169-174.
49. Böhler, C., P.E. Nielsen, and L.E. Orgel, *Template switching between PNA and RNA oligonucleotides.* Nature, 1995. **376**: p. 578-581.
50. Li, X., et al., *DNA-catalyzed polymerization.* J. Am. Chem. Soc., 2002. **124**(5): p. 746-747.
51. Mattes, A. and O. Seitz, *Mass-spectrometric monitoring of a PNA-based ligation reaction for the multiplex detection of DNA single nucleotide polymorphisms.* Angew. Chem. Int. Ed., 2001. **40**(17): p. 3178-3181.
52. Mattes, A. and O. Seitz, *Sequence fidelity of a template-directed PNA-ligation reaction.* Chem. Commun., 2001: p. 2050-2051.
53. Von Kiedrowski, G., *A self-replicating hexadeoxynucleotide.* Angew. Chem. Int. Ed., 1986. **25**(10): p. 932-935.
54. Von Kiedrowski, G., B. Wlotzka, and J. Helbing, *Sequence dependence of template-directed syntheses of hexadeoxynucleotide derivatives with 3'-5' pyrophosphate linkage.* Angew. Chem. Int. Ed., 1989. **28**(9): p. 1235-1237.
55. Von Kiedrowski, G., et al., *Parabolic growth of a self-replicating hexadeoxynucleotide bearing a 3'-5'-phosphoamidate linkage.* Angew. Chem. Int. Ed., 1991. **30**(4): p. 423-426.
56. Sievers, D. and G. Von Kiedrowski, *Self-replication of complementary nucleotide-based oligomers.* Nature, 1994. **369**: p. 221-224.
57. Sievers, D. and G. Von Kiedrowski, *Self-replication of hexadeoxynucleotide analogues: autocatalysis versus cross-catalysis.* Chem. Eur. J., 1998. **4**(4): p. 629-641.

58. Terfort, A. and G. Von Kiedrowski, *Self-replication by condensation of 3-aminobenzamidines and 2-formylphenoxyacetic acids.* Angew. Chem. Int. Ed., 1992. **31**(5): p. 654-656.
59. Nowick, J.S., et al., *Kinetic studies and modeling of a self-replicating system.* J. Am. Chem. Soc., 1991. **113**: p. 8831-8839.
60. Tjivikua, T., P. Ballester, and J. Rebek, *A self-replicating system.* J. Am. Chem. Soc., 1990. **112**: p. 1249-1250.
61. Menger, F.M., A.V. Eliseev, and N.A. Khanjin, *"A self-replicating system": new experimental data and a new mechanistic interpretation.* J. Am. Chem. Soc., 1994. **116**: p. 3613-3614.
62. Reinhoudt, D.N., D.M. Rudkevich, and F. de Jong, *Kinetic analysis of the Rebek self-replicating system: is there a controversy?* J. Am. Chem. Soc., 1996. **118**: p. 6880-6889.
63. Conn, M.M., E.A. Wintner, and J. Rebek, *New evidence for template effects in a self-replicating system.* J. Am. Chem. Soc., 1994. **116**: p. 8823-8824.
64. Wintner, E.A., M.M. Conn, and J. Rebek, *Self-replicating molecules: a second generation.* J. Am. Chem. Soc., 1994. **116**(20): p. 8877-8884.
65. Von Kiedrowski, G., et al., *Parabolische Selbstvermehrung und der Ursprung der Replikation.* Nachr. Chem. Tech. Lab., 1992. **40**: p. 578-588.
66. Issac, R. and J. Chmielewski, *Approaching exponential growth with a self-replicating peptide.* J. Am. Chem. Soc., 2002. **124**: p. 6808-6809.
67. Yao, S., et al., *A self-replicating peptide under ionic control.* Angew. Chem. Int. Ed., 1998. **37**(4): p. 478-481.
68. Hansson, L.O., M. Widersten, and B. Mannervik, *Mechanism-based phage display selection of active-site mutants of human glutathione transferase A1-1 catalyzing S_NAr reactions.* Biochemistry, 1997. **36**: p. 11252-11260.
69. Bachmann, P.A., et al., *Self-replicating reverse micelles and chemical autopoiesis.* J. Am. Chem. Soc., 1990. **112**: p. 8200-8201.
70. Bachmann, P.A., et al., *Self-replicating micelles: aqueous micelles and enzymatically driven reactions in reverse micelles.* J. Am. Chem. Soc., 1991. **113**: p. 8204-8209.
71. Li, T. and K.C. Nicolaou, *Chemical self-replication of palindromic duplex DNA.* Nature, 1994. **369**: p. 218-221.
72. Pieters, R.J., I. Huc, and J. Rebek, *Reciprocal template effects in a replication cycle.* Angew. Chem. Int. Ed., 1994. **33**(15/16): p. 1579-1581.

