NMR solution structure of DNA double helices with built-in polarity probes

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

d o c t o r r e r u m n a t u r a l i u m
(Dr. rer. nat.)

im Fach Chemie

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
Dipl.-Chem. Lars Dehmel

Präsident der der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät:
Prof. Dr. Elmar Kulke

Gutachter:
1. Prof. N. P. Ernsting
2. Prof. Dr. H.-A. Wagenknecht

Tag der mündlichen Prüfung: 20.05.2015

Für meine Familie und die Zukunft.

Abstract

The solution structures of three differently modified DNA double strands were solved by NMR spectroscopy. They all incorporate polarity probes in the center of the helix that are sensitive to the immediate environment. Their melting behavior was characterized by a new method that utilizes complete absorption spectra in combination with Singular Value Decomposition (SVD). The latter allows to analyze the spectra in their entirety, which is required to follow the blue shift of the probe signal that is caused by the aforementioned sensitivity to the environment. In this way the duplex melting process is characterized in local and global terms.

The first modification, 2-hydroxy-7-carboxyfluorene (HCF), is placed opposite an abasic site to avoid steric strain. NMR spectroscopy revealed two equally distributed conformations, since rotation of the HCF chromophore is only hindered by stacking interactions inside the helix. The second double strand comprises R-glycerol linked 6-hydroxyquinolinium (6HQ) opposite cytosine. The incorporation of 6HQ as glycol nucleic acid (GNA) mononucleotide is a unique structural feature. Until now, only crystal structures of full GNA backbone duplexes are known, so the solution structure of this double strand is of general interest. The small size of R-glycerol disturbs the backbone of the 6HQ strand, which causes a stacking axis that differs from the helical long axis for the three central bases. The last modification is an artificial base pair made of 4-aminophthalimide (4AP) and 2,4-diaminopyrimidine (DAP). Instead of the desired three hydrogen bonds, two structures containing either a single or two hydrogen bonds are observed that can be explained by the linkage of 4AP to 2'-deoxyribofuranose.

Inhaltsangabe

Die Strukturen in Lösung dreier unterschiedlich modifizierter DNA Doppelstränge wurden mittels NMR Spektroskopie gelöst. Sie alle besitzen polare Sonden im Zentrum der Helix, welche sensitiv für die nähere Umgebung sind. Ihr Schmelzverhalten wurde mit Hilfe einer neuen Methode charakterisiert, welche komplette Absorptionsspektren in Kombination mit Singularwertzerlegung (SVD) nutzt. Letztere erlaubt die Analyse der Spektren als Ganzes, die notwendig ist um der Blauverschiebung des Sondensignals zu folgen, welche durch die zuvor genannte Sensitivität zur Umgebung verursacht wird. Auf diese Weise kann der Schmelzprozess des Duplex lokal und global beschrieben werden.

Die erste Modifikation, 2-Hydroxy-7-Carboxyfluoren (HCF), wurde gegenüber einer abasischen Seite platziert, um sterische Spannungen zu vermeiden. Die NMR Spektroskopie deckte zwei gleichverteilte Konformationen auf, da die Rotation des HCF Chromophors nur durch die Stapelwechselwirkung innerhalb der Helix unterbunden wird. Der zweite Doppelstrang enthält ein über R-Glycerol gebundenes 6-Hydroxychinolinium (6HQ) gegenüber Cytosin. Der Einbau von 6HQ als Mononukleotid einer Glykolnukleinsäure (GNA) ist ein strukturelles Alleinstellungsmerkmal. Bisher sind nur Kristallstrukturen von vollständiger GNA bekannt, daher ist die Struktur in Lösung dieses Doppelstranges von generellem Interesse. Die geringe Größe von R-Glycerol stört das Rückgrat des 6HQ-Stranges, welche eine von der helikalen Achse abweichende Stapelachse für die drei zentralen Basen verursacht. Die letzte Modifikation ist ein künstliches Basenpaar bestehend aus 4-Aminophthalimid (4AP) und 2,4-Diaminopyrimidin (DAP). Anstatt der gewünschten drei Wasserstoffbrücken wurden zwei Strukturen, die entweder eine oder zwei Wasserstoffbrücken beinhalten, beobachtet, welche durch die Verbindung von 4AP zur 2'-Deoxyribofuranose erklärt werden können.

Contents

1 Introduction

1.1 The development of DNA structure determination

The structure of nucleic acids (DNA, RNA) and proteins forms the basis for our understanding of biological processes like gene expression or protein bio-synthesis. In 1953 Watson and Crick were the first to propose the correct double helical structure of what is now known as B-form DNA[1]. They deduced their model from X-ray fiber diffraction, which was the method of choice at the time, but it only provided an overall configuration with idealized parameters.

Although true on average it could not explain sequence specific effects, until Wing et al.[2] published in 1980 the first single-crystal structure analysis of a B-DNA dodecamer with a full helical turn. This was followed by a series of articles in which Dickerson and Drew[3–6] described sequence-dependent features like molecular bending[4] or the "spine of hydration"[5], thus making it famous as the "Dickerson-Drew dodecamer".

During the next few years more and more crystal structures were solved, showing a wide variety of B-DNA helical parameters[7–9] and also giving structural insights into A- and Z-DNA[10]. Different approaches were used to study A-DNA crystals. Shakked et al. (1981)[11] and Conner et al. (1982)[12] used short self-complementary octamers and tetramers, respectively, while Wang et al. (1982)[13] used a DNA–RNA hybrid double helix to prove the assumed A-type structure.

The first single-crystal duplex of Z-DNA was also solved by Wang et al. in 1979[14], and together with the aforementioned A- and B-forms a detailed molecular picture was developed steadily[10,15–17]. However questions remained in which way crystal-packing factors and crystallization conditions affect the conformation of DNA. It was found that all octanucleotides crystallize in the A-form[18], due to crystal-packing effects, while their counterparts in solution favor the B-conformer[18,19]. Single-crystals of longer duplexes tend to the A-form when having a very GC-rich sequence[18,20], thus explaining the B-form helix of the Dickerson-Drew dodecamer d(CGCGAATTCGCG) in contrast to the A-form found in the dodecamer d(CCCCCGCGGGGG) of Verdaguer et al. (1991)[21]. In addition, Jain et al.[22] and Shakked et al.[23] showed (both in 1989) that the mean values of local helix parameters depend on crystallization conditions and therefore contribute to their "unexpectedly large range of variation"[19,24]. With this in mind, it becomes clear that structural information based on X-ray crystallography of single crystals can be contradictory or limited, since the native environment to study biological problems is a water containing solution.

Nuclear Magnetic Resonance Spectroscopy (NMR) is a well-established method to study molecules in solution, especially after the development of two-dimensional experiments in the late 1970s (Jeener et. al[25], Freeman et al.[26–29], Ernst et al.[30]) which extended the field of applications to biological macromolecules. At the beginning of the 1980s, Ernst et. al. and in particular Wüthrich et al. were the driving forces in protein structure determination[31–42], which also forms the basis for the related research on DNA structures. First efforts were targeted towards sequential resonance assignment strategies and qualitative analyses of the structure[43–47], but rapidly rising computational power made it possible to develop structure determination software[48,49] that uses NMR-derived inter-proton distances as restraints.

In contrast to X-ray crystallography, where a gap of 20 years separates the first protein structure (Bluhm et al. 1958[50]) from its DNA counterpart[14], the first solution structures of proteins (Williamson et al.[51], Kaptein et al.[52]) and nucleic acids (Clore et al.[53]) were all published in 1985. The former gap originated from the requirements to obtain single crystals of short oligonucleotides and therefore demanded the development of synthesis strategies for oligonucleotides with a predefined primary sequence, which were lacking in 1958. However, subsequent development led to the solid-phase phosphoramidite method[54–57] that not only offered a payable way to produce large quantities, but it also allowed the modification and expansion of the genetic alphabet. This, in combination with automated oligonucleotide synthesis, made it feasible to investigate functionalized or completely new nucleotides at any position in the duplex. Consequently, artificial DNA duplex strands became the subject of NMR structure determination in the late 1980s[58–62].

A large number of modified compounds has been introduced (or observed in naturally occurring DNA[63–65]) over the last decades depending on the scope of application. They can be divided into several classes; these include backbone variations (e.g. PNA[66,67], LNA[68,69], TNA[70], GNA[71–74] etc.) or molecules covalently linked to natural bases[75–80]. A famous example for using both strategies is the DNA sequencing method by Sanger et al.[81] which utilizes di-deoxynucleotidetriphosphates (ddNTPs) to terminate the chain-reaction of DNA polymerase at the 3'-end and, in addition, fluorescent labeling of the nucleobases for detection in automated sequencing machines[82]. Other strategies focus on intercalation[83–88], full replacement of a nucleobase[89–96] or even of a complete base pair[97,98] by nucleobase analogues[99].

Most of these studies introduce chromophores as base analogues or tethered label, since their fluorescent properties can be used in a number of ways to study DNA and RNA. A common field of application is the detection of single nucleotide polymorphisms (SNP) via molecular beacons[100–103], base-discriminating fluorescent nucleosides[104,105], detection by electron transfer-controlled emission quenching (DETEQ)[106] or forced intercalation TO-PNA probes (FIT)[107]. Real-time quantitative PCR (qPCR)[108,109] is also a possible application for molecular beacons[110] or FIT probes[111,112] and can be used, for example, to quantify gene expression[113]. Other methods like the well-known fluorescence resonance energy transfer (FRET)[114–116], pulsed electron-electron double resonance (PELDOR[117,118], strongly emerging field in NMR) or fluorescent silver nanoclusters[119] can serve as distance measurement tools, working on a larger scale than the classic NOESY experiment (< 5 vs. 80 Å with PELDOR or FRET). All examples mentioned before employed steady-state fluorescence. Time-resolved measurements on the femtosecond (fs) to nanosecond (ns) time scale were first performed by Zewail et al.[120,121] in 2000. Using transient absorption, these authors investigated electron transfer and water solvation in duplex DNA.

The fluorophores that will be studied in this work all belong to the group of nucleobase replacing analogues. In contrast to tethered labels, these molecules are primarily designed to minimize disruption of the local structure and to maintain the biochemical or biological function, which may be affected negatively by other strategies, like tethering of a bulky group, addition of an intercalator, or a groove binding molecule. The underlying idea on one hand is to place a nucleobase at the "center of action", and on the other hand to achieve a rigid, well defined position and orientation in the duplex[94]. These conditions are crucial for the above stated distance measurements like FRET as well as for the time-resolved studies below. As we will see later, another condition is that the spectral position of the fluorescence band depends sensitively on polarity, so that the label can be used to investigate the microenvironment[96].

The basic idea which caused the present structural study is to observe the hydration dynamics and vibrational modes of biomolecules. This aim can be achieved, in principle, by detecting the time-dependent Stokes shift (TDSS) of fluorescence from a suitable reporter molecule, or probe. When incorporated into duplex DNA, the chromophore serves as a "local" molecular spectrometer in the THz-region. This kind of measurement, but without a polarity probe that has been optimized for spatial and energetic fits, would be disturbed by the absorption of unbound water and would also lack space-resolved information (due to structural fluctuations). The THz experiment starts when the charge distribution of the chromophore is suddenly altered by femtosecond optical excitation. By this excitation the electric field around the probe is changed instantly and is now affecting nearby water molecules and neighboring nucleobases. Their response to the reaction field $R(t)$, induced by the chromophore, can be measured as dynamic fluorescence Stokes shift on a ps- to ns-timescale. In this way the chromophore is used not only as THz light source but also as detector. The local THz absorption spectrum is then obtained by a suitable Laplace transformation of the time-dependent Stokes shift[122,123], a method similar to the Fourier transformation of the Free Induction Decay (FID) in NMR. The requirements of molecular THz spectroscopy, as outlined above, restrict the repertory of suitable chromophores severely. Clearly the probe must be sensitive to the microenvironment or, in other words, should show sizable fluorescence shifts between different solvents. Also the fluorescence lifetime should be long enough to measure the dynamic Stokes shift up to several nanoseconds; therefore lifetime shortening interactions in the excited state like FRET, photo-induced electron transfer or intersystem crossing should be avoided. Moreover the probe has to be free of internal vibrational modes below 300 cm^{-1}, otherwise they could mix with external modes of the environment that are the aim of the detection. But all of these photophysical properties will be useless when the probe molecule alters or perturbs the DNA helical structure. The following section will discuss the advantages and drawbacks of known base analogues. In doing so we will also see why there is still a strong need for further development of base analogues.

1.2 Fluorescent base analogues extend the nucleobase alphabet

One of the earliest and most studied[95,120,124–127] fluorescent nucleobase substitutes is 2-aminopurine (2AP). Ward et al.[128] reported in 1969 that 2-aminopurine riboside (together with 2,6-diaminopurine riboside and formycin), in contrast to adenine and other naturally occurring bases, is fluorescent under physiological conditions, and they explored its photophysical properties under different solvent polarities, pH values and temperatures. 2AP has a high fluorescence quantum yield as free base in solution (0.68), but when incorporated into nucleic acids a 100-fold decrease is observed. Unfortunately it is necessary to place 2AP between two adenines to avoid efficient electron transfer involving adjacent guanines or cytosines[120,127]. Dallmann et al.[129] demonstrated that the structural perturbations by 2AP, replacing adenine and paired with thymine in the middle of the duplex are small. The two structures differ only in the position of the amino group that was moved from the major to the minor groove in 2AP (Fig. 1.1). However, the base pair dynamics was found to be four times faster, and also the lifetime of the next three base pairs were lowered in both directions, thus explaining the lower melting point of the duplex with 2AP in the center.

The acridine derivative 9-amino-6-chloro-2-methoxyacridine (ACMA) was also investigated by Ernsting et al.[130]. It was incorporated opposite adenine and needed to be separated from guanine for the same reason as 2AP. The Stokes shift is small and reaches a constant value within 200 femtoseconds, which is considerably faster than in aqueous solutions, where the Stokes shift develops on the nanosecond time scale. Although this is good evidence that ACMA intercalated into DNA, a larger and longer evolving shift is favored.

A series of pteridine analogues (third row in Fig. 1.1) of adenine (6MAP, DMAP)[131] and guanine (3-MI, 6-MI)[132,133] were developed by Hawkins and co-workers. These compounds, which are commercially available, are characterized by intense fluorescence around 430 nm (fl. quantum yield between 0.39 and 0.88) and a relatively long lifetime of the excited state (3.8 to 6.5 ns). They are very sensitive to the microenvironment, lowering their fluorescence quantum yield in DNA strands ($<$ 0.01 to 0.3) depending on the neighboring bases[134]. But melting experiments indicate, except for 6-MI, a sequence-dependent destabilization[131] which is, in case of 3-MI, similar to that of a single base pair mismatch[132]. Nevertheless, the well-documented quenching effects, the high fluorescence quantum yield and the already mentioned commercial availability made them useful in numerous applications[96].

In contrast to 2AP and the pteridine analogues, Matteucci and co-workers[135] developed a tricyclic cytosine analogue tC, 1,3-diaza-2-oxophenothiazine (Fig. 1.1), that is nearly insensitive to the environment. Although this property prevents it from being used as polarity probe, the negligible influence of surrounding bases on its fluorescence quantum yield and lifetime[136] makes tC particularly interesting in fluorescence anisotropy and FRET measurements[137]. It should be noted that tC , as evidenced by NMR structure determination, is the first artificial and highly fluorescent DNA base that does not perturb the DNA conformation[138]. Also the oxo-homologue of tC named tC^{O} is, on average, the brightest nucleobase analogue among other commercially available base substitutes, like 2AP and the pteridine analogues[139]. Moreover, tC^{O} (energy donor) was paired

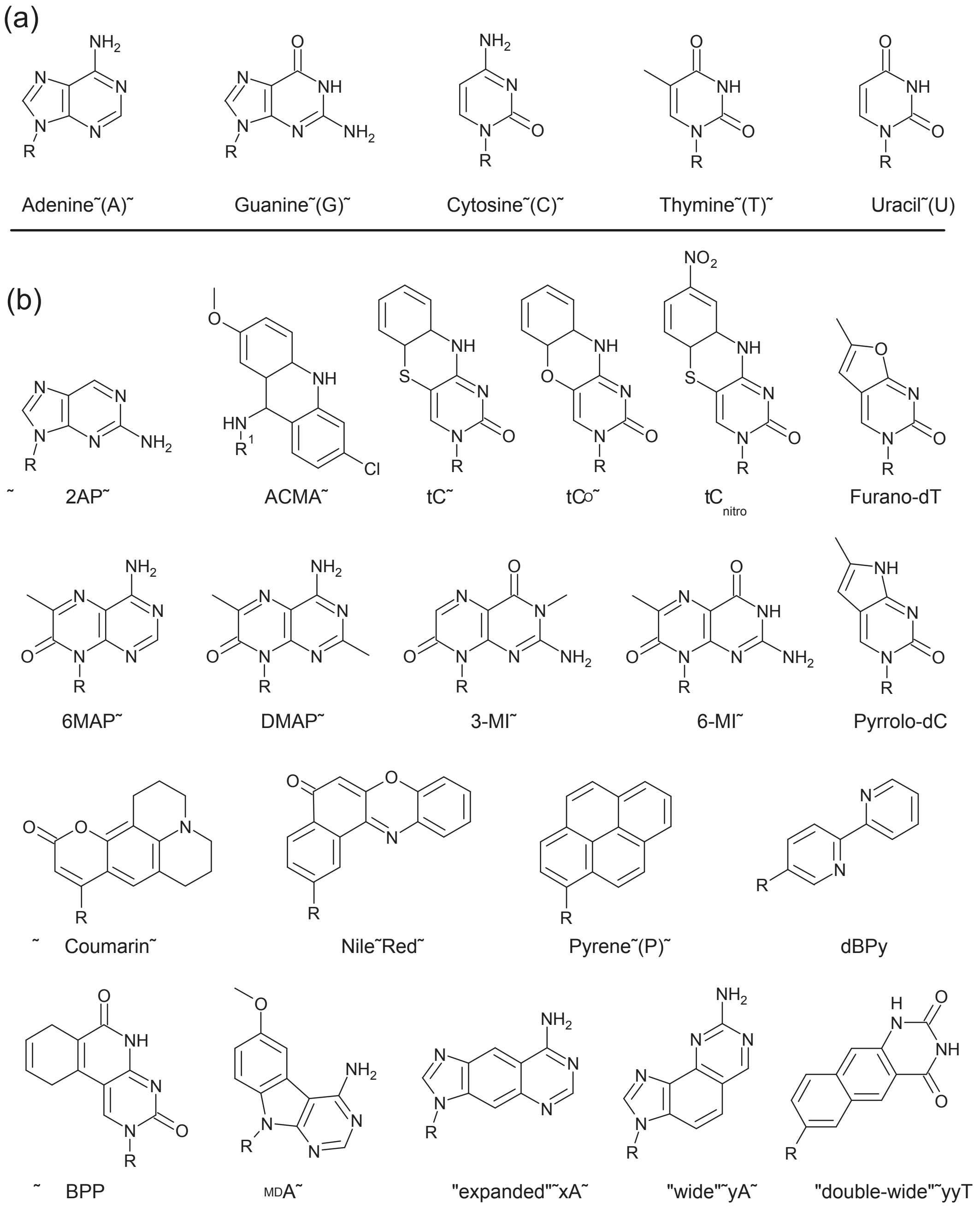


Figure 1.1: Structure of a) natural nucleobases and b) fluorescent base analogues. R = 2'-deoxyribofuranose; R^1 = alkyl-chain.

with tC_{nitro}, 7-nitro-1,3-diaza-2-oxophenothiazine (acceptor), to be the first nucleobase analogue FRET-pair and was then used to measure distances inside DNA of more than a full helix turn[140]. A recent work[141] offers strategies for a series of functionalized tricyclic cytosines in order to overcome the drawback of their insensitivity to the local environment.

A fluorescent nucleobase named furano-dT was introduced by Woo et al.[142] in 1996, but when incorporated into an oligonucleotide the final step of ammonia treatment in the solid-phase synthesis led to the C-analogue pyrrolocytosine (pyrrolo-dC, see Fig. 1.1). Fortunately, pyrrolo-dC pairs normally as a cytosine (selectively with guanine), does not disrupt the DNA helix, and is tolerated by DNA and RNA polymerases[143]. However the fluorescence quantum yield is reduced after incorporation into a single strand and decreases even more after hybridization with a complementary strand[143,144]. But pyrrolo-dC, since it is commercially available, has already served as a tool in several applications[145] like characterization of the transcription bubble in T7 RNA polymerase[146], detection of abnormal base pairing in DNA/RNA hybrid strands of the HIV-1 polypurine tract[144] or probing the kinetics of parts of damaged DNA by a human alkyltransferase[147]. Interestingly, pyrrolo-dC and 2AP can be used in a fashion similar to that of a molecular beacon[148]. Actual development of pyrrolocytosine is concentrated on the enhancement of the fluorescence quantum yield[149,150] by replacing the methyl group with an aromatic tether or the whole aromatic ring system[145].

Coleman and Madaras[151] introduced a coumarin 102 containing nucleoside opposite an abasic site (see Fig. 1.1). In the subsequent study, Brauns et al.[152] were the first to show the dynamic Stokes shift of fluorescence of a specially designed base-pair analogue (inside of a DNA double helix) and concluded that the interior of DNA is "a unique dynamic environment unlike either a fluid or a molecular crystal." The coumarin nucleoside was then used (also in the group of Ernsting) to explore the environment and dynamical features of DNA oligonucleotides[153–157] and "fraying" (5 ps timescale) at the end of the helix[155]. When placed into the center of an oligonucleotide, sequence-independent[158] dynamics was observed that is distributed over a time range covering six orders of magnitude (40 fs to 40 ns) and follows a power law with small exponent (0.15). However the power law dynamics, which lacks distinguishable subcomponents, indicates strong coupling of motions inside different parts of the DNA system[154], thus limiting its use as polarity probe in our TDSS studies.

The next chromophore, Nile Red (Fig. 1.1), also contains four fused aromatic rings and was introduced as base analogue by Okamoto et al.[159]. The nucleoside maintained high solvatochromicity comparable to the free Nile Red. The fluorescence, when incorporated opposite adenine, guanine or a missing base, was greatly shifted to shorter wavelength by the addition of β-cyclodextrin, but only slightly opposite cytosine, thymine and an abasic site. The same separation was found in the melting temperatures, where the latter group showed higher T_m values. A stabilization induced by a more tightly binding of Nile Red to the duplex was given as explanation. In case of placing Nile Red opposite the first group, Okamoto suggested the usage as probe for the microenvironment of DNA that can monitor polarity changes, caused by interactions between DNA and DNA-binding molecules. However, their results indicate that the observed effect of conformational change depending on polarity is driven by structural perturbations in the duplex.

Okamoto et al.[105] also developed a series of base-discriminating fluorescent nucleobases (e.g., BPP and ^{MD}A in Fig. 1.1). They have been designed for SNP discrimination, so they rely on quenching of fluorescence when paired with a certain native nucleobase, which stands in contrast to our design of a non perturbing polarity probe with high fluorescence quantum yield.

In 1998 Kool et al.[97] introduced pyrene (P, Fig. 1.1), a polycyclic aromatic hydrocarbon (PAH), that was incorporated into DNA by a Klenow fragment opposite an abasic site. The observed selectivity and efficiency were greater than those for the natural DNA triphosphates. Interestingly, the DNA synthesis stopped after incorporation of P, which makes it useful in detection of abasic mutations[160]. A point of more general interest is the observation that the replacement of a native base opposite an abasic site by pyrene, stabilized the duplex in a range of 18 to 23 °C[97]. This finding has been explained by restored π–π-stacking interactions in the strand with the abasic site, since P can cover a similar surface area as native base pairs. Use of hydrocarbons with less aromatic surface, for example, 2,2'-bipyridyl deoxynucleoside (dBPy, Fig. 1.1), led to a thermal stabilization of only 3.7 - 8.4 °C in comparison with nucleobases[161]. Although the larger porphyrin C-nucleoside (PP, Fig. 1.2) showed a stabilization of ca. 10 °C[162], it is only half of the enhancement induced by pyrene nucleoside P. Increased thermal stability (ΔT_m from 1.0 up to 8.4 °C) was also measured for bis-substituted and alkyl-linked phenanthrene (Phen), phenanthroline (PheN) and pyrene ($P_{1,8}$, see Fig. 1.2), depending on linker length and, again, in comparison with adenine as counterbase to the abasic site[163,164]. Additionally, it was demonstrated that the substitution of a native base by a pyrene (P) with flexible, acyclic linkage in the middle of the sequence destabilized the DNA duplex more (ΔT_m = -6.3 °C) than insertion close to the 3'- or 5'-ends (ΔT_m = -2.7 °C)[165].

An alternative strategy to the replacement of the nucleobase is the development of acyclic sugar analogs that are linked to an intercalator. The flexibility gained by the acyclic sugar linker allows to place large intercalating fluorophores, like the ethidium derivative Etd (see Fig. 1.2), in the center of DNA duplexes opposite native bases. These double strands showed not only similar Tm values in melting experiments, but also similar fluorescence spectra with emission maxima in a range of 622 and 625 nm under excitation at 520 nm[166]. A possible explanation for this unexpected insensitivity to the counterbase was given by suggesting an extrahelical position for them, due to the large space demand of Etd. Ethidium itself is known to prefer binding to duplexes of RNA, DNA and DNA-RNA-hybrids rather than to triplex or G-quadruplex DNA[167]. Upon binding into DNA/RNA a more than 10-fold increase in fluorescence signal is observed, since the exchange of amine hydrogens with the solvent is reduced[168]. A combination of Etd and 7-deazaguanine (as charge acceptor[169]) allows fluorescence detection of single base mismatches and abasic sites when these two pseudonucleotides are incorporated between two base pairs[170].

The aforementioned extrahelical position of the counterbase to the intercalating Etd is supported by further work, where a considerably larger perylene-diimide (PDI2, Fig. 1.2) was placed in the middle of the duplex opposite thymidine and then opposite an abasic site. Interestingly, the melting temperatures of both duplexes were identical[171], thus indicating same stacking interactions for PDI2 even in the presence of the thymi-

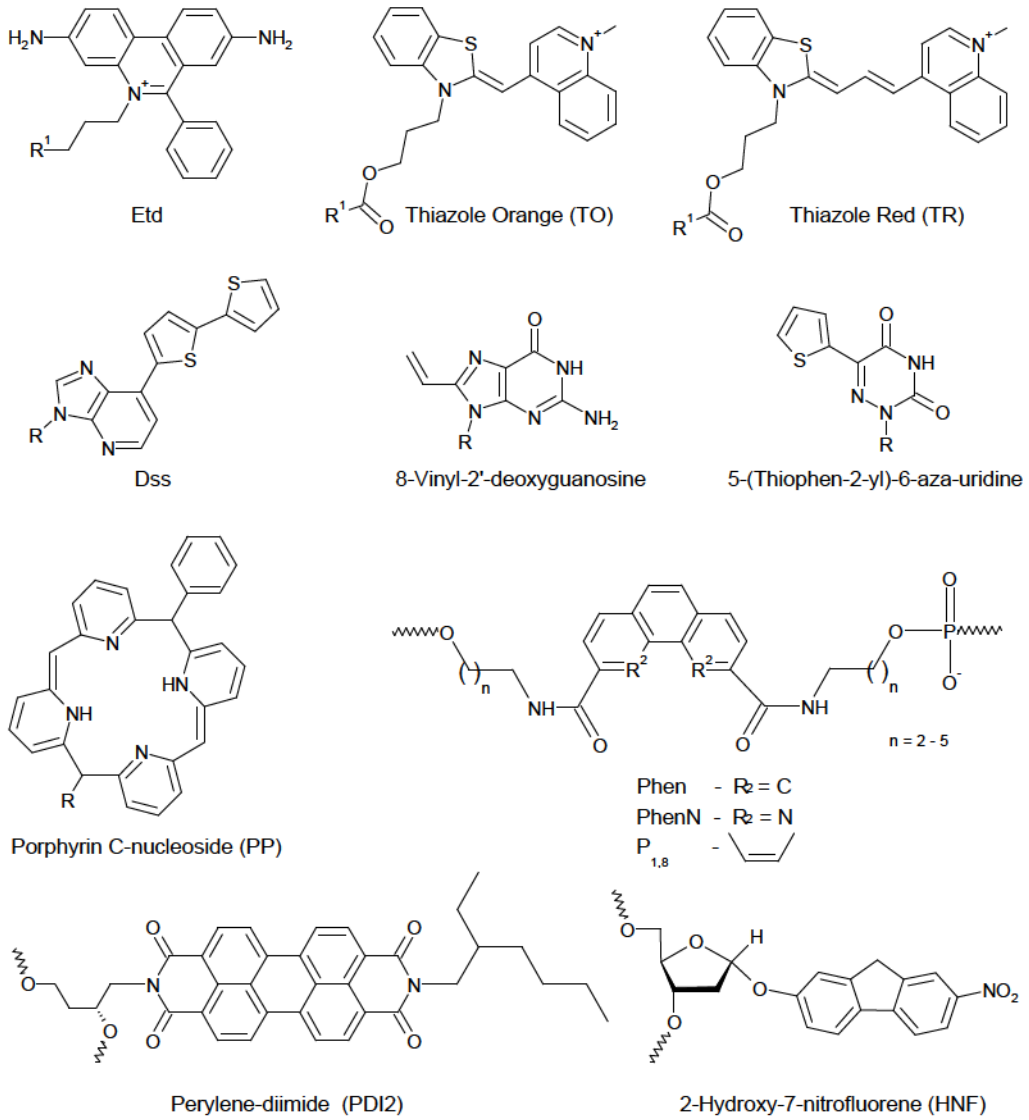


Figure 1.2: Structure of large fluorescent base analogues. R = 2'-deoxyribofuranose; R^1 = (S)-aminopropanediol.

dine counterbase. Other large chromophores, like 1-(Phenylethynyl)pyrene (PEPy) and 9,10-bis(phenylethynyl)anthracene (BPEA), were incorporated into DNA by using a 1,3-butanediol backbone[172]. Both were positioned opposite thymidine which led to a destabilization of the duplex (BPEA ΔT_m = -6.8 ℃, PEPy ΔT_m = -1.4 ℃). Compared to the initial pyrene, PEPy analogs have beneficial spectroscopic properties when used in molecular biology applications, since biomolecules will be excited at the same wavelength as pyrene[173]. A notable application of PEPy pairs was the detection of single polymorphisms in the gene fragment of 23S rRNA Helicobacter pylori[174].

The group of Hirao et al.[175] introduced a fluorescent purine analogue, 7-(2,2'-bithien-5-yl)-imidazo[4,5-b]pyridine (Dss in Fig. 1.2) that can be incorporated site-specifically into DNA and RNA by polymerases. Moreover, it functions as a universal base that pairs with all four natural bases with nearly equal thermal stabilities. An important drawback for the usage as polarity probe is the long extension of the base analogue by tethering two thienyl-groups in a row that are somehow located in the major grove. As a consequence, the fluorescence of the Dss chromophore is only slightly changed upon duplex formation. Such observations are typical for a class of nucleoside analogues that are based on extension of purine and pyrimidine moieties. This strategy was frequently used by Srivatsan et al.[176–180], but as already stated, tethering of bulky groups is not favored in the current thesis.

Recent activities by Wagenknecht et al. are centered around thiazole orange (TO) and thiazole red (TR, see Fig. 1.2) as "DNA traffic lights"[181–183]. The wavelength-shift of fluorescence upon duplex formation, due to aptamer target binding, is quantified as altered contrast ratio. Therefore, Wagenknecht and co-workers[181] suggested aptasensors as potential application.

The group of Diederichsen et al.[184,185] introduced 8-vinyl-2'-deoxyguanosine (1.2) as a DNA polymerase processable base analogue that is capable to detect different types of DNA quadruplex structures.

The group of Eric T. Kool[186,187] continued their work on multichromophoric DNA systems. This approach, where the DNA backbone offers a scaffold for an array of chromophores, has gained more and more interest during the last few years[188,189]. For this purpose, sets of size-expanded nucleobases were composed[190]. They are based on a series of earlier works by Kool et al.[94,191] and referred to as "expanded DNA" (xDNA), "wide DNA" (yDNA) and "double-wide DNA" (yyDNA, see Fig. 1.1). These size-expanded analogs were designed to extend the genetic alphabet and are able to form base pairs different (orthogonal) from those found in native DNA[95]. When a single expanded base pair is substituted into natural DNA, they are destabilizing the natural helix, due to their large size. Therefore, a single expanded nucleobase is not suitable as probe for the desired TDSS experiment. However, when all base pairs are expanded, xDNA and yDNA form a highly stable, sequence-selective and widened double helix[94].

A similar multichromophoric approach was used by Leumann et al.[192] who incorporated multiple 2-pyrenyl-C-nucleosides (each P replaced an adenine/thymine base pair) which then formed a stable excimer. In earlier studies, pyrene has been stacked by using $P_{1,8}$[163] or tethering to deoxyuridine[193,194], so that its position is defined upon duplex formation of the parent nucleobase. In an actual work, Häner et al.[195] switched from $P_{1,8}$ to a porphyrin nucleoside and were able to build double strands containing up to four free base porphyrins. They maintain duplex stability when placed pairwise in opposite positions, whereas a considerable destabilization is observed opposite to natural nucleobases. A model for H-aggregation of the porphyrins, which causes fluorescence quenching, is supported by UV/vis spectroscopy. Häner's group[196] also studied the stacking of electron-rich pyrene (P) and electron-poor perylene-diimide (PDI) and found that electrostatic complementarity is important for aromatic π-π-stacking interactions. P and PDI can stabilize the DNA duplex when incorporated into opposite strands with equal ratio of the chromophores.

Yitzhak Tor and his group extended their work (see pyrrolo-dC) on tethering,fusing thiophene and furan based moieties onto 6-aza-uridine[197,198] (Fig. 1.2) in order to build isomorphic fluorescent nucleosides; such molecules (including other nucleobases) were intensively studied by his group over the last years[96]. A possible application for 6-aza-uridine as chromophore in single molecule detection by two photon excitation has been suggested lately[199]. A contribution to the multichromophoric approach was also tested by demonstrating that three identical isomorphic fluorescent nucleosides in alternating or neighboring positions display enhanced, sequence-dependent signals for either duplex formation or dissociation[200]. Further extension of these molecules with substituted aryl rings increases the push-pull interactions yielding enhanced bathochromic shifts and solvatochromism[201] (but the price is a large non-rigid tether).Also worth mentioning is the design of an emissive RNA alphabet with bases that were all derived from thieno[3,4-d]pyrimidine[202]. These nucleobases exhibit visible emission, high quantum yield, and responsiveness to environmental perturbations.

At the end of this general overview about fluorescent nucleobase analogues we come to the direct predecessor of the chromophores that are studied in the current thesis. After coumarin, the group of Ernsting employed a 2-hydroxy-7-nitrofluorene (HNF, Fig. 1.2) opposite an abasic site[203]. In contrast to native DNA, the HNF chromophore is linked via an α-glycosidic-bond to the 2-deoxyribofuranose. NMR structure determination has revealed that the HNF moiety can intercalate into the duplex in two different orientations. In the “face-down” case the methylene group points towards the minor groove while in the “face-up” conformation towards the major groove. The magnitude of the time-resolved Stokes shift (2660 cm^{-1}) is large in comparison to coumarin (960 cm^{-1}), but the lifetime of the excited state is short, only 35 ps, due to intersystem crossing.

1.3 Aim of this work

Many fluorophores have been introduced into duplex DNA (or were attached) during the last few years, but most of them were designed for purposes (FRET, SNP etc.) that are not necessarily compatible with the envisioned TDSS experiments. The base analogues that Ernsting and coworkers have examined so far did not fulfill all desired requirements simultaneously. Short lifetimes of the excited state (ACMA, HNF), electron transfer with nearby bases like guanine (2AP, ACMA), or coupling with modes below 300 cm^{-1} (coumarin) all hampered the observation of the time-dependent Stokes shift. Continued development of suitable chromophores, backed up by femtosecond spectroscopy and NMR structure determination, is clearly required. Here we employ NMR spectroscopy to determine the structure of three modified oligonucleotides. The principle design and position of the fluorescent nucleobase analogues, including necessary modifications to their backbone, are given in Fig. 1.3. All modifications were studied in the same basic sequence (Fig. 1.3) to ensure comparability between strands.

The first modified duplex incorporates 2-hydroxy-7-carboxyfluorene (HCF, Fig. 1.3) opposite an abasic site; this construct will be abbreviated 13merHCF. It has been introduced as successor to the HNF chromophore, so structural similarities such as two chromophoric orientations are most likely[203], but this also allows discussing the effect of changing a functional group in a known environment. Furthermore it is necessary to investigate the protonation state of the carboxyl group, since the logarithmic acid dissociation constant (pK_a) of nucleobases is normally raised when incorporated into the duplex[19].

The second DNA duplex strand introduces 6-hydroxyquinoline (6HQ, Fig. 1.3) as base surrogate. A cytosine was chosen as counterbase and potential partner for hydrogen bonding. The resulting duplex strand will be abbreviated 13mer6HQ. The photophysical properties of the chromophore are well-known, since it was studied by Ernsting et al. as free N-methyl-6-quinoline[122] and as covalently linked tether to trehalose[204]. In contrast to the other duplex strands, the 6HQ base analogue has been incorporated as 2,3-dihydroxypropylnucleoside which is known as glycol nucleic acid (GNA[71–74]). The small size of this flexible acyclic linker will possibly affect the chromophore and the adjacent nucleotides, but it was the only synthesis method with sufficient product yield. Up to now, only crystal structures of duplex strands containing a full GNA backbone were published, so the solution structure of a DNA duplex with a GNA monomer in the center will be of general interest.

In the last duplex which is studied here, the central base pair is replaced by an artificial base pair analogue; this construct will be abbreviated 13mer4AP-DAP. Here 4-aminophthalimide (4AP, Fig. 1.3) is paired with 2,4-diaminopyrimidine (DAP) in the opposite strand. Both nucleobase analogues were synthesized in cooperative work[205] between the groups of Ernsting (DAP) and Wagenknecht (4AP). The 4AP chromophore is known to be highly sensitive to the polarity of the medium and it has been stated that the fluorescence properties (e.g. intensity, lifetime) are further enhanced due to the involvement of hydrogen bonding interaction with the solvent molecules[206]. In fact, the observed lifetime of the excited state is lowered from 14 - 15 ns in protic media to roughly 1 ns in water. In order to allow hydrogen bonding inside of the duplex, a second artificial nucleobase analogue (DAP) was necessary, since there is no native nucleobase that can provide the correct hydrogen bonding pattern. A problem that has to be addressed is the hydrolysis of the 4AP chromophore in water, especially under basic conditions. A weak acidic buffer extended the lifetime of the 4AP mononucleotide from hours to a few days, but in the duplex this could cause protonation of the DAP. Therefore, a whole series of questions about 13mer4AP-DAP has to be answered by NMR and UV/vis spectroscopy. These questions are centered around the duplex structure, hydrogen bonding pattern, protonation of DAP and finally, the chemical lifetime of the sample.

Even though the main topic of this work is the structure determination of oligonucleotides with embedded polarity probes, we will see that equal attention must be given to temperature-dependent measurements of absorption. The corresponding method will therefore be described first. Normally UV absorption measurements, as function of temperature, are used to obtain a hybridization curve and to determine the melting point T_m of a DNA duplex. For a modified strand the melting point indicates a degree of duplex stability (or destabilization) when compared to the native or unmodified reference duplex. In this work the measurement which is usually performed at a single wavelength (260 nm) will be replaced by the detection of full spectra at different temperatures, covering the UV/vis range (200-700 nm). The spectra are then analyzed in their entirety by Singular Value Decomposition (SVD). As will be shown, this procedure can uncover significantly more information about the binding properties of the chromophore/duplex system compared to a single melting point determination. Melting will be seen (by focussing on the UV absorption band around 260 nm) as a global and a local, sequence-dependent process of the duplex. This behavior can then be compared to that of the probe (monitoring the

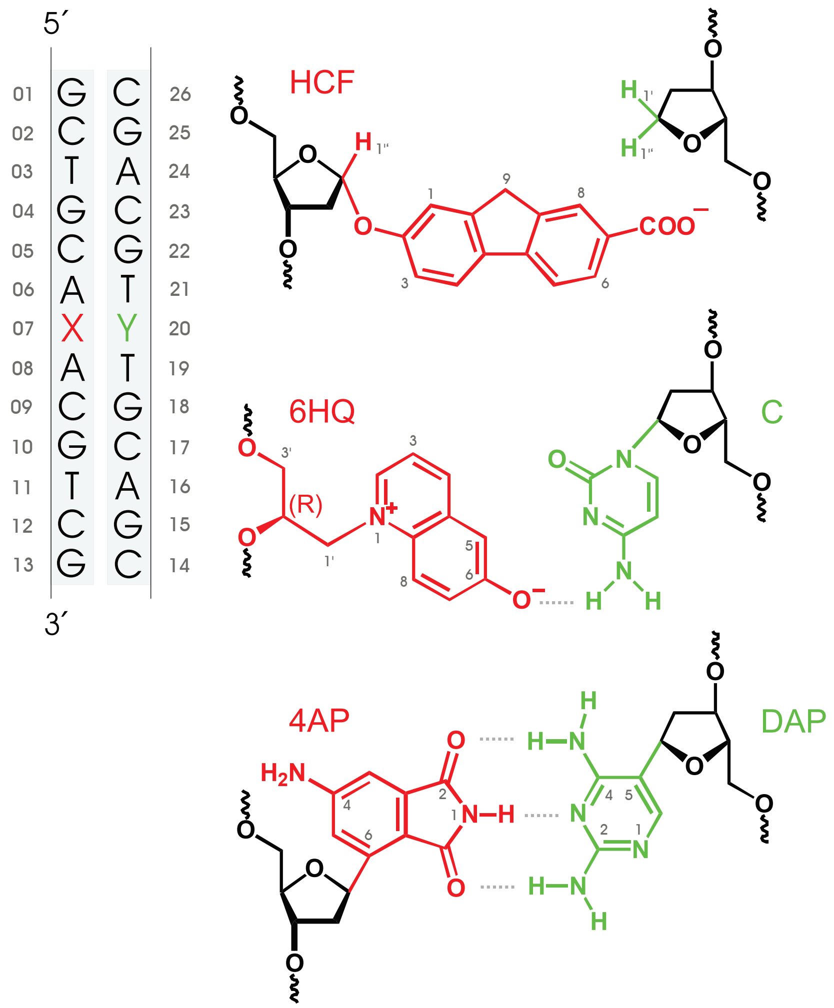


Figure 1.3: Structure of new fluorescent base and base pair analogues. Native DNA is marked black and all modifications including backbone are either red or blue colored. Red molecules are placed at position X and the green ones at Y, respectively.

corresponding absorption band in the visible) which reflects local changes only. In this way possible perturbations of the melting process, like bubble formation, can easily be explained or definitely excluded. Also concurrently evolving processes (e.g. peak shifts, change of active species) will be revealed that would otherwise affect the melting analysis. In combination with the solution NMR structure of the duplex, a deeper understanding is achieved of binding and solvation, structural fluctuations, and the melting process.

2 Conceptual background

The following sections will provide the background for the major topics of this work. At the beginning, the nomenclature and essential structural aspects of DNA (sec. 2.1) will be outlined to provide a common basis for discussion. The middle part will describe the methods to obtain structural restraints and how they are used to determine the structure of DNA (sec. 2.2 - 2.5). Finally, the last two sections are dedicated to the field of optical spectroscopy. They will introduce Singular Value Decomposition (sec. 2.6) as method to derive more information about the melting process of DNA (sec. 2.7).

2.1 Structural aspects of DNA

The nomenclature in Fig. 2.1 follows the recommendations of Markley et al.[207] for the presentation of NMR structures. Natural DNA is composed of four nucleobases and a phosphodiester bridged backbone of 2'-deoxy-β-D-ribose, also referred to as "sugar". The bases can be subdivided into purine and pyrimidine derivatives. Adenine (A) and guanine (G) belong to the purines, whereas thymine (T) and cytosine (C) represent the pyrimidines. They are shown in Fig. 2.1, forming the two Watson-Crick base pairs A:T and G:C.