73. Pieters, R.J., I. Huc, and J. Rebek, *Reciprocal template effects in bisubstrate systems: a replication cycle.* Tetrahedron, 1995. **51**(2): p. 485-498.
74. Mackenzie, J.A. and P.R. Strauss, *Oligonucleotides with bistranded abasic sites interfere with substrate binding and catalysis by human apurinic/apyrimidinic endonuclease.* Biochemistry, 2001. **40**: p. 13254-13261.
75. Hoehn, S.T., C.J. Turner, and J. Stubbe, *Solution structure of an oligonucleotide containing an abasic site: evidence for an unusual deoxyribose conformation.* Nucl. Acids Res., 2001. **29**: p. 3413.
76. Kool, E.T., J.C. Morales, and K.M. Guckian, *Mimicking the structure and function of DNA: insights into DNA stability and replication.* Angew. Chem. Int. Ed., 2000. **39**: p. 990-1009.
77. Kool, E.T., *Preorganization of DNA: design principles for improving nucleic acid recognition by synthetic oligonucleotides.* Chem. Rev., 1997. **97**(5): p. 1473-1487.
78. Jourdan, M., J. Garcia, and J. Lhomme, *Threading bis-intercalation of a macrocyclic bisacridine at abasic sites in DNA: nuclear magnetic resonance and molecular modeling study.* Biochemistry, 1999. **38**: p. 14205-14213.
79. Fkyerat, A., et al., *A new class of artificial nucleases that recognize and cleave apurinic sites in DNA with great selectivity and efficiency.* J. Am. Chem. Soc., 1993. **115**: p. 9952-9959.
80. Smirnov, S., et al., *Integrity of duplex structures without hydrogen bonding: DNA with pyrene paired at abasic sites.* Nucl. Acids Res., 2002. **30**(24): p. 5561-5569.
81. Groebke, K. and C. Leumann, *A method for preparing oligodeoxynucleotides containing an apurinic site.* Helv. Chim. Act., 1990. **73**: p. 608-617.
82. Strauss, P.R., et al., *Substrate binding by human apurinic/apyrimidinic endonuclease indicates a Briggs-Haldane mechanism.* J. Biol. Chem., 1997. **272**(2): p. 1302-1307.
83. Chou, K.-M. and Y.-C. Cheng, *An exonucleolytic activity of human apurinic/apyrimidinic endonuclease on 3' mispaired DNA.* Nature, 2002. **415**: p. 655-659.
84. Krotz, A.H., et al., *Preparation of oligonucleotides without aldehyde abasic sites.* Bioorg. Med. Chem. Lett., 2001. **11**: p. 1863-1867.
85. Lhomme, J., J.-F. Constant, and M. Demeunynck, *Abasic DNA structure, reactivity and recognition.* Biopolymers (Nucl. Acid Sci), 1999. **52**(2): p. 65-83.
86. Takeshita, M., et al., *Oligodeoxynucleotides containing synthetic abasic sites.* J. Biol. Chem., 1987. **262**(21): p. 10171-10179.

87. Talpaert-Borle, M. and M. Liuzzi, *Reaction of apurinic/apyrimidinic sites with [^{14}C]methoxyamine. A method for the quantitative assay of AP sites in DNA.* Biochim. Biophys. Acta, 1983. **740**(4): p. 410-416.

88. Kubo, K., et al., *A novel, sensitive, and specific assay for abasic sites, the most commonly produced DNA lesion.* Biochemistry, 1992. **31**: p. 3703-3708.

89. Ide, H., et al., *Synthesis and damage specificity of a novel probe for the detection of abasic sites in DNA.* Biochemistry, 1993. **32**: p. 8276-8283.

90. Boturyn, D., et al., *Synthesis of fluorescent probes for the detection of abasic sites in DNA.* Tetrahedron, 1997. **53**(15): p. 5485-8592.

91. Boturyn, D., et al., *A simple and sensitive method for in vitro quantitation of abasic sites in DNA.* Chem. Res. Toxicol, 1999. **12**(6): p. 479-482.

92. Manoharan, M., L.K. Andrade, and P.D. Cook, *Site-specific cross-linking of nucleic acids using the abasic site.* Org. Lett., 1999. **1**(2): p. 311-314.

93. Nguyen, R. and I. Huc, *Optimizing the reversibility of hydrazone formation for dynamic combinatorial chemistry.* Chem. Comm., 2003(8): p. 942-943.

94. Cline, S.D., et al., *DNA abasic lesions in a different light: solution structure of an endogenous topoisomerase II poison.* Biochemistry, 1999. **38**: p. 15500-15507.

95. Kalnik, M.W., et al., *NMR studies of abasic sites in DNA duplexes: deoxyadenosine stacks into the helix opposite the cyclic analogue of 2-deoxyribose.* Biochemistry, 1988. **27**: p. 924-931.

96. Singh, M.P., et al., *High-field NMR and restrained molecular modeling studies on a DNA heteroduplex containing a modified apurinic abasic site in the form of covalently linked 9-aminoellipticine.* Biochemistry, 1994. **33**: p. 10271-10285.

97. Beger, R.D. and P.H. Bolton, *Structures of apurinic and apyrimidinic sites in duplex DNAs.* J. Biol. Chem., 1998. **273**(25): p. 15565-15573.

98. Goljer, I., S. Kumar, and P.H. Bolton, *Refined solution structure of a DNA heteroduplex containing an aldehydic abasic site.* J. Biol. Chem., 1995. **270**(39): p. 22980-22987.

99. Coppel, Y., et al., *Solution conformation of an abasic DNA undecamer duplex d(CGCACXCACGC)·d(GCGTGTGTGCG): the unpaired thymine stacks inside the helix.* Biochemistry, 1997. **36**: p. 4817-4830.

100. Cuniasse, P., et al., *The abasic site as a challenge to DNA polymerase. A nuclear magnetic resonance study of G, C and T opposite a model abasic site.* J. Mol. Biol., 1990. **213**: p. 303-314.