By attaching A, G, T, C to the C1' of the sugar the nucleosides adenosine, guanosine, cytidine and thymidine are formed. Consequently, all atoms of the deoxyribose will be marked with " ' " to discern them from the nucleobase atoms. Five torsion angles ν_0 to ν_4 (Fig.2.1) specify the conformation of the sugar, but due to geometrical constraints in a five-membered ring, only two parameters are necessary to describe them. These are the pseudorotation P and the maximum torsion angle ϕ_m (pucker amplitude), so the angle ν_j

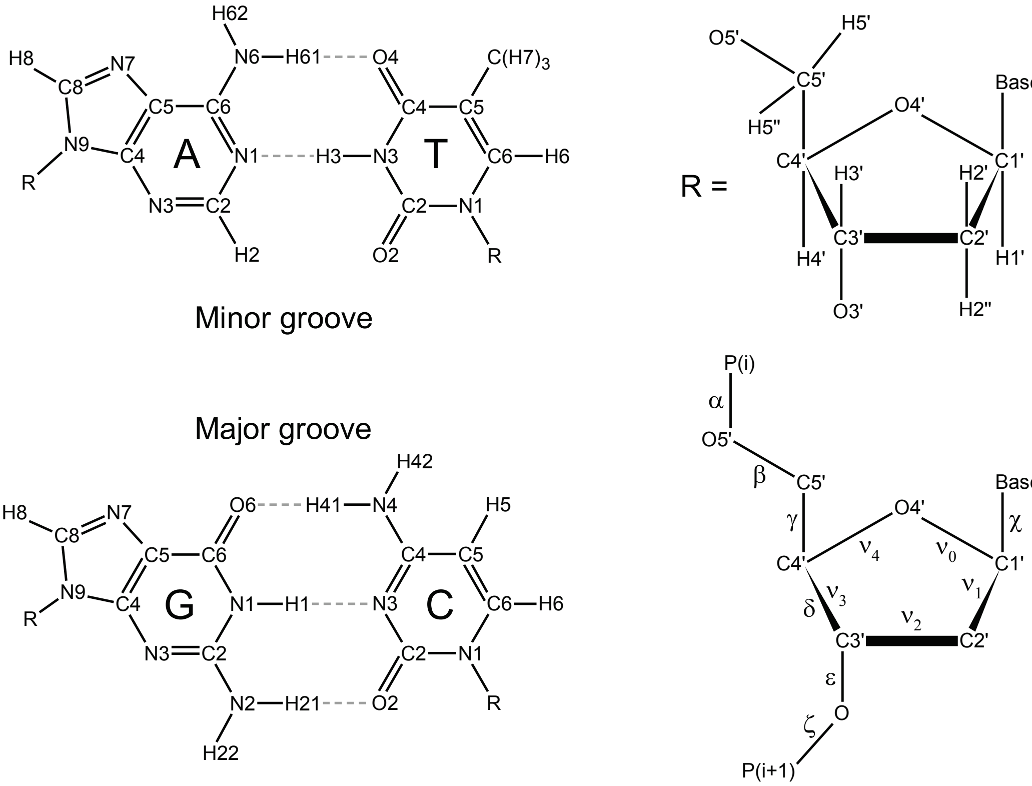


Figure 2.1: Structure and nomenclature of the Watson-Crick base pairs A:T and G:C are shown in the left panel. The upper right corner illustrates 2'-deoxy-β-D-ribose and the lower right denominates the torsion angles.

is given by

$$\nu_j = \phi_m cos[P + 144°(j-2)] \qquad (j = 0, 1, 2, 3, 4). \tag{2.1}$$

Please note that the concept of pseudorotation was originally introduced for cyclopentane[209], but Altona and Sundaralingam[208] extended the concept to the sugar ring of nucleosides and nucleotides, respectively. Although sugar conformations are not static, it is possible to define two regions where ribose and deoxyribose nucleotides are mainly found (black arrows in Fig. 2.2). One of these regions is centered around the C3'-endo conformation, also referred to as N-type (north), that dominates in A-DNA. As outlined in the introduction, the A-form is typically found in single crystals of short duplex strands, while in solution the B-form with C2'-endo conformation (or S-type, south) is predominant.

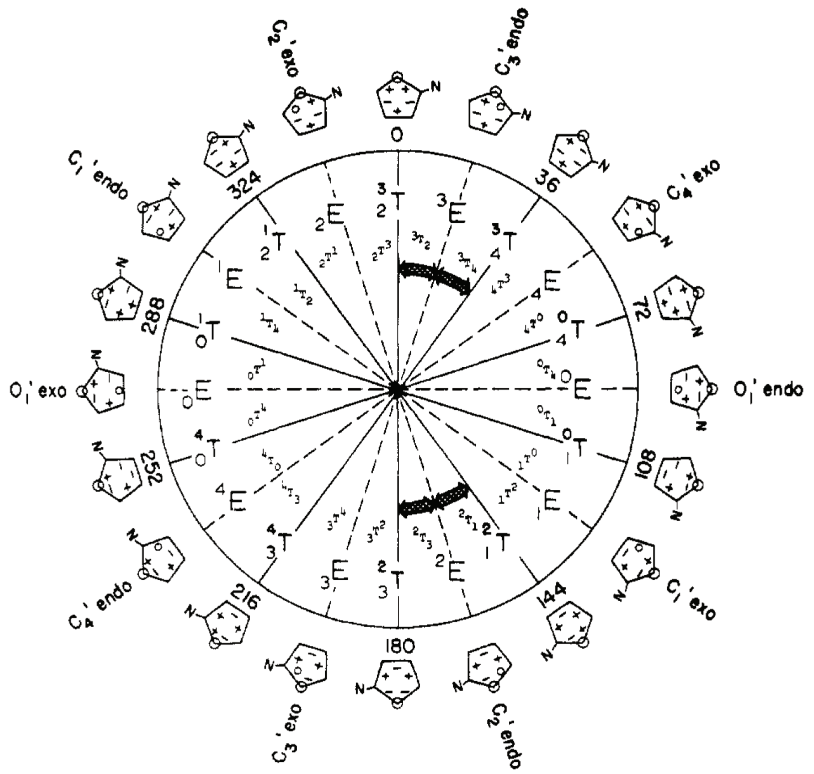


Figure 2.2: Pseudorotation phase angle P and related sugar conformations, picture taken from Altona and Sundaralingam[208]. Large black arrows indicate regions where mainly ribose and deoxyribose nucleotides are found.

The glycosidic bond angle χ characterizes, together with the backbone angles $\alpha - \zeta$, the helical structure of an oligonucleotide. Two ranges can be found for χ that are designated syn and anti. The more stable anti conformation is usually found in A- and B-form DNA. For the syn conformation, where the position of the nucleobase above the sugar causes steric interference, the assistance of some external force is required[19]. This can be either a high salt concentration to favor Z-DNA[17] or the attachment of a bulky group to the 8 position of purines (6 in pyrimidines).

The variation of the backbone angles $\alpha - \zeta$ allows to distinguish three major helix structures (A, B and Z). The A and B forms are right-handed helices, while Z is left-handed and occurs in alternating purine-pyrimidine sequences (mainly GC). Due to base

Table 2.1: Average structural parameters for different helical forms[19]

	A-DNA	B-DNA	Z-DNA
Helix handedness	Right	Right	Left
base pairs / reapeating unit	1	1	2
base pairs / helix turn	11.6	10	12
Helix twist (°)	32.7	36	-10[a], -50[b]
Rise / base pair (Å)	2.9	3.4	-3.9[a], -3.5[b]
Helix pitch (Å)	32	34	45
P distance from helix axis (Å)	9.5	9.4	6.2[a], 7.7[b]
Displacement of base pair to helix axis (Å)	-4.1	0.8	3.0
Glycosidic bond orientation	anti	anti	anti[c], syn[d]
Major groove depth (Å)	13.5	8.5	Convex
width (Å)	2.7	11.7	
Minor groove depth (Å)	2.8	7.5	9
width (Å)	11.0	5.7	4

[a]CpG or [b]GpC step. [c]Cytosine. [d]Guanine.

displacement from the helix axis with large inclination, the A-form is thick and compressed along the helix axis. In contrast to this, the nucleobases of B are in the center of the helix and inclined nearly perpendicular to it, thus leads to a smaller diameter and nearly equal depth of the grooves. The mean twist angle in B-DNA is 36°, though they vary in a wide range between 24 to 51°, giving roughly 10 base pairs per helix turn[19]. More helical properties that distinguish between A-, B- and Z-DNA are listed in Table 2.1, but note that averaged values are given for sequence-dependent parameters of a dynamic system.

2.2 Nuclear Overhauser Effect spectroscopy

NMR structure determination has become an important tool to study biomolecules in solution. The main reason for this success is the possibility to derive distance information from Nuclear Overhauser Effect spectroscopy (NOESY). The NOE was discovered by Overhauser in 1953[210] and originally described the interaction between the saturation of the electron spin resonance and the polarization of their nuclei. The same effect was then observed between different nuclei in decoupling experiments. Therein, a resonance line was selectively saturated prior to recording of the 1D spectrum. In such experiments, Anet and Bourn[211] (1965) found that the signal intensity for nuclei in close proximity to the saturated nucleus was enhanced by 17 to 45 % due to dipole-dipole cross-relaxation (see also Fig. 2.4). However, the application of this method to biological macromolecules suffered from the limited selectivity of preirradiation in crowded spectral regions and huge experimental effort[31]. The development of two-dimensional (2D) cross-relaxation spectroscopy in 1980 by Macura and Ernst[212] finally allowed the measurement of a complete NOE-network between all the protons in a macromolecule in only one experiment.

In Fig. 2.3 a basic NOESY pulse-sequence is shown. Transverse magnetization is created by the first non-selective 90° ($\pi/2$) pulse and allowed to precess freely for an evolution time t_1. The latter is varied in the course of the experiment, thereby frequency-labeling the magnetization components. The second pulse transfers the magnetization back along the (negative) z-axis and longitudinal cross-relaxation takes place for the length of mixing time t_m. In contrast to t_1, the mixing time remains constant during the measurement, but can be adjusted at the beginning to fulfill the needs of the sample. The last pulse finally generates transverse magnetization whose precession can be detected as a function of t_2.

The formal treatment of a multi-spin system, comprising g spin groups in a two-dimensional NOESY experiment, is presented in the following. Basic conditions are the presence of dipolar interactions to allow cross-relaxation and the absence of scalar spin-spin-interactions. Then cross-relaxation of longitudinal magnetization components M_{ij}

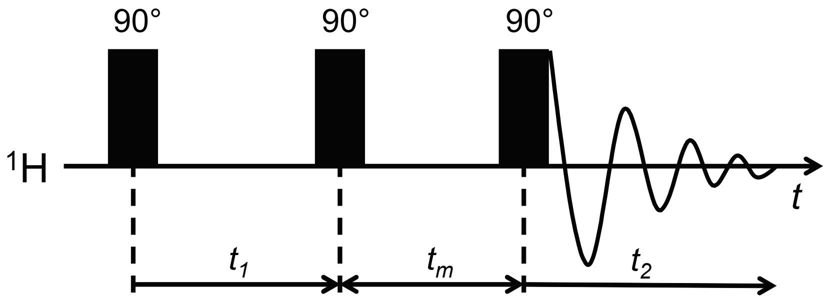


Figure 2.3: Basic NOESY pulse-sequence for two-dimensional cross-relaxation spectroscopy. Transverse magnetization created by the first 90° ($\pi/2$) pulse becomes frequency-labelled in the course of evolution period of length t_1. After the second pulse, longitudinal cross-relaxation takes place for the length of mixing time t_m. The last pulse finally generates transverse magnetization whose precession can be detected as a function of t_2.

can be described with the following system of equations[212]:

$$\dot{\mathbf{m}} = \mathbf{R} \times \mathbf{m}. \tag{2.2}$$

Here the vector $\mathbf{m}$ comprises the deviations of M_{zi} from thermal equilibrium for all g spin groups and the relaxation matrix $\mathbf{R}$ contains the cross relaxation rates R_{ij} as well as the external relaxation (leakage) rates R_i. A component m_i of $\mathbf{m}$ is defined by:

$$m_i = M_{zi} - \frac{n_i}{N} M_0 \qquad with\ \ N = \sum_i n_i \ \ and\ \ i = 1, 2, ..., g. \tag{2.3}$$

where M_0 is the total equilibrium magnetization of the N nuclei. At the beginning of the mixing period the initial z-magnetization components are encoded by the precession frequencies (ω_i) that are witnessed at t_1 of the evolution period.

$$m_i(0) = M_0 \frac{n_i}{N} [cos(\omega_i\, t_1)\, exp\left(-\frac{t_1}{T_{2i}}\right) - 1]. \tag{2.4}$$

The application of the second $\pi/2$ pulse marks the beginning of the mixing period during which cross-relaxation proceeds. The recovery of the magnetization towards equilibrium

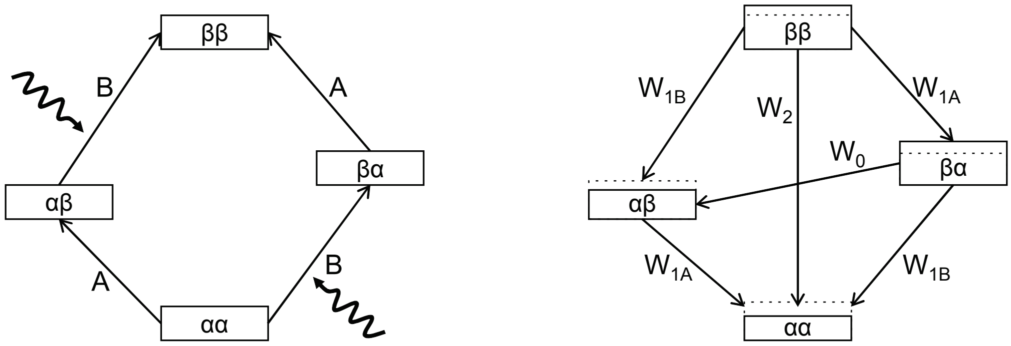


Figure 2.4: NOE effect in a two spin system. The left panel shows a system of two spins A and B, where the size of the boxes represent a simplified occupation ratio. Upon saturation of the B transitions (curly arrows) the ratio is changed to the scheme in the right panel, where solid lines mark the new ratio and dashed lines the old one. The arrows on the right now indicate the pathways for cross-relaxation and W the transition probability. The small indices denominate the type of transition (zero, single and double quantum). In small molecules with short correlation time W_2 dominates and hence enhances the intensity of A. The longer correlation time in macromolecules prefers W_0 for which a negative NOE effect is measured. The ratio between W_0 and W_2 decides whether enhancement, reduction or even no change of intensity is observed.

at time t_m can then be written as solution to eq. (2.2)[212]

$$\mathbf{m}(t_m) = exp[-\mathbf{R}\, t_m]\, \mathbf{m}(0), \tag{2.5}$$

where the matrix $\mathbf{m}(t_m)$ represents the magnetization components after the mixing time t_m and $\mathbf{m}(0)$ the intensities of the diagonal peaks (defined in eq. 2.4) at $t = 0$ of the mixing period. The diagonal (R_{ii}) and off-diagonal (R_{ij}) relaxation matrix elements are given as[213]:

$$R_{ii} = q_{ij} \sum_{i,j} (J_{0,ij}(\omega_i - \omega_j) + 3\,[J_{1,ij}(\omega_i) + J_{1,ij}(\omega_j)]\ + 6\, J_{2,ij}(\omega_i + \omega_j) + R_{1i}) \tag{2.6}$$

$$R_{ij} = q_{ij}[6\, J_{2,ij}(\omega) - J_{0,ij}(\omega)] \tag{2.7}$$

$$with\ \ q_{ij} = \frac{\hbar^2\, \gamma_i^2\, \gamma_j^2\, \mu_0^2}{160\, \pi^2}.$$

The term R_{1i} in eq. 2.6 represents the leakage rate, which can usually be neglected in the absence of paramagnetic nuclei. The factor q_{ij} collects all constant values, these are $\hbar$ as the reduced Planck constant, γ_i and γ_j as the gyromagnetic ratios for spins i and j, respectively and μ_0 as the magnetic constant or vacuum permeability. $J_{n,ij}(\omega)$ represents the spectral densities for the zero, single and double quantum transitions (n = 0,1,2). In Fig. 2.4 the pathways of these transitions are shown with a two-spin system as example and their spectral densities are defined as follows:

$$J_{n,ij}(\omega_0) = \frac{\tau_c^{ij}}{1 + \left(n\,\omega_0\,\tau_c^{ij}\right)^2}\,\frac{1}{r_{ij}^6}. \tag{2.8}$$

This equation correlates the mixing time t_m in eq. 2.5 with the rotational correlation time τ_c^{ij} of the vector between spin groups i and j. The corresponding internuclear distance is r_{ij}. One can now distinguish two cases in which similar spin groups are observed. In case of a small molecule with short τ_c^{ij} a relatively long mixing time has to be applied in order to allow full cross-relaxation of the longitudinal magnetization, but when compared to large biomolecules with long τ_c (e.g. DNA) then a short mixing time in the range of 100 - 200 ms has to be chosen. In addition, an approximation has been introduced in eq. 2.8 in which ω_i and ω_j are replaced by the center frequency ω_0, since the differences in resonance frequency for various spins are negligibly small compared to the value of the resonance frequency itself. The factor n in the denominator again marks the type of transition and can range between 0 and 2 (see Fig. 2.4). With the help of some additional assumptions the intensities of NOE cross-peaks can then directly related to distances in a molecule. This approach is called “isolated spin pair approximation” and will be introduced in the next section.

2.3 Isolated Spin Pair Approximation - From NOESY to distance

The usually employed method to derive distances from NOESY spectra is the "isolated spin pair approximation"[214,215] (ISPA), also known as "two-spin approximation"[216].

Several assumptions have to be made at the beginning. First, a single correlation time (τ_c) for the whole molecule is introduced in order to replace the individual τ_c^{ij} in eq. 2.8.

$$J_{n,ij}(\omega_0) = \frac{\tau_c}{1 + n^2\,\omega_0^2\,\tau_c^2}\,\frac{1}{r_{ij}^6}. \tag{2.9}$$

This assumption is valid, since Reid et al.[216] have shown that correlation times for base and sugar protons are comparable. Moreover, it can be assumed that oligonucleotides shorter than 15 base pairs are like isotropic rotors[217]. Sometimes one has to consider local mobility of residues or whole substructures. In such cases a modified spectral density function (eq. 2.10) has to be defined[218,219], where τ_e is introduced as the effective correlation time of the local mobility site and S^2 as the generalized order parameter. S^2 is a measure for the flexibility of the site with values ranging from 0 (unrestricted motion) to 1 (fully restricted motion).

$$J_{n,ij}(\omega_0) = \left(\frac{S^2\,\tau_c}{1 + n\,\omega_0^2\,\tau_c^2} + \frac{S^2\,\tau_e}{1 + n\,\omega_0^2\,\tau_e^2}\right)\frac{1}{2\,r_{ij}^6}. \tag{2.10}$$

In proteins, order parameters range from 1 to as low as 0.6 for flexible side chains[220], while S^2 in DNA is on the order of 0.8 for all proton pairs[221,222]. In the ISPA approach the contribution of local mobility will be canceled when the desired distances are referenced to a series of fixed and known distances. The next step in the ISPA approach is the expansion of eq. 2.5 into a Taylor series[212]:

$$exp[-\mathbf{R}\,t_m] = 1 - \mathbf{R}\,\tau_m + \frac{1}{2}\mathbf{R}^2\,\tau_m^2 + \ldots, \tag{2.11}$$

whereby the cross-peak intensities a_{ij} [212] are given by

$$a_{ij}(t_m) = (\delta_{ij} - R_{ij}\, t_m + \frac{1}{2}\sum_k R_{ik}\, R_{jk}\, t_m^2 + \ldots)\frac{n_j}{N}\, M_0. \tag{2.12}$$

The central assumption of the ISPA approach is that the Taylor series can be truncated after the linear term for short mixing times, thus cross-peak intensity ($i \neq j$) becomes a linear function of r_{ij}^{-6} in equation 2.13. For longer mixing periods, effects of spin diffusion may be taken into account, which is magnetization transfer through a third atom (represented by the quadratic term).

$$a_{ij}(t_m) = R_{ij}\, t_m = q_{ij}\, \tau_c\, \tau_m \left(\frac{6}{1 + 4\,\omega_0^2\, \tau_c^2} - 1\right)\frac{n_j}{N} M_0\, \frac{1}{r_{ij}^6} \tag{2.13}$$

The intensity a_{ij} can now be referenced to a known, fixed distance r_{ref} with a corresponding a_{ref}, thereby eliminating all constant terms.

$$\frac{a_{ref}}{a_{ij}} = \left(\frac{r_{ij}}{r_{ref}}\right)^6 \qquad or \qquad r_{ij} = r_{ref}\sqrt[6]{\frac{a_{ref}}{a_{ij}}} \tag{2.14}$$

Commonly used reference distances in nucleic acids are the C H5-H6 [216], T C7-$(H7)_3$ to account for fast rotation in methyl groups or solvent exchange with amino and imino protons in C H42-H5.

2.4 Residual Dipolar Couplings

Over the last years Residual Dipolar Coupling (RDC) measurements have evolved into an important source of structure information beside NOE distances[223–225]. They offer complementary information about biological macromolecules that compensate a major drawback of NOE data. Due to the r^{-6}-dependence of the NOE effect, the latter is limited to distances up to 5 Å (see sec. 2.2). Although it is possible to observe long-range cross-peaks for proteins, where two residues that are far distant in primary sequence can be folded in close proximity to each other, the rod-like shape of DNA provides cross-peaks between adjacent bases only. Consequently it was impossible to describe long range effects like bending of an A-tract in DNA prior to the development of RDC measurements[226]. RDCs, on the other hand, provide information about the orientation of the bond vector relative to the external magnetic field, which is (in case of rod-like shaped DNA) identical to the orientation of the helical long axis. Thus it is possible to compare orientations between residues along the whole strand.

The NMR spectroscopy in partially oriented media was discovered by Saupe and Englert[227] in 1963, followed by a theoretical description a year later[228]. One limit for the application to large biomolecules was overcome with the introduction of high resolution NMR and corresponding methods. Tolman et al.[229] were the first to present RDC measurements of cyanometmyoglobin, which has a very highly anisotropic paramagnetic susceptibility. Only two years later, in 1997, Bax and Tjandra[230] were able to measure the diamagnetic protein ubiquitin, which was dissolved in a very dilute solution of bicelles that adopted an ordered, liquid crystalline phase. The induced order by an external component allowed the measurement of residual dipolar couplings, while the high resolution of NMR can be retained. In the following, Tjandra and Bax[231] showed that the degree of solute alignment with the magnetic field can be tuned by variating the concentration of the bicelles. This marked a breakthrough in biological NMR, since dipolar coupling experiments were not limited anymore to samples with highly anisotropic paramagnetic susceptibility.

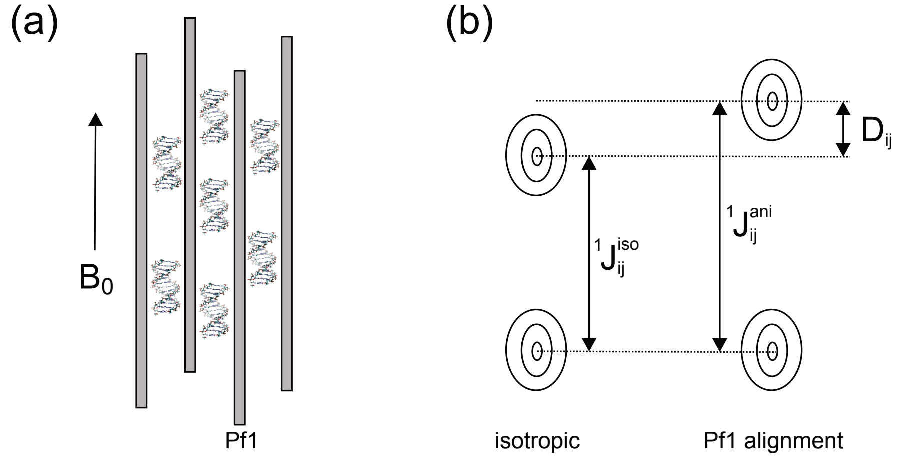


Figure 2.5: Acquisition of RDCs. (a) Showing the steric interaction between Pf1 phage and the DNA, preventing isotropically tumbling and inducing residual order. B_0 indicates a static magnetic field. (b) RDCs can be determined by measuring the difference of dipolar coupling in a bond vector with and without alignment.

Over the next years more alignment media were developed. Some were also based on bicelles to align in a liquid crystalline phase[232–235], others on filamentous phage[236], stretched gels[237,238], paramagnetic tagging[239] or DNA nanotubes[240]. In this work the bacteriophage Pf1 (see Fig. 2.5 a) was utilized to align oligonucleotides, since it is stable over a wide range of temperatures, but more important is that the interaction between DNA and Pf1 is minimized due to electrostatic repulsion of their negatively charged backbones[241].

RDCs are determined by measuring the difference of the dipolar coupling in the presence ($^1J_{ij}^{ani}$) and absence of molecular alignment ($^1J_{ij}^{iso}$), as is shown in Fig. 2.5 b.

$$^1J_{ij}^{ani} = {}^1J_{ij}^{iso} + D_{ij}, \tag{2.15}$$

where i and j are non-equivalent spins connected via a chemical bond. The dipolar contribution D_{ij} to the observed splitting between i and j derives from the secular part of the magnetic dipole–dipole interaction between the spins. In the high field limit we can write the effective Hamiltonian[225]

$$H_{ij}^{D}(t) = -\frac{\gamma_i\,\gamma_j\,\mu_0\,h}{8\,\pi^3\,r_{ij}^3(t)}\,I_{iz}\,I_{jz}\,\langle\frac{3\,cos^2\alpha_{ij}(t)-1}{2}\rangle. \tag{2.16}$$

Therein, r_{ij} is the distance between the nuclei, γ_i and γ_j are the gyromagnetic ratios, h is the Planck constant, μ_0 the permittivity of free space, I_{iz} and I_{jz} are angular momentum spin operators, and α_{ij} is the angle between the inter-nuclear vector of the two spins and the static magnetic field.

Note that the dipolar Hamiltonian $H_{ij}^{D}(t)$ depends on the orientation defined by the angle α_{ij}. The measurement of dipolar couplings D_{ij} represents a time and ensemble averaging of $H_{ij}^{D}(t)$ over all sampled orientations. This averaging is denoted by angular brackets in the following equation

$$D_{ij} = -\frac{\gamma_i\,\gamma_j\,\mu_0\,h}{8\,\pi^3}\left\langle\frac{3\,cos^2\alpha_{ij}(t)-1}{2\,r_{ij}^3(t)}\right\rangle. \tag{2.17}$$

For isotropically tumbling molecules D_{ij} would be zero, but a non-zero value is obtained when there is an anisotropic distribution of orientations relative to the static magnetic field. The averaging in eq. 2.17 contains information about orientation of the B_0 field and the inter-nuclear vector in the molecular frame (xyz-coordinates). For macromolecules like DNA, it is desirable to describe the orientation of the inter-nuclear vector in relation to the molecular frame rather than the magnetic field, since the gathering of information to support structure determination is the motivation for the RDC experiment. The time averaged α_{ij} are therefore written as convolution of the macromolecule tumbling with respect to the magnetic field vector (ξ_x, ξ_y, ξ_z) and the inter-nuclear vector moving inside the macromolecular frame $(\zeta_x, \zeta_y, \zeta_z)$. The convolution is illustrated graphically in Fig. 2.6 and mathematically as follows[225]

$$cos\,\alpha_{ij} = \begin{pmatrix} cos\,\xi_x \\ cos\,\xi_y \\ cos\,\xi_y \end{pmatrix} \begin{pmatrix} cos\,\zeta_x \\ cos\,\zeta_y \\ cos\,\zeta_y \end{pmatrix} = \sum_{k=}^{x,y,z} cos\,\xi_k\,cos\,\zeta_k. \tag{2.18}$$

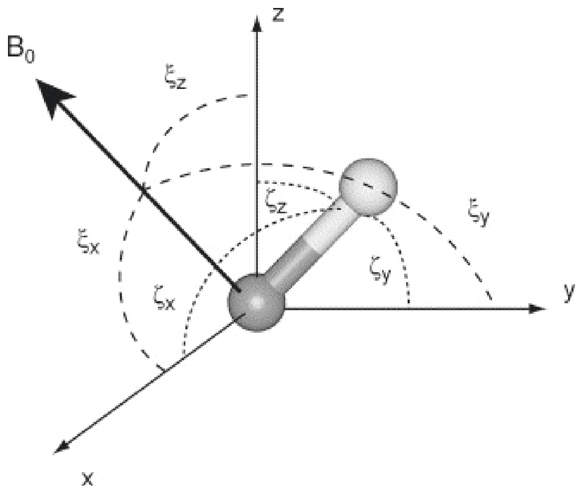


Figure 2.6: Orientation of the B_0 field and the inter-nuclear vector in the molecular frame. The angles ξ_x, ξ_y and ξ_z represent the orientation of the macromolecular frame (xyz-coordinates) relative to the magnetic field, while the orientation of inter-nuclear vector is defined by the angles ζ_x, ζ_y and ζ_z relative to the frame. Picture taken from Blackledge et al.[225]

The inter-nuclear vector between i and j is assumed to be rigid within the macromolecular frame. As consequence, the averaging of a_{ij} only acts on the orientation of the frame (ξ-angles) relative to the static magnetic field B_0. With this in mind, the preferential orientational averaging of the molecule can be defined as alignment tensor **A** whose units are dimensionless

$$A_{kl} = \frac{3}{2}\left\langle cos\, \xi_k \, cos\, \xi_l \right\rangle - \frac{1}{2}\delta_{kl}. \tag{2.19}$$

At this point, an effective inter-nuclear distance $r_{ij,eff}$ will be introduced to account for the averaging of r_{ij}[225]. With eq. 2.19 and $r_{ij,eff}$ at hand eq. 2.17 can be rewritten as

$$D_{ij} = -\frac{\gamma_i \, \gamma_j \, \mu_0 \, h \, S_{axial}}{8\, \pi^3 \, r^3_{ij,eff}} \sum_{k,l=}^{x,y,z} A_{kl} \, cos\, \zeta_k \, cos\, \zeta_l. \tag{2.20}$$

The scaling factor S_{axial} accounts for the local flexibility of the inter-nuclear vector. It is based on a model for axially symmetric motion that is called “diffusion in a cone” and depends on the amplitude of the motion but not on the position of the vector with respect to the alignment tensor[225]. In this model, the order parameter S_{axial} is related to the

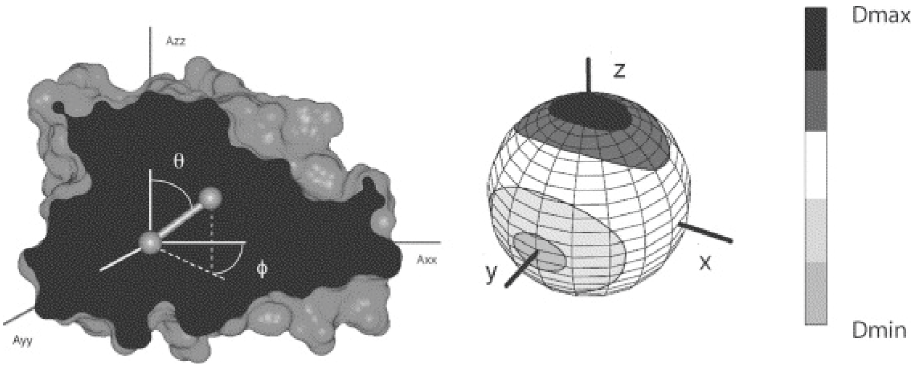


Figure 2.7: Dependence of RDC values on the orientation of the inter-nuclear vector (θ , ϕ) in the eigenframe of the alignment tensor with eigenvalues A_{xx}, A_{yy} and A_{zz} (left panel). The orientational degeneracy of RDCs is shown on the right. The surface of the sphere is shaded as function of equal couplings. In other words, a solely measured RDC can cover a whole range of orientations indicated by a single color. At least 5 RDCs are necessary to calculate a distinct orientation of a vector inside the molecule. Picture taken from Blackledge et al.[225]

generalized order parameter S, which scales down the measured RDCs linearly[242] and is again related to an effective correlation time[218].

In practice, the alignment tensor **A** has all elements non-zero. It would be preferable to find a specific molecular frame, the so called principal axis system (PAS), in which all off-diagonal elements of **A** are zero and only the diagonal terms A_{xx}, A_{yy} and A_{zz} remain. A three-dimensional Euler rotation of the current molecular frame with parameters α, β and γ[225] can be used to transform eq. 2.20 into the following equation, where the orientation of the inter-nuclear vector is defined by the polar angles θ and ϕ in the eigenframe of the alignment tensor:

$$D_{ij} = -\frac{\gamma_i \, \gamma_j \, \mu_0 \, h \, S_{axial}}{16 \, \pi^3 \, (r_{ij}^{eff})^3} \left[A_a \, (3 \, cos^2\theta - 1) + A_r \, sin^2\theta \, cos \, 2\phi\right] \tag{2.21}$$

By convention is $|A_{xx}| \leq |A_{yy}| \leq |A_{zz}|$[225]. The axial A_a and rhombic component A_r of the alignment tensor **A** are defined in relation to the eigenvalues A_{xx}, A_{yy} and A_{zz}

$$A_a = \frac{1}{2} A_{zz} \quad and \quad A_r = \frac{1}{3}(A_{xx} - A_{yy}) \quad . \tag{2.22}$$

It is easy to see from eq. 2.21 and eq. 2.22 that five parameters determine the orientation of any structure or sub-structure of interest, these are the eigenvalues A_{xx}, A_{yy} and A_{zz} of the alignment tensor $\mathbf{A}$ and the polar angles θ and ϕ. The necessary parameters can be determined directly via singular value decomposition[243] after the measurement of at least five RDCs.

Unfortunately there are number of orientations for a single inter-nuclear vector that are compatible with a solely measured RDC. The right panel in Fig. 2.7 illustrates the strong angular degeneracy as shades on a spherical surface, where only extreme values give nearly unambiguous orientations. The more common intermediate values lead to a large number of potential solutions for the orientation, thus limiting the value of RDCs in structure determination. But the angular degeneracy can be lifted either by measuring more couplings in structures of known conformation, or by measuring RDCs in the presence of liquid crystals that orient the molecule differently[225,242].

2.5 Simulated Annealing calculations

Simulated Annealing (SA) is an algorithm to overcome the problem hat the energy function in a Molecular Dynamics (MD) simulation converges to a local instead of the global minimum (see FIg. 2.8). The basic idea is to allow the molecule to leave a found local minimum by providing sufficient kinetic energy. The average kinetic energy for a given temperature can be calculated via Boltzmann statistics

$$< E_{kin,i} >=< \frac{1}{2} m_i \, \nu_i^2 >= \frac{3}{2} k_b \, T, \tag{2.23}$$

where k_b is the Boltzmann factor, m_i the atom mass and ν_i the atom velocity. The amount of kinetic energy $E_{kin,i}$, that is necessary to overcome kinetic barriers and allows access to the global minimum, can be provided as high temperature at the beginning of the simulation. In order to achieve temperature coupling of the MD (to a target temperature T_0), a friction coefficient $b_i = \beta_i(T_0/T - 1)$ is added[244] to the Newtonian equation of motion:

$$F_i(t) = m_i \frac{\partial^2 r_i}{\partial t^2} = -\frac{\partial V_i}{\partial r_i} + \beta_i(T_0/T - 1)\nu_i. \tag{2.24}$$

The acceleration $\frac{\partial^2 r_i}{\partial t^2}$ of each atom i at time t is related to the derivative of the potential energy V_i with respect to the atom position r_i. The force $F_i(t)$ at position r_i can act on the atoms for a given time-step (typically between 1-5 fs). A verlet algorithm determines then a new set of coordinates and velocities from the last and current values. This cycle is repeated until a convergence criterion (e.g. a minimum change in the gradient of the potential energy) is met. Initially, atom velocities are computed using a Gaussian or Maxwell distribution. The atom coordinates are derived from a starting structure. Since the initial coordinates and velocities determine all subsequent ones, it is important to start from a reasonable structure. In this work, where unknown structures of modified oligonucleotides are probed, a starting structure is obtained via MD hybridization of extended strands to a reasonable double strand.

The number of cartesian coordinates, that has to be calculated during each step of

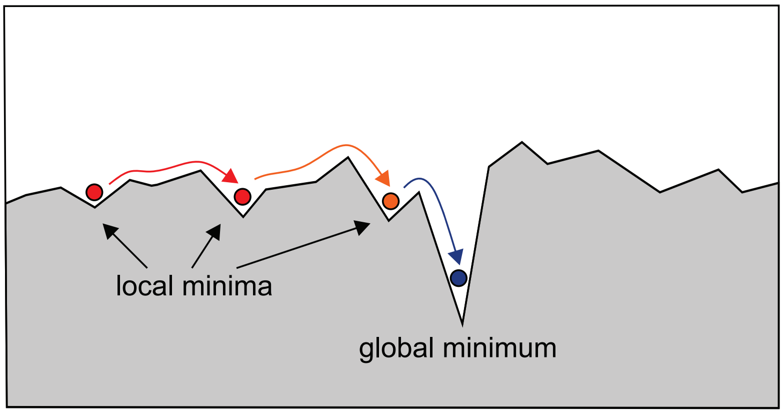


Figure 2.8: Find global minimum with Simulated Annealing. The red dot marks a molecule at high temperature. It has sufficient kinetic energy to reach the next local minimum on the energy surface (red arrow). Constant cooling is applied, but as a long as the temperature is high enough, the molecule can "jump" to the next minimum (orange arrow). Finally in the cold state (at room temperature), the molecule is expected to be trapped in the global minimum.

the Molecular Dynamics simulation, is usually three times larger than the atom counted in the molecule. Especially in macromolecules like DNA or proteins, this number easily reaches the order of thousands. Nevertheless it is possible to simulate macromolecules via Molecular Dynamics by predefinition of atom types. For these predefined atoms many parameters such as bond lengths, bond angles, dihedral angles, partial charges etc. are assumed to be fixed and are comprised in the force field. In the present work the program Xplor-NIH[245] was utilized which employs the CHARMM force field[246,247]. The total potential energy V_{tot} consists of two components[244]

$$V_{tot} = E_{emp} + E_{eff}, \tag{2.25}$$

where the empirical (E_{emp}) and the effective energy term (E_{eff}) are given as[244]

$$E_{emp} = E_{bond} + E_{angle} + E_{dihe} + E_{vdW} + E_{Coulomb}, \tag{2.26}$$

$$E_{eff} = E_{noe} + E_{rdc} + E_{plan} + E_{cdih}. \tag{2.27}$$

E_{bond}, E_{angle}, E_{dihe} and all energy terms of E_{eff} are calculated as the product of a force constant and the deviation of the observed value from the equilibrium one, e.g.

$$E_{bond} = k_{bond}\,(r_{ij}^{obs} - r_{ij}^{equ}) \tag{2.28}$$

The equilibrium values for E_{noe} and E_{rdc} are taken from experiment while these of E_{cdih} and E_{plan} are averages from the literature[244]. The corresponding scaling factor for each term are defined in the calculation input and thus can be used to increase the restraining power of selected energy terms. Equilibrium values and force constants of E_{bond}, E_{angle}, E_{dihe} constitute one part of the force field. E_{vdW} is given as

$$E_{vdW} = \sum_{i,j}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) \tag{2.29}$$

with $A_{ij} = 2\sqrt{\varepsilon_{ii}\,\varepsilon_{jj}}(\sigma_{ii} - \sigma_{jj})$ and $A_{ij} = 2\sqrt{\varepsilon_{ii}\,\varepsilon_{jj}}(\sigma_{ii} - \sigma_{jj})$. The terms for atomic permittivity ($\varepsilon_{ii/jj}$) and van-der-Waals radii ($r_{ii/jj}$) set up another part of the force field. The partial atomic charges (q_i, q_j) in $E_{Coulomb}$ constitute the last part of the force field.

$$E_{Coulomb} = \sum_{ij}\frac{q_i\,q_j}{\varepsilon_0\,r_{ij}} \tag{2.30}$$

Explicit treatment of water is not feasible due to restrictions on the calculation time, so the solvent screening effect is approximated by introducing a distance dependent permittivity of free space $\varepsilon_0(r_{ij})$.

Finally, force field parameters for the native bases and their 2'-deoxyribose backbone are derived from crystal structures, infrared spectroscopy data (force constants) and empirical testing (where no experimental source is available)[246]. Every non-native modification to our basic DNA sequence has to be added manually to the existing force field. The necessary parameters for bond lengths, bond angles, dihedral angles and in particular partial atomic charges are calculated using ab-initio methods.

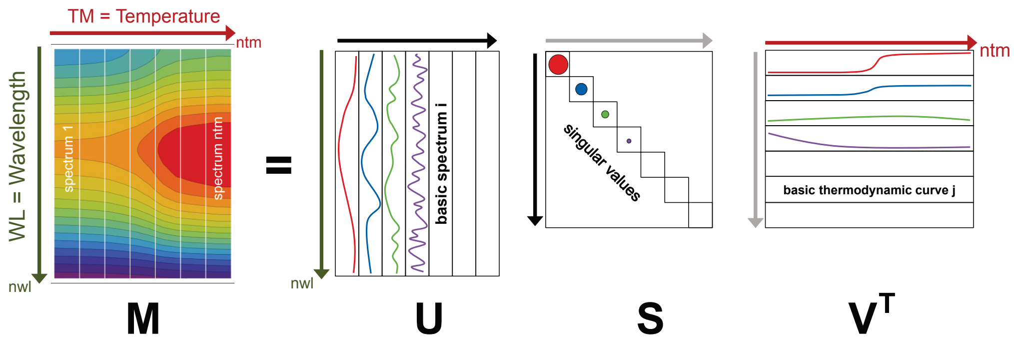


Figure 2.9: Data matrix **M** and Singular Value Decomposition (SVD) of DNA absorption spectra. The result is a product of three matrices named **U**, **S** and $\mathbf{V}^T$.

2.6 Singular Value Decomposition

The last two sections enter the field of optical spectroscopy, whereby the hybridization of DNA will be monitored. The underlying concepts are developed here in sections 2.6 and 2.7.

The Singular Value Decomposition (SVD), also referred to as Principal Component Analysis (PCA), decomposes a two-dimensional data matrix **M** into the product of three matrices named **U**, **S** and $\mathbf{V}^T$ (see Fig. 2.9). The SVD method serves here as a tool to analyse the absorption spectra of the modified oligonucleotides. Therefore, the absorption spectra will be written as columns to build up the data matrix **M** with dimensions nwl x ntm. Here nwl is the number of wavelengths which are stored on an array WL, and ntm is the number of temperatures which are stored on an array TM.

After decomposition the columns in **U** store a set of orthonormal "basic spectra" (depending on wavelength λ), while the rows of $\mathbf{V}^T$ describe "basic thermodynamic curves" of populations or concentrations as a function of temperature. Since **U** and $\mathbf{V}^T$ only consist of orthonormal functions, a list of factors is needed to fully describe our data in **M**. The last matrix **S** stores these factors on its diagonal axis, connecting the i-th column of **U** with the i-th row of $\mathbf{V}^T$, and furthermore sorts them by importance or weight, so the first one is the most important. They represent the singular values, but not all of them are necessary to describe the data in **M**. If not mentioned otherwise, the ns singular values

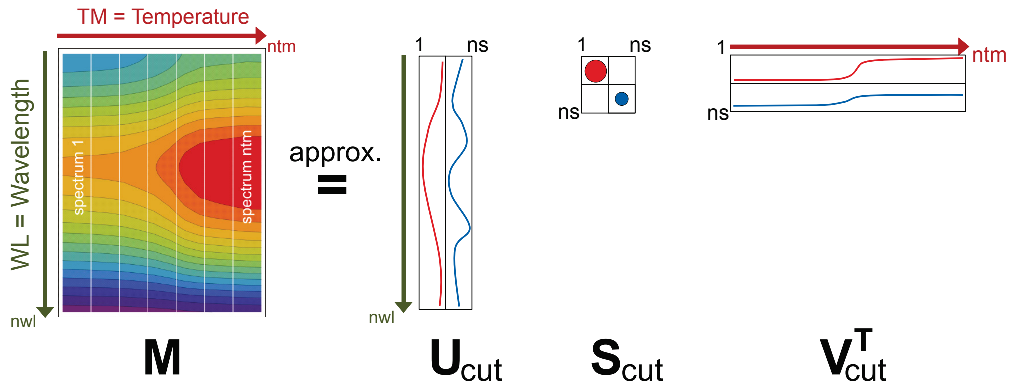


Figure 2.10: A Reduced set of matrices can fully describe the data matrix $\mathbf{M}$. For example, when only two singular values ($ns = 2$) of $\mathbf{S}$ are necessary to describe $\mathbf{M}$, one can build matrices with ns columns in $\mathbf{U}_{cut}$ and ns rows in $\mathbf{V}_{cut}^T$, respectively. In addition, the product of the new matrices should produce a dataset with better signal to noise ratio than $\mathbf{M}$.

higher than the 100th part of the first one are used to construct a cut set of matrices with ns columns in $\mathbf{U}_{cut}$ and ns rows in $\mathbf{V}_{cut}^T$, respectively(see Fig. 2.10). It is assumed that the omitted singular values only contribute to the noise in the spectra, so the product of the new matrices should have a better signal to noise ratio than the original dataset in $\mathbf{M}$.