101. Whitka, J.M., J.A. Wilde, and P.H. Bolton, *Characterization of conformational features of DNA heteroduplexes containing aldehydic abasic sites.* Biochemistry, 1991. **30**: p. 9931-9940.
102. Gelfand, C.A., et al., *Thermodynamic consequences of an abasic lesion in duplex DNA are strongly dependent on base sequence.* Biochemistry, 1998. **37**: p. 7321-7327.
103. McCann, J.A.B. and P.J. Berti, *Adenine release is fast in mutY-catalyzed hydrolysis of G:A DNA mismatches.* J. Biol. Chem., 2003. **278**(32): p. 29587-29592.
104. Heinemann, U. and C. Aling, *Crystallographic study of one turn of G/C-rich B-DNA.* J. Mol. Biol., 1989. **210**: p. 369.
105. Heinemann, U. and M. Hahn, *C-C-A-G-G-C-M5C-T-G-G; Helical fine structure, hydration, and comparison with C-C-A-G-G-C-C-T-G-G.* J. Biol. Chem., 1992. **267**: p. 7332.
106. Kielkopf, C.L., et al., *Structural basis for G.C recognition in the DNA minor groove.* Nat. Struct. Biol., 1998. **5**: p. 104.
107. Moran, S., R.X.-F. Ren, and E.T. Kool, *A thymidine triphosphate shape analog lacking Watson-Crick pairing ability is replicated with high sequence selectivity.* Proc. Natl. Acad. Sci. USA, 1997. **94**: p. 10506-10511.
108. Schweitzer, B.A. and E.T. Kool, *Hydrophobic, non-hydrogen-bonding bases and base pairs in DNA.* J. Am. Chem. Soc., 1995. **117**(7): p. 1863-1872.
109. Harnden, M.R. and R.L. Jarvest, *Pyrrolidine analogues of 2',3'-dideoxynucleosides: synthesis via 9-aminopurines and 1-aminopyrimidines.* J. Chem. Soc., Perkin Trans. 1, 1991: p. 2073-2079.
110. Imaizumi, M., F. Kano, and S. Sakata, *Novel uracil derivatives: newly synthesized centrally acting agents.* Chem. Pharm. Bull., 1992. **40**(7): p. 1808-1813.
111. Iino, T., Y. Yoshimura, and A. Matsuda, *Nucleosides and nucleotides. 134. Synthesis of 2'-C-alkynyl-2'-deoxy-1-β-D-arabinofuranosylpyrimidines via radical deoxygenation of Tert-progargyl alcohols in the sugar moiety.* Tetrahedron, 1994. **50**(35): p. 10397-10406.
112. Hassan, A.E.A., S. Shuto, and A. Matsuda, *Nucleosides and nucleotides. 124. Chemical reactivity of the sugar moiety of 2'-deoxy-2'-methylidene pyrimidine nucleosides: synthesis of 3'-amino-2',3'-dideoxy-2'-methylidene pyrimidine nucleosides via [2,3]-sigmatropic rearrangement of allylic selenides as potential antitumor agents.* Tetrahedron, 1994. **50**(3): p. 689-700.

113. Harnden, M.R. and R.L. Jarvest, *Synthesis of pyrrolidin-1-yl analogues of pyrimidine dideoxynucleosides.* Tetrahedron, 1991. **32**(31): p. 3863-3866.
114. Kohda, K., et al., *Syntheses and properties of N-aminopyrimidines.* Tetrahedron, 1993. **49**(19): p. 3947-3958.
115. Kohda, K., et al., *Synthesis and properties of N-aminoguanines.* Tetrahedron, 1989. **45**(20): p. 6367-6374.
116. Glemarec, C., et al., *N^2,3-Ethenoguanosine and IA'-metamorphosine: ^{15}N NMR spectroscopy and elucidation of physico-chemical properties by kinetic and equilibrium measurements.* Tetrahedron, 1991. **47**(33): p. 6689-6704.
117. Montgomery, J.A. and J.C. Temple, *Synthesis of potential anticancer agents. XXIII 9-aminohypoxanthine and related compounds.* J. Am. Chem. Soc., 1960. **82**: p. 4592-4596.
118. Dornow, A. and E. Hinz, *Uber ortho-Kondensationen heterocyclischer o-amino-carbonsäure-derivate.* Chem. Ber., 1958. **91**: p. 1834-1840.
119. Cheesman, G.W.H., *The synthesis and tautomerism of some 2-substituted pyrazines.* J. Chem. Soc., 1960: p. 242-247.
120. Ross, L.O., L. Goodman, and B.R. Baker, *Potential anticancer agents. XLIV. Some derivatives of uracil-5- and -6-carboxylic acid.* J. Org. Chem., 1960. **25**: p. 1950-1952.
121. Armarego, W.L.F., *Covalent hydration in 1,4,5,8-tetra-azanaphthalenes.* J. Chem. Soc., 1963: p. 4304-4311.
122. Nguyen, R., *Personal communication*. 2001.
123. Nguyen, R., *De l'utilisation de biomolecules comme matrices pour favoriser la formation de molecules complementaires*. 2002, Ecole Doctorale de l'Ecole Polytechnique: Palaiseau.
124. Aimi, K. and S. Ando, *Conformation analysis and molecular mobility of ethylene and tetrafluoroethylene copolymer using solid-state 19F MAS and 1H ->19F CP/MAS NMR.* Magn. Res. in Chem, 2004. **42**(7): p. 577-588.
125. Ellis, D.A., et al., *The use of 19F NMR to interpret the structural properties of perfluorocarboxylate acids: a possible correlation with their environmental disposition.* J. Phys. Chem. A, 2004. **108**: p. 10099-10106.
126. Lindahl, T. and A. Andersson, *Rate of chain breakage at apurinic sites in doublestranded deoxyribonucleic acid.* Biochemistry, 1972. **11**(3618-3623).
127. Lindahl, T. and B. Nyberg, *Rate of depurination of native deoxyribonucleic acid.* Biochemistry, 1972. **11**(19): p. 3610-3618.