At this point the number of independent spectra contained in the data is known. The chemical species which cause them must be ns or larger, because such independent spectrum may, accidentally, be a linear combination of two species. The matrix $\mathbf{V}_{cut}^T$ is then separated into a product of linear factors $\mathbf{F}$ and nonlinear parameters $\mathbf{P}$ (see Fig. 2.11). This allows scaling on purpose, for example, if the last column of P is scaled to be 1 in in the first row and 0 for the second one, one would know, that (for highest temperature) only the first basic spectrum is necessary to describe the last spectrum in the data matrix $\mathbf{M}$ (in figure 2.10). In addition, the resulting $\mathbf{F}$ can be multiplied with $\mathbf{U}_{cut}.\mathbf{S}_{cut}$ to give the product $\mathbf{U}_{cut}.\mathbf{S}_{cut}.\mathbf{F}$. In consequence of the rescaled $\mathbf{P}$, the first spectrum of $\mathbf{U}_{cut}.\mathbf{S}_{cut}.\mathbf{F}$ is equal to the last (high temperature) spectrum in $\mathbf{M}$. Also important, $\mathbf{P}$ (in combination with $\mathbf{U}_{cut}.\mathbf{S}_{cut}.\mathbf{F}$) can be inspected to assign species and their thermodynamic behaviour when changing temperature. For example, linear line shapes can be stacking interactions or peak shifts and sigmoidal line shapes denote melting curves. The

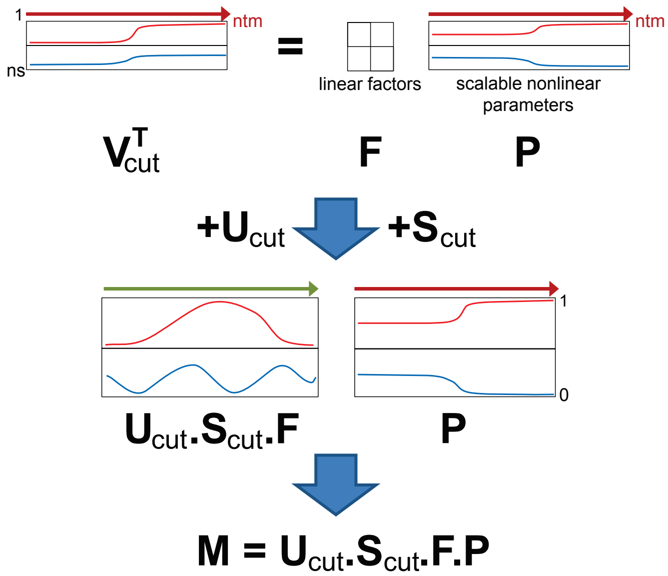


Figure 2.11: Separating $\mathbf{V}^T_{cut}$ into a product of linear factors $\mathbf{F}$ and nonlinear parameters $\mathbf{P}$.This allows scaling on purpose, but more important, $\mathbf{P}$ can be inspected to assign species and their thermodynamic behaviour when changing temperature. A possible way to present the dissection of $\mathbf{M}$ via SVD is shown in the middle of the figure by the two pictures of $\mathbf{U}_{cut}.\mathbf{S}_{cut}.\mathbf{F}$ and $\mathbf{P}$, respectively. Please note, the product $\mathbf{U}_{cut}.\mathbf{S}_{cut}.\mathbf{F}.\mathbf{P}$ still reproduces the data stored in $\mathbf{M}$.

reader should note that $\mathbf{V}^T_{cut}$ can contain more than one melting curve, which may show different melting characteristics, and thus indicates a non-uniform melting process of the DNA double strand. The next step is to find the parameters which describe the thermodynamic functions found in $\mathbf{P}$ and the expansion factors in $\mathbf{F}$. How to do this is topic of the following chapter 2.7, while the SVD, as mentioned in the beginning, has served as tool to access the thermodynamic behaviour like melting from a set of full absorption spectra.

2.7 Development of the double SVD assisted two-state model

The thermodynamics of duplex formation is usually examined by melting curves, which are measured at a single wavelength over different temperatures. However, such melting curves are directly affected by the errors of the measurement at the given wavelength [248]. To overcome this, one could think of repeated experiments or analysis of additional wavelengths nearby. In this work a more advanced approach is used to obtain precise results, especially when the complexity of the chromophore absorption change increases.

The basic idea is to measure full absorption spectra at different temperatures. The fact that such method needs considerable more time to record the desired spectra has two side effects. One is the larger step size between temperatures to limit the acquisition time of the whole experiment, the other is the avoidance of hysteresis or temperature gradients in the sample, since the temperature remains constant during the record of each spectrum.

Our recorded spectra are then combined into the two-dimensional data matrix $\mathbf{M}$ (see Fig. 2.9). By Singular Value Decomposition (SVD, see 2.6), the data matrix $\mathbf{M}$ is decomposed into a product of three matrices $\mathbf{U.S.V}^T$ containing "basic spectra", "singular values" and "thermodynamic curves". Usually, only the first 2-4 values (columns in $\mathbf{U}$, rows in $\mathbf{V}$)are necessary to describe the data, so a new product $\mathbf{U}_{cut}.\mathbf{S}_{cut}.\mathbf{V}_{cut}^T$ is made which results in a smoothed (noise reduced) $\mathbf{M}$. The desired melting curves of the full absorption spectra are stored in $\mathbf{V}_{cut}^T$. From here onwards, each row can be analyzed like curves obtained by single wavelength measurement.

In figure 2.12 a melting curve of 13mer4AP-DAP is shown (red line, obtained via SVD, see 4.3.4). A common way to check helix stability is the estimation of the melting point. It is, by definition, the temperature where one half of the strands is in duplex state and the other half in single strand state. Applied on the melting curve which is scaled to the amplitude change over temperature, one would note from the figure at half height 61 ℃, but it is easy to see that such reading depends on the measured temperature range, due to the slope of the linear parts of the melting curve [248]. Another common method [111,138,139,249], plotted as blue line in figure 2.12, uses the first derivative of the melting curve, thus the

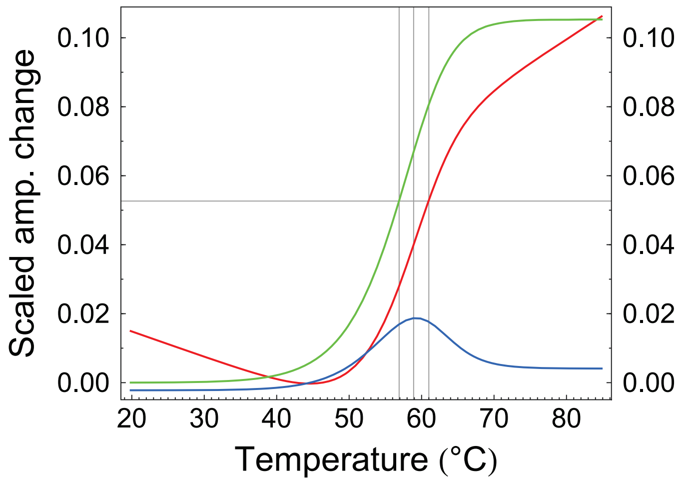


Figure 2.12: Melting Point Estimation from a double strand DNA melting curve is a common way to check DNA stability. The melting curve (red), its first derivative (blue) and the fitted two-state model (green) without assumed linear stacking interactions are shown and compared against each other. The estimated melting points are indicated by grey vertical lines, depending on the method and slope of the linear parts, three different melting points will be estimated.

melting point is indicated by its maximum at 58.9 °C, but this method is also affected by the linear parts of the melting curve and therefore is not reliable.

The most widely used method is a two-state model[250,251] with linear stacking interactions[19,252,253]. It assumes that the temperature (T) dependence of the extinction coefficients is linear ($\epsilon_{[ss]} = m_{[ss]}T + b_{[ss]}$) and, in addition, different for the single- (ss) and double-stranded (ds) forms. The two-state part defines the equilibrium constant K for hybridization of nonself-complementary DNA depending on c_T and α (molar fraction of single stranded form to the total ss strand concentration):

$$K = \frac{[ds]}{[ss_1][ss_2]} = \frac{2(1-\alpha)}{\alpha^2 c_T}. \tag{2.31}$$

After substituting K into $\Delta G^\circ = -RTln(K) = \Delta H^\circ - T\Delta S^\circ$ and rearranging, we obtain an equation for α depending on ΔH° and ΔS°.

$$\alpha = \frac{-1 + \sqrt{2\, c_T\, exp\left[-\frac{\Delta H^{\circ} - T\Delta S^{\circ}}{RT}\right] + 1}}{c_T\, exp\left[-\frac{\Delta H^{\circ} - T\Delta S^{\circ}}{RT}\right]}. \tag{2.32}$$

Equation (2.32) is then combined with the linear temperature dependence of the extinction coefficients, leading to an equation for the measured amplitude change $A(\alpha, T)$:

$$A(\alpha, T) = \alpha(m_{[ss]}T + b_{[ss]}) + (1 - \alpha)(m_{[ds]}T + b_{[ds]}). \tag{2.33}$$

After least-squares fitting with equation 2.33, the resulting $\alpha(T)$, which only describes the two-state model, can be plotted as green line in figure 2.12. Its melting point can be estimated to 58.9 °C. Unfortunately, this is the third melting point for only one measured melting curve, but the last one is not affected by the slopes of the assumed stacking interactions. Furthermore, if the slopes for the single strand state and the duplex state are either similar or small, then the melting points of the red and blue curves would converge towards that one of the two-state model (green line), which is so far the most reliable.

However, a two-state model, wherein only fully separated strands and the duplex exist, is an oversimplification which cannot cover all types of DNA/RNA double strands. Statistical models like the zipper model[19,254] or the extended type with bulge formation introduced by Ernsting[203] account for a more complex way of hybridization, but this means to assume a specific model for a given double strand, which may end in a different model for every modified oligonucleotide analyzed in this work. Also the number of parameters increases with the complexity of the model and consequently the number of possible solutions in the least-squares fit of a simple melting curve, which explains the relatively rare application of statistical models[19]. The last method, which should be mentioned here, is the widely used nearest-neighbor method by SantaLucia[255,256]. It focuses on the interactions between neighboring base pairs, but is based on experimental data of natural nucleobases and therefore not suitable for oligonucleotides with noncanonical modifications to the strands.

The new approach will use the basic idea of the two-state model in combination with the analysis of the spectral change via SVD. Remember, that full absorption spectra were measured covering the DNA band as well as the absorption of the built-in chromophore. The before mentioned two-state model (with linear stacking interactions) assumes temperature dependent extinction coefficients for the single strands and the duplex state at a given wavelength. The expansion to a full spectrum can then take additional effects into account like peak shifting, local melting or the rise of a new species. The way to do this is splitting the SVD into the analysis of duplex state at temperatures below the melting point and the analysis of the single strands at higher temperatures. Knowing the change of absorption with temperature for both, an extrapolation can be done from either side into the middle, i.e. in the direction where strand separation occurs. For example, the absorption spectrum of the double strand can be simulated for 70 °C where, in reality, the separation into single strands is almost complete. At any given temperature, the observed spectrum should be a linear combination of the two extrapolated spectra, i.e. of the separate strands $S_{[ss]}(T)$ and of the duplex $S_{[ds]}(T)$. The mathematical description of this linear combination is shown in equation 2.34, where α is the fraction of extrapolated single strand spectra $S_{[ss]}(T)$, which is used to build the measured spectra $M(\alpha, T)$. In other words, the linear equation terms of (2.33) are substituted with the extrapolated spectra $S_{[\dots]}(T)$.

$$M(\alpha, T) = \alpha S_{[ss]}(T) + (1 - \alpha) S_{[ds]}(T). \tag{2.34}$$

From the corresponding fit the degree of dissociation α is obtained. As result we obtain a dissociation curve (shown in Fig. 2.13) following in any case the characteristics of a two-state model, since it was constructed from a set of two spectra (measured or extrapolated) at a given temperature. The advantage of this approach is, as long as there are spectral changes dominated by either the duplex state or the single strands (which would allow a separate analysis via SVD), one can reduce the analysis of any complex melting process to correspond to the easy pattern of the two-state model. Additionally, the SVD parts can be investigated for the type of spectral change (e.g. amplitude change, peak shift). In

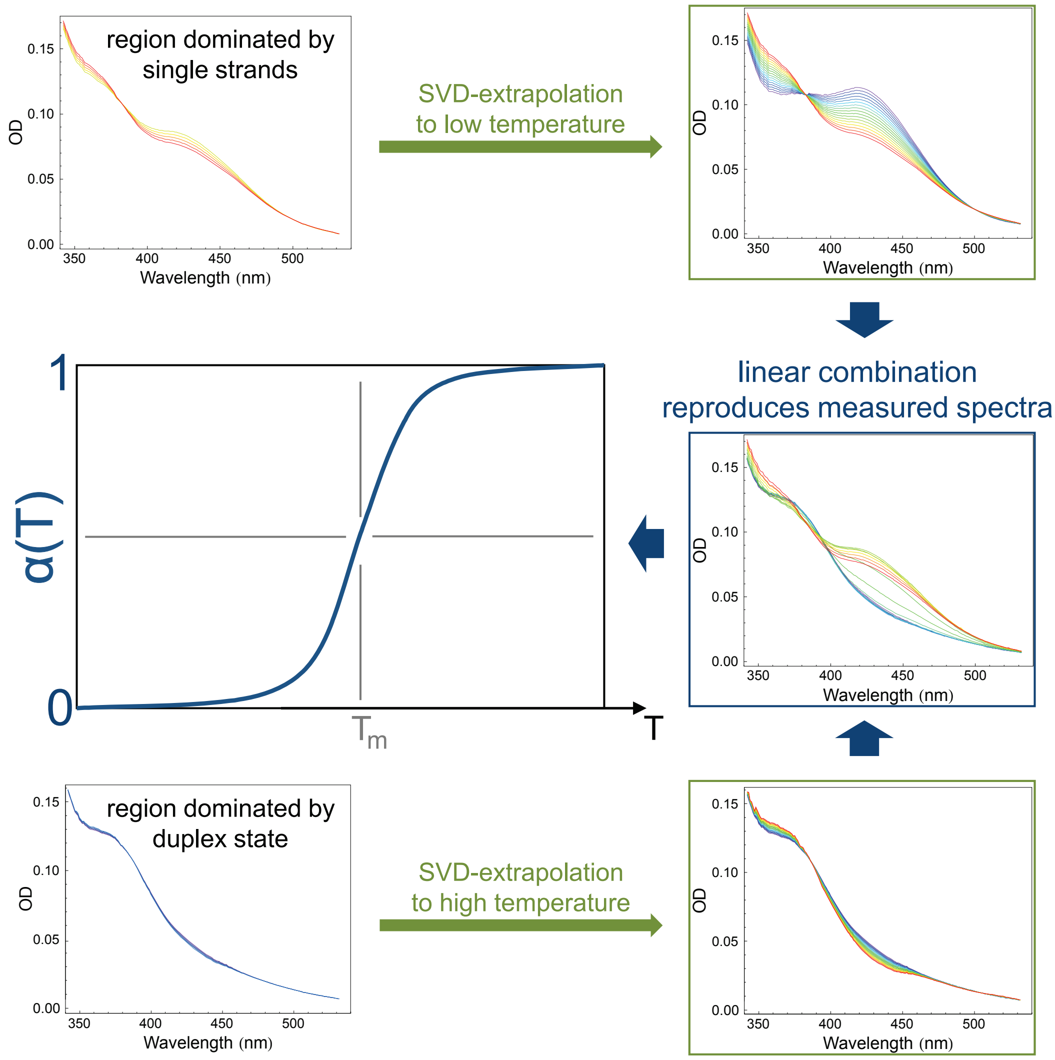


Figure 2.13: SVD assisted two-state model shown as schematic diagram. SVD analysis is performed on regions dominated by either separated strands or duplex state and then extrapolated to the missing temperature region. By fitting the degree of dissociation $\alpha(T)$ to the linear combination of the extrapolated spectra, a melting curve is obtained on which the two-state model can be applied.

the end, one can gain more knowledge about a modified double strand than from a single wavelength melting curve. A detailed demonstration of the method was performed on the 6-Hydroxy-quinolinium chromophore (13mer6HQ) in chapter 4.2.3.

3 Experimental section

All methods and preparations will be outlined in detail for 13merHCF. In order to avoid redundancy, the corresponding sections of 13mer6HQ and 13mer4AP-DAP will only comprise experimental details that differ from 13merHCF.

3.1 13merHCF

3.1.1 NMR sample preparation

2-Hydroxy-7-carboxyfluorene was synthesized by Matthias Pfaffe in the same way as HNF [203]. The 2'-deoxyriboside of HCF was prepared by reaction of 2-hydroxyfluorene-7-carboxylic acid methyl ester with 1'-α-Chloro-3',5'-di-O-toluoyl-2'-deoxy-D-ribose in the presence of activated molecular sieve. The corresponding phosphoramidite was reached in three steps by standard methods. The predominantly formed α-glycoside was then purified by column chromatography. Fixed-phase synthesis of the labeled strand was performed at BioTeZ (Berlin) with a small modification, since the coupling of HCF-2'-deoxyriboside required a fourfold increase over the normal reaction time. The reader should note that instead of the originally intended methyl ester, the free acid was incorporated into the strand, due to the preparation conditions in the last step of the fixed-phase synthesis. The HCF labeled strand and the abasic counterstrand, also from BioTeZ, were delivered already purified by reverse-phase high-pressure liquid chromatography (HPLC). After hybridization they were subjected to size exclusion chromatography (in Sephadex PD-10 column) and lyophilization with 3 % NH_3-solution to remove residual, low molecular weight impurities (mainly NEt_3-buffer from HPLC). Equivalent amounts of complementary single strands were hybridized by rapid heating to 90 ℃ and subsequent gradual cooling to room temperature at a rate of 0.5 ℃ per minute. The NMR samples were prepared in (D_2O matched) Shigemi tubes at 5 mM duplex concentration in D_2O (D_2O 99.98 %) and H_2O (H_2O:D_2O/90:10) buffer solutions at pH 7, containing 10 mM Na_2HPO_4/NaH_2PO_4 and 150 mM NaCl.

3.1.2 RDC sample preparation

The samples for the residual dipolar coupling (RDC) experiment were prepared in D_2O buffer as described before. The addition of 20 mg/ml Pf1 (obtained from Asla Biotech Ltd., Riga) requires exchange of the Pf1 buffer, since it is obtained in a non-deuterated solution. The exchange is achieved by ultracentrifuging 100 µl of Pf1 two times with 600 µl deuterated phosphate buffer at 60000 rpm for 2 hours (at 4 ℃). Afterwards the Pf1 sediment is joined with the DNA sample. The high viscosity of Pf1 complicates sample handling and thus the suspension has to be stirred until a viscose, clear, gel-like sample is obtained. After transfer into the Shigemi tube, bubbles have to be removed by slow centrifugation of the NMR tube (up to 500 rpm). The degree of orientation can be checked by measuring the quadrupole splitting of deuterium[257], which is expected to show a symmetric doublet with splitting in the range of 5 to 15 Hz. In case of degradation or non-complete suspension of Pf1, this peak doublet can be asymmetric, extremely broadened or even non-observable.

3.1.3 Duplex melting experiments

The UV/vis absorption experiments were performed on a Varian Cary 300 spectrometer in double beam mode. As stated in the introduction, full spectra (between 210 and 400 nm) were measured with 1nm step size, 0.6 s average time, 2 nm bandwidth and source changeover at 400 nm. Between each rising temperature step of 5 ℃, a time delay of 10 minutes was included to heat and equilibrate the sample with the heating block. A sample of double-stranded 13merHCF was prepared in a water/phosphate buffer at pH 7 with 10 mM NaH_2PO_4 and 150 mM sodium chloride. Measurements were performed in a closed non-degassed cuvette with 2 mm optical path length. A total concentration $c_T = 45$ µM of single strands was estimated from maximum absorption around 260 nm at 90 °C. 14 absorption spectra between 25 and 90 °C were recorded and corrected for density change.

3.1.4 Titration against pH experiments

The Varian Cary 300 spectrometer was used with the same settings as mentioned before. A sample containing 150 mM sodium chloride and 20 µM (c_T) of double-stranded 13merHCF was prepared at pH = 1.15 (through addition of 1 M HCl). A set of 29 spectra were measured up to pH = 12.07 by adding increasing concentrations of NaOH (0.01 - 1 M) into a cuvette with 10 mm optical path length.

3.1.5 NMR experiments

All NMR experiments were carried out on a Bruker Avance 600 MHz spectrometer with inverse probehead. Dallmann and the author determined in an earlier work[129] that 298 K is the most suitable temperature for monitoring the imino proton signal intensity of duplex strands, and this temperature was therefore chosen as standard for all types of NMR experiments. For the 13merHCF duplex, dissolved in the aforementioned D_2O buffer, a DQF-COSY- (Double Quantum Filtered Correlated Spectroscopy), a TOCSY- (Total Correlation Spectroscopy) and a NOESY-spectrum were measured. For DQF-COSY and TOCSY, 2048 x 512 points (F2 x F1 dimension) were acquired with 16 and 56 transients, respectively. The more important NOESY-spectrum employs 4096 x 2048 points with 16 transients and 100 ms mixing time. The NOESY in H_2O was recorded using the same parameters in a WATERGATE pulse-sequence to suppress the HOD signal. The RDC experiment requires two HMQC-spectra (Heteronuclear Multiple Quantum Coherence) in D_2O before Pf1 is added, one ^{1}H decoupled spectrum for assignment and a coupled spectrum as basis to calculate the coupling. After addition of Pf1, the sample became highly viscous and the standard shim procedure failed to produce reasonable linewidths. As a workaround, 1D spectra were measured after each shimming step in order to iteratively optimize the lineshape of the HOD signal. The alternative worked well, but is much more time-consuming than the standard method. The two ^{1}H coupled HMQC- spectra with and without Pf1 were acquired with 8192 x 512 points and 192 transients. The high number of points in the F2 dimension is necessary to obtain RDC-values with a precision below 1 Hz. All spectra were processed with the Bruker TopSpin software.

3.1.6 Force field parametrization for HCF

Density functional theory (DFT) calculations of the HCF moiety were performed in Gaussian03 using the b3lyp method and triple zeta valence plus polarization (TZVP) as basis set. Partial charges were derived with the of ChelpG-algorithm by Breneman et al.[258]. Although the sugar moiety was simulated as well, only the partial charges for the HCF residue were integrated into the force field. A comparison between the neutral HCF (with COOH) and the negatively charged (with COO^-) confirmed that the effect on the partial charges of the sugar is negligible, so the same parameter set as for the sugar of the native nucleobases could be used.

3.1.7 Distance restraints

The assigned NOE cross-peaks were converted to distance restraints by referencing their integrals to the integrals of known distances employing the Isolated Spin Pair Approximation (ISPA). The NOE cross-peaks were integrated with the program Cara[259] using the sum-over-rectangle method. As reference distances Methyl-H6 T (3.09 Å) for all NOE cross-peaks involving methyl protons, H42-H5 C (2.4 Å) for all NOE cross-peaks involving exchangeable protons and H5-H6 C (2.48 Å) for the remaining NOE cross-peaks were used (bond lengths adapted from the force field parameters). For the purpose of exporting the integral values obtained by Cara[259] to an Xplor-NIH[245] restraints file, a LUA script is used that was written by Dallmann[260] (see section 3.1). This script classifies the integrated peaks according to the overlap with other peaks and scales their volume integrals accordingly. Additionally, uncertainties for the NOE restraints are automatically calculated from the standard deviation of the reference peaks' volume integrals. The estimated uncertainty is then increased according to the classification of each peak. Furthermore, this classification is printed into a separate file, which can be used to assess whether or not peak overlap might prevent a reliable estimation of the peak volume.

3.1.8 Residual Dipolar Coupling restraints

The RDC values of 13merHCF are listed in table 3.1. As outlined in section 2.4, the difference in the coupling constant of C-H bond vectors in the presence and absence of Pf1 phage was measured. The orientation of the corresponding inter-nuclear vector was defined in eq. 2.21 in section 2.4. All occurring constants can be joined in the factor D_a, which can be written as

$$D_a = -\frac{\gamma_i\,\gamma_j\,\mu_0\,h\,S_{flex}}{16\,\pi^3\,(r_{ij}^{eff})^3}. \tag{3.1}$$

Only one value of D_a may be used throughout the calculation. RDC values of different vectors like C-C bonds need to be scaled to C-H values. Scaling is achieved by introducing a prefactor to the D_a-term, which is simply defined as the ratio of the two D_a involved. An example for C-C RDCs is given in the next equation

$$D_a^{pre}(CC) = \frac{D_a(CH)}{D_a(CC)} = \frac{\gamma_H}{\gamma_C}\left(\frac{r_{CC}}{r_{CH}}\right)^3 = \frac{42.576}{10.705}\left(\frac{1.496}{1.090}\right)^3 \approx 10.28. \tag{3.2}$$

where $\frac{\gamma_H}{\gamma_C}$ is the change in gyromagnetic ratio between C-C and C-H, while r_{CC} and r_{CH} are the lengths of the corresponding inter-nuclear vector. Although only C-H RDCs were measured in this work, it was necessary to implement the experimentally determined C-H RDCs of the T methyl groups as C-C RDCs. Due to the fast rotation of the methyl group, only a single averaged value can be measured for the three C-H bond vectors, but it can be scaled as follows to a single C-C bond vector

$$P_2(cos\beta) = \frac{3}{2}\,cos^2\beta - \frac{1}{2} \tag{3.3}$$

where β is the C5-C7-H7[1-3] angle. It has been revealed by Ottiger and Bax[261] that the usually assumed ideal tetrahedral angle of 109.5° for methyl groups has to be replaced with experimentally determined 110.9°. With eq. 3.2 and eq. 3.3 the scaling of methyl

C-H to methyl C-C can be written as

$$\frac{D_{CH\_Me}}{D_{CC\_Me}} = P_2(cos\beta)\, D_a^{pre}(CC) \quad . \tag{3.4}$$

A value of -3.17 was determined by measuring the correlation of experimentally determined C-H and C-C methyl RDCs[261]. For the ideal tetrahedral angle one would obtain a factor of -3.42. In order to implement the methyl C-H RDCs into the structure calculations, they have to be converted into the corresponding C5-C7 RDCs. This is done in a second input file, where all experimentally determined methyl C-H RDC values were converted by hand to the corresponding C-C values with the factor 1 / -3.17 = -0.3155. In that case, the prefactor D_a^{pre}(CC) can be used to scale the C-C RDC input file to the C-H RDC input. However, the direct scaling of the the RDC values is not equivalent to scaling via D_a-factor, since the energy for the RDC potential term is given by[244]

$$E_{RDC} = k_{RDC}\,(D_{calc} - D_{obs})^2 \tag{3.5}$$

where k_{RDC} is the scale factor for the RDC energy term, which has to be modified for different sets of RDCs with a weighting factor ω_{ij}. When using the prefactor D_{Me}^{pre} for implementation of methyl RDCs, the energy of the C-C methyl RDCs has to be scaled by

$$\omega_{CC} = \frac{1}{(-3.17)^2}\,\omega_{CH} \approx 0.0995\,\omega_{CH} = 0.1\,\omega_{CH}. \tag{3.6}$$

The dependence of the energy on the square of the difference between calculated (D_{calc}) and observed (D_{obs}) would otherwise allow methyl RDCs to have a stronger restraining effect than the C-H RDCs.

Table 3.1: Experimentally determined RDCs used in the structure determination of 13merHCF. The RDCs were measured with a precision of ±0.6 Hz.

Res	**Vector**	**$J_{(CH)}$ (Hz)**	**$J_{(CH)}$(aligned) (Hz)**	**RDC (Hz)**
A6	C2-H2	202.8	214.8	12.0
A8	C2-H2	200.4	209.4	9.0
A16	C2-H2	202.2	211.8	9.6
A24	C2-H2	201.0	213.6	12.6
C2	C5-H5	166.8	171.6	4.8
C12	C5-H5	169.2	174.6	5.4
C14	C5-H5	168.0	173.4	5.4
T11	C5-C7	127.2	121.2	-2.0
T19	C5-C7	127.2	121.2	-2.0
T21	C5-C7	126.6	121.2	-1.8
HCF	C1'-H1"	171.0	171.6	0.6
C14	C1'-H1'	168.0	166.2	-1.8
C17	C1'-H1'	161.4	118.8	-42.6
G25	C1'-H1'	159.0	171.6	12.6
C2	C6-H6	175.2	183.6	8.4
T3	C6-H6	174.6	182.4	7.8
C9	C6-H6	174.0	182.4	8.4
T11	C6-H6	176.4	183.6	7.2
C12	C6-H6	174.0	179.4	5.4
C14	C6-H6	171.6	167.4	-4.2
C17	C6-H6	174.0	182.4	8.4
T19	C6-H6	175.2	183.6	8.4
T21	C6-H6	175.8	183.6	7.8
G1	C8-H8	214.8	223.8	9.0
A8	C8-H8	214.8	223.2	8.4
G13	C8-H8	214.2	222.6	8.4
G15	C8-H8	214.2	223.8	9.6
A16	C8-H8	214.8	223.2	8.4
A24	C8-H8	214.2	225.0	10.8
G25	C8-H8	214.2	229.2	15.0
HCF	C1-H1	151.8	156.6	4.8
HCF	C3-H3	151.2	153.0	1.8
HCF	C4-H4	150.0	155.4	5.4
HCF	C5-H5	152.4	152.4	0.0
HCF	C6-H6	150.6	156.6	6.0
HCF	C8-H8	152.4	157.8	5.4
HCF	C9-H91	129.0	135.0	6.0
ABA	C3'-H3'	151.2	147.0	-4.2

3.1.9 Structure calculation

Calculation input All structure calculations were performed with Xplor-NIH v2.20[245]. A total of 334 (330) NOE distance restraints and 38 Residual Dipolar Couplings were used in calculations for the face-up (-down) orientation. The experimental data were supplemented with 124 backbone dihedral restraints, 72 hydrogen bond distance restraints and 27 planarity restraints (see Tab. 3.2).

The initial molecular dynamics calculations of extended strands were performed with dihedral restraints allowing both A-form and B-form conformations (with error bars of $\pm 50\,°$). B-form conformation was experimentally confirmed by ^{3}J coupling constants for H1'-H2' derived from DQF-COSY and NOESY-cross-peak intensities characteristic for B-DNA. Consequently, regular dihedral values from the literature[213] were included in the calculations.

Table 3.2: Overview of structural statistics for 13merHCF in face-up and face-down orientation.

	face-up	**face-down**
RDC restraints	38	38
NOE restraints		
- total	375	362
- interresidue	110	102
- intraresidue	265	260
Dihedral angle restraints	124	124
H-bond restraints	72	72
Base pair planarity restr.	27	27
NOE viol. ($> 0.5\,Å$)	0	0
RDC viol. ($> 0.4\,Hz$)	0	0
Dihedral viol. ($> 5\,°$)	0	0
RMSD to ave. struct. in Å	0.41	0.45

The structures were calculated in two steps. First, a reasonable starting structure with well defined local conformation was computed. To ensure that no bias is introduced towards local energy minima, the calculation started from an elongated and equilibrated structure. The resulting structure, which is mainly defined by NOE restraint data, was used as input for Simulated Annealing calculations including RDC data. The need for locally well defined starting structures in order to calculate reasonable structures which satisfy NOE as well as RDC data is documented in the literature[262,263].

Simulated Annealing protocol The complete MD protocols used for structure determination are given in the Appendix, section 2. The same protocols were used for both (orientational) forms of 13merHCF. The input scripts are based on the example files of the Xplor-NIH package (refine_full.py and sa.inp) but were substantially modified.

The first protocol is used to generate a reasonable starting structure. It starts with two extended strands which are then hybridized to form a duplex strand. This protocol only uses the NOE restraints as experimental input and starts with an initial minimization (50 steps) followed by 48 ps of high-temperature cartesian coordinate dynamics at 3000 K, subsequent gradual cooling to 25 K in 120 steps of 0.05 ps length and a final minimization (3000 steps).

The second MD protocol, which utilizes in addition the experimental RDC restraints and the starting structure, consisted of an initial cartesian coordinate minimization (1000 steps) followed by 50 ps of high-temperature torsion angle dynamics at 20000 K, subsequent gradual cooling to 25 K in 154 steps of 0.5 ps length (34 steps to cool down to 3000 K, followed by 120 steps to reach the end temperature) and a final cartesian coordinate minimization (3000 steps). The alignment tensor values were allowed to float during the calculations, as implemented in Xplor-NIH (v2.20)[245].

For each run an ensemble of 100 structures was computed. The 10 minimum energy structures without violation of restraints were chosen to compute an averaged structure which was energy-minimized to yield the final structure. The root-mean-square deviation (RMSD) of the 10 minimum energy structures to the average structure is a measure for the precision of the calculation.

Structure validation Several methods were used to check the accuracy of the average structures. Back-calculation of NOESY-spectra along the Full Matrix Relaxation Approach[213,264,265] were performed with Xplor-NIH (v2.20)[245]. The back-calculated spectra were then visualized with the help of a Mathematica script to allow comparison with experimental spectra. The script was written by the author on the basis of his diploma thesis[266] (see sec. 3.2). It uses the same input files like Gifa[267], a program that was used prior to the Mathematica script. However, the latter offers more user comfort and has the advantage of running on actual hard- and software. RDCs were predicted from the average structure using the program Pales[268] and were also compared to the experimental data.

3.2 13mer6HQ

3.2.1 NMR sample preparation

The DNA-oligonucleotide, that incorporates 6-hydroxyquinoline (6HQ) linked to R-Glycerol, was assembled in 24 µmol scale synthesis on an ÄKTA Oligopilot at Noxxon Pharma GmbH, Berlin. The necessary phosphoramidite was synthesized by Felix Hövelmann in four steps, using 6-hydroxyquinoline and S-glycidol as precursors. It should be noted that the latter will lead to the R-glycerol phosphoramidite. The counterstrand with cytosine as complementary base to 6HQ was ordered from BioTeZ, Berlin. Purification and hybridization were carried out the same way as for13merHCF (sec. 3.1.1). Interestingly, the final product already contained the deprotonated quinolinium, which is optically indicated by the yellow color of the solution. The deprotonation was caused by the 3 % NH_3-solution that was used in the lyophilization step (see sec. 3.1.1). The double strand was then diluted in pH 7 buffer solution(10 mM Na_2HPO_4/NaH_2PO_4 and 150 mM NaCl), but the quinolinium retained its deprotonated state. However, the reason for this contradiction is a very slow reaction rate, so that roundabout four weeks had passed until the reaction was complete. The concentration of the sample was 3 mM.

3.2.2 RDC sample preparation

Several months separate the first NMR experiment from the RDC measurement, so the sample was colorless, due to protonation of the quinolinium. In order to regain the deprotonated state, the sample was treated again with 3 % NH_3-solution between desalting and lyophilization. All other preparation steps were performed as stated in sec. 3.1.2.

3.2.3 Duplex melting experiments

A single- and double-stranded sample of 13mer6HQ was prepared in a water/ammonia mixture with $pH = 8.5$ and 150 mM NaCl. Measurements were performed in a double cuvette (closed but not evacuated) with 1 and 10 mm optical path length. A total concentration $c_T = 131$ µM of single strands was estimated for the double strand from maximum absorption around 260 nm at 90 ℃, and the single strand was concentrated to give a comparable absorption signal. A set of 17 (15) spectra between 10 and 90 (80) ℃ were recorded for the double strand (single) and corrected for density change. The configuration of the Varian Cary 300 spectrometer was the same as in sec. 3.1.3.

3.2.4 NMR experiments

13mer6HQ utilized the same experiments as 13merHCF in sec. 3.1.5. Only experimental parameters like the number of transients were adjusted to the needs of the sample.

3.2.5 Residual Dipolar Coupling restraints

Table 3.3: Measured RDCs of 13mer6HQ (±0.6 Hz). For more details see sec. 3.1.8.

Res	Vector	$J_{(CH)}$ (Hz)	$J_{(CH)}$(aligned) (Hz)	RDC (Hz)
A8	**C2-H2**	201.6	245.4	43.8
A16	**C2-H2**	201.6	238.2	36.6
A24	**C2-H2**	204.6	241.8	37.2
C2	**C5-H5**	165.6	223.2	57.6
C14	**C5-H5**	167.4	184.2	16.8
T3	**C7-H7**	126.6	108.6	-6.0
T11	**C7-H7**	126.6	108.6	-6.0
T19	**C7-H7**	127.2	109.8	-5.8
T21	**C7-H7**	126.6	108.6	-6.0
G1	**C1'-H1'**	165.0	184.2	19.2
C2	**C1'-H1'**	166.8	176.4	9.6
G4	**C1'-H1'**	157.8	180.6	22.8
G10	**C1'-H1'**	165.0	184.2	19.2
T11	**C1'-H1'**	163.2	194.4	31.2
C12	**C1'-H1'**	164.4	183.6	19.2
G13	**C1'-H1'**	160.2	174.6	14.4
C14	**C1'-H1'**	165.0	177.0	12.0
C17	**C1'-H1'**	162.6	175.2	12.6
T19	**C1'-H1'**	163.2	194.4	31.2
C20	**C1'-H1'**	164.4	179.4	15.0
G22	**C1'-H1'**	161.4	190.8	29.4
A24	**C1'-H1'**	166.8	182.4	15.6
G25	**C1'-H1'**	161.4	190.8	29.4
A24	**C8-H8**	214.8	255.0	40.2
6HQ	**C2-H2**	180.0	194.4	14.4

3.2.6 Structure calculation

The structure calculation and validation was performed analogue to 13merHCF in sec. 3.1.9, therefore only the calculation input is given.

Calculation input Hydrogen bonding restraints for the central AT base pairs were omitted, due to a lack of evidence in the H_2O-NOESY spectrum. Consequently, 60 restraints remain for the other 10 base pairs (13merHCF 72). Moreover, the lower concentration of 13mer6HQ (3 mM) in comparison to 13merHCF (5 mM) reduces the number of determinable RDCs (25 6HQ vs. 38 HCF).

Table 3.4: Overview of structural statistics for 13mer6HQ.

	13mer6HQ
RDC restraints	25
NOE restraints	
- total	418
- interresidue	148
- intraresidue	270
Dihedral angle restraints	124
H-bond restraints	60
Base pair planarity restr.	23
NOE viol. (> 0.5 Å)	0
RDC viol. (> 0.4 Hz)	0
Dihedral viol. ($> 5°$)	0
RMSD to ave. struct. in Å	0.36

3.3 13mer4AP-DAP

Some significant changes were made for 13mer4AP-DAP. First of all a weak acidic buffer (pH 6.35) was used to avoid hydrolysis under basic conditions of the 4AP. Furthermore, only a small amount of sample was available, so that all experiments were carried out with the same sample and the optical experiments were measured before NMR.

3.3.1 NMR sample preparation

Subsequently to the optical experiments, the whole sample was desalted via size exclusion chromatography (in Sephadex PD-10 column) and lyophilization, again without 3 % NH_3-solution. The NMR samples were prepared in (D_2O matched) Shigemi tubes at 1.2 mM duplex concentration in D_2O (D_2O 99.98 %) and afterwards in H_2O (H_2O:D_2O/90:10) buffer solutions at pH 6.35, containing 10 mM Na_2HPO_4/NaH_2PO_4 and 150 mM NaCl.

3.3.2 Duplex melting experiments

Michael Weinberger (Wagenknecht group) was responsible for the synthesis of 4AP and Falko Berndt (Ernsting group) correspondingly for DAP. Both synthetic routes[205] utilized stereoselective Heck-type palladium-catalysed cross-coupling with 2'-deoxyribofuranoside glycal followed by stereoselective reduction with $NaBH(OAc)_3$. Both nucleosides were further processed to the corresponding phosphoramidites and subsequently incorporated via automated DNA synthesis. The 4AP single strand, provided by the Wagenknecht group, and the DAP counterstrand, obtained from BioTeZ, were delivered already purified by reverse-phase high-pressure liquid chromatography (HPLC). After hybridization they were subjected to size exclusion chromatography (in Sephadex PD-10 column) and lyophilization, but this time without 3 % NH_3-solution. Equivalent amounts of complementary single strands were hybridized by rapid heating to 90 °C and subsequent gradual cooling to room temperature at a rate of 0.5 °C per minute. The whole sample of double-stranded 13mer4AP-DAP was then prepared in a water/phosphate buffer (10 mM NaH_2PO_4, pH = 6.35) with 150 mM sodium chloride. Measurements were performed in a double cuvette (closed but not evacuated) with 1 and 10 mm optical path length. A total concentration $c_T = 20.5$ µM of single strands was estimated from maximum absorption around 260 nm at 85 °C. A set of 14 spectra between 20 and 85 °C were recorded and corrected for density change. The configuration of the Varian Cary 300 spectrometer was the same as in sec. 3.1.3.

3.3.3 Titration against pH experiments

The Varian Cary 300 spectrometer was used with the same settings as mentioned before. A solution of 1.717 mg 2,4-diaminopyrimidine in 10 g pure water (Millipore) was prepared. A set of 15 samples were measured between pH 3 and 11 by adding 0,3 g of aforementioned sample to 2.7 mg of the corresponding pH buffer solutions.

3.3.4 NMR experiments

All NMR experiments were carried out on a Bruker Avance 600 MHz spectrometer with inverse probehead. In order to raise the signal-to-noise ratio, the NOESY experiments were performed with lower FID size of 4096 x 1024, but huge number of scans in a range of 160 to 240. HMQC experiments were omitted, due to the low concentration of the sample.

3.3.5 Structure calculation

The structure calculation and validation was performed analogue to 13merHCF in sec. 3.1.9. The RDC part in the Simulated Annealing algorithm was not included, due to missing RDC values. This is simply done by adding a comment tag to the line where the RDC term is added to the target function. For later discussion a second set of calculations with RDCs from 13merHCF were performed and compared to the original without RDCs.

Calculation input The data in following table were the basis for both calculation types, so that the only difference is the additional application of RDCs in the second set. The

RDC data for the second set was taken from 13merHCF (Tab. 3.1), therein the RDCs of the HCF chromophore were omitted to avoid interference with 4AP.

Table 3.5: Overview of structural statistics for 13mer4AP-DAP in 1H-bond and 2H-bond orientation.

	1H-bond	**2H-bond**
NOE restraints		
- total	357	420
- interresidue	134	137
- intraresidue	223	283
Dihedral angle restraints	124	124
H-bond restraints	72	72
Base pair planarity restr.	27	27
NOE viol. (> 0.5 Å)	0	0
RDC viol. (> 0.4 Hz)	0	0
Dihedral viol. ($> 5\,°$)	0	0
RMSD to ave. struct. in Å	0.52	0.85

4 Results and discussion

4.1 2-Hydroxy-7-carboxyfluorene - 13merHCF

The HCF chromophore is the direct successor of HNF[203] and only differs in the carboxyl group (HNF: NO_2) at position 7 of fluorene (see Fig. 4.1). So both molecules have many common properties. First of all, both were linked via an α-glycosidic-bond to 2-deoxyribofuranose, the only form that yielded sufficient amount and purity. They also share incorporation into the center of the strand opposite an abasic site, thus making HCF also a base pair surrogate. The resulting duplex strand is named 13merHCF. As stated in the introduction (see 1.3), the HNF chromophore suffers from a short lifetime of the excited state (35 ps), due to intersystem crossing, and for this reason the HCF derivative was designed.

The introduction of the carboxyl group makes it necessary to investigate the protonation state of HCF. For this purpose, a sample of double-stranded 13merHCF was measured at different pH values (see sec. 3.1.4). The change in absorption of the HCF band around 320 nm (see Fig. 4.9) was fitted to a pH dependent two-state model. A pK_a value of 3.9 ± 0.1 has been determined which is similar to other organic compounds like benzoic acid. Therefore, 2-hydroxy-7-carboxyfluorene is present in the deprotonated form (shown in Fig. 4.1) when buffered at pH = 7. This finding is also supported by Manoharan and Dogra[269] who measured absorption spectra at different pH values and found that 2-carboxyfluorene is deprotonated at pH = 8 and exists in neutral form at pH = 2.

A symmetric and non-palindromic sequence was chosen to minimize mispairing, loop formation and fraying effects[155]. A length of 13 base pairs allows to use the central base

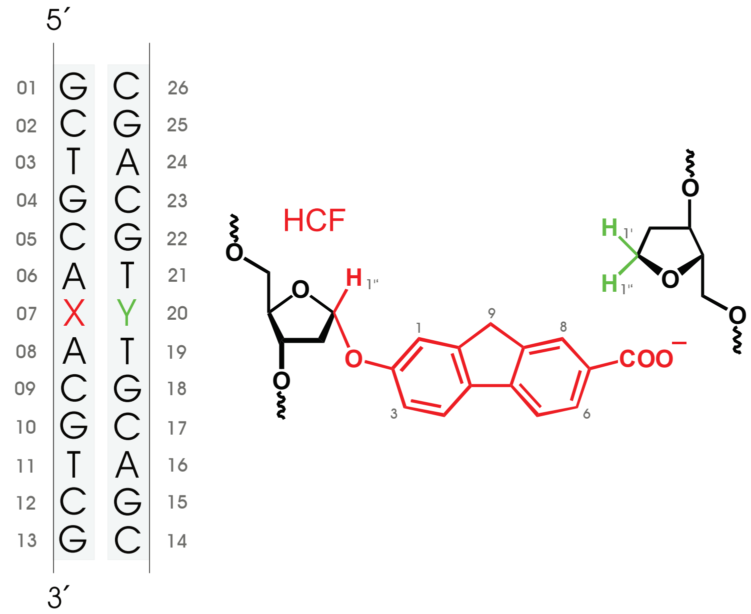


Figure 4.1: Structure of 13merHCF. Native DNA is marked black and all modifications including backbone are colored either red or green. The HCF moiety is placed inside the duplex at position X and the abasic site (in green) at Y, respectively.

pair as modification site, which also ensures that perturbations of the structure can be directly assigned to the incorporated chromophore. Furthermore, it should be noted that in contrast to native DNA, both strands incorporate a H1" hydrogen in the center, due to the α-glycosidic linkage of HCF and the omitted base in the complementary strand. This will interrupt the stepwise assignment of DNA, because under the assumption that B-DNA is the dominant form in solution, one can "walk" in alternating steps between intra- and interresidual cross-peaks from the 5'-end down to the 3'end and vice versa. The method is known as "NOE-walk" and will be described in more detail in the next section.

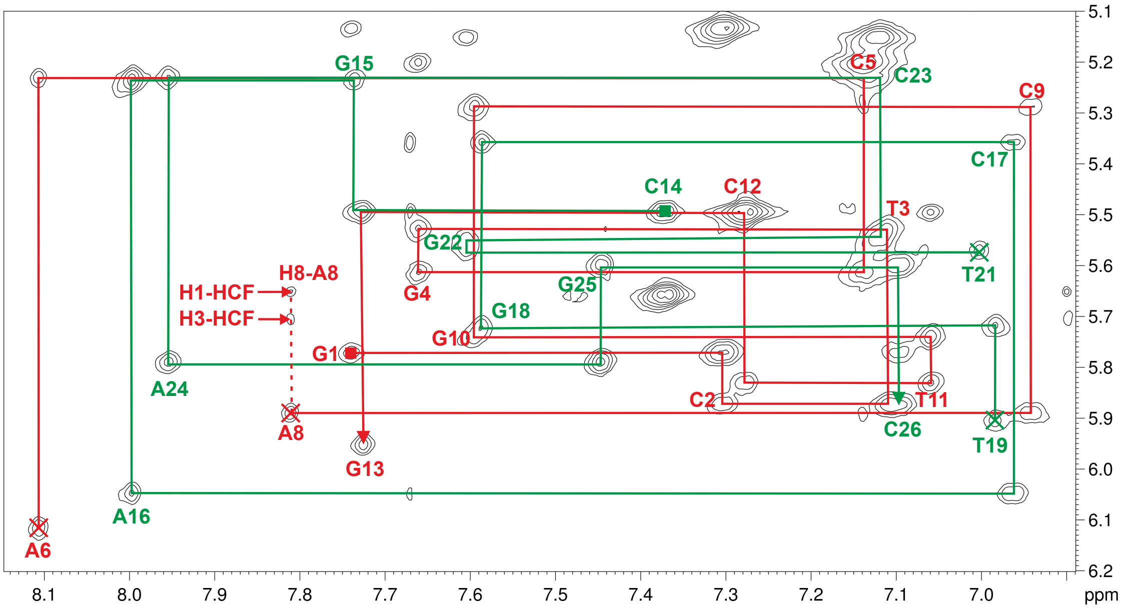


Figure 4.2: NOE-Walk in 13merHCF. The assignment of the H6/H8 (abscissa) and H1' (ordinate) region in the D_2O-NOESY spectrum is shown. Red lines mark the NOE-walk for the HCF containing strand and green lines for the abasic strand. The starting point (5'-end) of both walks was denoted by a square, while the endpoint (3'-end) is indicated by an arrowhead. Note that both walks are interrupted at the modification site (marked by X), due to the α-glycosidic linkage of HCF and the omitted base in the other strand. For clarity only the positions of intraresidual H6/H8-H1' cross-peaks were labeled.

4.1.1 Chemical shift analysis

The assignment of the spectra followed the guidelines described by Roberts[213] and Bloomfield et al.[19]. The "NOE-walks" of the H6/H8 to H1' region are in case of 13merHCF interrupted at the central strand positions 7 (HCF) and 20 (abasic site), but as stated before, this was expected and agrees well with HNF. In Figure 4.2 the NOESY spectrum measured in D_2O is shown and the corresponding walks are marked red for the HCF strand and green for the abasic strand, respectively. The starting point (5'-end) of both walks was denoted by a square, while the endpoint (3'-end) is indicated by an arrowhead. The crosses mark the positions, where the alternating walk between H1' sugar protons and H6/H8 nucleobase protons is interrupted by a H1' to H1" step. Even though the corresponding cross-peak can be found in the spectrum, it cannot be depicted in Fig. 4.2, because it is part of a completely different spectral region. The other sugar protons (H2', H2", H3' etc.) were subsequently assigned by additional NOE-walks in combination with other spectra (COSY,TOCSY). Chemical shift tables of all assigned hydrogens and carbons are given in the Appendix (sec. 1.1).

Exchangeable hydrogens were assigned in H_2O after completion of the D_2O-NOESY spectrum. The separation of the amino hydrogen (H41, H42) signals of C and, in addition, the large low-field shift of the imino proton signals (H1 in G, H3 in T) in the range of 12 - 15 ppm[213] clearly indicate a proper hydrogen bonding pattern for the native base pairs. Even though ppm values were not used directly in the MD simulations, the knowledge of correct hydrogen bonding allows to introduce a set of hydrogen bonding restraints (see Tab. 3.2 in 3.1.9) that were derived from high-resolution X-ray structure determination[270].

Comparison of ^{1}H Chemical Shift Deviations (CSDs) were made between 13merHCF, 13merHNF and 13merRef[260]; the results are presented in Fig. 4.3. Here the chemical shift differences of all protons belonging to a single residue but different DNA samples are summarized. Note that absolute values are given to prevent canceling of deviations. Three comparisons are made, 13merHCF to 13merRef with A-T in the center (blue), the corresponding 13merHNF to 13merRef (red) and the modified double strands against each other (green). The fluorene containing duplexes (blue,red) show a similar pattern of deviations when compared to 13merRef, so one can expect that the deviations between them would be rather small. This is true except for the fluorene residues (7), where the deviation nearly reaches 0.5 ppm. The difference can be explained readily by the incorporation of the carboxyl group, which leads to higher ppm values for the neighboring H6 and H8 hydrogens (see Fig. 4.1). Moreover it can be concluded for both duplexes, 13merHCF and 13merHNF, that the effect of the chromophore is limited roughly to the next two base pairs in both directions around the center.

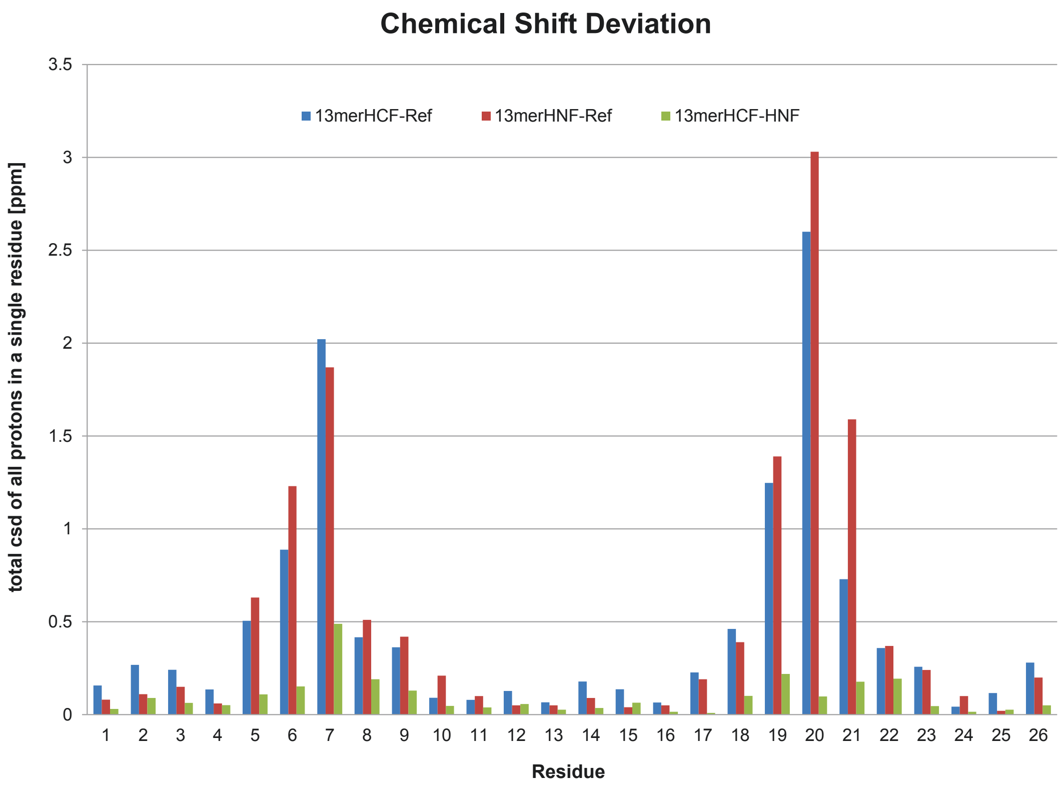


Figure 4.3: CSD comparison between 13merHCF and 13merHNF with reference strand (13merRef). Chemical shift differences of all protons belonging to a single residue but different DNA samples were summed and are given as absolute values.

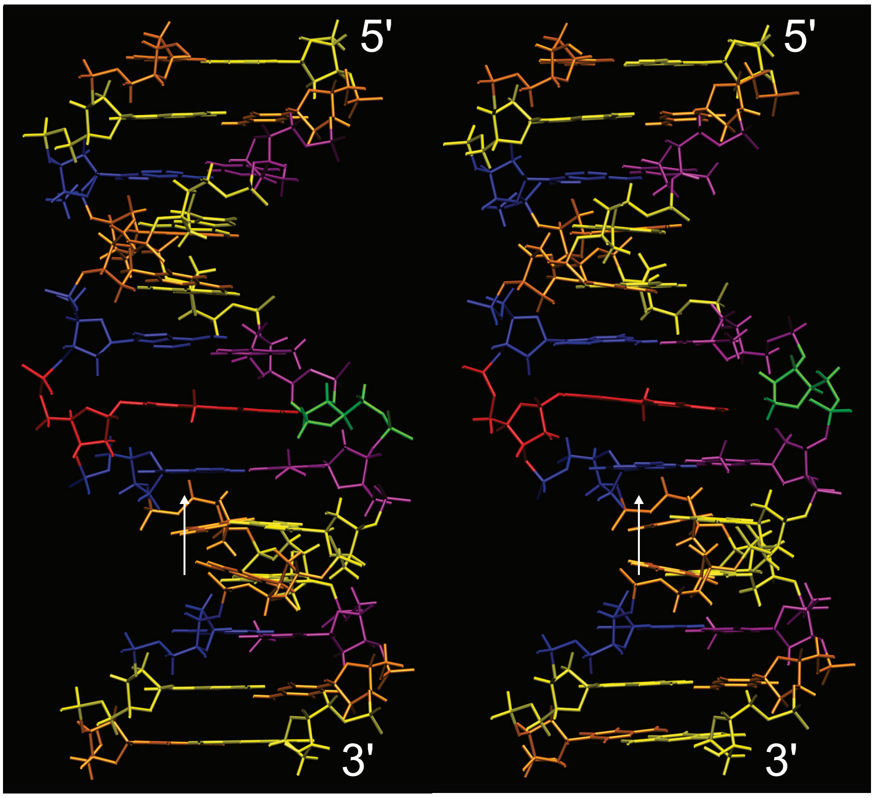


Figure 4.4: Averaged structures of 13merHCF. The HCF is marked red and the abasic site green, while the corresponding colors for A, G, C and T are blue, yellow, orange and violet. The HCF methylene group faces up to the reader in the left panel and down in the right panel; the orientations are named correspondingly. RMSD among all hydrogens besides methyl protons is 0.41 Å for face-up and 0.45 Åfor face-down, respectively. The white arrows indicate the point of view in Fig. 4.6.

4.1.2 NMR solution structure

Two NMR solution structures of 13merHCF had to be determined using RDC and NOE distance restraints. As in case of 13merHNF [203], a single structure was not sufficient to describe the data found in the NOESY spectra. Therefore an ensemble of 100 structures was calculated for each orientation and the ten energy-lowest, violation-free structures were chosen to obtain averaged structures for both. Further details about the Simulated Annealing simulations can be found in section 3.1.9 and 3.1.9.

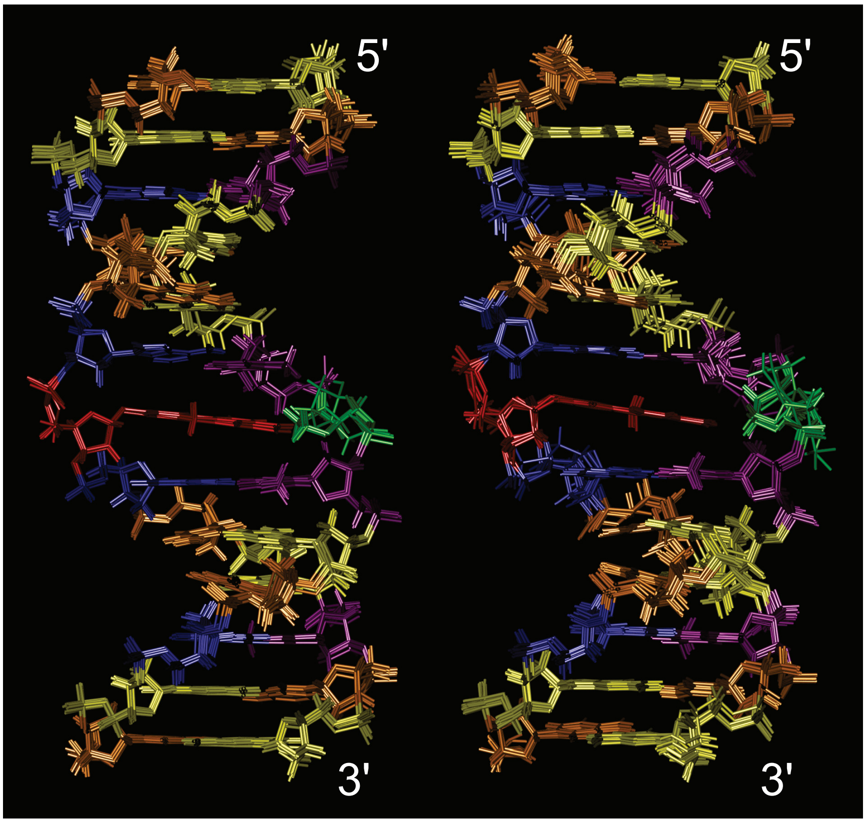


Figure 4.5: Overlay of the 10 minimum-energy, violation-free structures. The left side corresponds to the face-up orientation of the average structures (left in Fig. 4.4) and the right to face-down.

In Fig. 4.4 the averaged structures for the face-up (left) and the face-down (right) orientation of 13merHCF are shown. The names were derived from the side where the HCF methylene group faces to the reader. The chromophore in both structures fits nearly parallel between the surrounding base pairs. A perturbing factor that limits a fully parallel placing of HCF is the additional oxygen atom between the chromophore and the sugar. On the other hand is the α-glycosidic-bond unproblematic to the structure, even in the presence of the added oxygen. The 2-deoxyribofuranose and HCF are able to retain a roughly perpendicular orientation to each other, in which H1' and H1" would be in a plain area with the chromophore, thereby allowing the sugar to flip from the minor to the major groove side without inducing perturbations.

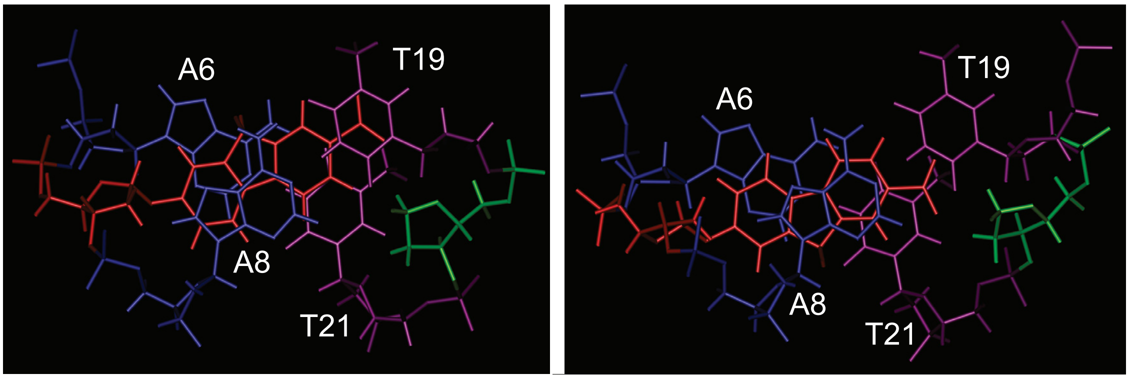


Figure 4.6: Close view on the three central base pairs of 13merHCF. The face-up orientation is again at left and face-down at right. In order to fit the HCF into the center of the helix, the 2'-deoxyribose compensates the α-glycosidic bond by switching to a sugar conformation between O_1'-endo and C_4'-exo for face-up and O_1'-exo for face-down, respectively.

The structures that were combined to averaged structures are depicted as overlay in Fig. 4.5. The set of face-up structures on the left sum up to a root mean square deviation (RMSD) of 0.41 and to 0.45 for the face-down set, respectively. The biggest contribution to the RMSD value comes from the abasic site. Single structures with different placement of the abasic site in comparison to their corresponding averaged structures can be found in both orientations. Without a nucleobase attached to the 2-deoxyribofuranose the NOE-walk along the abasic strand is interrupted, which means that interresidual distance NOEs are missing and hence the flexibility of the abasic site cannot be restrained by experimental data. As a second consequence of the missing base, a preferred orientation of the chromophore through potential hydrogen bonding or sterical interactions is also missing.

The only interactions that actually limit the rotation around the C2-O2-bond are stacking interactions with the neighboring bases. In Fig. 4.6 a close view of the three central base pairs is shown. For both orientations stacking with adjacent bases was found. In face-up position the chromophore stacks mainly with adenine residue 6, while the face-down conformer favors adenine 8. Additional stacking to the corresponding thymine T21 and partly T19 of the abasic strand suggests a possible contribution to duplex stability, which will be topic of the melting experiments in the next section. Moreover, it becomes clear how the 2'-deoxyribose compensates the α-glycosidic bond in the HCF nucleotide. For example, the sugar of adenine A6 retains the usual C2'-endo conformation of B-DNA, while the 2'-deoxyribose of HCF switches to a conformation between O_1'-endo and C_4'-exo for face-up and O_1'-exo for face-down, respectively.

Up to this point 13merHCF clearly resembles 13merHNF, and due to the fact that this was already expected at the beginning of the NMR experiments, a small detail besides the functional group was changed during sample preparation. More precisely, the concentration of duplex DNA had been raised from 3 mM in 13merHNF to 5 mM in 13merHCF. The intention was to verify the RDC measurements, which may have suffered from one of two major drawbacks. The first is hidden in the name Residual Dipolar Couplings. The word *residual* points out that in RDC measurements only a small portion of the sample contributes to the measured anisotropy, which is induced by weak alignment of the sample. This adds to the demand for high precision (1 Hz) in a highly viscous sample, thus making RDC measurements a challenging task. Is in consequence only a limited number of suitable RDCs available, a second drawback has to be considered. As outlined in section 2.4, at least five RDCs are necessary to determine a single angle in the molecular frame. The general degeneracy of a single RDC value limits the usability as restraint in molecular dynamics simulations. In case of the 13merHNF sample only 19 RDC values were part of the structure refinement. The question which arose at that time was, is the compatibility of the measured RDC values to both orientations a consequence of the parallel orientation of the chromophore or caused by degeneracy of the few restraints? In the latter case, it follows that potential bending of the duplex has not been detected, due to a lack of experimental data.

In the highly concentrated 13merHCF solution, 38 RDCs were found which is twice the number for 13merHNF. The experimental values were then plotted against predicted values of the averaged structures in Fig. 4.7. The results agree well with both structures and support the previous finding, according to which the RDC values are compatible with both orientations. However, a high concentration of duplex DNA is neither always

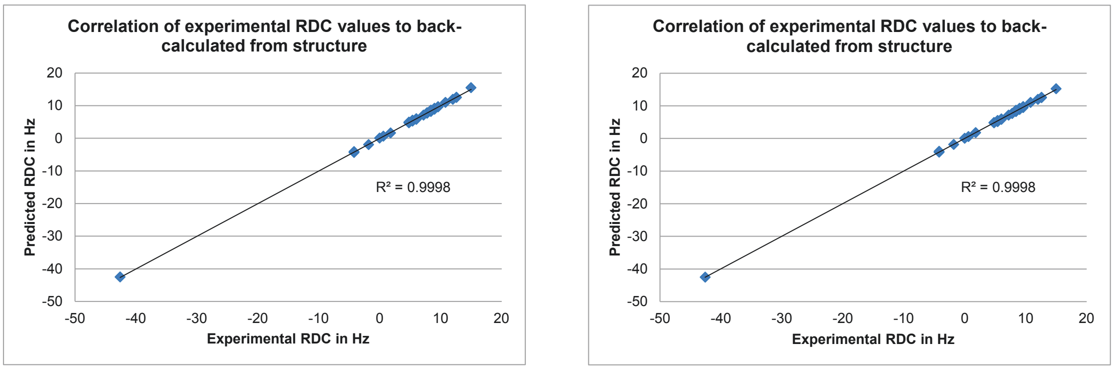


Figure 4.7: Pales[268] plots of experimental RDCs against predicted values of the 13merHCF structures (face-up left, face-down right). Both plots look nearly identical, due to degeneracy of RDCs (see sec. 2.4). In case of 13merHCF, both conformations have nearly the same orientation relative to the helical axis (roughly perpendicular), so the same set of values can describe both conformations.

available nor desirable for a unknown structure. A side effect that can be observed in such samples are unusual cross-peaks in the NOESY spectrum. At first sight they might be misidentified as spin diffusion signals, but in that case the intensity of their signal would be sensitive to the mixing time. With the help of back-calculated NOESY spectra, depicted in Fig. 4.8, those cross-peaks can be explained as well as the necessity for two orientations of the chromophore. The upper panel of Fig. 4.8 presents the region where H1 and H3 of HCF meet the H1" of their backbone. The black lines mark the experimental NOESY data with equal integrals of intensity for both peaks. The different line shape of the H3 signal is caused by coupling with the neighboring H4 proton. The back-calculated face-up spectrum (green) can only cover the H1/H1" cross-peak, while the face-down orientation (red) only covers the H3/H1" signal, so both conformers are required to complete the experimental NOESY spectrum. The reason for this is the strong distance change upon rotation. In close proximity the distance is roughly 2 Å which results in a strong signal, while after rotation the weak intensity of a 4.5 Å distance is observed.

The lower panel of Fig. 4.8 allows to compare the whole H6/H8 to H1' region between all NOESY spectra. It is not surprising that they describe most of the signals in the same fashion, but in this picture the two differences are of greater interest. The first deviating signal is the (red) HCF-H8/H1'-T21 cross-peak that is only present in the face-up orientation. In contrast to the H1/H3 pair the corresponding H6 atom is in both orientations more than 5 Å away from H1'-T21 and therefore not visible, due to the missing symmetry of HCF around the long axis. However the most interesting signal is the (black) C26-H6/G1-H1' cross-peak which is only present in the experimental NOESY spectrum. The distance corresponding to this assignment would cover 11 Å in both orientations. An explanation via spin diffusion would involve too many atoms and is not reasonable. The real reason for this very unusual peak is the high concentration of 13merHCF as mentioned before, since it is not an interresidual cross-peak between two strands of a single duplex. The effect is known as "end-to-end" stacking and was first described by Nakata et al.[271], who observed long rods of stacked oligonucleotides. In fact, C26-H6/G1-H1' assigns atoms of two different duplexes and therefore should be named "interduplex" cross-peak. Such peaks raise the complexity of the spectrum and one should be aware of it when assignment reaches the helix ends.

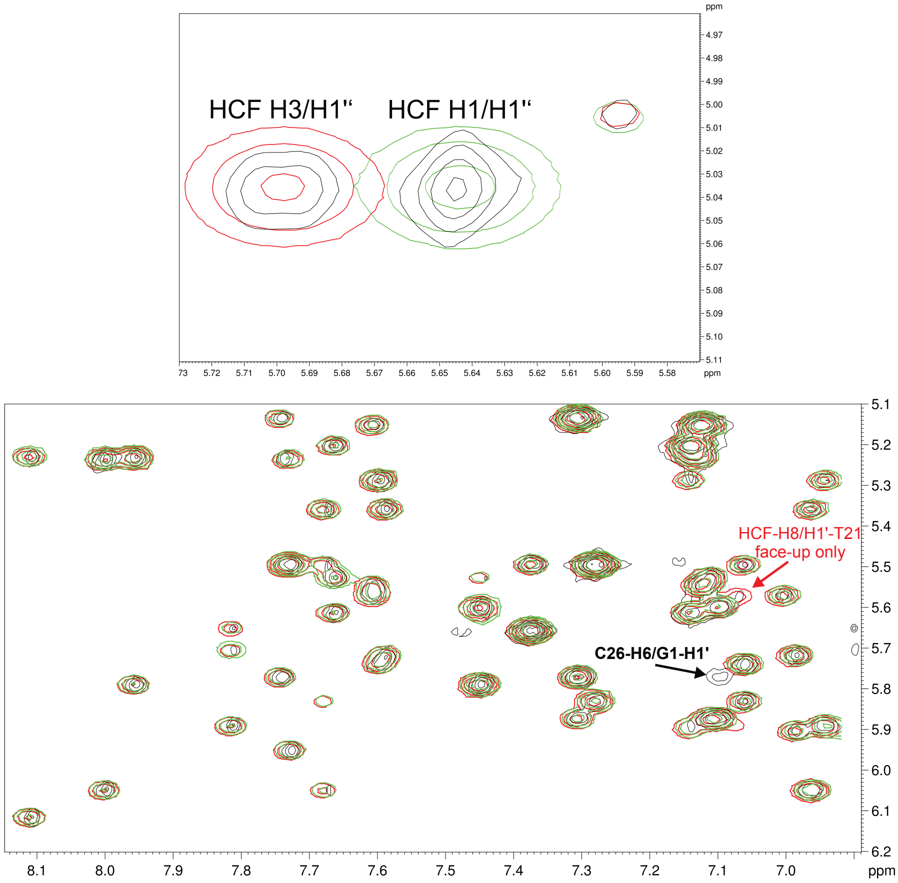


Figure 4.8: NOESY back-calculations of both conformers overlaid with experimental data. The face-up spectrum is colored red, the face-down spectrum green, and the experimental spectrum black. Top: spectra of both conformers are needed to simulate the cross-peaks of the experimental spectrum. Bottom: overlay of the H6/H8 to H1' region. The cross-peak marked with the read arrow is only visible in the face-up orientation. An interduplex cross-peak that indicates "End-to-end" stacking is marked with a black arrow.

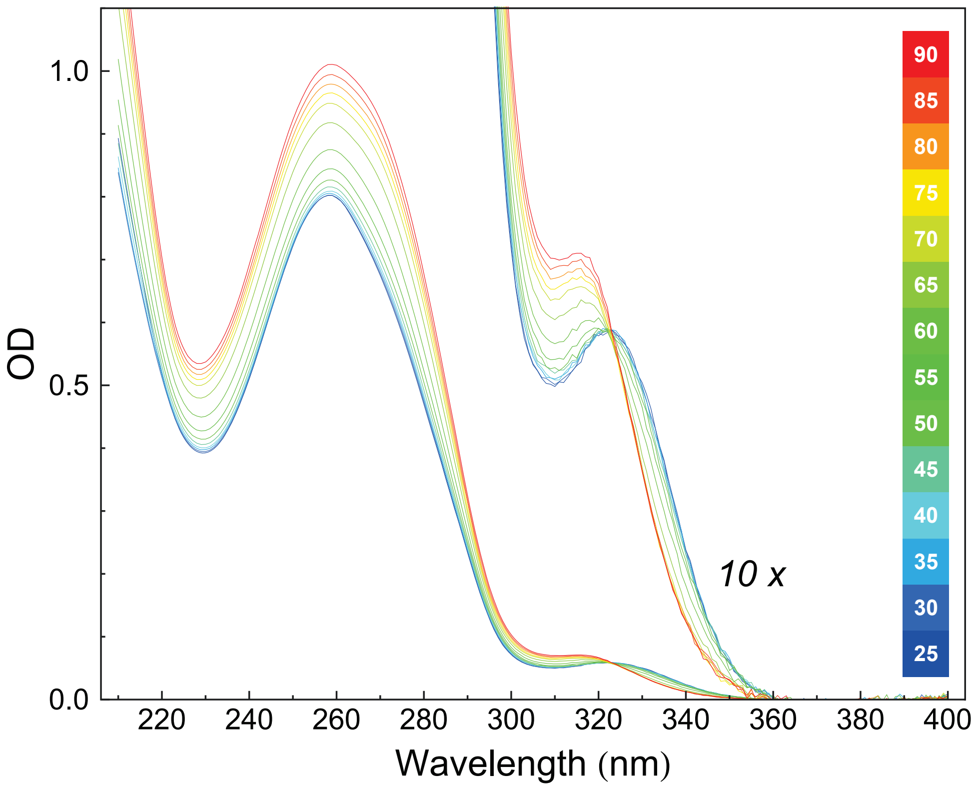


Figure 4.9: Absorption change of ds13merHCF upon melting, when raising temperature from 25 °C (blue) to 90 °C (red line; c_T = 45 µM). Spectra are shown for an optical path length of 2 mm. For better comparison they are also shown on an expanded ordinate scale (labeled "10x").

4.1.3 Duplex melting experiments

A sample of double-stranded 13merHCF (having the sequence composition shown in Fig. 4.1) was prepared in a water/phosphate buffer (10 mM NaH_2PO_4, $pH = 7$) with 150 mM sodium chloride. Measurements were performed in a cuvette with 2 mm optical path length. A total concentration c_T = 45 µM of single strands was estimated from maximum absorption around 260 nm at 90 °C. 14 absorption spectra between 25 and 90 °C were recorded and corrected for density change. The resulting spectra are shown in Fig. 4.9, including a ten-times magnification of them (labeled "10x").

The thermodynamics of duplex formation is usually examined by "melting curves" measured at a single wavelength. In this work, however, full absorption spectra were measured

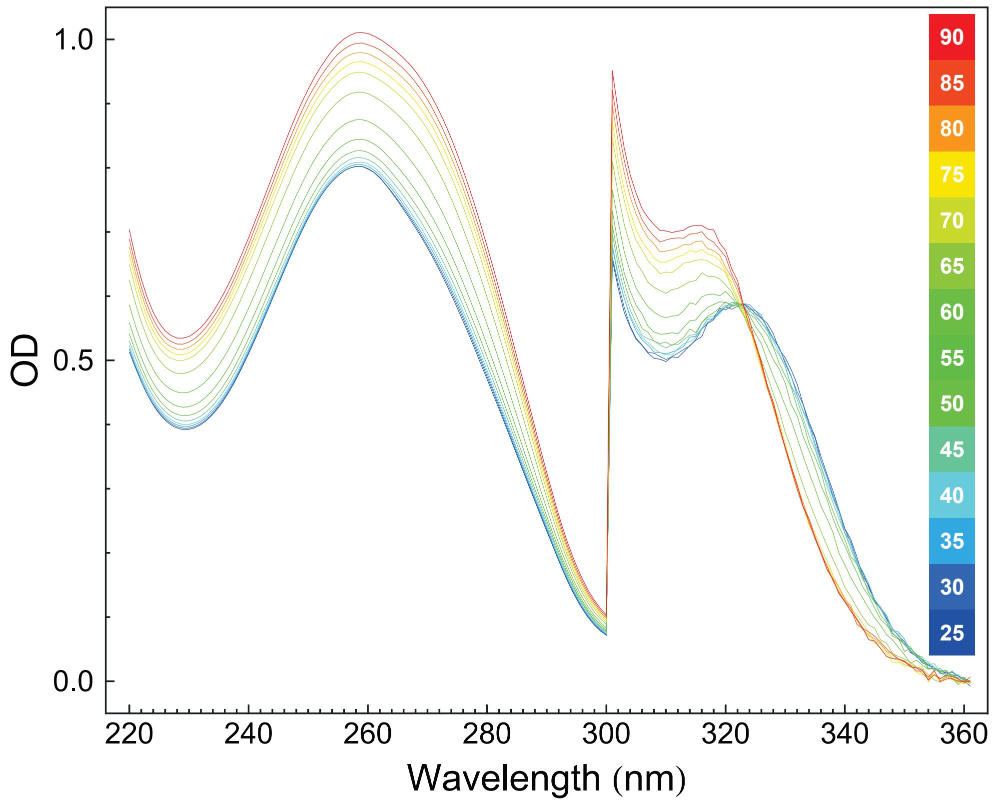


Figure 4.10: Weighted spectra of 13merHCF, as used in SVD. At wavelengths longer than 300 nm, absorbance values were multiplied by a factor ten. In this way the absorption change of DNA (around 260 nm) and of the HCF chromophore (around 330 nm) are given equal importance; otherwise the DNA band would dominate the analysis.

at different temperatures to allow a more detailed description of the melting process. The basic idea is to take advantage of the full spectral window shown in Fig. 4.10. Here we see the absorption change of the DNA band on the left side and the corresponding change of HCF on the right. Such weighted spectra are needed to treat the absorption change of the DNA (around 260 nm) and of the chromophore (around 320 nm) on the same level; otherwise the DNA band could dominate in the Singular Value Decomposition (SVD). This method searches for principal components in the entire spectral window (see chapter 2.6). In other words, "melting curves" for every single wavelength are analyzed in order to find temperature dependent components which describe the melting of the double strand and of the chromophore. A quick inspection by eye shows already that the HCF

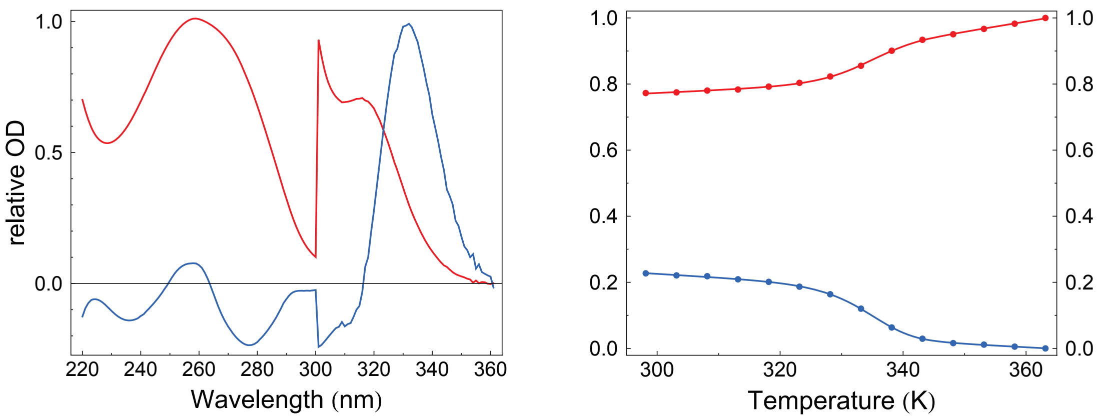


Figure 4.11: Basic spectra are shown at left ($\mathbf{U}_{cut}.\mathbf{S}_{cut}.\mathbf{F}$ in 2.6). The first component (red) was set to be the 90 °C spectrum, i.e. it represents the fully separated single strands. The right panel shows the thermodynamic curves corresponding to the basic spectra ($\mathbf{P}$ in 2.6). Upon hybridization, the amplitude of the first-component spectrum (red) decreases from 1 to 0.77 while that of the second component increases from 0 to 0.23.

probe combines a blue shift with hyperchromism upon melting. Interestingly, only two components are needed to describe the hyperchromism of the bands due to DNA and to the built-in chromophore at the same time. Fig. 4.11 shows the basic spectra on the left ($\mathbf{U}_{cut}.\mathbf{F}$ in 2.6) and the corresponding thermodynamic curves on the right ($\mathbf{P}$ in 2.6). The first component (red) was set to be the 90 °C spectrum. Note that the thermodynamic curves were multiplied with a scale factor so that the amplitudes fall into the range from 0 to 1 (the corresponding basic spectra were scaled with the inverse factor). Upon duplex formation, the one shown in red experiences an amplitude decrease of 0.23. By design, the blue curve has the same amount as amplitude increase (Fig. 4.11).

A SVD performed only on the DNA absorption band also reveals two components. But in that picture (which is not presented here) both components describe the same melting behaviour and melting point, due to the fact that the absorbance shape around 260 nm is a mixture of absorption bands from the four natural nucleotides[19]. Therefore the effect of hyperchromism is also a mixture: one component is related to the lineshape of the single strands at high temperature, and the other transforms it to the lineshape of the duplex state upon hybridization. But let us return to the combined analysis in figure 4.11. Red

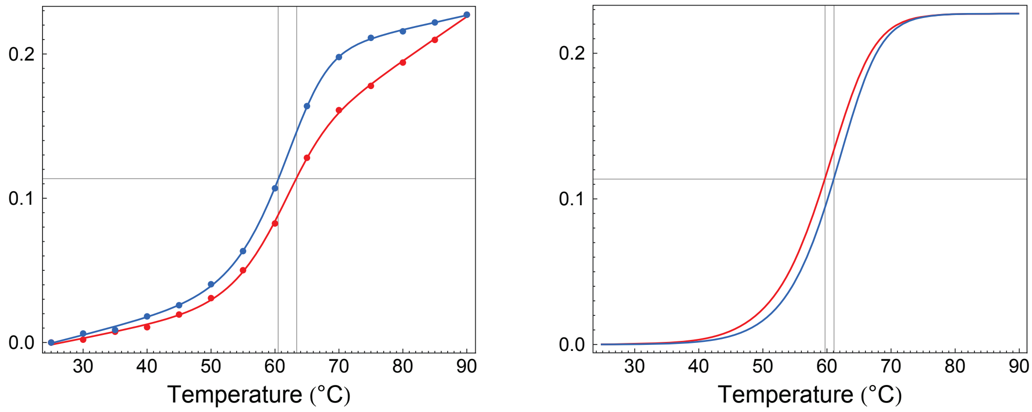


Figure 4.12: Thermodynamic curves of 13merHCF, scaled to the measured amplitude change of 0.23. The left panel shows SVD data (points) and fits with a two-state model including stacking (lines). The effect of stacking, assumed to be linear, was combined with the sigmoidal line shape of the two-state model. The two-state model part is plotted separately (right panel), representing a melting curve of hyperchromism without the influence of stacking (i.e. the degree of dissociation). The grey vertical lines indicate the melting points (derived from the right panel). The strands as a whole (red) melt at 59.7 °C and the HCF region (blue) at 61.1 °C.

lines denote the hybridization of the whole strand, while the blue lines are connected to the chromophore and upon duplex formation affect the lineshape of the DNA absorption band. Keep in mind that the latter is a mixture of bands, and their carriers are not equally distributed over the double strand. More specifically, the HCF chromophore is surrounded by A-T base pairs. The NMR solution structure indicates additional stacking interactions with T21 and partly T19 of the opposite strand (see Fig. 4.6). Therefore an interaction between HCF and all adjacent base pairs, strong enough to have an observable effect on the DNA absorption band, seems likely.

Based on equation 2.33, a two-state model with stacking interactions was applied to the melting curves in figure 4.11. The resulting fits with stacking interactions (left) and the degree of dissociation (i.e. with the optical effects of stacking removed, right) are shown in figure 4.12. As stated before, the red lines are related to the whole strand and the blue ones to the HCF part. The melting points are marked by grey vertical lines and were derived from the right part of figure 4.12. The parameters for DNA (c_T = 45 µM) are T_m

$= 59.7 \pm 0.1\,°C$, $\Delta H° =$ -340 ± 30 kJ/mol, $\Delta S° =$ -920 ± 80 J/(K mol) . The parameters for HCF are $T_m = 61.1 \pm 0.1\,°C$, $\Delta H° =$ -360 ± 20 kJ/mol, $\Delta S° =$ -1000 ± 40 J/(K mol).

The 1.4 °C higher melting point of HCF suggests that the middle of the strand melts last. The slightly steeper slope of the duplex state part (blue line on the left of Fig. 4.12) indicates that DNA stacking is more cooperative, even though the overall melting point is lower compared to the 63.3 °C of 13merRef[129] ($c_T = 23.7$ µM) with a central A-T base pair. These findings agree with the NMR structure as well as with Kool et al., who found that large aromatic hydrocarbons like the Pyrene nucleoside[97] or later the β-C-Porphyrinyl nucleoside[162] help to maintain stacking in both strands of a duplex. In the case of Pyrene opposite an abasic site ($T_m = 41.6\,°C$), they observed a thermal stabilization between 18 and 23 °C compared to natural nucleobases in the same position (T,C,A or G opposite abasic site), furthermore, it fits perfectly in the melting point gap between their natural control duplex with central A-T pair ($T_m = 43.2\,°C$) and a shortened duplex in which the central base pair is deleted ($T_m = 39.9\,°C$). The last point is not surprising, since the stabilizing effect depends on stacking with adjacent base pairs. When the ongoing melting process causes them to unstack, the additional stabilization will be lost.

4.2 6-Hydroxy-quinolinium - 13mer6HQ

13mer6HQ incorporates N-methyl-6-quinolone (MQ) as artificial nucleobase in the center (Fig. 4.13). MQ was introduced by Ernsting et al.[122] as polarity probe for femtosecond solvation experiments. In this first study, where pure water and methanol were used as benchmark solvents, it was shown that the time-resolved Stokes shift of fluorescence reflects the infrared spectrum of the surrounding liquid. MQ is a betaine that has several advantages over other probes. Betaines are not outwardly charged, so they do not polarize the environment more than is needed for the experiment. Additionally, MQ is free of internal vibrational modes that could interfere with solute motion (see sec. 1) and the solvent polarity couples mainly through dipole moment change. More interesting for this work is the property that the molecule is small enough to replace natural bases in nucleic acids or tryptophan in proteins.

This goal was reached in several steps. The first one was the measurement of the disaccharide trehalose in a mixture with water and MQ as free probe[272]. This was followed by the attachment of the MQ moiety to trehalose in order to allow measurement of water molecules in the vicinity of the disaccharide[123,204]. The study uncovered problems that occur when a betaine is linked to a biomolecule. The originally intended linkage via N-carboxymethyl ester failed, due to unexpected high instability, but was successfully replaced by direct N-alkylation of the quinoline derivative with trehalose triflate. Finally, MQ was incorporated into DNA by Felix Hövelmann, who continued previous works of Lucas Bethge in the group of Oliver Seitz.

6-Hydroxyquinoline (6HQ) was chosen as precursor for the synthesis of 13mer6HQ, but two critical issues need to be considered. The first one is the linkage of the quinoline moiety to the backbone. The native linkage via 2-deoxyribose would be too labile due to the positively charged chromophore. The second issue is the limited choice of hydroxyl protective groups. It was discovered that any protecting group placed on the hydroxyl group became unstable upon alkylation on the nitrogen, prohibiting the use of acyl or silyl groups.

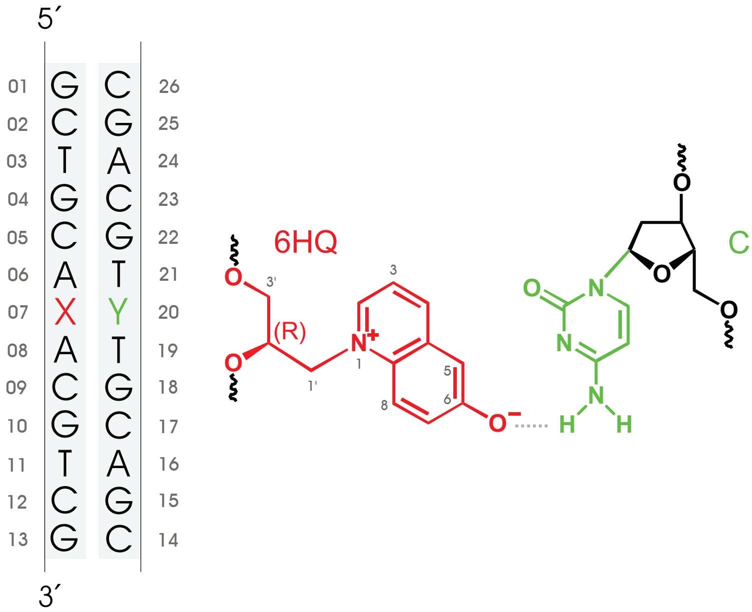


Figure 4.13: Structure of 13mer6HQ. For convenience , the 6HQ moiety including R-glycerol backbone is depicted in red and the complementary cytosine in green. The correspondingly colored letters X and Y mark their positions in the duplex strand.

As first alternative to 2-deoxyribose a L-Serinol linkage was tested, but proved to be too labile, so the product could not be isolated. Then the strategy was changed towards a carbacycle, which is missing the endocyclic oxygen, and could therefore overcome the stability problem. However the low yield and poor solubility of intermediate compounds demanded a second change of the strategy. In order to finally overcome the problem of labile linkers, the tethering of 6HQ via a short alkyl chain was chosen, resulting in R-glycerol-6HQ after four steps. The glycerol backbone is advantageous for the following reasons: First of all, it is known to produce intact duplex structures (GNA[73]). Secondly, the alkyl linkage in R-glycerol-6HQ should be sufficiently stable to survive the conditions during DNA-synthesis, cleavage and purification. Finally, the allyl proctection group on the hydroxyl group will be sufficiently stable and can be cleaved on the CPG after DNA synthesis.

R-glycerol-6HQ was incorporated as central nucleotide of the sequence shown in Fig.