128. Stuart, G.R. and R.W. Chambers, *Synthesis and properties of oligodeoxynucleotides with an AP site at a preselected position.* Nucl. Acids Res., 1987. **15**: p. 7451-7462.

129. Péoc'h, D., et al., *Efficient chemical synthesis of oligodeoxynucleotides containing a true abasic site.* Tetrahedron Lett., 1991. **32**(2): p. 207-210.

130. Godde, F., J.-J. Toulmé, and S. Moreau, *Benzoquinazoline derivatives as substitutes for thymine in nucleic acid complexes. Use of fluorescence emission of benzopg[quinazoline-2,4-(1H,3H)-dione in probing duplex and triplex formation.* Biochemistry, 1998. **37**: p. 13765-13775.

131. Shishkina, I.G. and F. Johnson, *A new method for the postsynthetic generation of abasic sites in oligomeric DNA.* Chem. Res. Toxicol, 2000. **13**(9): p. 907-912.

132. Conn, M.M., et al., *Convergent functional groups. 13. High-affinity complexation of adenosine derivatives within induced binding pockets.* J. Am. Chem. Soc., 1993. **115**: p. 3548-3578.

133. Tamura, Y., J. Minamikawa, and M. Ikeda, *O-Mesitylenesulfonylhydroxylamine and related compounds - powerful aminating reagents.* Synthesis, 1977. **1977**(1): p. 1-17.

134. Epple, C. and C. Leumann, *Bicyclo[3.2.1]-DNA, a new DNA analog with a rigid backbone and flexibly linked bases: pairing properties with complementary DNA.* Chemistry & Biology, 1998. **5**: p. 209-216.

135. Tener, G.M. and H.G. Knorana, *Phosphorylated sugars. VI. Syntheses of α-D-ribofuranose 1,5-diphosphate and α-D-ribofuranose 1-pyrophosphate 5-phosphate.* J. Am. Chem. Soc., 1958. **80**: p. 1999-2004.

136. Greene, T.W. and P.G.M. Wuts, *Protective groups in organic synthesis*. 3 ed. 1999, New York: Wiley Interscience.

137. Ates, A., et al., *Remarkably efficient deprotection of cyclic acetals and ketals.* Tetrahedron Lett., 1999. **40**: p. 1799-1802.

138. Marko, I.E., et al., *Cerium(IV)-catalyzed deprotection of acetals and ketals under mildly basic conditions.* Angew. Chem. Int. Ed., 1999. **38**(21): p. 3207-3209.

139. Cereda, E., et al., *The preparation of five metabolites of feprazone. Stereoselective Wittig synthesis of new hemiterpenoid synthons.* Il Farmaco - Ed. Sc., 1983. **38**(6): p. 429-446.

140. Dang, H.-S. and B.P. Roberts, *Radical-chain cyclisation of unsaturated acetals and thioacetals in the presence of thiols as polarity-reversal catalysts.* Tetrahedron Lett., 1999. **40**: p. 8929-8933.

141. Choi, J.-R., S. Han, and J.K. Cha, *An enantioselective synthesis of (-)-slaframine.* Tetrahedron Lett., 1991. **32**(45): p. 6469-6472.

142. De Nino, A., et al., *Stereoselective complete reduction of α-alkyl-β-ketonitriles to anti γ-amino alcohols.* Eur. J. Org. Chem, 2002. **2002**(17): p. 2924-2927.

143. Vik, J.E., *Studies on intermediates involved in the syntheses of pentaerythritol and related alcohols. III. Syntheses of α-hydroxymethyl-substituted aldehydes.* Acta Chem. Scand., 1973. **27**: p. 239-250.