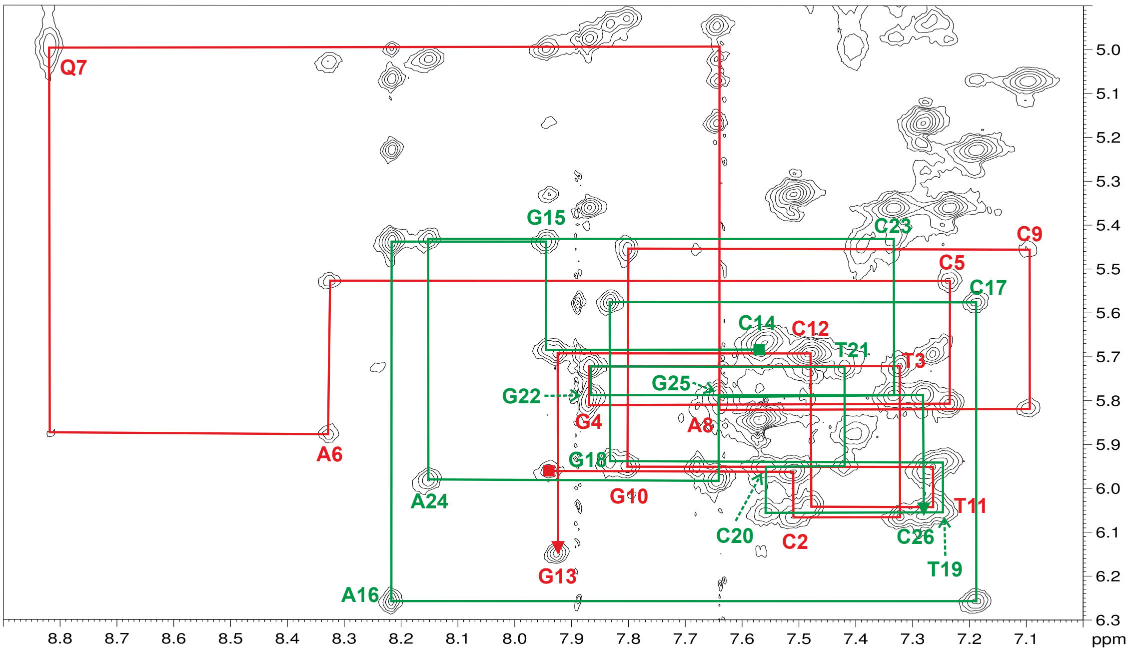


Figure 4.14: NOE-Walk in 13mer6HQ. The assignment of the H6/H8 (abscissa) and H1' (ordinate) region in the D_2O-NOESY spectrum is shown. The red line depicts the NOE-walk for the 6HQ containing strand and the green line for the complementary strand. The starting point (5'-end) of both walks was denoted by a square, while the endpoint (3'-end) is indicated by an arrowhead. For clarity only the positions of intraresidual H6/H8-H1' cross-peaks were labeled. The introduction of R-glycerol, depicted as Q7, made it necessary to extend the spectral region in both dimensions.

4.13 and Cytosine was chosen as potential hydrogen bonding partner to allow base pair formation. R-glycerol for the backbone was selected after simulations of both configurations using Hyperchem 7.5, which revealed a slightly better energy for R. Up to now, only crystal structures of duplex strands containing a full GNA backbone are known[273,274], so the solution structure of a DNA duplex with a GNA monomer in the center will be of general interest.

4.2.1 Chemical shift analysis

The assignment of the spectra followed the guidelines described by Roberts[213] and Bloomfield et al.[19] as before. In contrast to 13merHCF, complete "NOE-walks" for both strands of 13mer6HQ could be achieved. The D_2O-NOESY spectrum of 13mer6HQ is shown in Figure 4.14. The walks are marked red for the 6HQ strand and green for the complementary strand, respectively. The H2/H1' signal of R-glyerol-6HQ (denoted as Q7) can be found in the upper left corner, which made it made it necessary to extend the spectral region in both dimensions. The strong low-field shift of the H2 hydrogen (8.82 ppm, see Tab. 4) can be explained by deshielding through the adjacent, positively charged nitrogen, which adds to the already observed ring current deshielding in aromatic systems.

The other unusual shift, that should be noted, is the heavily high-field shifted cross-peak A8, which is now in the center of the H6/H8 to H1' region. The signals of adenine H8/H1' are usually found on the left, which is true for the remaining adenines A6, A16 and A24. The explanation for the shift can be derived from Fig. 4.17 of the solution structure, where a close view on the three central base pairs is shown. Here one can see that the five-membered ring of the adenine A8 stacks very well with 6HQ, which shields the magnetic field in the vicinity of the adenine H8, causing a high-field shift of the A8 cross-peak. In contrast to this, a usual chemical shift is observed for the H8 hydrogen of adenine A6. The reason for this is that the stacking with 6HQ is centered around the six-membered ring of adenine and hence far away from the H8 hydrogen.

The chemical shift deviations were calculated the same way as for 13merHCF, but this time compared to native strands, which contain either GC or AT as central base pair. One has to consider that the complementary strand is now a native strand, which is furthermore identical to the corresponding strand in the reference DNA with GC in the center (13merRef(GC)). It is then not surprising that the CSDs between the complentary strand and 13merRef(GC) are relatively small and the only larger difference to the AT reference strand (for convenience 13merRef(AT)) occurs at residue 20, where the cytosine of 13mer6HQ is compared to a thymine in 13merRef(AT). The CSD pattern for the 6HQ strand is similar to 13merHCF (see Fig. 4.3). When taking into account that large CSD at position 7 is mainly caused by the completely different backbone (GNA, DNA) and the different nucleobases (6HQ, G and A), one can see that 13mer6HQ shows similar behaviour in the CSDs for the residues 5,6 and 8 . Since the reference strands are known to comprise

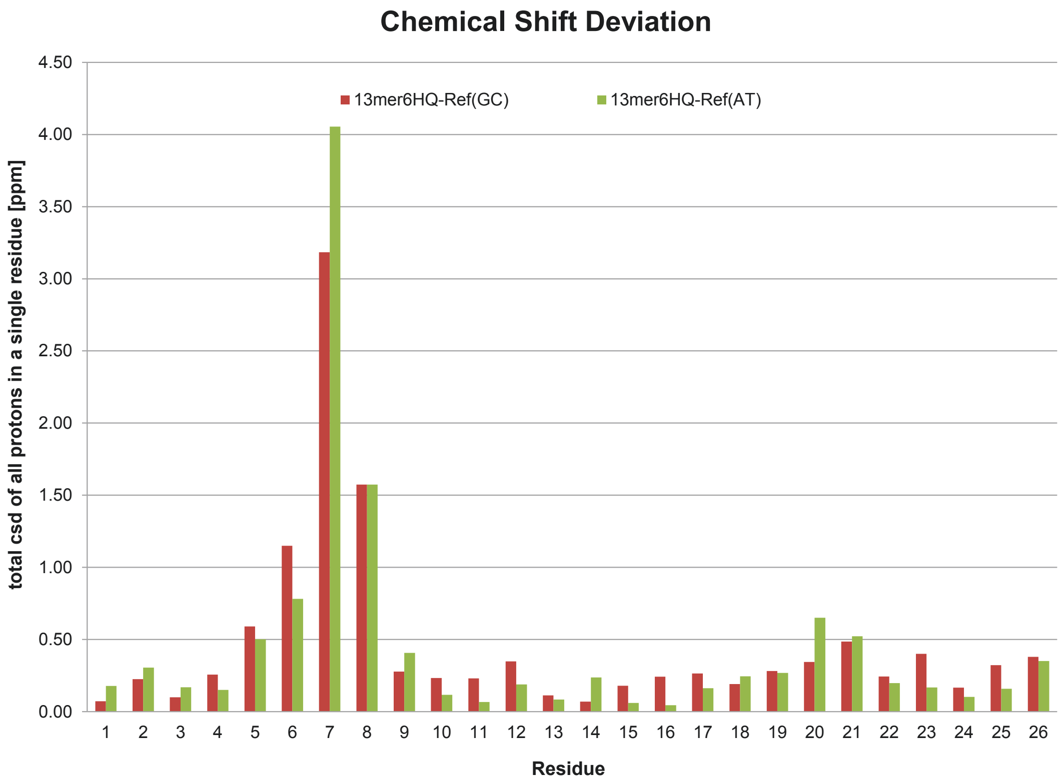


Figure 4.15: CSD comparison between 13mer6HQ and different reference strands, which contain either GC or AT as central base pair. Chemical shift differences of all protons belonging to a single residue but different DNA samples were summed and are given as absolute values.

usual B-DNA helices, this may indicate structural deviation in the 6HQ strand but not in the complementary strand.

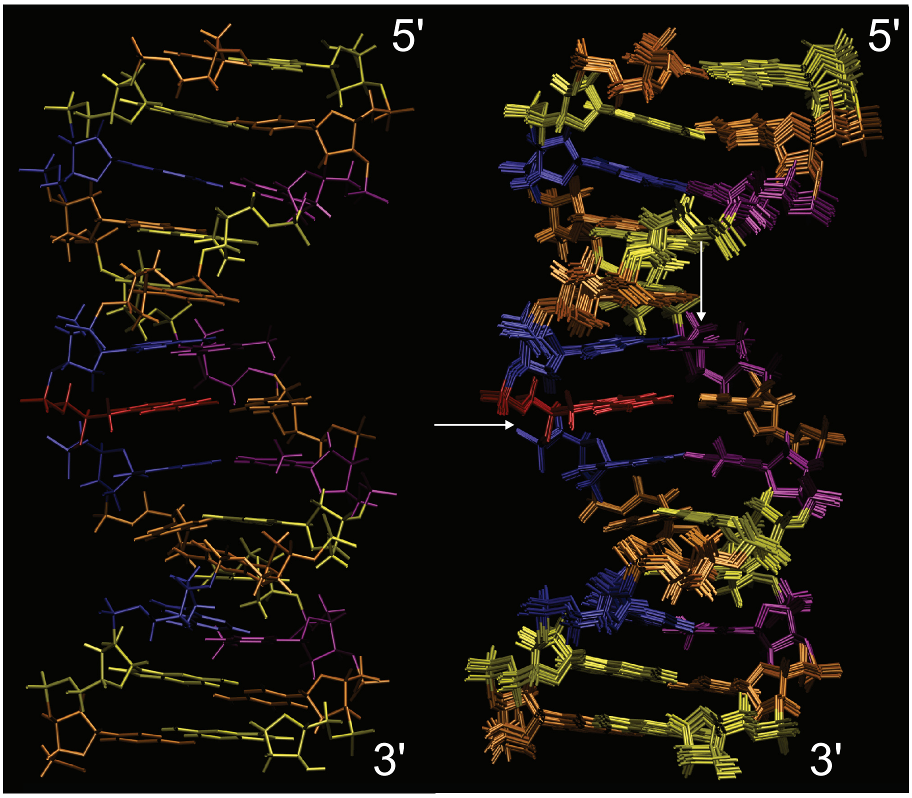


Figure 4.16: Averaged structure of 13mer6HQ. The averaged structure on the left was constructed from the 10 best-energy and violation-free structures on the right. The 6HQ is marked red, while the corresponding colors for A, G, C and T are blue, yellow, orange and violet.The vertical arrow indicates the point of view in Fig. 4.6 and the horizontal one in Fig. 4.19. RMSD among all atoms besides methyl protons is 0.36 Å.

4.2.2 NMR solution structure

The NMR solution structure was determined from experimental NOE and residual dipolar coupling data (see sec. 3.1.9). All NMR resonances could be assigned with the exception of some severely overlapped H5'/H5" signals. Integration and conversion yielded 418 distance restraints, which were used in a first step to generate a start structure. Afterwards, a total of 25 $^1J_{CH}$ RDC restraints were included into the refinement using a single floating alignment tensor. The 10 best-energy, violation-free structures out of 100 calculated were

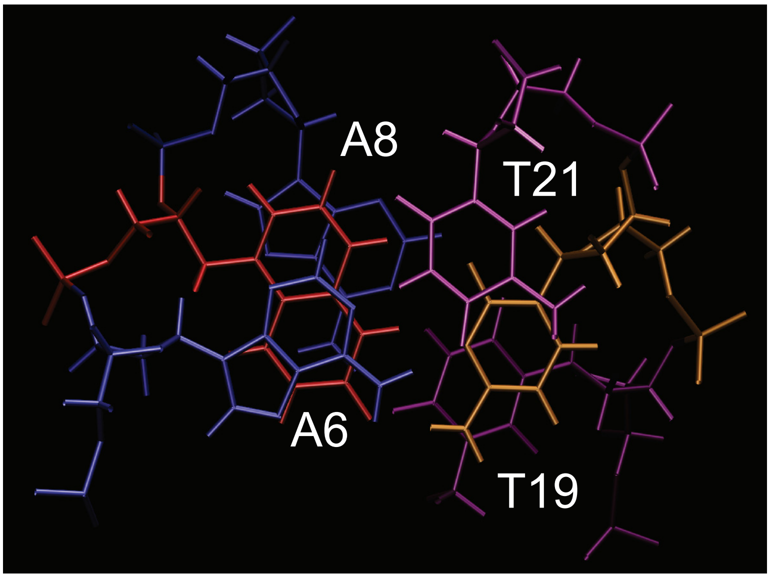


Figure 4.17: Close view on the three central base pairs of 13mer6HQ. The 6HQ moiety stacks well with the five-membered ring of adenine A8 and the six-membered ring of A6.

used to construct the averaged structure, which is shown in Figure 4.16. The root-mean-square-deviations (RMSD) among all hydrogens besides methyl protons are 0.36 Å(see sec. 3.2.6).

The deprotonated 6-Hydroxyquinolinium fits perfectly into the helical fold (see Fig. 4.16 and 4.17), other conformations, like in 13merHCF, were not observed. The distance of 1.79 Å between the 6HQ oxygen and amino H42 of cytosine indicates the desired hydrogen bond (Fig. 4.18), thus forming an artificial base pair. Interestingly, 6HQ does not face the cytosine with its short side (along carbons C6-C7), like the purines (A,G) do, it presents instead the long side (carbons C3 to C6). Two reasons are conceivable, which may complement each other. One is the linkage to the backbone, the position of hydrogen N9 in the five-membered rings of purines is more compatible to the C2 position in 6HQ than to nitrogen N1. Moreover, glycerol is missing a bond in comparison to the size of 2-deoxyribose, which could be denoted in terms of sugar nomenclature as attachment of 6HQ to C2' and not C1'. However, the missing bond is balanced by the size and direction

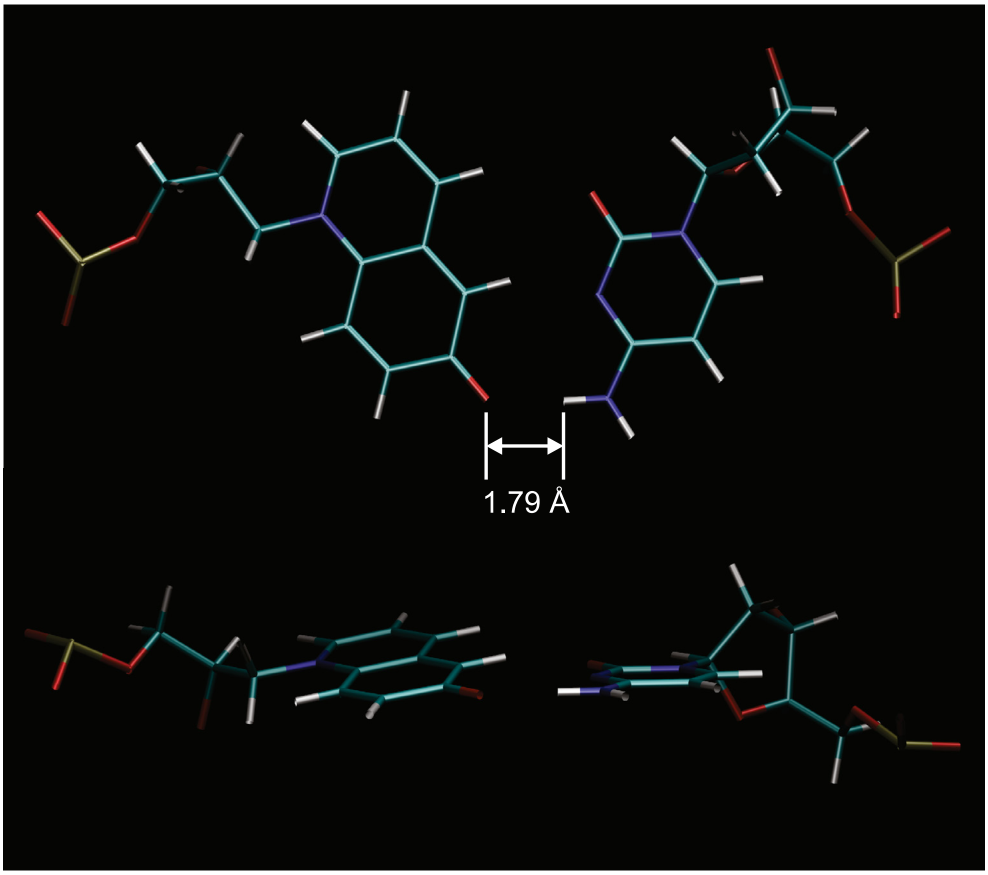


Figure 4.18: Close view on the central base pair of 13mer6HQ. 6HQ forms a base pair with cytosine and a single hydrogen bond. Interestingly, 6HQ presents its long side to the cytosine and not the short one along carbons C6-C7.

of 6HQ.

Figure 4.19 is a side view of the three central base pairs as indicated by the vertical arrow in Fig. 4.16. The white arrow marks the helical long axis of the double strand, while the red arrow indicates the deviating direction of the stacking in adenine A6, 6HQ and A8 (in front). The nucleobases in the back (T19, C20, T21) are not affected and stack along the helical axis. It is known that double stranded GNA shows a strong inclination between backbone and nucleobase, which might explain the observation that neither the R- nor S-enantiomer of GNA cross-pairs with DNA [275]. However, the 6HQ strand only comprises a single GNA monomer and it has been demonstrated that 6HQ forms a base pair with

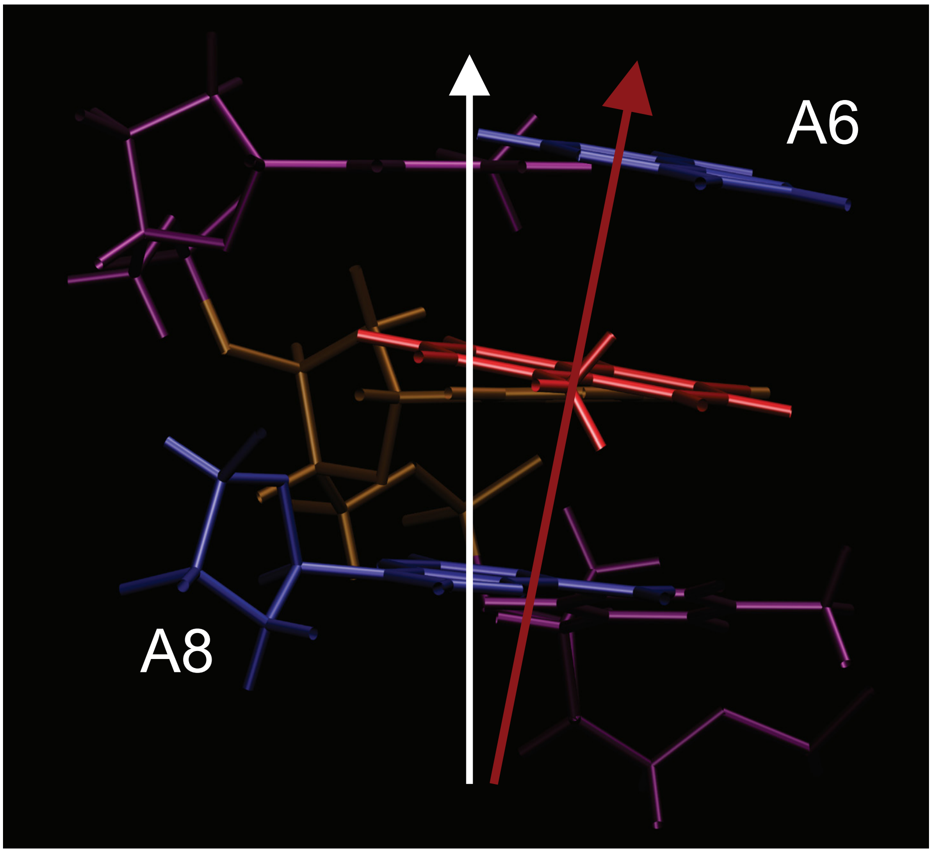


Figure 4.19: Side view on the three central base pairs of 13mer6HQ. The face-up orientation is again at left and face-down at right. The base A6 - 6HQ - A8 are in front and the corresponding T21 - C20 - T19 in the back. The short length of glycerol in comparison to a 2-deoxyribose backbone leads to a differing stacking axis (red arrow), which is in usual B-DNA identical to the helical axis (white arrow).

cytosine, so this cannot be the explanation. Unfortunately, the glycerol backbone provides only three bonds between O3' and O2', which means that a second bond compared to 2-deoxyribose is missing. This time, the distance is covered by a small inclination between the the sugar of the adjacent adenines and the helical long axis. Consequently, since the nucleobases retain their roughly perpendicular angle relative to the sugar, they form the differing stacking axis with 6HQ. A comparison between 13mer6HQ and our reference DNA 13merRef[129], where the central base pair is A:T, shows that the distances between

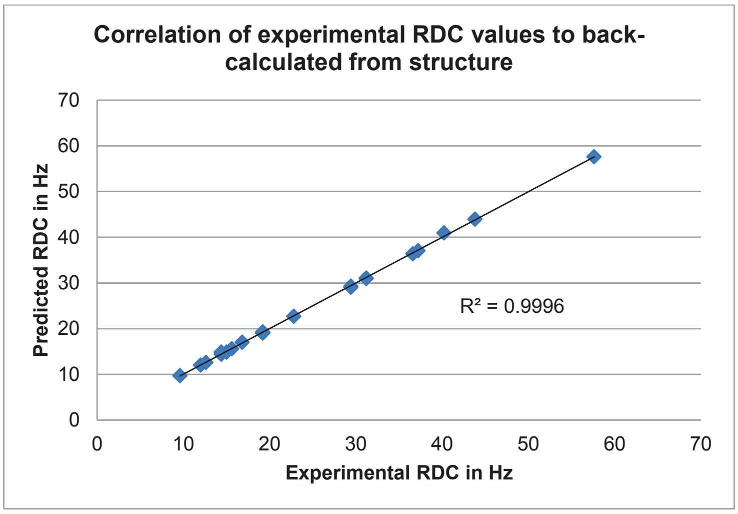


Figure 4.20: Pales[268] plots of experimental RDCs against predicted values of the 13mer6HQ structure. The predicted values, which were directly derived from the averaged structure, agree well with the experimental RDCs.

C3' of adenine A6 and C5' of adenine A8 are 10.18 Å in 13merRef and 8.98 Å in 13mer6HQ, respectively. The resulting difference of 1.2 Å and the differing stacking axis prove that the missing bond length is crucial for the B-DNA backbone.

The Pales[268] plot in Fig. 4.20, which compares experimental RDCs with predicted RDCs derived from the average structure, confirms that the obtained structure is part of the experimental data.

Finally for this section, the validity of the 13mer6HQ average structure will be checked using the back-calculated spectrum in Fig. 4.21. The experimental NOESY spectrum (D_2O) is depicted in black and the back-calculated in red, respectively. On the left half, a series of back-calculated cross-peaks can be found that appear not to be present in the experimental spectrum. In fact, the signals are present, but suffer from a low signal-to-noise ratio. More interesting is the second C26-H6/H5 cross-peak on the right, which is shifted to higher field in both dimensions, due to unstacking from the helix. The occurrence of this cross-peak indicates fraying at the helix ends. The other cross-peaks of the back-calculated spectrum agree well with the data.

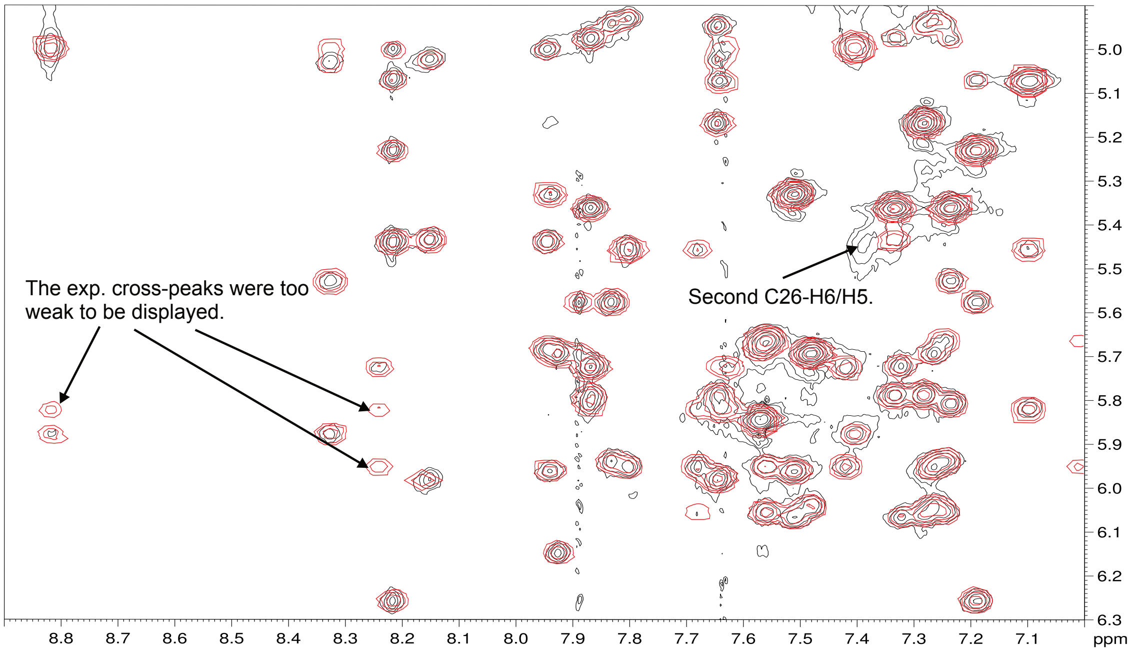


Figure 4.21: NOESY back-calculation (red) overlaid with experimental data (black). The left side shows back-calculated cross-peaks, which are apparently missing in the experimental data. In fact, weak signals are present in the data, but suffer from a low signal-to-noise ratio. The cross-peak on the right marks the unstacked H6/H5 signal of C26, which indicates fraying of the helix ends.

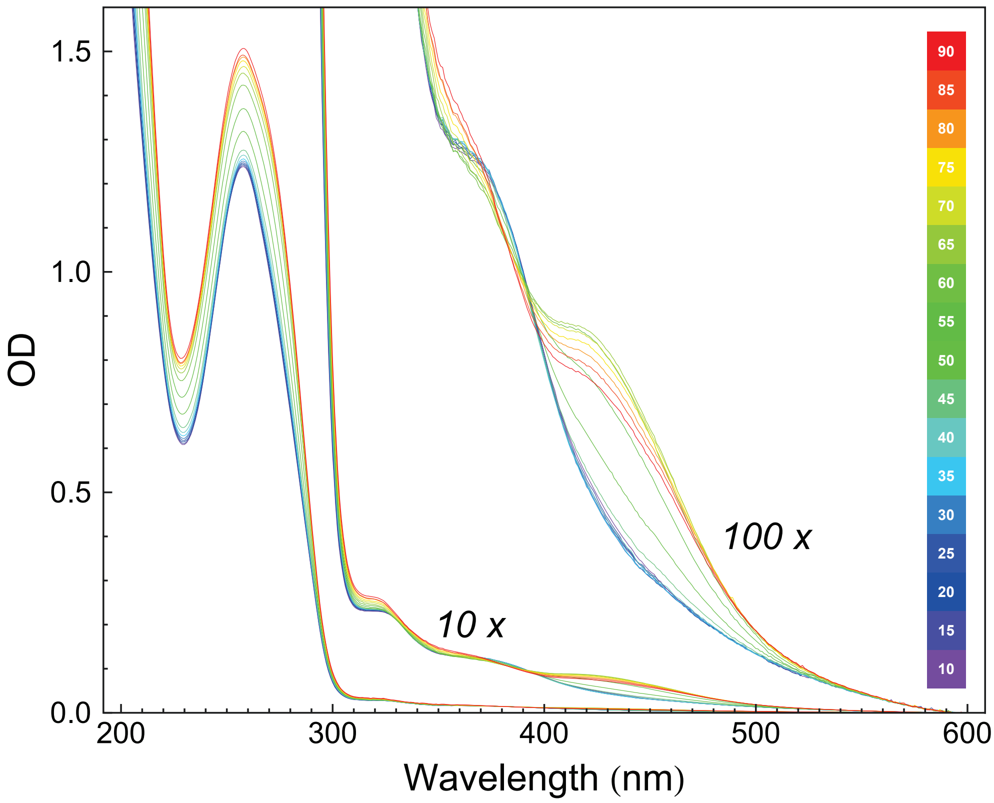


Figure 4.22: Absorption change of ds13mer6HQ upon melting, when raising temperature from 10 °C (violet) to 90 °C (red line; c_T = 131 µM). Data from the 1 and 10 mm cuvettes are labelled “1x” and “10x”, respectively. A ten-times magnification of the latter is labelled “100x”.

4.2.3 Duplex melting experiments

A sample of double-stranded 13mer6HQ was prepared in a water/ammonia mixture with $pH = 8.5$ and salt concentration of 150 mM sodium chloride. Measurements were performed in a double cuvette (closed but not evacuated) with 1 and 10 mm optical path length. A total concentration $c_T = 131$ µM of single strands was estimated from maximum absorption around 260 nm at 90 °C. A set of 17 spectra between 10 and 90 °C were recorded and corrected for density change. Results are shown in Fig. 4.22, where the data from the 1 and 10 mm cuvettes are labelled “1x” and “10x”, respectively. A ten-times magnification of the latter is labelled “100x”.

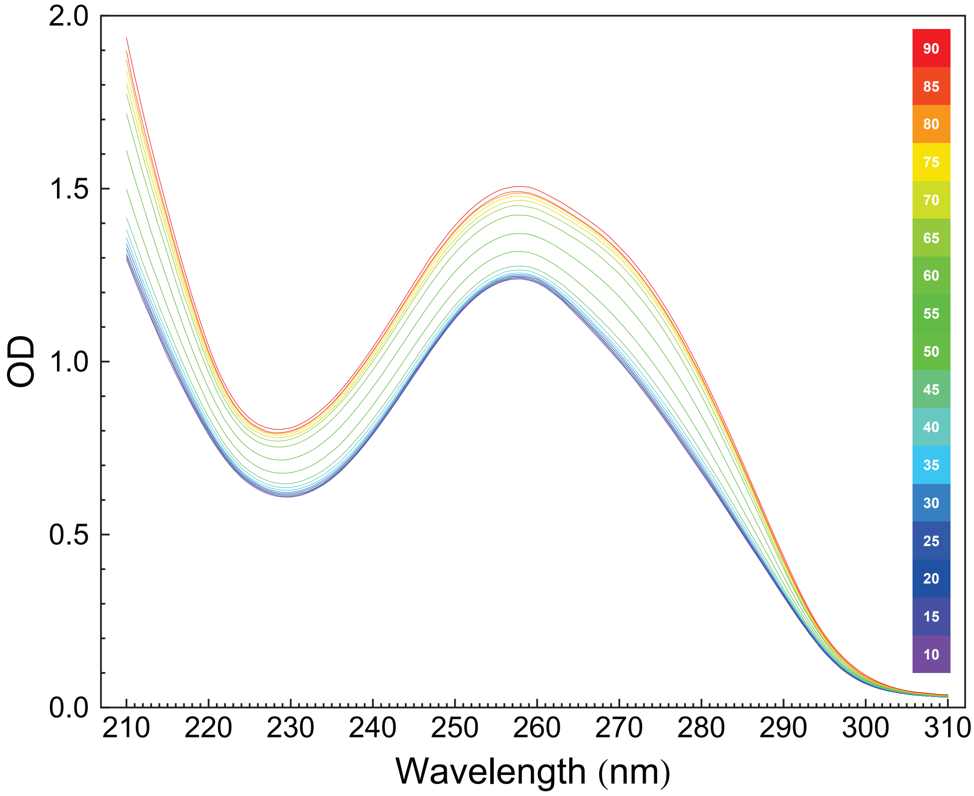


Figure 4.23: The UV part of the spectrum was used for SVD of the DNA absorption peak. The first spectrum is violet (10 ℃) and the last one is red (90 ℃).

Analysis of the 13mer6HQ DNA band

The changes of the DNA peak with increasing temperature (10 - 90 ℃) were quantified in a spectral window from 300 to 210 nm (see Fig. 4.23). The UV absorption spectra in this range were analyzed for principal components via Singular Value Decomposition (SVD). Two components are needed to describe the hyperchromism. Fig. 4.24 shows the basic spectra on the left and the corresponding thermodynamic curves on the right. The first component (red) was set to be the 90 ℃ spectrum. Note that the thermodynamic curves were multiplied with a scale factor so that the amplitudes fall into the range from 0 to 1 (the corresponding basic spectra were scaled with the inverse factor.) Upon duplex formation, the red one shows an amplitude decrease of 0.24. By design, the blue curve shows the same amount as amplitude increase (Fig. 4.24).

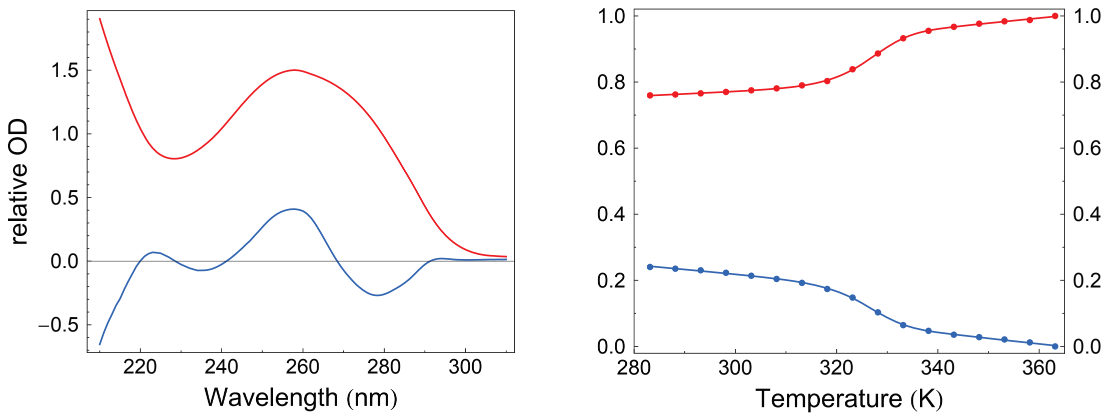


Figure 4.24: Basic spectra are shown left. The first component (red) was set to be the 90 ℃ spectrum, i.e. it represents the fully separated single strands. The right part shows the thermodynamic curves related to the basic spectra. Upon hybridization, the amplitude of the first-component spectrum (red) decreases from 1 to 0.76 while that of the second component increases from 0 to 0.24.

Like before (chapter 4.1.3) a two-state model with stacking interactions[19] was applied to the thermodynamic data (Fig. 4.25). It assumes that the temperature (T) dependence of the extinction coefficients is linear ($\epsilon_{[ss]} = m_{[ss]}T + b_{[ss]}$) and, in addition, different for the single- (ss) and double-stranded (ds) forms. It is not necessary to fit the linear parts separately, therefore Equation 2.33 can be directly applied (as introduced in 2.7) to the data (dots) shown in Fig. 4.25. The thermodynamic parameters for the red line are ΔH° = -320 ± 20 kJ/mol, ΔS° = -900 ± 50 J/(K mol) and ΔH° = -330 ± 30 kJ/mol, ΔS° = -930 ± 70 J/(K mol) for the blue one, respectively. The values for both lines are equal in the range of their errors. A unique melting point of 53 ± 0.1 ℃ is noted at the total concentration c_T = 131 µM of single strands. Remember that in case of 13merHCF a second melting point was observed, reflecting the HCF chromophore in its environment. The 53 ℃ of the 13mer6HQ DNA absorption band can be compared with the 63.3 ℃ of 13merRef[129] (c_T = 23.7 µM, central A-T). The observed thermal destabilization, corresponding to a melting-point decrease of 10.3 K, is in fact a minimum value. This is because the total strand concentration of 13mer6HQ (owing to the weak absorbance of the 6-Hydroxy-quinolinium chromophore) is more than 5 times higher than

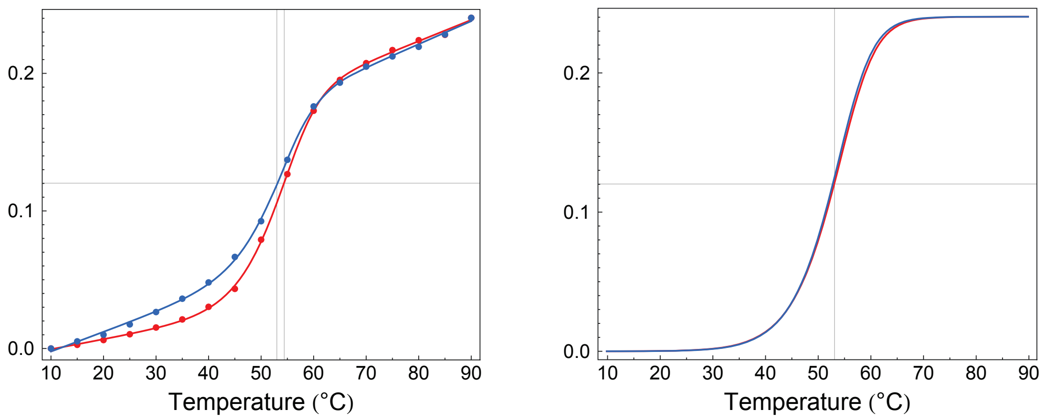


Figure 4.25: Thermodynamic curves, scaled to the measured amplitude change of 0.24. The left panel shows SVD data (points) and fits with a two-state model including stacking (lines). The two-state model part is plotted separately (right panel), representing a melting curve of hyperchromism without the influence of stacking. In contrast to the initial curves on the left, a unique melting point of 53 °C for both principal components is estimated (grey vertical lines).

was used for 13merRef. A melting point of 67 °C can be estimated for 13merRef at c_T = 131 µM (of 13mer6HQ) by using the Nearest Neighbor Model[255], the corresponding melting-point decrease is then in a range of 14 K.

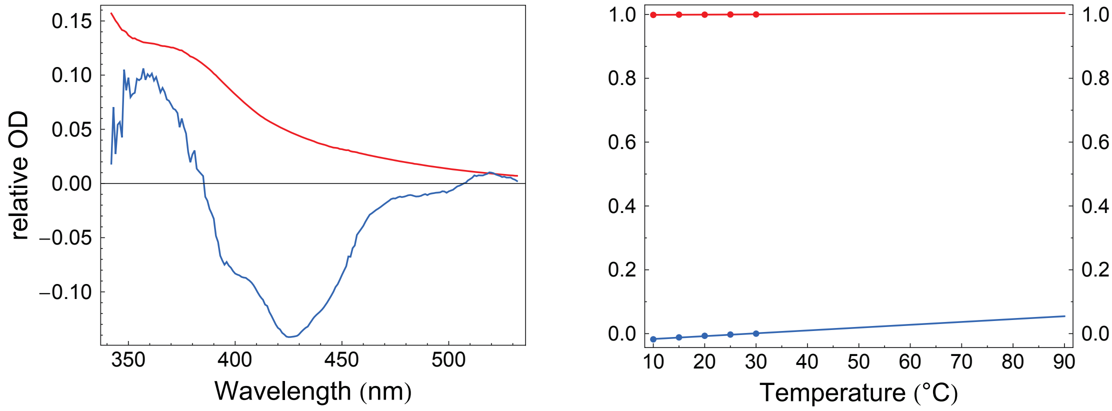


Figure 4.26: Basic spectra for the duplex part are shown left. The first is arbitrarily set to the absorption spectrum at 30 °C (red). The right part shows the corresponding amplitude curves. Dots represent the experimental data and lines the fit and extrapolation up to 90 °C.

Dissection of the 6-Hydroxy-quinolinium band

The approach for this part is similar to the previous description, and the main goal is to observe the hybridization of the chromophore in the visible spectral region. For this purpose the spectral range, which was subjected to analysis, was restricted to $\lambda > 340$ nm where the absorption by the quinolinium chromophore should be dominant. The first examination via SVD revealed at least three principal components. They all show a melting curve with negative slopes for the linear parts. Inspection of the basic spectra led to the conclusion that two different species are involved in the melting process, since the change of the chromophore absorption could not be explained by simple hyperchromism with blue shift of the peak, as observed before with 13merHCF. At this point the new approach of the double SVD-assisted two-state model was developed. The theoretical background was already described in 2.7. Since melting/hybridization occurs in a small temperature range around 50 °C, one can try to understand the behaviour of the duplex separately at low temperature, and that of the fully separated single strands at high temperature. In each of these two limiting temperature intervals, a SVD analysis of the absorption spectra in Fig. 4.22 is made. Knowing the change of absorption with temperature for both, an extrapolation can be done from either side into the middle, i.e.

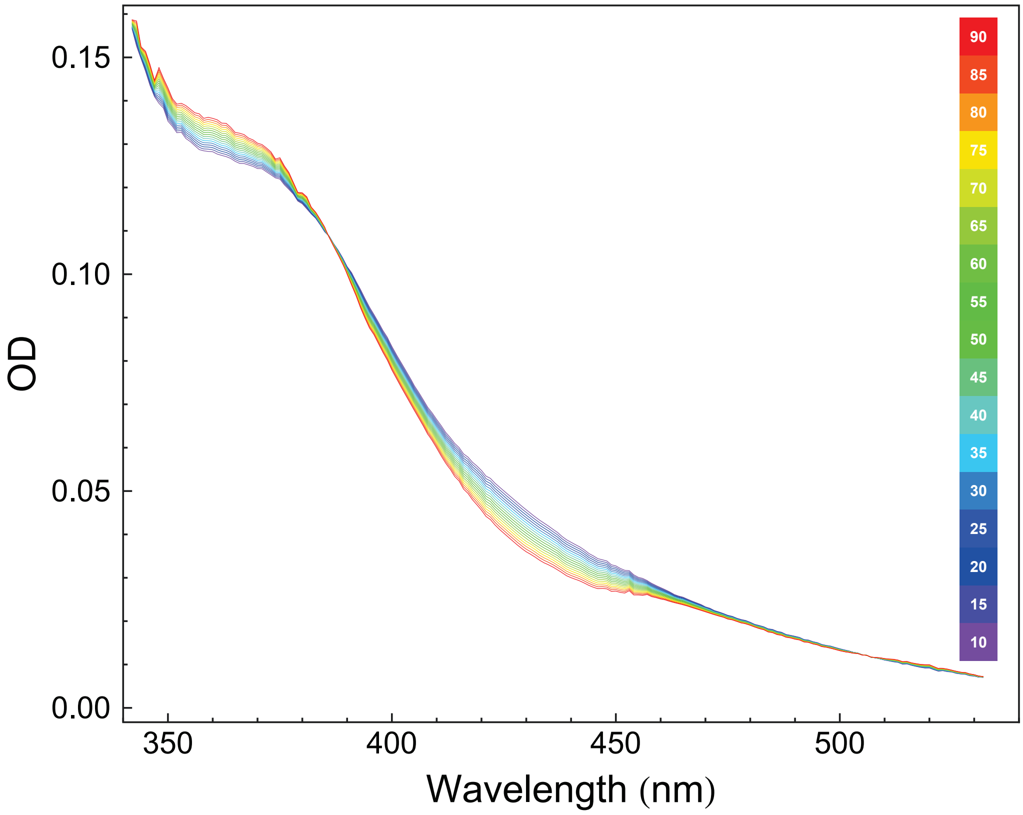


Figure 4.27: From the known duplex region (10-30 °C) onwards, the next twelve spectra were extrapolated (35 up to 90 °C) to simulate the behaviour of a non-melting double strand.

in the direction where strand separation occurs. For example, the absorption spectrum of the double strand can be simulated for 60 °C where, in reality, the separation into single strands is almost complete. At any given temperature, the observed spectrum should be a linear combination of the two extrapolated spectra, i.e. of the separate strands and of the duplex. From the corresponding fit the degree of dissociation α is obtained. In other words, extrapolations with the help of basic spectra (from SVD) replace the linear parts of the two-state model used above. As a result we obtain a dissociation curve which is shown in Fig. 4.31 below.

The first step on that way is the extrapolation of the duplex part. The spectra from 10 to 30 °C were used, and they could be described by two principal components in the

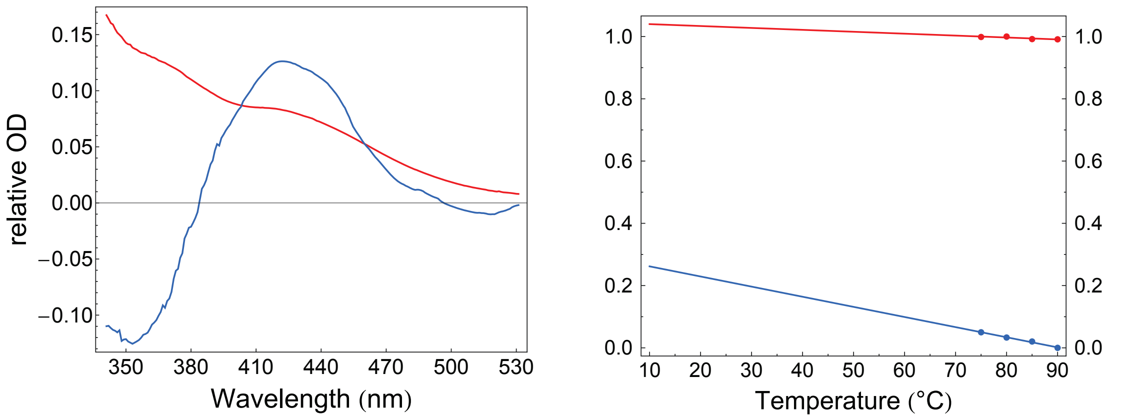


Figure 4.28: Basic spectra and amplitudes as in Fig. 4.26, but for the separated single strands. The first basic spectrum is taken to be the absorption spectrum at 70 °C (red).

SVD. Both thermodynamic curves can be described by straight lines, which are then used to extrapolate the thermodynamic behaviour up to 90 °C (Fig. 4.26). The second principal spectrum in Fig. 4.26 was smoothed (by a Wiener filter [276]) to reduce noise when extrapolating to high temperature. Fig. 4.27 shows the resulting change of absorption for the duplex, as extrapolated for the whole temperature range.

The same procedure was then performed for the single strands, using the last 4 spectra, i.e. for temperature from 70 up to 90 °C. In this way Figs. 4.28 and 4.29 are obtained (analogues to Figs. 4.26 and 4.27, respectively). A look at the extrapolated spectra (Fig. 4.27 and 4.29) reveals a similar behaviour compared to the HCF chromophore, which underwent a blue shift and hyperchromism on the blue side of the spectrum upon melting. In addition, the missing absorption band around 430 nm in the duplex state spectra (Fig. 4.27) is now present in the extrapolated single strand spectra (Fig. 4.29). The band indicates the deprotonated state of 6HQ, due to the basic buffer (pH = 8.5), and disappears when the 6HQ-C base pair and hence the hydrogen bond is formed.

As last step before the two-state model can be applied, a linear combination based on equation 2.34, where α is the fraction of extrapolated single strand spectra $S_{[ss]}(T)$, is used to build the measured spectra $M(\alpha, T)$ around 400 nm. In other words, the linear equation terms of (2.33) are substituted with the extrapolated spectra $S_{[...]}(T)$.

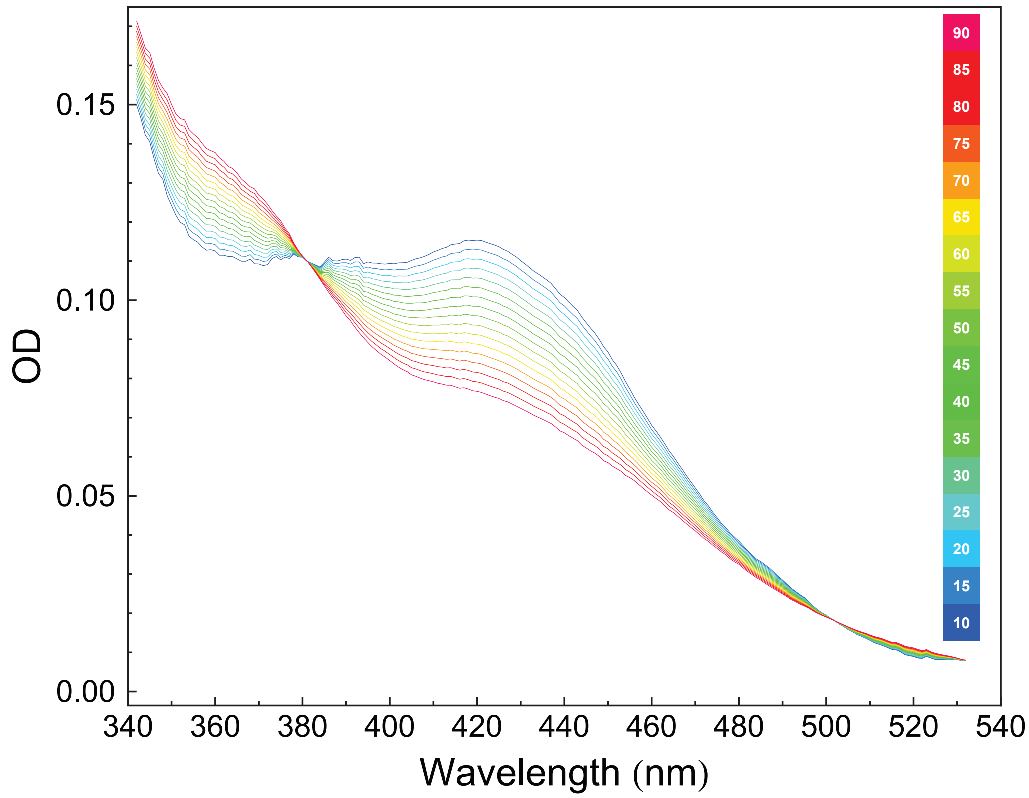


Figure 4.29: The known region is now the high temperature part (70-90 °C), where fully separated single strands are assumed. In this case, extrapolation from 65 °C down to 10 °C simulates two strands which cannot hybridize at lower temperatures.

The α values produced, using equation (2.34), are plotted as dots in Fig. 4.31. Hence the extrapolated spectra are already scaled to the measured data; the two-state model defined in equation (2.32) can be directly applied (line in Fig. 4.31). Best fit parameters are ΔH° = -365 $\pm$ 5 kJ/mol and ΔS° = -1030 $\pm$ 20 J/(K mol) for the hybridization. The melting point is 0.1 °C higher (53.1 $\pm$ 0.1 °C) than the 53 °C for the complete strand, which is a negligible difference. The fact that the melting points for the whole strand and the built-in chromophore are the same and more than 10 K lower than in 13merRef gives no evidence for local melting (e.g. bubble formation) and is in good agreement with the NMR solution structure. Figure 4.19 shows the differing stacking axis in the middle of the modified strand, caused by the short glycerol linker. In combination with the limited

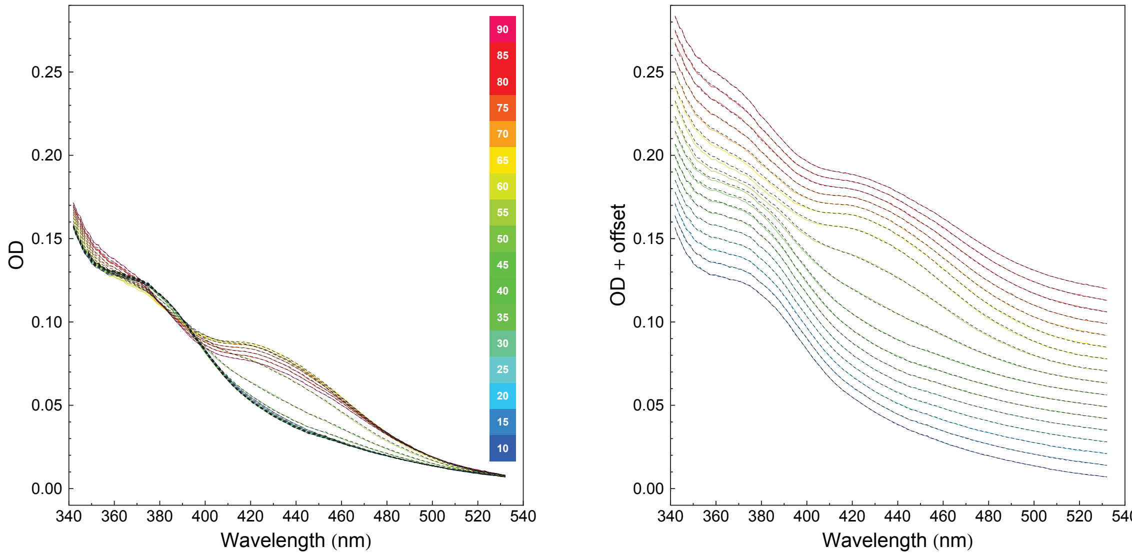


Figure 4.30: Overlay of linearly combined spectra (dashed black lines, eq. (2.34)) and measured data (blue 10 - 90 °C red) around 400 nm (10 mm path length). On the right, for a better pairwise comparison, each data curve and its linear combined spectrum is plotted with an offset.

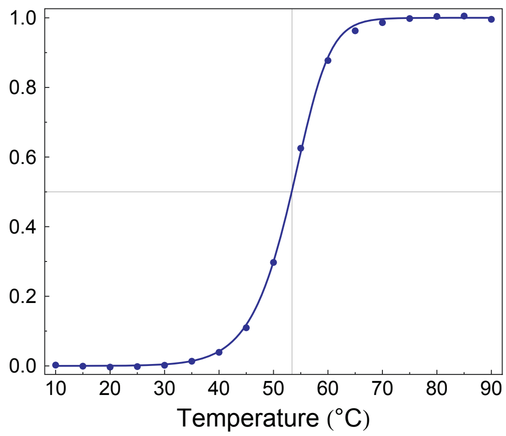


Figure 4.31: The α values of the linear combined spectra are shown as points, whereas the line represents the applied two-state model for α. The resulting melting point is 53.1 °C.

flexibility of the small linker, this promotes melting of the strand as a whole at lower temperature. Nevertheless, the designated 6HQ-C base pair was formed and proven by the missing absorption band at 430 nm in duplex state, which in turn became accessible by the new analysis method.

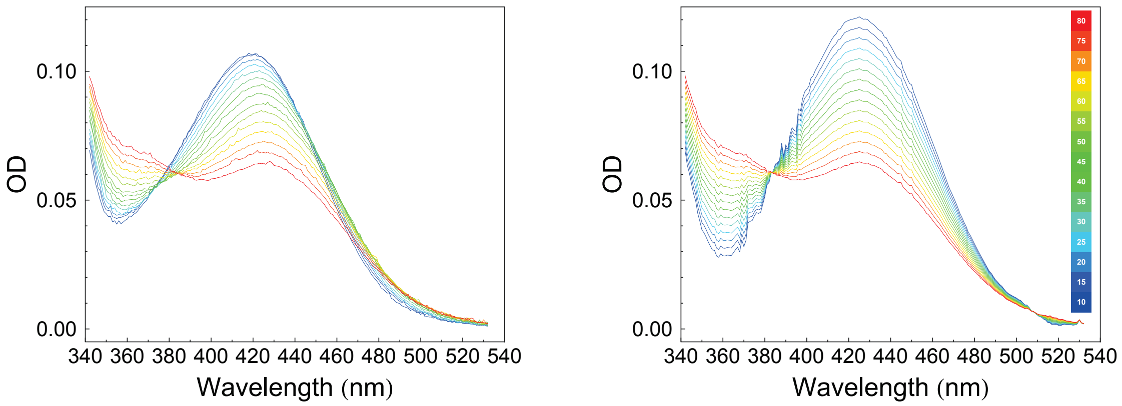


Figure 4.32: The absorption spectrum of single-stranded 13mer6HQ, measured from 10 (blue) to 80 °C (red), is shown on the left hand side. The extrapolation on the right was performed in the same way as the separated strands form of ds 13mer6HQ. Therefore, the 70 °C spectrum (orange) was set to be the first basic spectrum and all spectra with higher temperature were used as dataset for the extrapolation to lower temperatures.

Comparison of the measured 13mer6HQ single strand with its extrapolated spectrum

A single strand of 13mer6HQ was measured prior to the double strand. The data can be used to compare our SVD-assisted extrapolation on the isolated single strand (see right part of Fig. 4.32) with its real spectra at lower temperatures. The aim is to get an idea of the robustness and possible limits of the extrapolation. The extrapolation was performed the same way as on the single strand form of ds 13mer6HQ. Therefore, the 70 °C spectrum (orange) was set to be the first basic spectrum, which is in addition the first data point where the fully single stranded form of ds 13mer6HQ can be assumed (see Fig. 4.31). In contrast to the five high temperature data points used in Fig. 4.28, only three spectra were available in this case, due to end of data acquisition at 80 °C.

Figure 4.32 shows measured (left) and extrapolated data (right) next to each other. It is easy to notice that the extrapolation increasingly differs from the measurement at lower temperatures. Due to the fact that in an eventual application the duplex form would be dominant at lower temperatures, the extrapolation only needs to be exact for spectra which are largely dominated by the single stranded form. The overlay in figure 4.33 allows a direct comparison between the measured and extrapolated data. From 80 down to 55

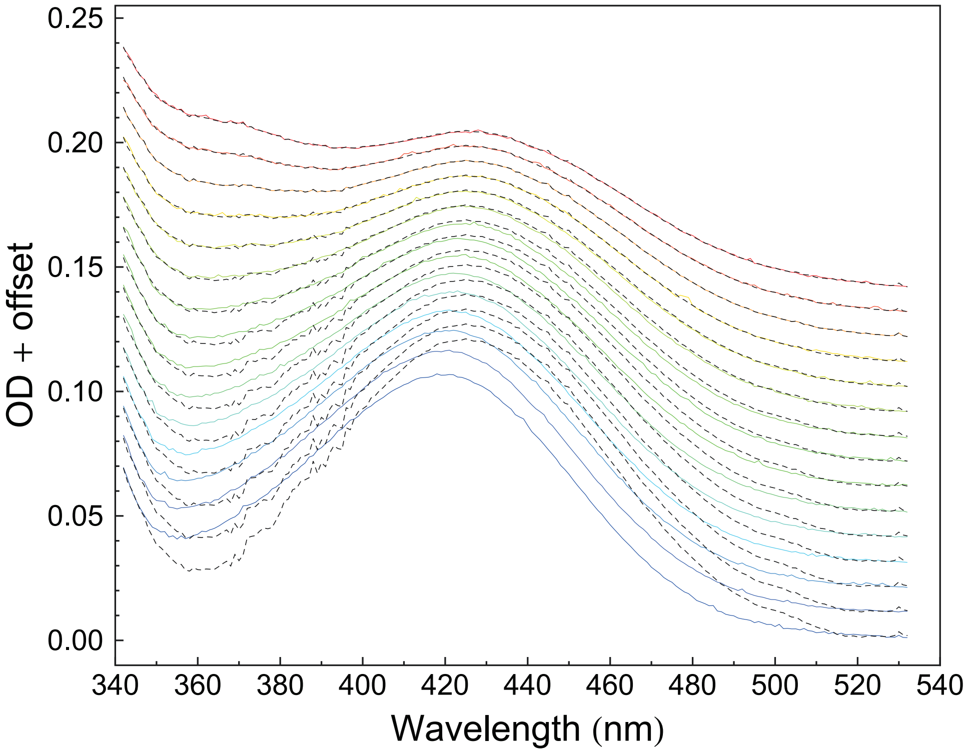


Figure 4.33: Overlay of the extrapolated spectra (dashed black lines) and measured data (blue 10 °C to red 80 °C). For a better pairwise comparison, each data curve and its extrapolated spectrum is plotted with an offset. From 80 down to 55 °C (6th line from top) there is a good agreement. Deviations become relevant at 50 °C (7th from top) and rise more and more when going downwards. However, the effect on the linear combined spectra (in Fig. 4.30) is small, when taking the melting point of the double strand (53.1 °C, Fig. 4.31) into account, which states that for 50 °C and below the duplex form is dominant.

°C (6th line from top) there is a good agreement. The deviations become visible in the 50 °C spectrum and rise more and more when going downwards. However, at 50 °C the duplex form is dominant and the upcoming deviations from the measurement cannot have a visible effect on the linear combination of extrapolated spectra (see Fig. 4.30, the same should be observed for the duplex part going to high temperature). In other words, the transition of the DNA eliminates the rising error of the extrapolations as long as they are both correct around the melting point.

4.3 4-Aminophthalimide and 2,4-Diaminopyrimidine - 13mer4AP-DAP

The 13mer4AP-DAP double strand was developed in cooperation with the group of H.-A. Wagenknecht (Karlsruhe Institute of Technology). The artificial base pair in the center comprises 4-aminophthalimide (4AP) and 2,4-diaminopyrimidine (DAP) as base surrogates. The 4AP chromophore is comparable in size to natural purines and shows remarkable solvatochromicity, red-shifted fluorescence in polar solvents, and hydrogen bonding capabilities[205]. Therefore DAP was designed as a potential counterbase, offering three potential hydrogen bonding sites. Michael Weinberger (Wagenknecht group) was responsible for the synthetic route of 4AP and Falko Berndt (Ernsting group) correspondingly for DAP. Both synthetic routes[205] utilized stereoselective Heck-type palladium-catalysed cross-coupling with 2'-deoxyribofuranoside glycal followed by stereoselective reduction with $NaBH(OAc)_3$. Both nucleosides were further processed to the corresponding phosphoramidites and subsequently incorporated via automated DNA synthesis.

The finally obtained double strand of 13mer4AP-DAP required some alterations in the process of structure determination. As stated in the introduction (sec. 1.3), the hydrolysis of the 4AP chromophore in water had to be addressed, especially under basic conditions. A weak acidic buffer was chosen (pH 6.35), which extended the lifetime of the 4AP mononucleotide from hours to a few days. As a consequence of this, it was discussed that DAP might be protonated by the acidic buffer. A subsequent measurement of 2,4-diaminopyrimidine absorption at different pH values revealed a pK_a of 7.4 ± 0.1. Hence the picture of the 4AP-DAP base pair presented in the introduction (Fig. 1.3) was changed to that of Fig. 4.34. Furthermore it should be noted that only a small amount of the 4AP strand was available, so that the same sample had to be used in all experiments. In order minimize the time and sample preparation effort, a closed double cuvette was used that offers 1 and 10 mm optical path lengths at the same time. In addition to the customized cuvette, the order of experiments was reversed, so that the optical experiments (absorbance, fluorescence etc.) were measured first, since their overall

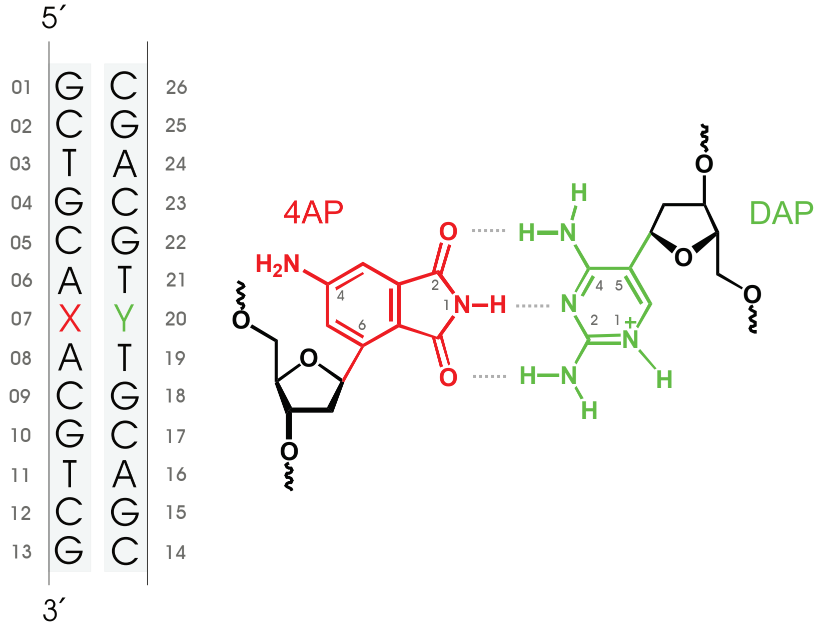


Figure 4.34: Structure of 13mer4AP-DAP. The 4AP moiety (red) was placed at position X and the DAP (green) at Y, respectively. Please note the additional hydrogen in DAP.

acquisition time was shorter than that of a single NOESY experiment. Moreover, it was doubted that the sample amount would be sufficient for structure determination, since the theoretically achievable maximal concentration was only 1.5 mM in a Shigemi NMR tube (250 µl). In contrast to this, all previous samples were measured with at least 3 mM concentration. The limiting factor for the sample concentration is the RDC measurement, which requires much more nuclei in the center of the magnetic field than the NOESY experiment, due to the principal concept of measuring *residual* dipolar couplings. Even though that RDCs became a standard tool in structure determination, they are not required to perform restrained Molecular Dynamics. As outlined in the introduction, this method was invented to use only NOE distance restraints which is, of course, still possible. So it was decided to omit the RDC experiment for the price of some quality, due to the fact that RDCs can offer structural information without suffering from the limited range of NOEs (< 5 Å). In the first cycle only a minimal set of NMR spectra was measured at 10 °C directly after the last optical experiment, just in case that 13mer4AP-DAP is not

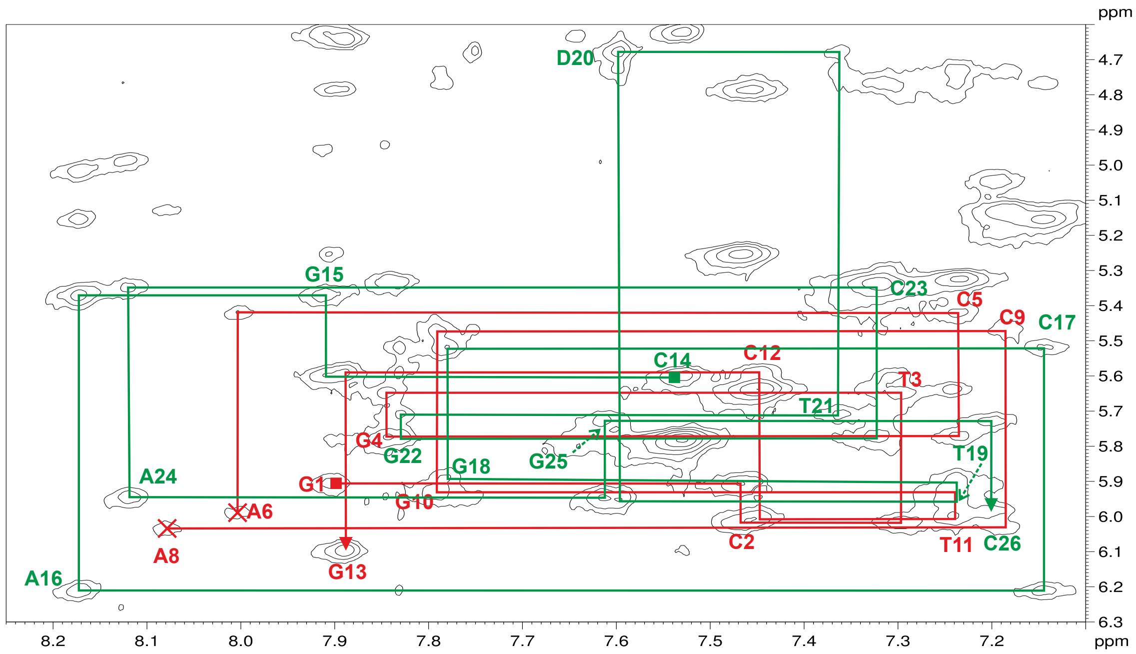


Figure 4.35: NOE-Walk in 13mer4AP-DAP. The assignment of the H6/H8 (abscissa) and H1' (ordinate) region in the D_2O-NOESY spectrum is shown. Red lines mark the NOE-walk for the 4AP-DAP containing strand and green lines for the DAP strand. The starting point (5'-end) of both walks was denoted by a square, while the endpoint (3'-end) is indicated by an arrowhead. Note that the 4AP walk is interrupted at the modification site (marked by X), while DAP (D20) is high-field shifted. For clarity only the positions of intraresidual H6/H8-H1' cross-peaks were labeled.

stable in the long term. Fortunately the sample remained stable, but the finally achieved concentration was only 1.2 mM. However, the gained time window allowed for extensive optimization of the pulse sequence and a high number of transients during the NOESY experiment.

4.3.1 Chemical shift analysis

The assignment was achieved by standard methods[19,213] that were already described for 13merHCF and 13mer6HQ. In Figure 4.35 both NOE walks are shown. The walk of the 4AP strand (red) is interrupted at the 4AP site (like the HCF strand), while the complementary strand resembles the 6HQ strand with shifted signals at the DAP

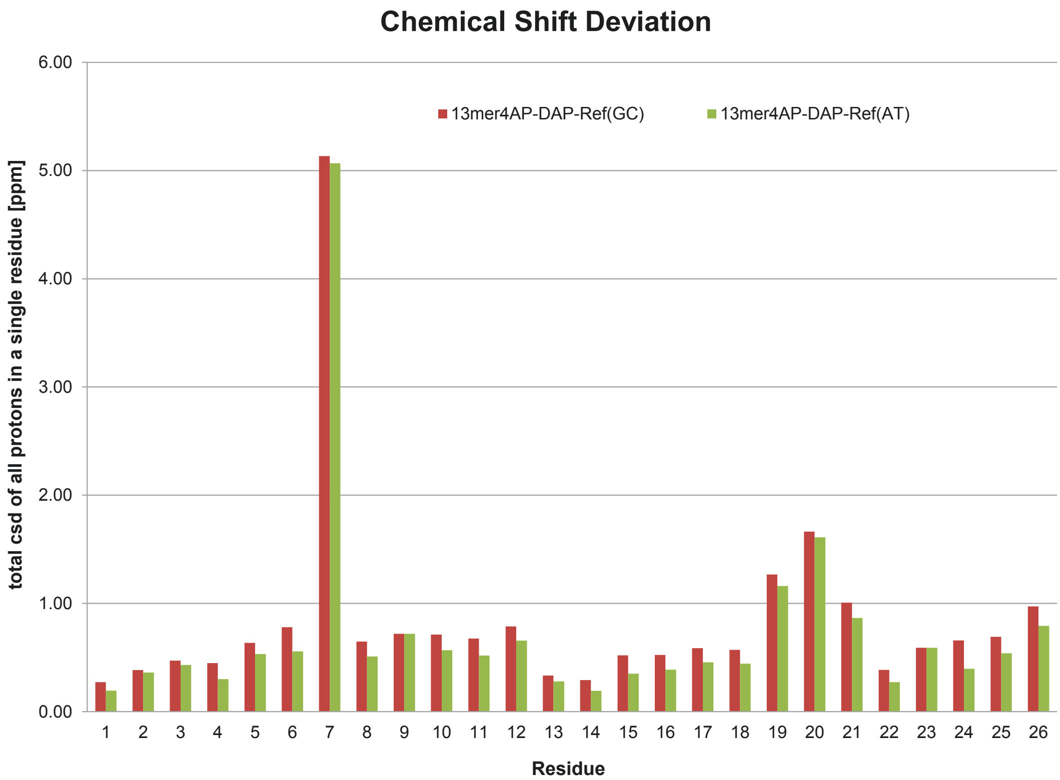


Figure 4.36: CSD comparison between 13mer4AP-DAP and different reference strands, which contain either GC or AT as central base pair. Chemical shift differences of all protons belonging to a single residue but different DNA samples were summed and are given as absolute values.

modification. The interruption may indicate deviations from B-DNA stacking at the 4AP site, although the chromophore is linked to a conventional DNA backbone and its H5 resembles the H6 of pyrimidine bases, whereas the interruptions in 13merHCF were caused by the α-glycosidic bond.

In Fig. 4.34 the artificial base pair was depicted with three hydrogen bonds. Even without knowing the structure, it is possible to analyze the hydrogen bonding pattern. Imino proton signals in DNA double strands are usually observed at 12 ppm or higher, when the proton is part of a hydrogen bond. Otherwise they are shifted high-field to values around 10 ppm[213]. In case of 13mer4AP-DAP, a sole signal was observed in the latter region which indicates that a base pair is not correctly hydrogen bonded. Moreover, 4AP contains the only new imino hydrogen and its principle design differs from purines in the sense that the five- and six-membered ring switched positions, so 4AP was suspected to lack a hydrogen bond. It should be noted that a missing hydrogen bond in the center of 4AP-DAP (see Fig. 4.34) also questions the existence of the other desired hydrogen bonds.

The chemical shift deviations of 13mer4AP-DAP were calculated in comparison to native strands (Fig. 4.36), which contain either GC or AT as central base pair. It must be considered that the spectra were measured at 10 °C instead of 25 °C , therefore an overall shift in all residues is observed, due to stronger stacking at lower temperature. A prominent example is cytosine at position 26 (C26), which shows the largest CSD among all residues far from modification sites. As stated in the end of section 4.2.2 of 13mer6HQ, C26 is sensitive to fraying at the helix end, which is in case of 13mer4AP-DAP reduced, due to the lower sample temperature. The deviations around DAP (20) are higher than the average inside the double strand, but among all nucleobase analogues in this work they seem rather small and this maybe related to the fact that DAP is also a pyrimidine like thymine and cytosine. In contrast to this, 4AP shows the largest CSD to which all its hydrogens contribute, especially the aforementioned H1. So it seems more and more likely that 4AP will not take the desired position inside the helix.

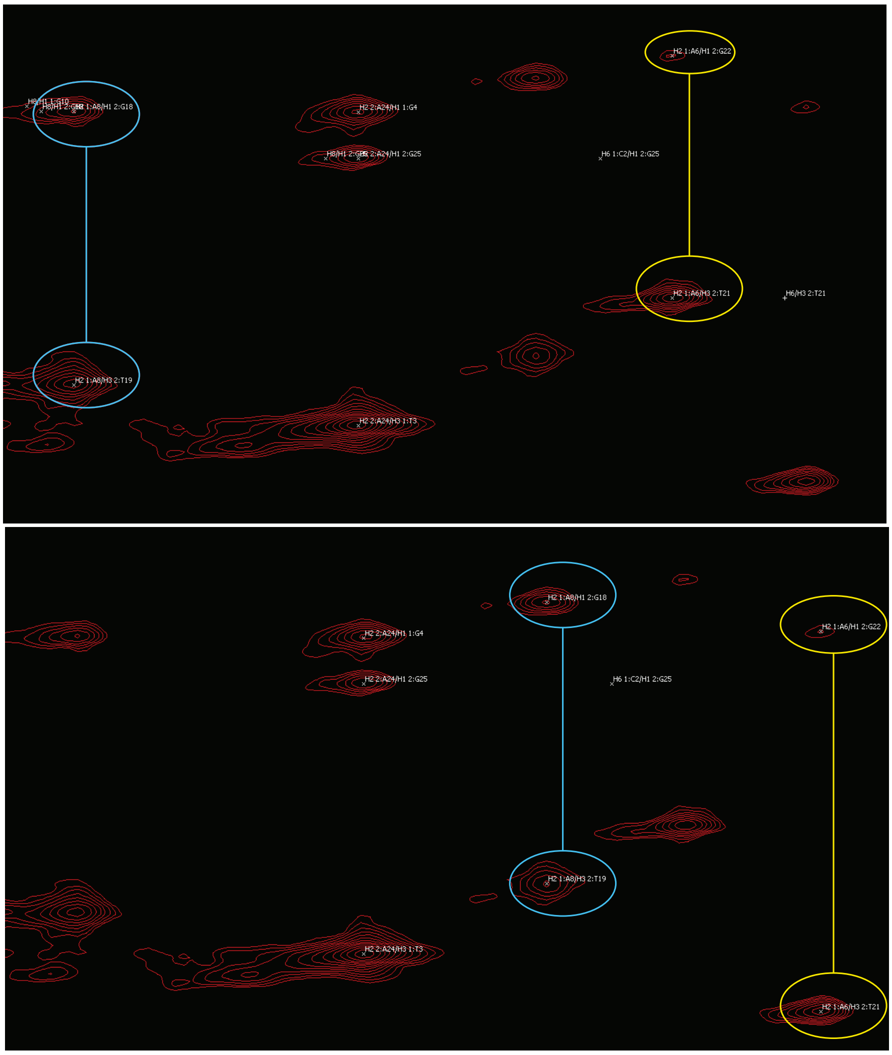


Figure 4.37: Assignment of two imino signal sets in 13mer4AP-DAP. Cross-peaks with the same assignment are circled either yellow or blue. Top and bottom picture represent the two conformations.

4.3.2 NMR solution structure

Two solution structures were found for 13mer4AP-DAP. In Figure 4.37 the reason for this is depicted. Two sets of imino proton signals can be assigned for the central AT base pairs. The region of the signals indicate correct hydrogen bonding pattern of these base pairs in both conformations. Furthermore, both pairs are adjacent to the 4AP-DAP pair, so that the conformational change should be found here. Two conformations with

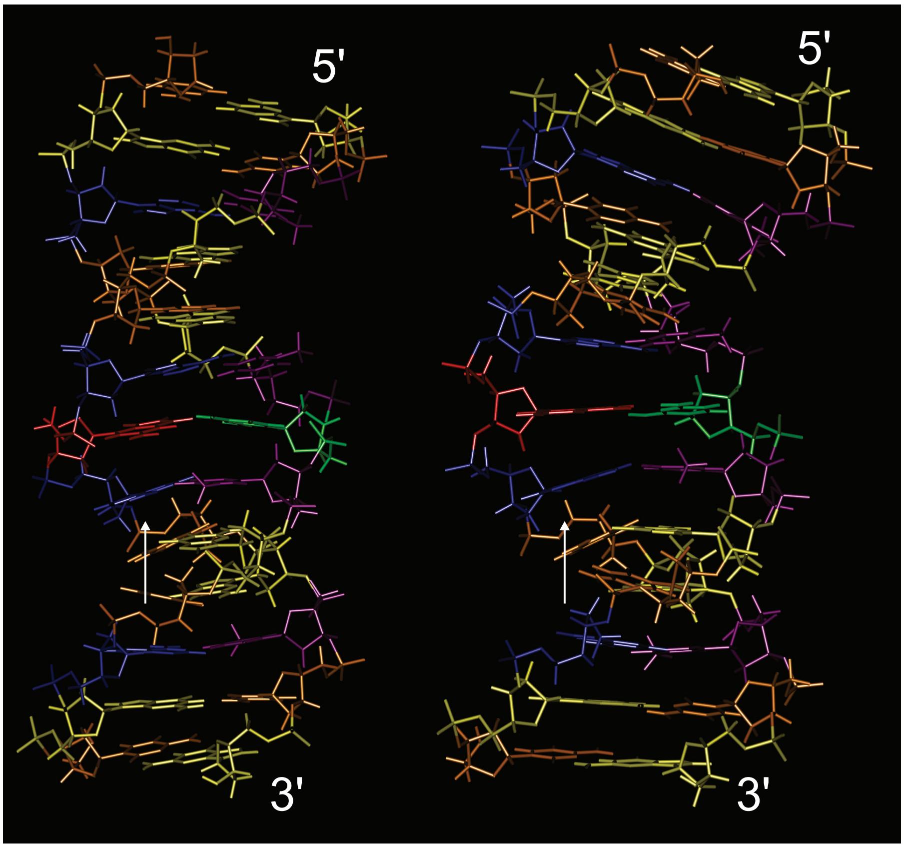


Figure 4.38: Averaged structures of 13mer4AP-DAP. The 4AP is marked red and the DAP green, while the corresponding colors for A, G, C and T are blue, yellow, orange and violet. The conformations were named after the number of hydrogen bonds for the 4AP-DAP base pair. Therefore is on the left the "1H-bond" conformer and on the right the "2H-bond", respectively. RMSD among all hydrogens besides methyl protons is for 1H-bond 0.52 Å and 2H-bond 0.85 Å, respectively. The white arrows indicate the point of view in Fig. 4.40.

different assignment patterns raise the complexity of the structure determination, since the intensity of all NOEs around the central base pairs depends, in addition to the distance, in such a case on the relation between both conformations. All details about the Simulated Annealing simulations can be found in sections 3.1.9 and 3.3.5.

None of the conformers shown in Fig. 4.38 comprises the three desired hydrogen bonds. However, they were named in accordance to the number of observed hydrogen bonds in the 4AP-DAP base pair, therefore is on the left the "1H-bond" conformer and on the right

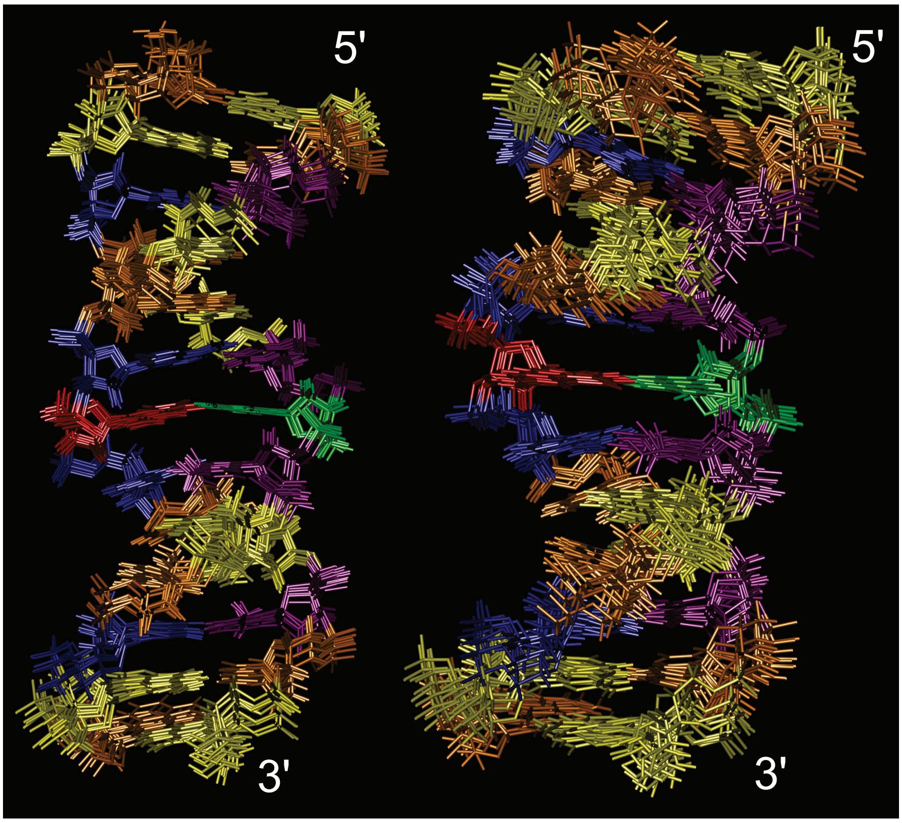


Figure 4.39: Overlay of the 10 minimum-energy, violation-free structures. The left side corresponds to the 1H-bond orientation of the average structures (left in Fig. 4.38) and the right to 2H-bond. The deviations alongside the backbone, especially for the 2H-bond, are larger than in comparison to the other structures (13merHCF, 13mer6HQ). The problem which arises here is that without RDCs only short distance information from base to base is available. So a weakly bent helix can be described with the same NOEs as a straight helix.

the corresponding “2H-bond” structure. Therein, the 1H-bond conformer is characterized by a straight well-defined B-DNA , while 2H-bond is wider in the center and starts to bend above the 4AP chromophore. The problem with a bent structure is that there is no evidence for it in the NOE data, since they only provide information from one residue to the next residue, thus limiting the quality of the structure. RDC measurements could validate or falsify the bending, since they provide long range information which describes

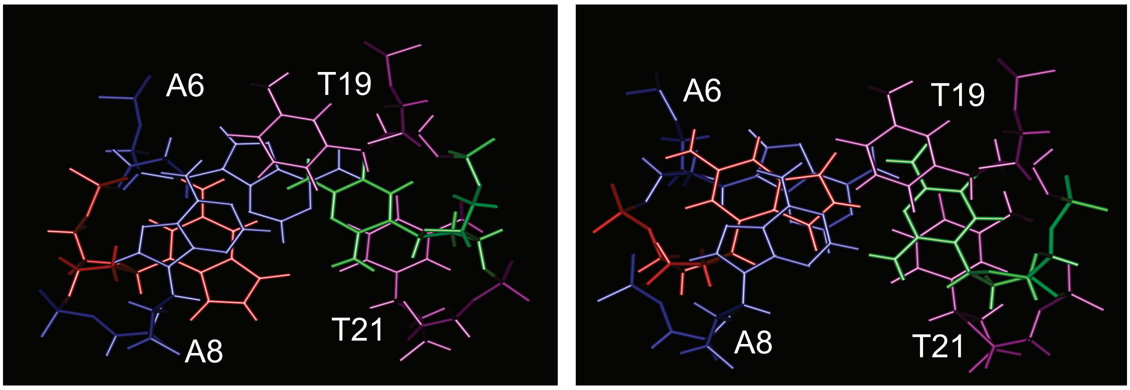


Figure 4.40: Close view on the three central base pairs of 13mer4AP-DAP. The 1H-bond orientation is again at left and 2H-bond at right. Interestingly, the 1H-bond 4AP (left) only stacks to adenine A8, while the five-membered ring faces to the minor groove instead of the center of the helix. Also worth to mention, the six-membered ring in the 2H-bond conformation is stacked to C1'-N9 bond of adenine A6.

the orientation of the bases in relation to the helical axis, but cannot be measured due to the low concentration of the sample. As a consequence, a second set of structures was calculated, which take advantage of the way restraints are used in Molecular Dynamics. The background and the results will be discussed in the next section (4.3.3).

The overlay of the 10 minimum-energy, violation-free structures in Fig. 4.39 shows more variation along the backbone than the previous structures (13merHCF, 13mer6HQ). Especially the 2H-bond overlay varies at the helical ends, which is usually compensated by the aforementioned RDCs. It should be noted that the RMSD value can also indicate high flexibility in parts of the structure or the presence of different conformers like in 13merHCF.

In Figure 4.40 the three central base pairs are shown as indicated by the white arrows in Fig. 4.38. The 1H-bond orientation is again at left and 2H-bond at right. One can see that the 4AP-DAP base pair fits into the helical structure for both conformations. Interestingly, the 1H-bond 4AP (left) only stacks to adenine A8, while the five-membered ring faces to the minor groove instead of the center of the helix. Remember, in case of 13mer6HQ the glycerol linker placed 6HQ in a position where stacking with both adjacent

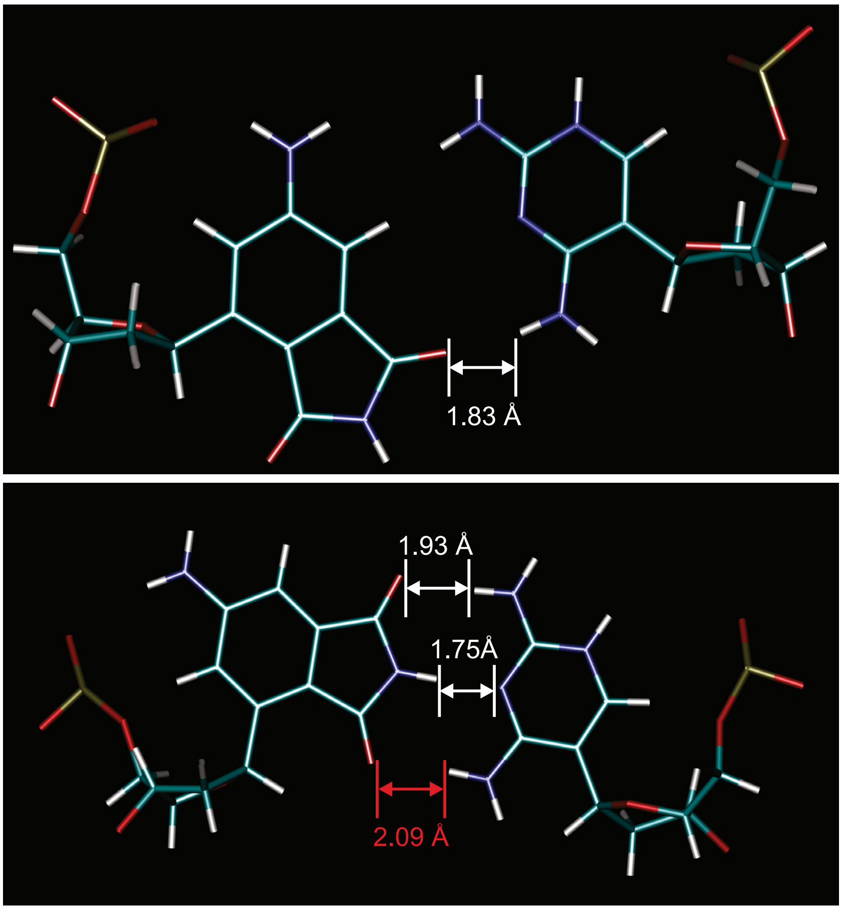


Figure 4.41: Close view on the central base pair of 13mer4AP-DAP. In the 1H-bond conformation (top), the 4AP forms a base pair with DAP that clearly resembles the 6HQ-C pair in Fig. 4.18. Reason for this is that both chromophores, 4AP and 6HQ, were linked via the same atom position in their six-membered ring to the backbone. In the bottom picture the 2H-bond conformer is shown. Although at first sight the bases seem to be in the right position for triple hydrogen bonding, the distances clearly indicate the presence of only two bonds.

adenines was possible. Also worth mentioning is that the six-membered ring in the 2H-bond conformation is stacked to C1'-N9 bond of adenine A6. So both conformations are somewhat displaced inside the the helix, which might affect the stacking interactions.

Figure 4.18 provides a close view to the central base pairs of both conformations. In the 1H-bond conformation (top), the 4AP forms a base pair with DAP that clearly resembles

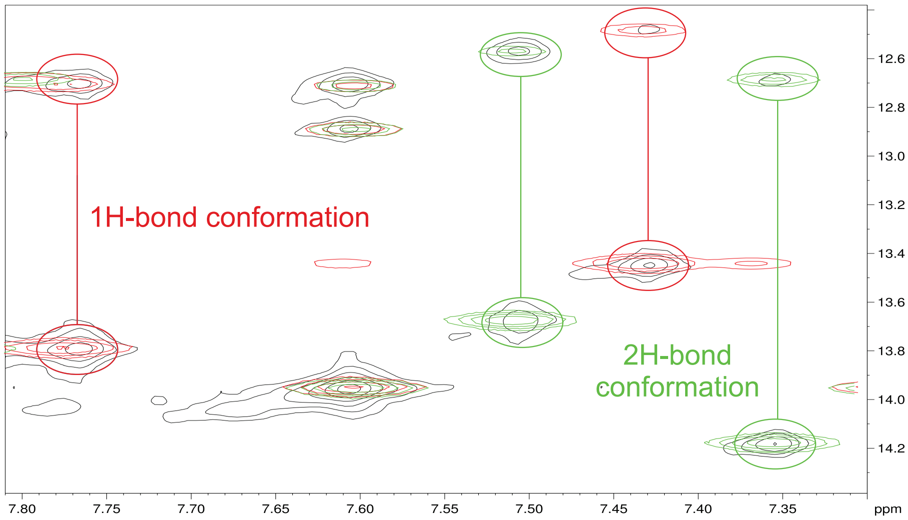


Figure 4.42: NOESY back-calculations of both conformers overlaid with experimental data. The 1H-bond spectrum is colored red, the 2H-bond spectrum green, and the experimental spectrum black. The circles depict the same cross-peaks as in Fig. 4.37.

the 6HQ-C pair in Fig. 4.18. Reason for this is that both chromophores, 4AP and 6HQ, were linked via the same atom position in their six-membered ring to the backbone. The only difference is that the hydrogen bond faces to the minor groove, while the bond in 13mer6HQ faces the major groove, due to the changed perspective indicated in Fig 4.16. In the bottom picture the 2H-bond conformer is shown. Although at first sight the bases seem to be in the right position for triple hydrogen bonding, the distances clearly indicate the presence of only two bonds. Furthermore, the different placement of 4AP is handled by the 2'-deoxyribose by switching between 2'-endo conformation (top) and 3'-endo (bottom). The latter is typical for A-DNA, thereby explaining the wider center, the interruption of the NOE-walk and the chemical shift deviations of the backbone in the previous section. Also interesting , the backbone of DAP has also changed in the bottom picture to an intermediate state between C2'-exo and C3'-endo.