144. Fuson, R.C., W.E. Parham, and L.J. Reed, *Alkylation of ethyl malonate with diethoxymethyl acetate.* J. Org. Chem., 1946. **11**: p. 194-198.

145. Späth, E. and L. Raschik, *Uber die konstitution der dimeren glykolaldehyde.* Monatsh. Chemie, 1946. **76**(2): p. 65-76.

146. Manzo, E., et al., *Reaction of cyclic ketals with ceric ammonium nitrate in acetonitrile/water.* Tetrahedron, 2002. **58**: p. 129-133.

147. Sumaoka, J., Y. Azuma, and M. Komiyama, *Enzymatic manipulation of the fragments obtained by cerium(IV)-induced DNA scission: characterization of hydrolytic termini.* Chem. Eur. J., 1998. **4**(2): p. 205-209.

148. Takasaki, B.K. and J. Chin, *Cleavage of the phosphate diester backbone of DNA with cerium(III) and molecular oxygen.* J. Am. Chem. Soc., 1994. **116**: p. 1121-1122.

149. Komiyama, M., et al., *Efficient and oxygen-independent hydrolysis of single-stranded DNA by cerium(IV) ion.* J. Chem. Soc., Perkin Trans. 2, 1995(2): p. 269-274.

150. Nishiguchi, T., et al., *Conversion of ketals to ketones by nitrogen dioxide in the presence of silica gel.* J. Chem. Soc., Chem. Commun., 1995. **11**: p. 1121-1122.

151. Jung, M.E., W.A. Andrus, and P.L. Ornstein, *Non-aqueous conversion of ketals to ketones via treatment with trimethylsilyliodide.* Tetrahedron Lett., 1977. **18**(48): p. 4175-7178.

152. Roush, W.R. and R.J. Sciotti, *Enantioselective total synthesis of (-)-chlorothricolide.* J. Am. Chem. Soc., 1994. **116**(14): p. 6457-6458.

153. Lipshutz, B.H. and D.F. Harvey, *Hydrolysis of acetals and ketals using $LiBF_4$.* Synth. Commun., 1982. **12**(4): p. 267-277.

154. Bonin, M., et al., *Asymmetric synthesis VIII: Biogenetically patterned approach to the chiral total synthesis of (-)-pumiliotoxin-C.* Tetrahedron Lett., 1986. **27**(14): p. 1569-1572.

155. Lipshutz, B.H., et al., *Pd(II)-catalyzed acetal/ketal hydrolysis/exchange reactions.* Tetrahedron Lett., 1985. **26**(6): p. 705-708.
156. Coppola, G.M., *Amberlyst-15, a superior acid catalyst for the cleavage of acetals.* Synthesis, 1984. **1984**(12): p. 1021-1023.
157. Sterzycki, R., *Pyridinium tosylate, a mild catalyst for formation and cleavage of dioxolane-type acetals.* Synthesis, 1979. **1979**(9): p. 724-725.
158. Colvin, E.W., R.A. Raphael, and J.S. Roberts, *The total synthesis of (±)-trichodermin.* J. Chem. Soc., Chem. Commun., 1971. **15**: p. 858-859.
159. Zatsepin, T.S., et al., *Synthesis of peptide-oligonucleotide conjugates with single and multiple peptides attached to 2'-aldehydes through thiazolidine, oxime and hydrazine linkages.* Bioconjugate. chem., 2002. **13**: p. 822-830.
160. Zatsepin, T.S., et al., *Synthesis and properties of modified oligodeoxyribonucleotides containing 2'-O-(2,3-dihydroxypropyl)uridine and 2'-O-(2-oxoethyl)uridine.* Russ. J. Bioorg. Chem., 2001. **27**(1): p. 39-44.
161. Sproat, B.S., et al., *New synthetic routes to protected purine 2'-O-methylriboside-3'-O-phosphoramidites using a novel alkylation procedure.* Nucleosides & Nucleotides, 1991. **10**: p. 25-36.
162. Fessenden, R.J. and M.D. Coon, *Silicon heterocyclic compounds. I. Synthesis of the 3-sila-1-heterocyclohexanes.* J. Org. Chem., 1964. **29**(6): p. 1607-1610.
163. Jackson, D.T. and C.G. King, *Synthetic glycerides. IV. Esters of aromatic and aliphatic acids.* J. Am. Chem. Soc., 1933. **55**: p. 678-680.
164. Efimtseva, E.V., et al., *Synthesis and conformational properties of O-β-D-Ribofuranosyl-(1"-2')-guanosine and (adenosine)-5"-phosphate.* Nucleosides, Nucleotides and Nucleic acids, 2003. **22**(5-8): p. 1109-1111.
165. Tadano, K., et al., *A novel transformation of four aldoses to some optically pure pseudohexopyranoses and a pseudopentofuranose, carbocyclic analogs of hexopyranoses and pentofuranose. Synthesis of derivatives of (1S,2S,3R,4S,5S)-, (1S,2S,3R,4R,5S)-, (1R,2R,3R,4R,5S)-, (1S,2S,3R,4S,5R)-2,3,4,5-tetrahydroxy-1-(hydroxymethyl)cyclohexanes and (1S,2S,3S,4S)-2,3,4-trihydroxy-1(hydroxymethyl)cyclopentane.* J. Org. Chem., 1987. **52**(10): p. 1946-1956.
166. Gathergood, N., K.R. Knudsen, and K.A. Jorgensen, *Enantioselective synthesis of optically active carbocyclic sugars.* J. Org. Chem., 2001. **66**(3): p. 1014-1017.
167. Elliot, R.P., et al., *A reductive aldol strategy for the synthesis of very highly substituted cyclopentanes from sugar lactones.* Tetrahedron Lett., 1991. **32**(42): p. 6227-6230.