As last point of this section the back-calculation of the NOESY spectrum will illustrate the two conformations for the spectral region in Fig. 4.37. In order to cover all cross-peaks in the experimental spectrum, both back-calculated spectra are needed.

4.3.3 What if RDCs were included?

In this section it is argued that the lack of RDC information can be overcome by a single assumption. First of all, some important information will be given, how Xplor-NIH[245] handles restrained Molecular Dynamics. The theoretical background introduced the total potential energy V_{tot} in equation 2.25, which is composed of an effective V_{eff} and an empirical energy term V_{emp}. All experimental restraints are part of the effective energy, while the empirical term contains force field components like Coulomb or Van-der-Waals interaction.

When a structure calculation has found a supposed global minimum then the total potential energy V_{tot} only contains V_{emp}, since the contributions of all restraints in V_{eff} are zero as long as the structure is free of violations. It is important to understand that the restraints do not define the structure in the minimum, instead they *limit the possible pathways* on the potential energy surface in order to find a reasonable minimum and hence a possible solution structure. The great advantage of this concept is that faulty restraints cannot affect the structure as long as a pathway to a minimum in energy is described. In this case they show up as violated restraints and one has to check the restraint and the structure for the reason of the difference. Usually one has to exclude NOE restraints which are heavily overlapped or near the limit of detection. Other reasons include errors in the assignment or that the structure, when everything else is ruled out, is only in a local minimum and therefore not correct. The latter is the reason, why so many calculations have to be performed and so much time invested to find the solution structure. Additional validity checks like back-calculation of NOESY spectra, Pales plots and RMSD values are helpful in the decision.

In the end, structure determination via restrained Molecular Dynamics means to find the best way to the potential energy minimum. At this point one could ask if the bent H2-bond conformation is stuck in a local minimum, since the 1H-bond conformer and also the other structures (13merHCF, 13mer6HQ) were rod shaped. The next question is then, how to describe a rod shape as a route on the potential energy surface, which can be scanned for a minimum. The headline of this section has already answered the last question. RDCs

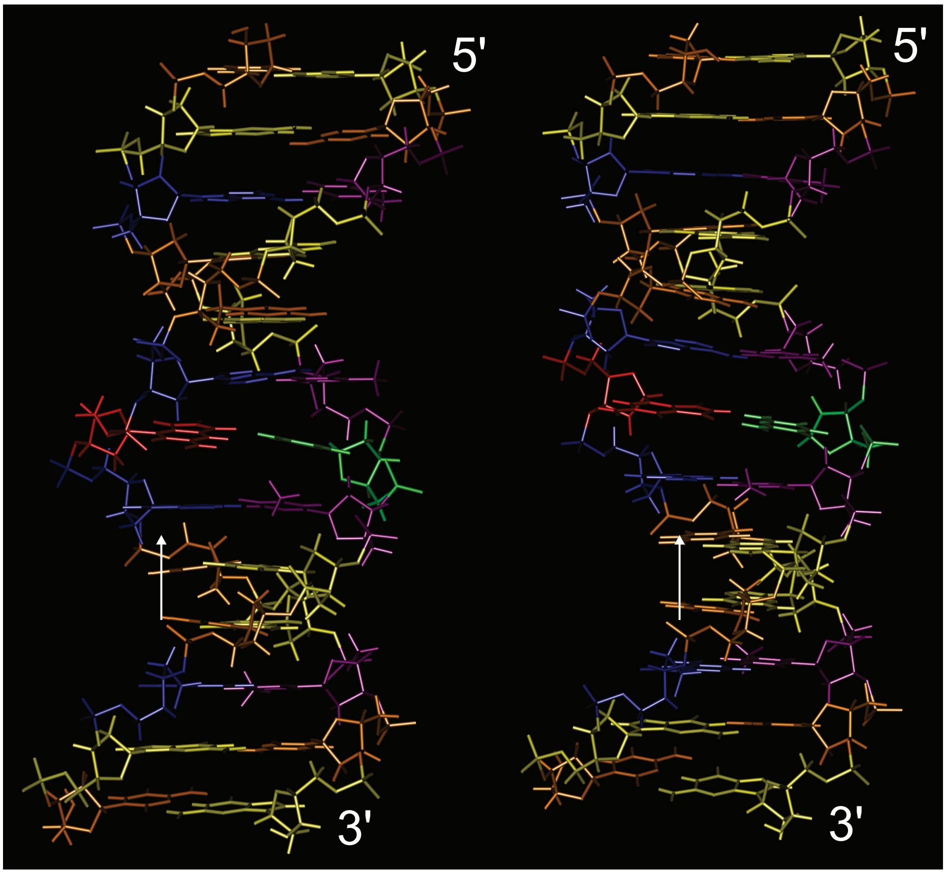


Figure 4.43: Averaged structures of 13mer4AP-DAP with RDCs of 13merHCF. The 4AP is marked red and the DAP green, while the corresponding colors for A, G, C and T are blue, yellow, orange and violet. Still on the left side is the "1H-bond" conformer and on the right the "2H-bond", respectively. RMSD among all hydrogens besides methyl protons is for 1H-bond 0.26 Å and 2H-bond 0.28 Å, respectively. The white arrows indicate the point of view in Fig. 4.45.

can provide the long range information that is necessary to describe a straight double strand, but measured RDCs are not available for 13mer4AP-DAP. However, the required information can be taken from set of RDCs that have proven to describe a straight strand, in this case from 13merHCF. The only necessary modification prior to their application is that all values that belong to the HCF chromophore will be omitted, since all double strands of this work differ only in their central base pair.

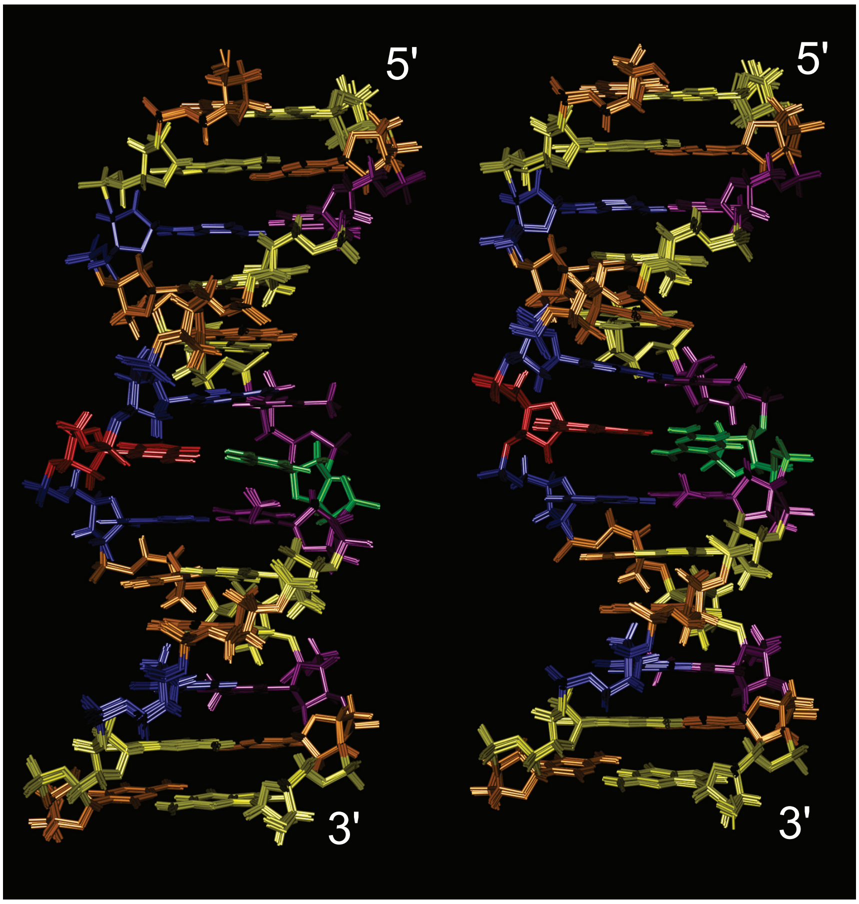


Figure 4.44: Overlay of the 10 minimum-energy, violation-free structures of 13mer4AP-DAP with RDCs of 13merHCF. The left side corresponds to the 1H-bond orientation of the average structures (left in Fig. 4.38) and the right to 2H-bond.

Figure 4.43 presents the averaged structures of 13mer4AP-DAP that were obtained by simply adding RDCs from 13merHCF (Tab. 3.1) to the simulations described in section 3.3.5. Two straight double strands were obtained. It is now easy to see, how the change of the 4AP backbone to C3'-endo affects the 2H-bond structure (right panel). The enlarged distance between the adenines A6 and A8, which surround the 4AP chromophore , appears as an uplift of A6 (located above 4AP). The effect of the 4AP-DAP introduction is now limited to the next two base pairs in both directions.

A closer look on the central base pair in Figure 4.45 reveals that the relative positions

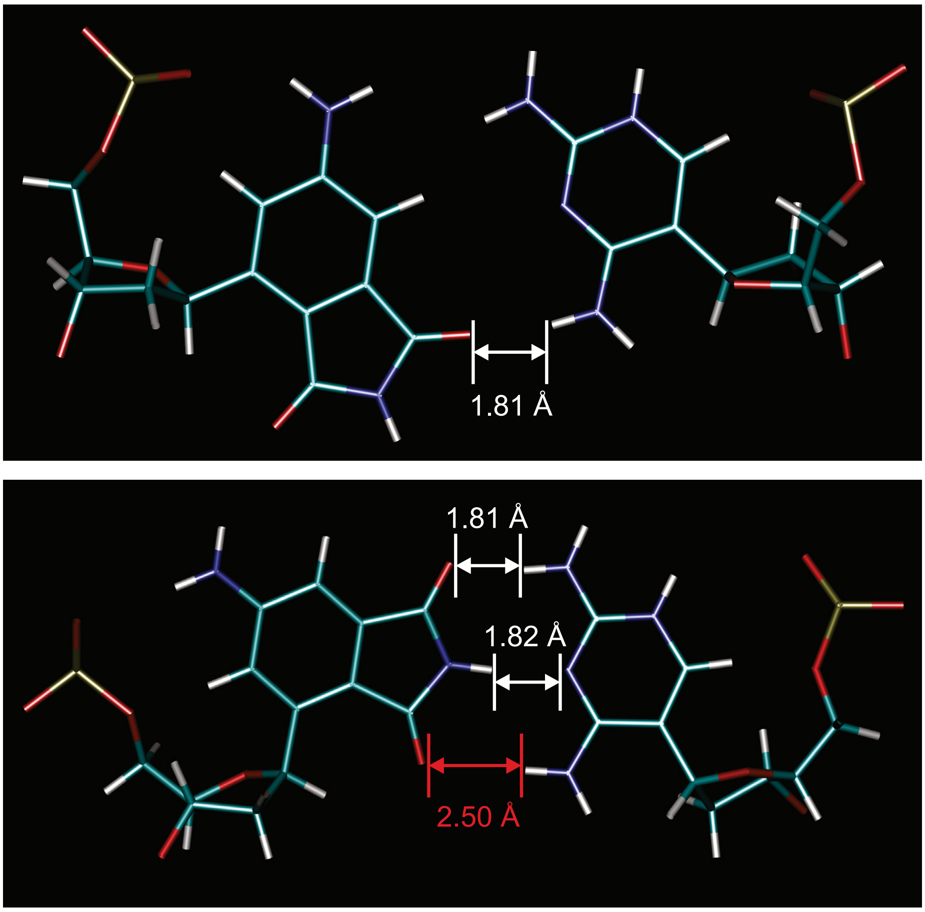
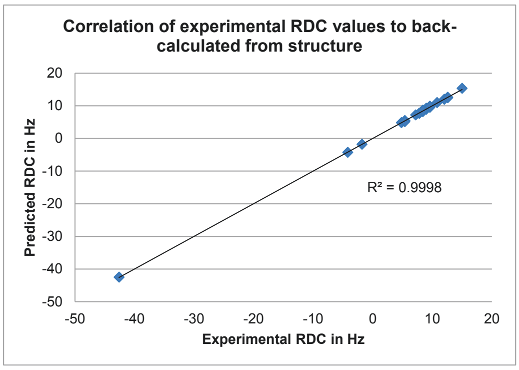


Figure 4.45: Close view on the central base pair of 13mer4AP-DAP with RDCs of 13mer-HCF. The only notable difference to the previous picture of the central base pair conformations (Fig. 4.18) is the distance of the missing hydrogen bond (red), which is raised by 0.41 Å , a value below the error margin of the involved NOE distances. Neither the number of hydrogen bonds nor the relative positions of 4AP and DAP have changed. The deviating conformations of the 2'-deoxyribose in the 2H-bond structure (bottom), for example C3'-endo for 4AP, were also retained.

of 4AP and DAP to each other did not change, which is not surprising. As outlined in the beginning of the section, the RDCs affect mainly the overall shape of the double strand. The only notable difference to the previous picture of the central base pair conformations (Fig. 4.18) is the distance of the missing hydrogen bond (red), which is raised by 0.41 Å , a value below the error margin of the involved NOE distances. The reason for the rise



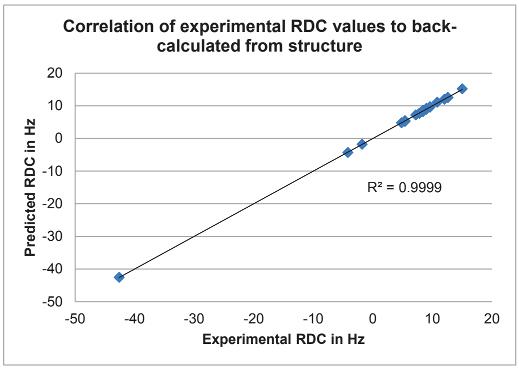


Figure 4.46: Pales[268] plots of experimental 13merHCF RDCs against predicted values of the of 13mer4AP-DAP structures (1H-bond, 2H-bond right). It is easy to see that the RDC values of 13merHCF work as well with 13mer4AP-DAP.

is the induced propeller twist between 4AP and DAP in contrast to the nearly planar arrangement of the NOE only structure. A base pair twist is not special, in fact planarity is, since it was a side effect of the first crystal structures that offered only averaged values for the helix parameters and was later revised by the first single crystal structures[2,14] . Note that the backbone conformation of 4AP in the 2H-bond structure is still C3'-endo and that DAP also retained its intermediate state between C2'-exo and C3'-endo.

Pales plots for both RDC structures will prove that 13mer4AP-DAP is as compatible as 13merHCF to the RDC data (Fig. 4.46).

Finally for this section, it can be summarized that the solution structures, which were found under the assumption of linear double strands, exhibit better RMSD values and less perturbations. So it seems very likely that the RDC structures are closer to the real structures, even though the RDCs came from 13merHCF. At this point it should be repeated that restraints were used to find a way to the minimum in potential energy, where the structures are defined by the effective energy V_{eff} and the restraints sum to zero. This clearly shows, how important the RDC measurement for structure determination is. Moreover, it was demonstrated that it is possible to benefit from RDC data of other double strands.

4.3.4 Duplex melting experiments

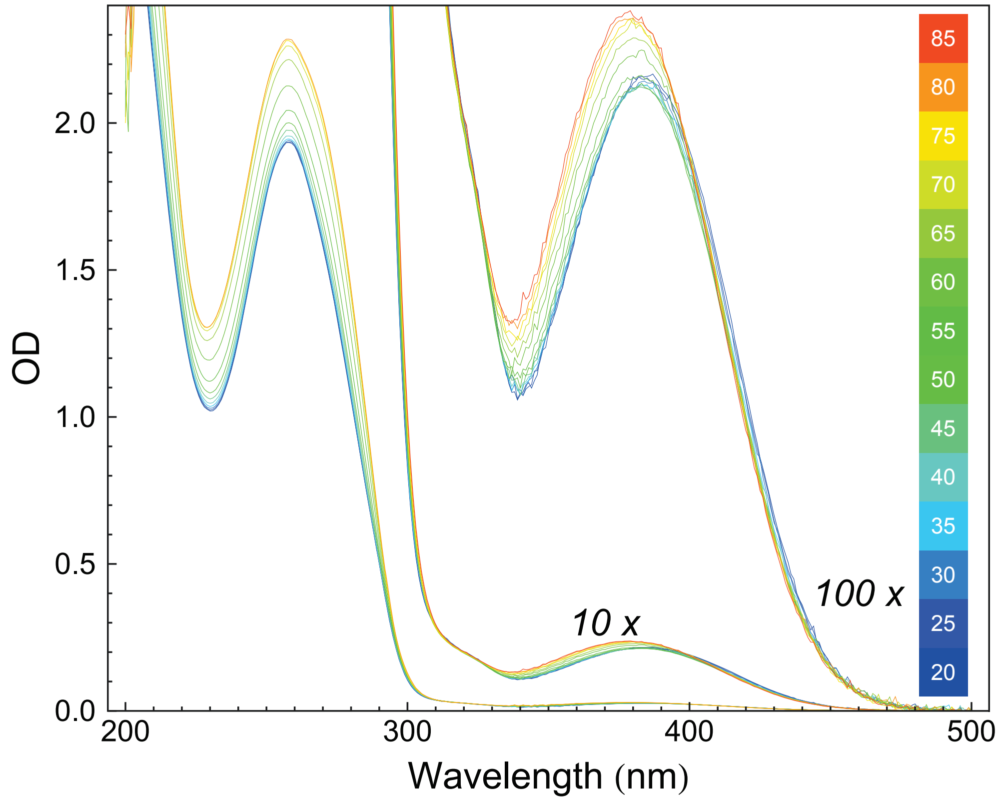


Figure 4.47: Absorption change of ds13mer4AP-DAP upon melting, when raising temperature from 20 °C (blue) to 85 °C (red line; $c_T = 20.5$ µM). Data from the 1 and 10 mm cuvettes are labelled "1x" and "10x", respectively. A ten-times magnification of the latter is labelled "100x".

A sample of double-stranded 13mer4AP-DAP was prepared in a water/phosphate buffer (10 mM NaH_2PO_4, $pH = 6.35$) with 150 mM sodium chloride (see sec. 3.3). Measurements were performed in a double cuvette (closed but not evacuated) with 1 and 10 mm optical path length. A total concentration $c_T = 20.5$ µM of single strands was estimated from maximum absorption around 260 nm at 85 °C. A set of 14 spectra between 20 and 85 °C were recorded and corrected for density change. Results are shown in Fig. 4.47, where the data from the 1 and 10 mm cuvettes are labelled "1x" and "10x", respectively. A ten-times magnification of the latter is labelled "100x".

Analysis of the 13mer4AP-DAP DNA band

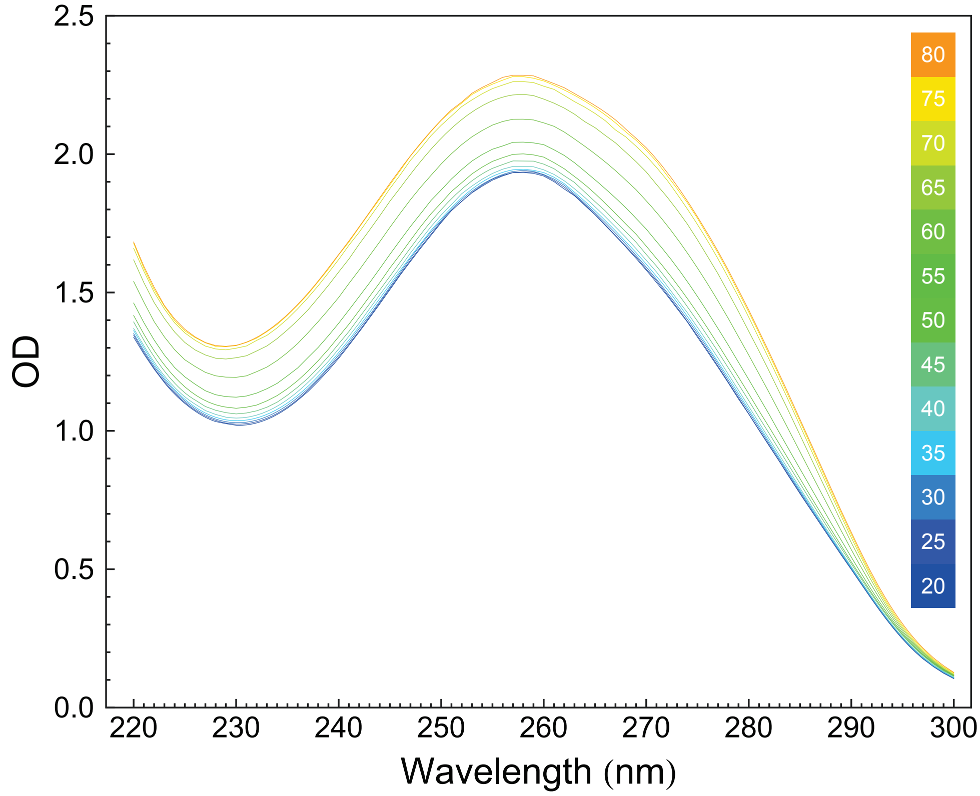


Figure 4.48: The UV part of the spectrum was used for SVD of the DNA absorption peak. The first spectrum is blue (20 °C) and the last one is orange (80 °C). The 85 °C measurement (1 mm cuvette for DNA peak) had to be omitted due to an error in the cuvette holder while heating the sample.

The strategy here starts similar to the one in 13mer6HQ by separate examination of the DNA absorption band. The changes of the DNA peak with increasing temperature (20 - 80 °C) were quantified in a spectral window from 300 to 220 nm (see Fig. 4.48). The UV absorption spectra in this range were analyzed for principal components via Singular Value Decomposition (SVD). Two components are needed to describe the hyperchromism. Fig. 4.49 shows the basic spectra on the left and the corresponding thermodynamic curves on the right. The first component (red) was set to be the 80 °C spectrum. As before, the thermodynamic curves were multiplied with a scale factor so that the amplitudes fall into the range from 0 to 1 (the corresponding basic spectra were scaled with the inverse factor.)

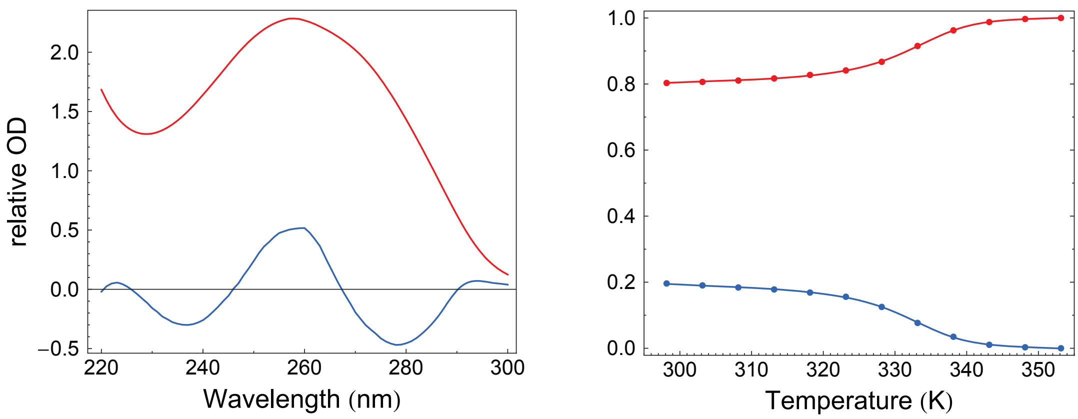


Figure 4.49: Basic spectra are shown left. The first component (red) was set to be the 80 °C spectrum, i.e. it represents the fully separated single strands. The right part shows the thermodynamic curves related to the basic spectra. Upon hybridization, the amplitude of the first-component spectrum (red) decreases from 1 to 0.8 while that of the second component increases from 0 to 0.2.

Upon duplex formation, the red one shows an amplitude decrease of 0.2. By design, the blue curve shows the same amount as amplitude increase (Fig. 4.49).

Again, the two-state model with stacking interactions was applied to the thermodynamic data shown in figure 4.50. When comparing the basic spectra of 13mer4AP-DAP in figure 4.49 with the DNA part of 13merHCF (Fig. 4.11) or 13mer6HQ (Fig. 4.24 one can see that they all look the same. Assuming the same assignment, introduced with 13merHCF, the red line is related to the whole strand, while the blue line describes the basepairs around the two artificial nucleobases in the center. In contrast to 13merHCF, the order of the melting points in the right (two-state only) part of figure 4.50 (grey vertical lines) has changed. The blue one comes first at 59.1 ± 0.1 °C and the whole strand (red) follows at 59.5 ± 0.1 °C. In principle, this indicates a weak local melting or bubble formation in the center of the duplex, but when taking the other parameters into account (red: ΔH° = -330 ± 20 kJ/mol, ΔS° = -880 ± 50 J/(K mol) and blue: ΔH° = -340 ± 20 kJ/mol, ΔS° = -930 ± 60 J/(K mol), they are possibly equal within the margin of errors. In order to clarify whether or not there is evidence of local melting, the analysis of the 4AP part of the absorption spectrum is necessary. A lower estimated melting point would support this

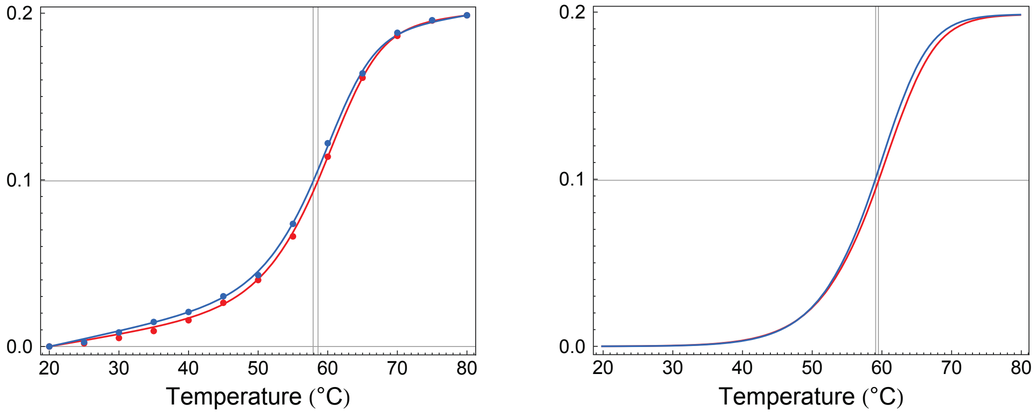


Figure 4.50: Thermodynamic curves, scaled to the measured amplitude change of 0.2. The left panel shows SVD data (points) and fits with a two-state model including stacking (lines). The two-state model part is plotted separately (right panel), representing a melting curve of hyperchromism without the influence of stacking. The grey vertical lines indicate the melting points (derived from the right panel), therein the whole strand (red) melts at 59.5 ℃ and the center (blue) at 59.1 ℃.

theory. A second possible cause, why the blue line has a lower melting point, would be the absorption of 2,4-diaminopyrimidine around 270 nm, but then the effect on the lineshape of the DNA peak is rather small, since there is no evidence of alteration (additional hidden peak) in the basic spectra (Fig. 4.49) and in the end, the theory of local melting would still remain.

Dissection of the 4-Aminophthalimide band

The same procedure was then performed on the spectral window of 4-Aminophthalimide ($\lambda > 480$ nm). Two components are needed in the SVD to describe the hyperchromism in figure 4.52. The red component describes the rising amplitude of hyperchromism, but compared to the other molecules it has a rather small amplitude change of 0.1, less than a half of the usually observed value. In addition, it is the only sample with linear lineshape of the blue shift over the full temperature range (blue component). Both observations indicate weaker base-stacking for the 4AP chromophore.

While the red curve was fitted with equation 2.33, representing the two-state model with

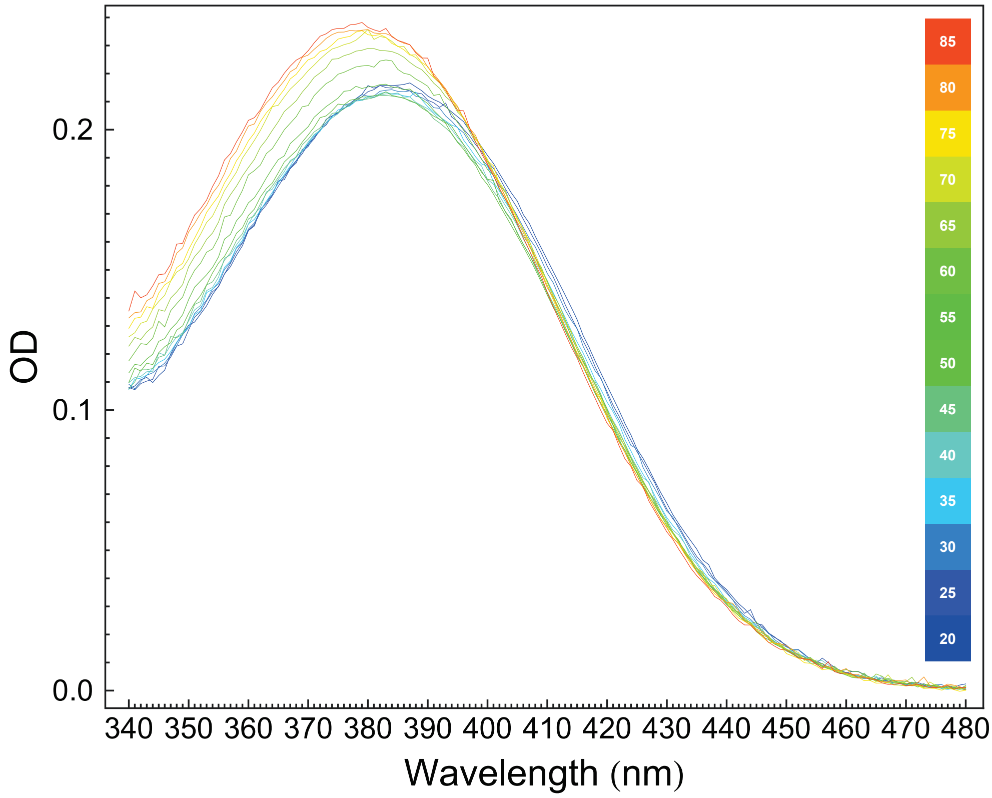


Figure 4.51: Spectrum was used for SVD of the 4AP absorption band. The first spectrum is blue (20 °C) and the last one is red (85 °C).

linear stacking, the blue line was fitted by a simple linear least-squares fit, which describes a linear blue shift from 385 nm at 20 °C down to 379 nm at 85 °C. Best fit parameters for the melting curve are $\Delta H°$ = -360 ± 40 kJ/mol and $\Delta S°$ = -1000 ± 100 J/(K mol) upon hybridization. Interestingly, when looking closely on the duplex part of figure 4.53, we notice an initial decrease of absorption up to 45 °C. This observation suggests that stacking first increases, even though the hypsochromic effect is small (< 2 %). This finding can be explained when taking the NMR solution structures into account. At low temperatures the weakly stacked 1H-bond conformer dominates which changes with rising temperature. Shortly before the melting point more and more molecules are in the 2H-bond state, which shows a better stacking interaction (see Fig. 4.40) and therefore would explain the initial absorption decrease. From the right panel (Fig. 4.53) a melting point of 56.9 °C was estimated, which is as supposed lower than the 59.5 °C (red component) or 59.1 °C (blue)

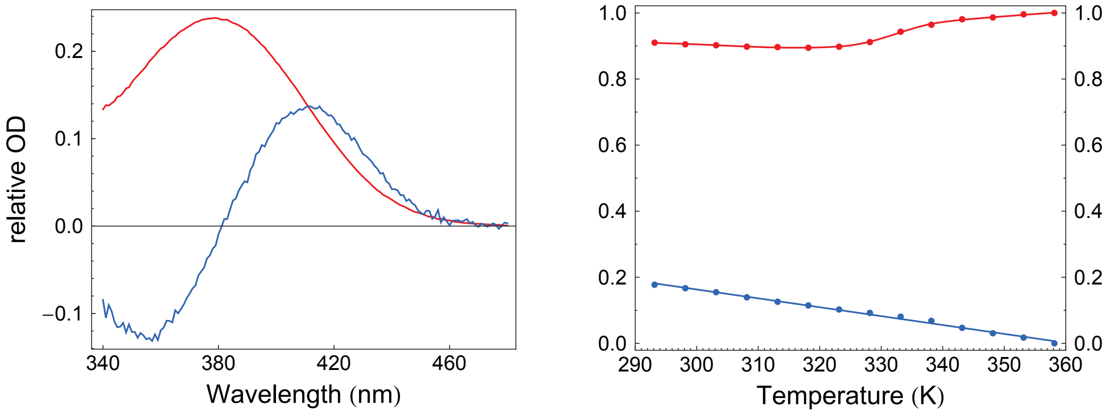


Figure 4.52: Basic spectra are shown left. The first component (red) was set to be the 85 °C spectrum, i.e. it represents the fully separated single strands. The right part shows the thermodynamic curves related to the basic spectra. Upon hybridization, the amplitude of the first-component spectrum (red) decreases from 1 to 0.9.

obtained for the DNA peak. In summary, the theory of local melting in the center of the duplex is supported by its lower melting point, the two melting points of the DNA peak, the linear blue shift and the weak base-stacking of the 4AP chromophore. Additional evidence comes from NMR, where the signal of the 4AP proton H1 was assigned in the region of non-hydrogen-bonded imino protons around 10 ppm (see sec. 4.3.2). Moreover, the 1H-bond conformation supports the findings of the melting experiments, due to the placing of the five-membered ring into the minor groove region. In case of 2-aminopurine, which contains an amino group facing towards the minor groove, it has been demonstrated that the base pair dynamics is much faster and thereby lowering the melting temperature[129] in comparison to an adenine containing reference strand (13merRef). It should be noted that both double strands only differed in the position of a single amino group in the center.

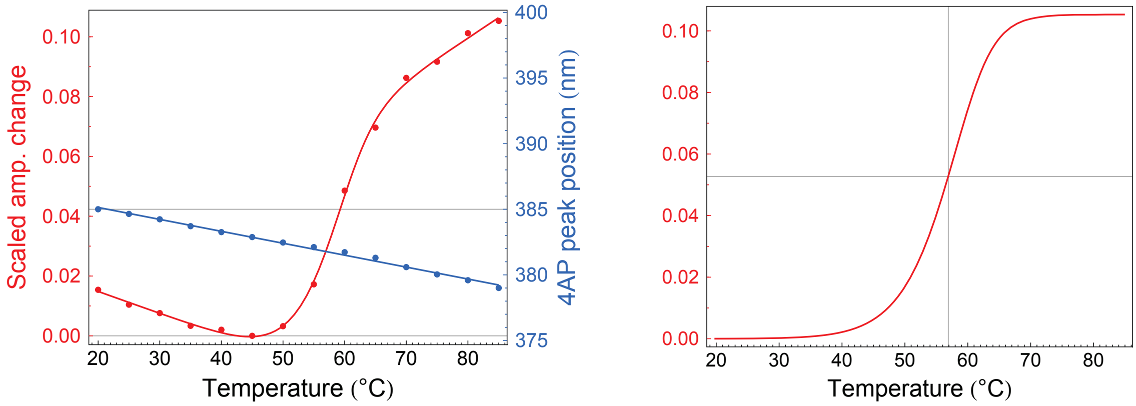


Figure 4.53: Melting curve (red), scaled to the measured amplitude change of 0.1. The left panel shows SVD data (points) and fit with a two-state model including stacking (red line) for the red component and, in addition, the peak shift (blue) with a separate axis to allow a direct reading of wavenumbers. The right panel shows the red component alone and without stacking, since it is the only one with melting behaviour. The grey vertical line indicates the melting point of the 4AP at 56.9 °C. Best fit parameters are $\Delta H° = -360 \pm 40$ kJ/mol and $\Delta S° = -1000 \pm 100$ J/(K mol) for the hybridization.

5 Summary

The same basic sequence of 13 base pairs with chemical modifications only in the middle position was investigated by NMR structure determination. With this conservative approach it becomes possible to study perturbations which are induced by the artificial nucleobase. Especially the influence of base stacking and of the linker can be detailed. Structural differences are also associated with thermodynamic variations. The latter are observed with the help of entire UV/Vis-spectra as function of temperature. In this way the duplex melting process is characterized in local and global terms.

UV/Vis-spectra upon melting were analyzed in their entirety by Singular Value Decomposition (SVD). Thus the spectral shift of the probe absorption band is followed, being caused by the solvatochromicity of the incorporated chromophore. The band is located in the visible range of the spectrum and provides *direct* information about local melting. In contrast, the strong absorption of the native nucleobases is located in the UV region and contains two kinds of information. One refers to the usually observed global melting of the double strand, the other gives *indirect* information on local melting, as seen by the nucleobases in the immediate neighborhood of the modification. Altogether three melting points T_m are obtained as specified in Tab. 5.1. Melting can now be understood in terms of global and local processes. The latter describe effects near the modified center, like bulge formation upon temperature rise.

NMR structure determination in solution is a valuable instrument to investigate structural perturbations, base stacking, and linker effects. Three chemical modifications were studied in the center of the basic 13mer sequence, representing different types of perturbations. These are (i) the large 2-hydroxy-7-carboxyfluorene (HCF) replacing a full base pair, (ii) the betaine 6-Hydroxyquinolinium (6HQ) incorporated as glycol nucleic acid (GNA), and (iii) 4-aminophthalimide (4AP) together with 2,4-diaminopyrimidine (DAP) forming an artificial base pair. The NOESY spectra of the HCF and 4AP-DAP double strands gave in each case evidence for the presence of a second conformer. Instead of three structures, two additional conformations were required to fully describe the found cross-peak patterns. As a result, altogether five structures were determined in the course of this study.

Table 5.1: Global and local melting points of the three investigated DNA double strands.

13mer denoted by central base pair	HCF	6HQ	4AP-DAP	Ref[129]
Global T_m of DNA (°C)	59.7	53	59.5	63.3
Local T_m of DNA (°C)	61.1	53	59.1	
Local T_m of Chromophore (°C)	61.1	53.1	56.1	

The HCF chromophore is the largest of all fluorescent base analogues in this work and was placed opposite an abasic site, thus making it a full base pair surrogate. HCF adopts two conformations inside the double helix which are distinguished by the methylene group, in the sense that this group faces either towards the major groove or towards the minor groove. The HCF moiety fits well into the helical fold for both conformations. The unusual α-glycosidic linkage of HCF is compensated by small deformations of the 2'-deoxyribose moiety. The sugar switches to conformations other than C2'-endo of B-DNA to allow optimal chromophore incorporation. Because of the sugar flexibility, an equal distribution for both conformations is observed. The abasic site has no effect on HCF, but the lack of a nucleobase raises the flexibility in the center of the counterstrand, since its position is no longer restricted by the stacking of an attached nucleobase. On the other hand, stacking interactions of HCF to the neighboring adenines are accompanied by additional interactions with thymine T21 (and partly T19) of the opposite strand (see Fig. 4.6). The local melting points of the chromophore and of the DNA (61.1 °C, in Tab. 5.1) agree, and they are 1.4 °C higher than the global melting point of the double strand (59.7 °C). Even though the local melting indicates a stabilizing effect of HCF, the global melting is lower than for the native double strand 13merRef with AT in the center (63.3 °C). Both findings agree well with earlier results of Matray and Kool[97], who used pyrene as base pair surrogate.

The second base analogue, 6HQ , was placed opposite cytosine with which it forms a single hydrogen bond. The unique feature of this double strand is the incorporation of 6HQ as GNA. Up to now, structures of nucleic acids with full GNA backbone are known only from X-ray diffraction of crystals[274,275], therefore the solution structure is of general interest. Our linker R-Glycerol is small compared to 2'-deoxyribofuranose. For insertion of the chromophore into the helix one C-C bond length is missing, and another C-C bond length is missing for spacing the adjacent adenines. The first deficiency is compensated by the size of 6HQ and its linkage to R-glycerol via N1, allowing a more diagonal placement of 6HQ relative to cytosine to reach the distance needed for hydrogen bonding (Fig. 4.13). A side effect of this placement is strong π-π-stacking interaction to adenines A6 and A8 (Fig. 4.17), leading to a huge chemical shift of the H8 hydrogen in adenine A8. The need to compensate the second missing bond length causes structural perturbations. In order to minimize the distance between 3'-end of adenine A6 and the 5'-end of A8, both bases are inclined towards the major groove; thus, the backbone distance is shortened by 1.2 Å in comparison to 13merRef. As a consequence, a stacking axis of A6-6HQ-A8 is observed that leans towards the major groove, stabilized by strong stacking interactions in between. For the melting analysis of 13mer6HQ it was necessary to develop a variant of the new method for UV/vis spectral analysis, to treat adequately the complex behavior of the 6HQ chromophore. The double SVD assisted two-state model revealed that 6HQ not only undergoes a blue shift and hyperchromism upon melting like HCF, but also a change of the absorption spectrum that resembles the alteration from the protonated to the deprotonated state. A closer look on global and local melting shows that only a single melting point within error margins ($\pm$ 0.1 °C) is observed (Tab. 5.1), which is otherwise only found for pure native strands. However this does not prove the native melting character of 13mer6HQ, since it has the lowest melting point of all double strands. The differing stacking axis of the three central bases destabilizes the 6HQ strand, but the strong stacking interactions also prevent bulge formation. Stability of the duplex is only maintained by the remaining base pairs far from the center. This competes with the fraying at the helix ends upon temperature rise. Altogether this double strand shows the melting behavior of a shorter duplex.

The last double strand introduces an artificial base pair composed of 4AP and DAP. First evidence that refuted the desired pattern of three hydrogen bonds was found in the NOESY spectrum. There the chemical shift of the H1 proton (in 4AP) was assigned in the non-hydrogen bonded region of the spectrum. In addition, the H_2O-NOESY also proved the presence of the protonated form of DAP (Fig. 4.34). Instead of a structure with three hydrogen bonds, two conformations were found that comprise either a single hydrogen bond or two hydrogen bonds. The single-bonded form is similar to the structure of 13mer6HQ, with a diagonal placement of the 4AP moiety so that its long side faces towards DAP. The reason for this is the linkage via C6 which corresponds to N1 in 6HQ and is well-known from pyrimidine bases. However, 6HQ and 4AP are in size and shape more comparable to purines which are linked via the five-membered ring, resulting in a different angle of the nucleobase to its counterpart in the opposite strand. The conformer that comprises two hydrogen bonds tries to overcome this by switching the sugar of 4AP to the C3'-endo conformation of A-DNA. Although at first sight all sites which are capable of hydrogen bonding now face each other, distance measurements revealed (in Fig. 4.18) that only two bonds are possible. Interestingly, the melting experiments show that the second form, which exhibits better π-π-stacking interactions, is favored upon rising temperature, causing an initial decrease in absorption. But finally, when comparing local and global melting in 13mer4AP-DAP (Tab. 5.1), bulge formation caused by premelting of the central 4AP-DAP pair is clearly indicated. The melting point of the adjacent base pairs is also lower than the global one, supporting the bulge formation in the center. On the other hand the difference is only 0.4 ℃, so the bulge seems to be mostly limited to the 4AP-DAP base pair.

In summary, all investigated double strands suffer from structural perturbations, either caused by the linker or by the chromophore itself. Further development of base analogues and their linkage is clearly needed in future; some general lessons can be already learned from the set of structures in this work. From 13merHCF it can be concluded that a α-glycosidic linkage can be balanced by the 2'-deoxyribofuranose, and that a large aromatic surface can replace a complete base pair, but at the price of increased flexibility. 13mer6HQ demonstrates that R-glycerol is too small in comparison to 2'-deoxyribofuranose. A possible solution is the addition of at least one bond length in order to connect the 3'-end of adenine A6 with the 5'-end of A8, but this would raise flexibility. The problem of high flexibility inside the double strand is faster base pair dynamics and hence a lowered duplex stability. Moreover, this effect propagates to the surrounding base pairs in both directions[129]. From this point of view the 4AP-DAP base pair linked via the native 2'-deoxyribofuranose looks more promising. However, the weak point of this design is the linkage of 4AP to the backbone. The connection to the sugar moiety should be changed from the current C6 atom of 4AP to the neighboring C5. This should enable 4AP to form a triple hydrogen bonded base pair with DAP. In such a case one can expect the triple form to be the only present conformer, since both actual conformers require the connection via C6.

The SVD analysis of UV/Vis spectra in their entirety revealed measurable local melting in the DNA absorption band. The local melting is induced by perturbations of the double strand and is only reported by native bases in close proximity to the modification. Native bases as simultaneous detector of global and local melting offer the possibility to study modifications that do not contain separate absorption bands like the chromophores which were studied here. A second possibility could be melting experiments of native double strands comprising an intercalator, for example the groove binding Hoechst 33258. In that case, perturbations of the melting process will be reported by the strand and can then be compared to the signal of the Hoechst dye. The absence of a local melting component in native strands like 13merRef can serve as indicator of undisturbed melting, when observed in modified double strands. Finally, it should be noted that the method can be automated experimentally and analytically. The fully automated variant would allow application as a replacement for the conventional DNA melting experiment at a single wavelength.