168. Joyce, G.F., *Booting up life.* Nature, 2002. **410**: p. 278-279.

169. Orgel, L.E., *The origin of life - a review of facts and speculations.* Reviews, 1998. **23**: p. 491-495.

170. Eschenmoser, A., *Chemical etiology of nucleic acid structure.* Science, 2001. **284**(5423): p. 2118.

171. Nielsen, P.E., *Peptide nucleic acid. A molecule with two identities.* Acc. Chem. Res., 1999. **32**: p. 624-630.

172. Singh, S.K., et al., *LNA (locked nucleic acid): synthesis and high-affinity nucleic acid recognition.* Chem. Commun., 1998: p. 455-456.

173. Nielsen, K.E., et al., *Solution structure of an LNA hybridized to DNA: NMR study of the d(CT^{L}GCG^{L}T^{L}CT^{L}GC):D(GCAGAAGCAG) duplex containing four locked nucleotides.* Bioconjugate. chem., 2000. **11**: p. 228-238.

174. Singh, S.K. and J. Wengel, *Universality of LNA-mediated high-affinity nucleic acid recognition.* Chem. Commun., 1998: p. 1247-1248.

175. Koshkin, A.A., et al., *LNA (locked nucleic acid): an RNA mimic forming exceedingly stable LNA:LNA duplexes.* J. Am. Chem. Soc., 1998. **120**: p. 13252-13253.

176. Herdewijn, P., *TNA as a potential alternative to natural nucleic acids.* Angew. Chem. Int. Ed., 2001. **40**(12): p. 2249-2251.

177. Froeyen, M., et al., *Molecular-dynamics studies of single-stranded hexitol, altritol, mannitol, and ribose nucleic acids (HNA, MNA, ANA, and RNA, resp.) and of the stability of HNA·RNA, ANA·RNA, and MNA·RNA duplexes.* Helv. Chim. Act., 2000. **83**: p. 2153-2182.

Word of Thanks / Remerciément / Dankwoord

To bring a thesis to a good end, one needs many people around for support. Not only on the scientific part, knowledge and physical help of people, but also on the mental part. To continue albeit things that are not working the way you want is not easy, as I noticed last year. Happily there was always someone beside me when I needed support.

The first person I want to acknowledge is Ivan Huc. Four years ago you were willing to welcome me in your lab, before you had even seen me. Thank you for the confidence you had in me! As a thesis director you led me a lot of freedom. I have really appreciated to be able to develop my own work and discover things on my own. In addition you were always there for an open discussion when I did no longer see a way out of my troubles. Your fresh look always helped me to go in the right direction. And although the information you threw at me always gave me something to ponder over in the evening (and work the following day / weekend), I could not complain I didn't learn anything from those meetings.

A word of thanks goes also to the members of the PhD. commission. Especially to Bernard Rayner and Sijbren Otto as 'rapporteurs', but also to Joelle Prunet as president of the commission. The latter two had a long journey to come here.

I would like to thank the Ecole Polytechnique for financing the first two years of my thesis and the CNRS for the third year.

I would like to thank Léon Ghosez, Jean-Jacques Toulmé and Jean-Marie Schmitter as directors of the "Institut Européen de Chimie et Biologie" for giving me room to work in the building. I would like to apologize to Jean-Marie Schmitter for squatting in an office of his students for too long.

Une très grande Merci pour les secrétaires de cet institut. Merci a Véronique, Stéphanie, Annie et Sandra, et bien sur Stéphanie comme directrice de l'administration. Sans vous, personne dans cet institut peut faire de la recherche!

From a research perspective I would like to thank Fred for helping me out with the oligonucleotides whenever that was necessary again and for the synthesis of the difluorobenzene compound. Axelle, thanks for always being able to find me five minutes for a quick proton NMR when it was most necessary. Thanks to Carmelo for doing the SPR measurements, albeit that it did not work as we planned. But that was not your faults.
Thanks to Katell for the rapid measurements of all my samples. Thanks to Michel and Julien for helping me with my modeling experiments, and sorry for asking stupid (PIBCAK) questions. Merci à Julie et Nathalie pour la synthèse des oligonucleotides. Bernard Rayner, merci pour les discussions sur les sites abasiques qui m'ont toujours fait mieux comprendre pourquoi ça ne marchait pas.

I would like to thank all my lab-mates for the pleasant time I had doing my experiments. I cannot name everyone, but your presence was appreciated!
A few groups of VSP (very special people) that I want to thank:
Thanks to 'The Belgians', Fred, Laurence et Eric, with whom it was nice to take a beer after work and who were always willing to listen to again another blague belge.
Thanks to 'The Chinese', Eric, Jiang (or Hua actually but Jiang somehow sounds much more familiar) and Xiao-Nan, who come from pingpong country number one. That none of you played at a level I hoped for was no problem at all.
Thanks to the 'Miscellaneous Foreigners', Ralf et Cristina, Ana, Sofi, Silvina, Peter, Jez, for the multicultural experience, cinema, beers, skiing and other small things that make life good.
Merci aux 'Français'; Victor, Thomas, et Damien, avec qui c'était toujours bien de jouer au basket, de faire la fête, ou de regarder une coupe du monde de foot. Merci Reiko (hein, mais tu are not French!... Well, half now. OK, I leave ton nom ici) pour ta gentillesse et ta bonne humeur. Merci Anne pour ta bonne humeur, ton courage et ta volonté de te battre. Merci Erick pour me mettre encore une branlé au squash. Merci Jean pour le soutien moral et technique.

Christiaan en Edward bedankt voor het nakijken van allerlei Nederlandse teksten. Zoals jullie gemerkt hebben (meer dan ik waarschijnlijk) gaat je grammatica drastisch achteruit als je zo'n tijd in het buitenland verblijft. Bovendien bedankt voor morale steun in moeilijke tijden, zoals met name in september 2004.

Merci du fond de mon coeur à tous les pongistes des Coqs Rouges. La confiance que vous avez eu en moi, dès le début, ça fait vraiment plaisir. Pour tenir le choque, il y avait toujours le ping le soir ou le week-end pour me détendre. Pour organiser des trucs, pour parler, pour boire un coup, pour apprendre le français, tous. Un petit investissement qui m'a remporté cent fois plus.

Mijn ouders bedankt voor de morele support, ondanks dat jullie het in het begin niet makkelijk vonden dat ik niet meer elke dag thuis kwam geloof ik. Jullie vonden het "wel erg ver weg", hoorde ik later via via… Ik geloof dat jullie er toch aan gewend geraakt zijn.