6 Zusammenfassung

Die gleiche Grundsequenz von 13 Basenpaaren mit chemischen Modifikationen nur in der mittleren Position wurde mittels NMR Strukturbestimmung untersucht. Mit diesem konservativen Ansatz ist es möglich strukturelle Störungen zu studieren, welche durch die künstliche Nukleobase verursacht wurden. Insbesondere der Einfluss der Basenstapelung sowie des Linkers kann so genau beschrieben werden. Strukturelle Unterschiede gehen auch mit thermodynamischen Änderungen einher. Letztere wurden mit Hilfe vollständiger UV/Vis-Spektren als Funktion der Temperatur erfasst. Auf diese Weise kann der Schmelzprozess des Duplex lokal und global beschrieben werden.

UV/Vis-Spektren beim Schmelzen wurden in ihrer Gesamtheit mittels Singularwertzerlegung (SVD) analysiert. Somit kann die spektrale Verschiebung der Sonden-Absorptionsbande verfolgt werden, welche durch die Solvatochromie des eingebauten Chromophors verursacht wird. Die Bande befindet sich im sichtbaren Bereich des Spektrums und liefert *direkt* Informationen über das lokale Schmelzen. Im Gegensatz dazu absorbieren die natürlichen Nukleobasen stark im UV-Bereich und enthalten zwei Arten von Informationen. Eine gehört zum normalerweise beobachteten globalen Schmelzen des Doppelstranges, die andere liefert *indirekte* Informationen über das lokale Schmelzen, wie es von den Nukleobasen in unmittelbarer Umgebung der Modifikation gesehen wird. Insgesamt drei Schmelzpunkte T_m wurden erhalten, so wie in Tab. 6.1 angegeben. Das Schmelzen kann nun als globaler und lokaler Prozess verstanden werden. Letzteres beschreibt Effekte in der Nähe des modifizierten Zentrums, beispielsweise Blasenbildung bei steigender Temperatur.

NMR Strukturbestimmung in Lösung ist ein wertvolles Instrument um strukturelle

Störungen, Basenstapelung und Linker-Effekte zu untersuchen. Drei chemische Modifikationen, welche unterschiedliche Störungen verursachen, wurden hierfür im Zentrum der 13mer Grundsequenz studiert. Diese sind (i) das große 2-Hydroxy-7-carboxyfluoren (HCF), welches ein komplettes Basenpaar ersetzt, (ii) das Betain 6-Hydroxychinolinium (6HQ), welches als Glykolnukleinsäure (GNA) eingebaut wurde, und (iii) 4-Aminophthalimid (4AP) zusammen mit 2,4-Diaminopyrimidin (DAP), welche ein künstliches Basenpaar bilden. Die NOESY-Spektren der HCF und 4AP-DAP Doppelstränge geben in beiden Fällen Hinweise auf das Vorhandensein einer zweiten Konformation. Anstelle von drei Strukturen, waren zwei zusätzliche Konformationen notwendig, um die gefundenen Kreuzsignal-Muster vollständig zu beschreiben. Im Ergebnis wurden insgesamt fünf Strukturen während der Untersuchung bestimmt.

Table 6.1: Globale und lokale Schmelzpunkte der drei untersuchten DNA Doppelstränge.

13mer benannt nach zentralem Basenpaar	**HCF**	**6HQ**	**4AP-DAP**	**Ref[129]**
Globaler T_m der DNA (°C)	59.7	53	59.5	63.3
Lokaler T_m der DNA (°C)	61.1	53	59.1	
Lokaler T_m des Chromophors (°C)	61.1	53.1	56.1	

Das HCF Chromophor ist das größte von allen fluoreszierenden Basenanaloga in dieser Arbeit und wurde gegenüber einer abasischen Stelle eingebaut, was es zu einem vollen Basenpaar-Ersatz macht. HCF nimmt zwei Konformationen innerhalb des Doppelstranges ein, welche durch die Methylengruppe unterschieden werden, in dem Sinne, dass diese Gruppe entweder in Richtung der großen Furche oder der kleinen Furche zeigt. Das HCF-Gerüst passt für beide Konformere sehr gut in die helikale Faltung. Die ungewöhnliche α-glykosidische Bindung des HCFs wird kompensiert durch kleine Deformationen des 2'-Deoxyribose-Teils. Der Zucker geht dabei zu Konformationen über, die von C2'-endo für B-DNA abweichen, um einen optimalen Einbau des Chromophors zu ermöglichen. Aufgrund der Flexibilität des Zuckers wird eine Gleichverteilung beider Konformere beobachtet. Die abasische Stelle hat keinen Effekt auf das HCF, aber das Fehlen einer Nukleobase erhöht die Flexibilität im Zentrum des Gegenstranges, da dessen Position nicht durch die Stapelung einer angehängten Nukleobase eingeschränkt wird. Auf der anderen Seite werden Stapelwechselwirkungen des HCF zu den benachbarten Adeninen begleitet von zusätzlichen Wechselwirkungen mit Thymin T21 (und teilweise T19) des Gegenstranges (siehe Fig. 4.6). Der lokale Schmelzpunkt des Chromophors und der DNA (61.1 °C, in Tab. 6.1) stimmen überein, und sie sind 1,4 °C höher als der globale Schmelzpunkt des Doppelstranges (59,7 °C). Obwohl das lokale Schmelzen einen stabilisierenden Effekt des HCFs anzeigt, ist der globale Schmelzpunkt niedriger als der des natürlichen Doppelstranges 13merRef mit AT im Zentrum (63,3 °C). Beide Ergebnisse stimmen sehr gut mit früheren Resultaten von Matray und Kool[97] überein, welche Pyren als Basenpaar-Ersatz verwendeten.

Das zweite Basenanalog, 6HQ, wurde gegenüber Cytosin platziert, mit dem es eine einzelne Wasserstoffbrücke bildet. Die einzigartige Eigenschaft dieses Doppelstranges ist der Einbau von 6HQ als GNA. Bisher sind nur Röntgen-Kristallstrukturen[274,275] von Nukleinsäuren mit vollem GNA-Rückgrat bekannt, daher ist die Struktur in Lösung von generellem Interesse. Unser Linker R-Glycerol ist klein im Vergleich zu 2'-Deoxyribofuranose. Eine C-C-Bindungslänge fehlt für den Einbau des Chromophors in die Helix und eine andere C-C-Bindung fehlt im Raum zwischen den Adeninen. Das erste Manko wird durch die Größe des 6HQ und dessen Bindung via N1 ausgeglichen, diese erlaubt eine mehr diagonale Positionierung von 6HQ relativ zu Cytosin, um die nötige Distanz für die Wasserstoffbrücke zu erreichen (Fig. 4.13). Ein Nebeneffekt dieser Positionierung ist eine starke π-π-Stapelwechselwirkung zu den Adeninen A6 und A8 (Fig. 4.17), welche zu einer starken chemischen Verschiebung des H8 Wasserstoffs in Adenin A8 führen. Die Notwendigkeit die zweite fehlende Bindungslänge auszugleichen verursacht strukturelle Störungen. Um die Distanz zwischen dem 3'-Ende von Adenin A6 und dem 5'-Ende von A8 zu minimieren, neigen sich beide Basen in Richtung der großen Furche, wodurch der Abstand im Vergleich zu 13merRef um 1,2 Å kürzer ist. Infolgedessen wird eine Stapelachse für A6-6HQ-A8 beobachtet, die sich zur großen Furche hin neigt, stabilisiert durch starke Stapelwechselwirkungen dazwischen. Für die Schmelzanalyse von 13mer6HQ war es notwendig eine Variation der neuen Methode zur UV/Vis-Spektralanalyse zu entwickeln, um das komplexe Verhalten des 6HQ Chromophors angemessen zu berücksichtigen. Das doppelt SVD gestützte Zwei-Zustände-Modell offenbarte, dass 6HQ beim Schmelzen nicht nur eine Blauverschiebung und Hyperchromie erfährt, sondern auch eine Änderung des Absorptionsspektrums, welche dem Übergang von der protonierten zur deprotonierten Form ähnelt. Ein genauer Blick auf das globale und lokale Schmelzen zeigt, dass innerhalb der Fehlergrenzen ($\pm$ 0,1 °C) nur ein Schmelzpunkt beobachtet wird, welcher sonst nur in reinen natürlichen Strängen vorkommt. Dennoch beweist dies nicht, dass 13mer6HQ auf natürliche Weise schmilzt, da es von allen Doppelsträngen den niedrigsten Schmelzpunkt besitzt. Die abweichende Stapelachse destabilisiert den 6HQ-Strang, aber die starken Stapelwechselwirkungen verhindern auch eine Blasenbildung. Die Stabilität im Duplex wird nur durch die verbliebenen Basenpaare fern vom Zentrum gewährleistet. Dies konkurriert wiederum mit dem Ausfransen an den Helix-Enden bei steigender Temperatur. Insgesamt zeigt dieser Doppelstrang das Schmelzverhalten eines kürzeren Duplexes.

Der letzte Doppelstrang führt ein künstliches Basenpaar bestehend aus 4AP und DAP ein. Das erste Anzeichen, welches den drei gewünschten Wasserstoffbrücken widersprach, wurde im NOESY-Spektrum gefunden. Darin wurde die chemische Verschiebung des H1 Wasserstoffs (in 4AP) in der nicht über Wasserstoffbrücken gebundenen Region des NOESY-Spektrums zugeordnet. Außerdem bewies das H_2O-NOESY das Vorhandensein der protonierten Form von DAP (Fig. 4.34). Anstelle einer Struktur mit drei Wasserstoffbrücken wurden zwei Konformere gefunden, die entweder eine einzelne oder zwei Wasserstoffbrücken enthalten. Die Form mit einer Brücke ähnelt der Struktur von 13mer6HQ, bei der 4AP-Teil diagonal liegt, so dass dessen lange Seite in Richtung DAP zeigt. Der Grund dafür ist die Bindung über C6, welche dem N1 in 6HQ entspricht und wohlbekannt ist von den Pyrimidin-Basen. Jedoch sind 6HQ und 4AP in Größe und Form eher vergleichbar mit den Purinen, welche über den Fünfring angebunden sind, was wiederum zu einem anderen Winkel zwischen Nukleobase und dem Gegenstück im gegenüberliegenden Strang führt. Das Konformer, welches zwei Wasserstoffbrücken trägt, versucht dies zu umgehen indem der Zucker von 4AP zur C3'-endo Konformation von A-DNA wechselt. Obwohl sich auf den ersten Blick nun alle zu Wasserstoffbrücken fähigen Stellen gegenüberstehen, zeigen Abstandsmessungen (in Fig. 4.18), dass nur zwei Bindungen möglich sind. Interessanterweise zeigen die Schmelzexperimente, dass die zweite Form, welche bessere π-π-Stapelwechselwirkungen aufweist, mit steigender Temperatur bevorzugt wird, wodurch die Absorption zu Beginn sinkt. Letztlich aber zeigt sich, wenn man das globale und lokale Schmelzen von 13mer4AP-DAP miteinander vergleicht (Tab. 6.1), dass sich eine Blase durch vorzeitiges Schmelzen des zentralen 4AP-DAP-Paares bildet. Der Schmelzpunkt der benachbarten Basenpaare ist ebenfalls niedriger als der globale, was die Blasenbildung im Zentrum unterstützt. Andererseits beträgt der Unterschied nur 0,4 °C, somit scheint die Blase hauptsächlich auf das 4AP-DAP Basenpaar limitiert zu sein.

Zusammengefasst leiden alle untersuchten Doppelstränge unter strukturellen Störungen, welche entweder durch den Linker oder das Chromophor selbst verursacht werden. Die weitere Entwicklung von Basenanaloga und ihrer Anbindung ist daher auch in Zukunft zwingend notwendig. Jedoch lassen sich einige allgemeine Lehren aus der Auswahl an Strukturen in dieser Arbeit ziehen. Aus dem 13merHCF kann geschlussfolgert werden, dass eine α-glykosidische Bindung durch die 2'-Deoxyribofuranose ausgeglichen werden kann, und dass eine große aromatische Oberfläche ein komplettes Basenpaar ersetzen kann, aber zum Preis erhöhter Flexibilität. 13mer6HQ zeigt, dass R-Glycerol im Vergleich zu 2'-Deoxyribofuranose zu klein ist. Eine potentielle Lösung wäre der Einbau wenigstens einer zusätzlichen Bindungslänge, um das 3'-Ende von Adenin A6 mit dem 5'-Ende von A8 zu verbinden, aber das würde die Flexibilität erhöhen. Das Problem erhöhter Flexibilität innerhalb des Doppelstranges ist eine schnellere Basenpaardynamik und somit eine geringere Duplexstabilität. Außerdem pflanzt sich dieser Effekt über die umliegenden Basenpaare in beide Richtungen fort[129]. Aus dieser Perspektive sieht das 4AP-DAP Basenpaar angebunden über die natürliche 2'-Deoxyribofuranose vielversprechender aus. Jedoch ist der Schwachpunkt dieses Designs die Anbindung von 4AP an das Rückgrat. Die Verbindung zum Zucker-Teil sollte geändert werden vom aktuellen C6-Atom in 4AP zum benachbarten C5. Das sollte es 4AP erlauben ein dreifach wasserstoffverbrücktes Basenpaar mit DAP zu bilden. In diesem Fall kann man erwarten, dass die Dreifach-Form das einzig vorhandene Konformer sein wird, da die beiden gegenwärtigen Konformere die Anbindung über C6 benötigen.

Die SVD-Analyse der UV/Vis-Spektren in ihrer Gesamtheit offenbarte ein messbares lokales Schmelzen in der DNA-Absorptionsbande. Das lokale Schmelzen wird durch Störungen in den Doppelsträngen verursacht und wird nur durch die natürlichen Basen in unmittelbarer Umgebung der Modifikation wiedergegeben. Natürliche Nukleobasen als simultaner Detektor für globales und lokales Schmelzen ermöglichen es Modifikationen zu untersuchen, welche im Gegensatz zu den hier genutzten Chromophoren keine separate Absorptionsbande enthalten. Eine zweite Möglichkeit könnten Schmelzexperimente natürlicher Doppelstränge sein, denen ein Interkalator hinzugefügt wurde, wie z.B. der Furchenbinder Hoechst 33258. In jenem Fall würden Störungen des Schmelzprozesses vom Strang angezeigt werden und könnten anschließend mit dem Signal des Hoechst-Farbstoffs verglichen werden. Die Abwesenheit der lokalen Schmelzkomponente in natürlichen Strängen wie 13merRef kann wiederum als Indikator für ungestörtes Schmelzen in modifizierten Doppelsträngen dienen. Schließlich sollte noch angemerkt werden, dass die Methode sowohl experimentell als auch analytisch automatisiert werden kann. Die voll automatisierte Variante würde es erlauben das konventionelle DNA-Schmelzexperiment bei einer Wellenlänge zu ersetzen.

Appendix

1 Chemical shift tables

1.1 13merHCF shift tables

Table 1: ^{1}H chemical shifts of the nucleobases in 13merHCF. Reference[277] is the signal of HOD at 4.77 pm (298 K).

Res	H1	H2	H3	H4	H41	H42	H5	H6	H7	H8
1	-	-	-	-	-	-	-	-	-	7.946
2	-	-	-	-	8.269	6.56	5.334	7.51	-	-
3	-	-	13.937	-	-	-	-	7.316	1.62	-
4	12.688	-	-	-	-	-	-	-	-	7.668
5	-	-	-	-	8.189	6.351	5.402	7.344	-	-
6	-	7.105	-	-	-	-	-	-	-	8.313
7	5.851	-	5.905	6.662	-	-	6.94	7.124	-	7.273
8	-	7.346	-	-	-	-	-	-	-	8.018
9	-	-	-	-	7.967	6.373	5.091	7.148	-	-
10	12.705	-	-	-	-	-	-	-	-	7.802
11	-	-	13.747	-	-	-	-	7.266	1.38	-
12	-	-	-	-	8.591	7.002	5.695	7.485	-	-
13	-	-	-	-	-	-	-	-	-	7.932
14	-	-	-	-	8.16	6.947	5.857	7.578	-	-
15	12.893	-	-	-	-	-	-	-	-	7.934
16	-	7.882	-	-	-	-	-	-	-	8.203
17	-	-	-	-	8.07	6.413	5.201	7.168	-	-
18	12.546	-	-	-	-	-	-	-	-	7.792
19	-	-	13.213	-	-	-	-	7.191	1.438	-
20	-	-	-	-	-	-	-	-	-	-
21	-	-	-	-	-	-	-	7.209	1.512	-
22	12.545	-	-	-	-	-	-	-	-	7.811
23	-	-	-	-	8.253	6.334	5.351	7.328	-	-
24	-	7.654	-	-	-	-	-	-	-	8.161
25	12.908	-	-	-	-	-	-	-	-	7.653
26	-	-	-	-	8.095	6.446	5.201	7.303	-	-

In addition to the tabulated spins HCF contains a H91 and H92 at 3.140 and 2.770 ppm, respectively. The corresponding carbon atom is C9 at 39.3 ppm.

Table 2: ^{1}H chemical shifts of the backbone in 13merHCF

Res	H1'	H1"	H2'	H2"	H3'	H4'	H5'	H5"
1	5.971	-	2.641	2.766	4.837	4.262	-	-
2	6.074	-	2.115	2.514	4.834	4.252	-	-
3	5.727	-	2.113	2.427	4.868	4.095	-	-
4	5.813	-	2.619	2.654	4.978	4.355	-	-
5	5.429	-	1.937	2.275	4.794	4.130	-	-
6	6.317	-	2.755	2.805	5.036	4.377	4.123	4.044
7	-	5.238	2.269	2.352	4.846	4.591	-	-
8	6.092	-	2.51	2.793	4.955	4.394	4.228	4.161
9	5.487	-	2.001	2.32	4.799	4.129	-	-
10	5.94	-	2.587	2.762	4.922	4.349	-	-
11	6.031	-	2.072	2.451	4.856	4.214	-	-
12	5.695	-	2.015	2.359	4.835	4.116	-	-
13	6.152	-	2.62	2.384	4.685	4.181	-	-
14	5.695	-	1.834	2.338	4.673	4.047	3.71	3.692
15	5.435	-	2.71	2.786	4.993	4.303	-	-
16	6.25	-	2.699	2.909	5.056	4.476	-	-
17	5.559	-	1.934	2.306	4.800	4.14	-	-
18	5.918	-	2.575	2.736	4.94	4.342	-	-
19	6.104	-	2.318	2.500	4.902	4.231	-	-
20	4.133	4.109	2.212	2.231	4.761	4.127	-	-
21	5.771	-	1.967	2.406	4.855	4.289	-	-
22	5.746	-	2.586	2.76	4.948	4.316	-	-
23	5.431	-	1.947	2.288	4.798	4.130	-	-
24	5.990	-	2.722	2.876	5.026	4.370	-	-
25	5.798	-	2.454	2.634	4.948	4.339	4.178	4.215
26	6.075	-	2.129	2.196	4.443	4.023	-	-

Table 3: ^{13}C chemical shifts of 13merHCF

Res	C1'	C2'	C3'	C4'	C1	C2	C3	C4	C5	C6	C8
1	82.1	-	76.6	-	-	-	-	-	-	-	135.8
2	83.7	-	74.0	-	-	-	-	-	95.7	140.1	-
3	82.7	-	75.0	-	-	-	-	-	-	136.6	-
4	81.4	-	76.6	-	-	-	-	-	-	-	135.4
5	83.6	-	73.6	-	-	-	-	-	95.4	139.7	-
6	82.4	-	76.7	-	-	151.5	-	-	-	-	139.1
7	102.0	-	76.4	85.9	109.7	-	112.8	120.3	118.0	125.6	122.3
8	83.1	-	75.0	-	-	150.8	-	-	-	-	138.3
9	83.3	-	73.7	-	-	-	-	-	95.1	139.2	-
10	82.1	-	76.3	-	-	-	-	-	-	-	135.4
11	82.7	-	75.2	-	-	-	-	-	-	135.8	-
12	83.7	-	75.2	-	-	-	-	-	96.0	141.0	-
13	81.9	-	70.5	-	-	-	-	-	-	-	136.5
14	85.0	-	75.0	-	-	-	-	-	96.7	140.3	-
15	81.2	-	76.6	-	-	-	-	-	-	-	135.5
16	82.2	-	76.9	-	-	152.4	-	-	-	-	138.3
17	83.1	-	73.7	-	-	-	-	-	95.1	139.1	-
18	81.9	-	76.4	-	-	-	-	-	-	-	135.3
19	82.7	-	74.6	-	-	-	-	-	-	135.9	-
20	67.3	32.9	77.3	-	-	-	-	-	-	-	-
21	83.7	-	74.8	-	-	-	-	-	-	135.8	-
22	81.2	-	76.3	-	-	-	-	-	-	-	135.2
23	83.5	-	73.9	-	-	-	-	-	95.4	139.7	-
24	82.3	-	76.9	-	-	151.5	-	-	-	-	138.7
25	81.1	-	76.2	-	-	-	-	-	-	-	134.4
26	84.0	-	68.5	-	-	-	-	-	95.4	140.3	-

1.2 13mer6HQ shift tables

Table 4: ^{1}H chemical shifts of the nucleobases in 13mer6HQ. Reference[277] is the signal of HOD at 4.77 pm (298 K).

Res	H1	H2	H3	H4	H41	H42	H5	H6	H7	H8
1	-	-	-	-	-	-	-	-	-	7.94
2	-	-	-	-	8.296	6.593	5.329	7.511	-	-
3	-	-	13.93	-	-	-	-	7.323	1.626	-
4	12.683	-	-	-	-	-	-	-	-	7.87
5	-	-	-	-	8.27	6.372	5.36	7.234	-	-
6	-	6.717	-	-	-	-	-	-	-	8.327
7	-	8.82	7.623	8.242	-	-	7.009	-	6.596	7.403
8	-	7.68	-	-	-	-	-	-	-	7.64
9	-	-	-	-	8.045	6.422	5.071	7.097	-	-
10	12.722	-	-	-	-	-	-	-	-	7.803
11	-	-	13.758	-	-	-	-	7.265	1.38	-
12	-	-	-	-	8.607	7.011	5.691	7.481	-	-
13	-	-	-	-	-	-	-	-	-	7.926
14	-	-	-	-	8.149	6.979	5.845	7.57	-	-
15	12.937	-	-	-	-	-	-	-	-	7.945
16	-	7.889	-	-	-	-	-	-	-	8.216
17	-	-	-	-	8.113	6.445	5.229	7.19	-	-
18	12.684	-	-	-	-	-	-	-	-	7.832
19	-	-	-	-	-	-	-	7.246	1.442	-
20	-	-	-	-	-	-	5.663	7.559	-	-
21	-	-	-	-	-	-	-	7.419	1.619	-
22	12.5	-	-	-	-	-	-	-	-	7.868
23	-	-	-	-	8.296	6.339	5.361	7.333	-	-
24	-	7.645	-	-	-	-	-	-	-	8.151
25	12.911	-	-	-	-	-	-	-	-	7.645
26	-	-	-	-	8.108	6.527	5.166	7.282	-	-

Table 5: ^{1}H chemical shifts of the backbone in 13mer6HQ

Res	H1'	H1"	H2'	H2"	H3'	H4'	H5'	H5"
1	5.96	-	2.639	2.76	4.836	4.267	3.736	3.728
2	6.067	-	2.118	2.512	4.838	4.161	4.135	4.093
3	5.721	-	2.116	2.435	4.873	4.137	4.091	4.05
4	5.808	-	2.606	2.666	4.974	4.353	4.132	4.058
5	5.528	-	1.803	2.279	4.846	4.117	-	-
6	5.876	-	2.706	2.802	5.028	4.364	4.111	4.043
7	4.993	4.833	4.299	-	4.199	-	-	-
8	5.82	-	2.365	2.681	4.719	4.161	3.899	3.524
9	5.454	-	1.959	2.297	-	-	-	-
10	5.952	-	2.586	2.762	4.927	4.351	4.112	4.053
11	6.041	-	2.077	2.451	4.86	4.219	4.128	4.115
12	5.693	-	2.013	2.358	4.837	4.217	4.12	4.068
13	6.147	-	2.623	2.383	4.683	4.18	4.117	4.083
14	5.681	-	1.821	2.33	4.669	4.047	3.713	3.689
15	5.436	-	2.714	2.791	4.998	4.309	4.081	3.962
16	6.257	-	2.714	2.917	5.067	4.482	4.225	4.153
17	5.575	-	1.939	2.314	4.816	4.483	4.266	4.151
18	5.938	-	2.597	2.74	4.939	4.359	4.142	4.066
19	6.055	-	2.152	2.512	4.848	4.256	4.207	4.124
20	5.951	-	2.019	2.471	4.843	4.216	4.125	4.096
21	5.724	-	2.163	2.403	4.873	4.239	4.122	4.089
22	5.788	-	2.608	2.648	4.972	4.348	4.126	4.044
23	5.433	-	1.966	2.294	4.815	4.137	4.109	4.083
24	5.978	-	2.711	2.864	5.022	4.357	4.109	3.967
25	5.787	-	2.444	2.624	4.944	4.337	4.17	4.127
26	6.061	-	2.121	2.193	4.436	4.261	4.018	4.021

Table 6: ^{13}C chemical shifts of 13mer6HQ

Res	C1'	C2'	C3'	C4'	C1	C2	C3	C4	C5	C6	C8
1	82.1	-	76.5	86.2	-	-	-	-	-	-	135.5
2	84.0	-	74.0	-	-	-	-	-	95.7	140.2	-
3	83.0	-	75.3	-	-	-	-	-	-	136.5	-
4	81.4	37.6	76.6	-	-	-	-	-	-	-	135.4
5	83.3	-	-	-	-	-	-	-	95.5	139.7	-
6	81.5	-	76.9	-	-	151.1	-	-	-	-	138.9
7	-	-	-	-	-	144.1	121.3	145.3	110.6	-	117.4
8	82.3	36.4	-	-	-	151.6	-	-	-	-	137.9
9	83.3	36.9	76.3	-	-	-	-	-	95.2	139.3	-
10	82.1	-	76.6	-	-	-	-	-	-	-	135.5
11	82.8	36.4	75.2	-	-	-	-	-	-	135.8	-
12	83.7	36.5	75.2	-	-	-	-	-	96.1	140.9	-
13	82.0	39.1	70.6	85.1	-	-	-	-	-	-	136.4
14	85.2	37.1	75.1	85.5	-	-	-	-	96.7	140.3	-
15	81.4	-	76.8	-	-	-	-	-	-	-	135.5
16	82.2	-	77.0	84.8	-	152.5	-	-	-	-	138.3
17	83.2	37.0	74.1	-	-	-	-	-	95.5	139.1	-
18	82.0	37.6	76.6	-	-	-	-	-	-	-	135.5
19	82.7	36.3	-	-	-	-	-	-	-	135.8	-
20	84.3	37.4	-	-	-	-	-	-	95.1	141.3	-
21	82.9	-	-	83.3	-	-	-	-	-	136.6	-
22	81.3	-	76.8	-	-	-	-	-	-	-	135.5
23	83.4	36.9	74.2	-	-	-	-	-	95.5	139.7	-
24	82.2	37.6	76.9	84.5	-	151.9	-	-	-	-	138.9
25	81.5	39.0	75.2	-	-	-	-	-	-	-	134.5

1.3 13mer4AP-DAP shift tables

Table 7: ^{1}H chemical shifts of the nucleobases in 13mer4AP-DAP. Reference[277] is the signal of HOD at 4.94 pm (283 K).

Res	H1	H2	H3	H41	H42	H5	H6	H7	H8
1	12.697	-	-	-	-	-	-	-	7.904
2	-	-	-	8.255	6.585	5.251	7.462	-	-
3	-	-	13.952	-	-	-	7.293	1.569	-
4	12.704	-	-	-	-	-	-	-	7.846
5	-	-	-	8.231	6.365	5.324	7.234	-	-
6	-	7.424	-	-	-	-	-	-	8.001
		7.502							
7	11.228	-	-	-	-	6.452	-	-	-
8	-	7.347	-	-	-	-	-	-	8.079
		7.760						-	
9	-	-	-	7.974	6.459	5.128	7.185	-	-
10	12.675	-	-	-	-	-	-	-	7.79
11	-	-	13.789	-	-	-	7.241	1.324	-
12	-	-	-	8.587	7.06	5.636	7.448	-	-
13	-	-	-	-	-	-	-	-	7.888
14	-	-	-	8.212	6.972	5.782	7.528	-	-
15	12.92	-	-	-	-	-	-	-	7.911
16	-	7.831	-	-	-	-	-	-	8.172
17	-	-	-	8.084	6.486	5.153	7.144	-	-
18	12.697	-	-	-	-	-	-	-	7.781
	12.682							-	
19	-	-	14.179	-	-	-	7.239	1.425	-
			13.791						
20	6.795	-	-	-	-	-	7.598	-	-
21	-	-	13.443	-	-	-	7.361	1.565	-
			13.670						
22	12.56	-	-	-	-	-	-	-	7.831
23	-	-	-	8.271	6.374	5.332	7.321	-	-
24	-	7.591	-	-	-	-	-	-	8.12
25	12.885	-	-	-	-	-	-	-	7.615
26	-	-	-	8.086	6.469	5.046	7.198	-	-

Table 8: ^{1}H chemical shifts of the backbone in 13mer4AP-DAP

Res	H1'	H2'	H2"	H3'	H4'	H5'	H5"
1	5.903	2.607	2.712	4.785	4.233	3.692	3.667
2	6.015	2.073	2.454	4.787	4.191	4.107	4.054
3	5.647	2.089	2.399	4.807	4.189	4.086	4.014
4	5.77	2.57	2.603	4.941	4.314	4.083	4.004
5	5.418	1.844	2.203	4.732	4.181	4.089	4.045
6	5.987	2.549	2.89	4.788	4.285	4.066	3.962
7	6.241	1.534	2.004	4.605	4.191	3.964	-
8	6.033	2.622	2.739	4.824	4.318	4.045	3.967
9	5.469	1.99	2.286	4.732	4.117	-	-
10	5.927	2.558	2.735	4.92	4.323	4.098	4.024
11	6.002	2.05	2.412	4.764	4.184	4.073	3.965
12	5.589	1.981	2.312	4.785	4.147	4.074	3.997
13	6.094	2.594	2.327	4.644	4.143	4.05	4.024
14	5.597	1.787	2.287	4.621	3.998	3.67	3.65
15	5.366	2.67	2.726	4.958	4.26	4.028	4.001
16	6.21	2.663	2.868	5.01	4.443	4.176	4.1
17	5.517	1.928	2.279	4.774	-	4.11	4.211
18	5.888	2.572	2.706	4.905	4.314	4.101	4.033
19	5.952	2.074	2.446	4.806	4.319	4.182	4.085
20	4.676	1.955	2.294	4.723	4.186	4.083	4.015
21	5.709	2.116	2.387	4.832	4.149	4.082	4.04
22	5.771	2.575	2.617	4.923	4.305	4.08	3.986
23	5.348	1.956	2.263	4.765	4.084	-	-
24	5.943	2.669	2.834	4.982	4.327	4.078	3.936
25	5.726	2.392	2.579	4.925	4.298	3.998	3.961
26	5.981	2.095	2.144	4.389	4.227	3.967	3.956

2 Input files for Molecular Dynamics calculations

In this chapter the protocols of the MD simulations are given using 13merHCF as example. All samples were using the same set of protocols, so they are not shown explicitly for each one. In case of 13mer4AP-DAP, a modification was necessary for the final MD script. All parts that are necessary to implement RDC values were commented, since they were not measured for this duplex.

2.1 Input file to generate extended strands

This file is used to generate extended single strands of the duplexes that were used as input to calculate the start-structure.

```
remarks  file  lars_13merHC-.inp
remarks  Sequence 13merHC-,lars 5/2011

 topology  @lars_nucleic.top      {*Read topology file for chldna.*}

 end

 parameter
 @lars_nucleic.par

 hbonds
     acce=true don=5.5 doff=6.5
     dcut=7.5 aon=60.0 aoff=80.0
     acut=100.0
end

   nbonds
      atom cdie shift eps=1.0  e14fac=0.4
      cutnb=7.5 ctonnb=6.0 ctofnb=6.5
      nbxmod=5 vswitch
   end

end                                         {*We are generating one strand*}
                                         {*at a time.                *}
  segment
   name="A   "                         {*This name has to match the    *}
                                       {*four characters in columns 73-*}
                                       {*76 in the coordinate          *}
                                       {*file; in XPLOR this name is   *}
   chain

 LINK NUC  HEAD - *  TAIL + *  END

 LAST 3TER  HEAD - * END
```

```
FIRST  5TER  TAIL + * END


Sequence GUA CYT THY GUA CYT ADE HC- ADE CYT GUA THY CYT GUA end

end

end

for $1 in ( 1 2 3 4 5 6 7 8 9 10 11 12 13) loop main
    patch deox reference=nil=( resid $1 ) end
end loop main

                                             {*at a time.                 *}
  segment
   name="B  "                             {*This name has to match the    *}
                                           {*four characters in columns 73-*}
                                           {*76 in the coordinate         *}
                                           {*file; in XPLOR this name is   *}
   chain

 LINK NUC  HEAD - *  TAIL + *  END

 LAST 3TER  HEAD - * END

FIRST  5TER  TAIL + * END


Sequence CYT GUA ADE CYT GUA THY ABA THY GUA CYT ADE GUA CYT end


end

end

for $1 in ( 1 2 3 4 5 6 7 8 9 10 11 12 13) loop main
    patch deox reference=nil=( segid b AND resid $1 ) end
end loop main





end

vector do ( resid = encode ( decode ( resid ) + 13 ) ) (segid "B   ")
vector do (segid = "    ") (segid "A   " or segid "B   ")



vector ident (x) ( all )
vector do (x=x/3.) ( all )
vector do (y=random(0.5) ) ( all )
vector do (z=random(0.5) ) ( all )
```

```
vector do (fbeta=50) (all)                {*Friction coefficient, in 1/ps.*}
vector do (mass=100) (all)                        {*Heavy masses, in amus.*}



parameter
   nbonds
      cutnb=5.5 rcon=20. nbxmod=-2 repel=0.9  wmin=1. tolerance=0.5
      rexp=2 irexp=2 inhibit=0.25
   end
end

flags exclude * include bond angle vdw end

minimize powell nstep=500  nprint=10 end

flags include impr dihedral end

minimize powell nstep=500 nprint=10 end

dynamics  verlet
   nstep=500  timestep=0.001 iasvel=maxwell  firsttemp= 300.
   tcoupling = true  tbath = 300.   nprint=50  iprfrq=0
end

parameter
   nbonds
      rcon=2. nbxmod=-3 repel=0.75
   end
end

minimize powell nstep=100 nprint=25 end

dynamics  verlet
   nstep=1500  timestep=0.001 iasvel=maxwell  firsttemp= 300.
   tcoupling = true  tbath = 300.   nprint=100  iprfrq=0
end

flags exclude vdw elec end
hbuild selection=( hydrogen ) phistep=360 end
hbuild selection=( hydrogen ) phistep=4 end
flags include vdw elec end

minimize powell nstep=4000 nprint=50 end

                                                        {*Write coordinates.*}
write coordinates output=start_13merHC-.pdb end
write structure output=start_13merHC-.psf end

print threshold=0.02 bonds
print threshold=3.0 angles
print threshold=3.0 dihedrals
print threshold=3.0 impropers

stop
```

2.2 Input file to generate start-structure

This protocol is used to generate the start-structure and only uses NOE distance restraints.

```
remarks file  nmr/sa.inp
remarks  Simulated annealing protocol for NMR structure determination.
remarks  The starting structure for this protocol can be any structure with
remarks  a reasonable geometry, such as randomly assigned torsion angles or
remarks  extended strands.
remarks  Author: Michael Nilges

{====>}
evaluate ($init_t = 3000 )       {*Initial simulated annealing temperature.*}
{====>}
evaluate ($high_steps= 48000 )         {*Total number of steps at high temp.*}
{====>}
evaluate ($cool_steps = 6000 )      {*Total number of steps during cooling.*}

parameter                                        {*Read the parameter file.*}
{====>}
   @lars_nucleic.par
end

{====>}
structure @start_13merHC-.psf end                     {*Read the structure file.*}
{====>}
coordinates @start_13merHC-.pdb                  {*Read the coordinates.*}



noe
{====>}
   nres=3000             {*Estimate greater than the actual number of NOEs.*}
   class all
{====>}
   @NOE_13merHCF_xplor.tbl                            {*Read NOE distance ranges.*}
   @hbond_13mer_HCF.tbl
end

{====>}
restraints dihedral
   nass = 1000
   @dihedral_13mer_HCF_cut_ABDNA.tbl                       {*Read dihedral angle
   restraints.*}
end
@plane_13mer_HCF.inp

{* Reduce the scaling factor on the force applied to disulfide            *}
{* bonds and angles from 1000.0 to 100.0 in order to reduce computation instability. *}
parameter
      bonds ( name SG ) ( name SG ) 100. TOKEN
      angle ( name CB ) ( name SG ) ( name SG ) 50. TOKEN
end


flags exclude * include bonds angle impr vdw elec noe cdih plan end
```

```
                        {*Friction coefficient for MD heatbath, in 1/ps.   *}
vector do (fbeta=10) (all)
                        {*Uniform heavy masses to speed molecular dynamics.*}
vector do (mass=100) (all)

noe                             {*Parameters for NOE effective energy term.*}
  ceiling=1000
  averaging  * cent
  potential  * soft
  scale      * 50.
  sqoffset   * 0.0
  sqconstant * 1.0
  sqexponent * 2
  soexponent * 1
  asymptote  * 0.1                         {*Initial value--modified later.*}
  rswitch    * 0.5
end

parameter                       {*Parameters for the repulsive energy term.*}
    nbonds
      repel=1.                   {*Initial value for repel--modified later.*}
      rexp=2 irexp=2 rcon=1.
      nbxmod=3
      wmin=0.01
      cutnb=4.5 ctonnb=2.99 ctofnb=3.
      tolerance=0.5
   end
end

restraints dihedral
      scale=5.
end

{====>}
evaluate ($end_count=100)          {*Loop through a family of 100 structures.*}

coor copy end

evaluate ($count = 0)
evaluate ($count2 = 0)
while ($count < $end_count ) loop main

   evaluate ($count=$count+1)
   evaluate ($count2=$count2+1)

   coor swap end
   coor copy end

     {* =========================================== Initial minimization.*}
  restraints dihedral   scale=5.   end
   noe asymptote * 0.1  end
   parameter  nbonds repel=1.   end end
   constraints interaction
           (all) (all) weights * 1  vdw 0.002 end end
   minimize powell nstep=50 drop=10.  nprint=25 end
```

```
  {* ======================================= High-temperature dynamics.*}
 constraints interaction (all) (all)
             weights * 1  angl 0.4  impr 0.1 vdw 0.002 end end

 evaluate ($nstep1=int($high_steps * 2. / 3. ) )
 evaluate ($nstep2=int($high_steps * 1. / 3. ) )

 dynamics  verlet
    nstep=$nstep1   timestep=0.003   iasvel=maxwell   firstt=$init_t
    tcoupling=true  tbath=$init_t  nprint=50  iprfrq=0
 end


  {* ============== Tilt the asymptote and increase weights on geometry.*}
 noe asymptote * 1.0  end

 constraints interaction
            (all) (all) weights * 1  vdw 0.002  end end

 {* Bring scaling factor for S-S bonds back *}
 parameter
   bonds ( name SG ) ( name SG ) 1000. TOKEN
   angle ( name CB ) ( name SG ) ( name SG ) 500. TOKEN
end

 dynamics  verlet
    nstep=$nstep2   timestep=0.001    iasvel=current   tcoupling=true
    tbath=$init_t  nprint=50  iprfrq=0
 end

  {* =================================================  Cool the system.*}

restraints dihedral   scale=200.   end

 evaluate ($final_t = 100)      { K }
 evaluate ($tempstep = 50)      { K }

 evaluate ($ncycle = ($init_t-$final_t)/$tempstep)
 evaluate ($nstep = int($cool_steps/$ncycle))

 evaluate ($ini_rad  = 0.9)        evaluate ($fin_rad  = 0.75)
 evaluate ($ini_con=  0.003)       evaluate ($fin_con=  4.0)

 evaluate ($bath  = $init_t)
 evaluate ($k_vdw = $ini_con)
 evaluate ($k_vdwfact = ($fin_con/$ini_con)^(1/$ncycle))
 evaluate ($radius=    $ini_rad)
 evaluate ($radfact = ($fin_rad/$ini_rad)^(1/$ncycle))

 evaluate ($i_cool = 0)
 while ($i_cool < $ncycle) loop cool
    evaluate ($i_cool=$i_cool+1)

    evaluate ($bath  = $bath  - $tempstep)
```

```
      evaluate ($k_vdw=min($fin_con,$k_vdw*$k_vdwfact))
      evaluate ($radius=max($fin_rad,$radius*$radfact))

      parameter  nbonds repel=$radius   end end
      constraints interaction (all) (all)
                     weights * 1. vdw $k_vdw end end

      dynamics  verlet
         nstep=$nstep time=0.001 iasvel=current firstt=$bath
         tcoup=true tbath=$bath nprint=$nstep iprfrq=0
      end


{====>}                                                  {*Abort condition.*}
      evaluate ($critical=$temp/$bath)
      if ($critical >  10. ) then
         display  ****&&&& rerun job with smaller timestep (i.e., 0.003)
         stop
      end if

   end loop cool
   {* ================================================= Final minimization.*}

   constraints interaction (all) (all) weights * 1. vdw 1. end end

   parameter                       {*Parameters for the repulsive energy term.*}
     nbonds
      repel=0.                   {*Initial value for repel--modified later.*}
      SWITCH
      VSWITCH
      RDIE
      cutnb=11.5
      nbxmod=5
      wmin=0.01
      ctofnb=10.5
      ctonnb=9.5
      tolerance=0.5
    end
   end
   flags exclude * include bonds angle impr vdw elec noe cdih plan end

   minimize powell nstep=5000 drop=10.0 nprint=25 end

   {* ================================== Write out the final structure(s).*}
   print threshold=0.5 noe
   evaluate ($rms_noe=$result)
   evaluate ($violations_noe=$violations)
   print threshold=5. cdih
   evaluate ($rms_cdih=$result)
   evaluate ($violations_cdih=$violations)
   print thres=0.05 bonds
   evaluate ($rms_bonds=$result)
   print thres=5. angles
   evaluate ($rms_angles=$result)
   print thres=5. impropers
   evaluate ($rms_impropers=$result)
```

```
   remarks ===============================================================
   remarks            overall,bonds,angles,improper,vdw,noe,cdih,elec
   remarks energies: $ener, $bond, $angl, $impr, $vdw, $noe, $cdih, $elec
   remarks ===============================================================
   remarks            bonds,angles,impropers,noe,cdih
   remarks rms-d: $rms_bonds,$rms_angles,$rms_impropers,$rms_noe,$rms_cdih
   remarks ===============================================================
   remarks               noe,  cdih
   remarks violations.: $violations_noe, $violations_cdih
   remarks ===============================================================
   remarks enviol: $ener $$violations_noe $violations_cdih
   remarks ===============================================================


{====>}                        {*Name(s) of the family of final structures.*}
   evaluate ($filename="z13merHC-"+encode($count)+".pdb")

   write coordinates output =$filename end

   evaluate ($filename2="z13merHC-"+encode($count2)+".noe")
   set display=$filename2 end
   @@picktbl_13merHCF
   close $filename2 end
   set display=OUTPUT end

end loop main

stop
```

2.3 Input file to finally calculate the NMR solution structure

The final MD protocol uses the start-structure, NOE distance restraints and RDC values to calculate the NMR solution structure. In case of 13mer4AP-DAP all lines that implement RDC values were set as comments to omit them in the calculation.

```
# Das ist das zur Zeit gueltige Skript mit allen Neuerungen!
# Hier wird das modifizierte python-file protocol.py benutzt!

seed                    = 10
numberOfStructures      = 100
startStructure          = 1



# User-specific which has to be adjusted for each new sample
lsdSampleName  = "13merHC-"                                    # specify sample name
outFilename     = lsdSampleName+"_STRUCTURE.pdb"          # pdb output filename
lsdNOEexp      = "NOE_13merHCF_xplor.tbl" # file for reading in experimental NOE constraints
lsdDipoInp     = "RDC_13merHCF.inp.xplor"
# file for reading in experimental RDC constraints
lsdDipoInpMe   = "RDC_13merHCF_Me.inp.xplor"
```

```
# file for reading in experimental Me-RDC constraints
lsdInitCoord   = "start_"+lsdSampleName+".pdb"
# file created with initial extended structure coordinates
lsdInitPSF     = "start_"+lsdSampleName+".psf"       # file created with initial extended structure
lsdPlan        = "plane_13merHCF.inp"                      # file for reading in planar constraints
lsdHbond       = "hbond_13merHCF.tbl" # file for reading in Hbond constraints
lsdDihe        = "dihedral_13mer_HCF_cut_ABDNA.tbl"
# file for reading in ideal dihedral constraints
lsdNOEthresh   = 0.5                                          # threshold for NOE error reports
lsdRDCthresh   = 2.5                                          # threshold for RDC error reports
lsdRDCscale    = 5.0                                          # scaling factor for RDCs
lsdMeRDCscale  = 0.5
# scaling factor for methyl RDCs (should usually be 0.1*lsdRDCscale)
lsdFixedRDC    = 1
# if lsdFixedRDC is 0, fixed values are used, else floating ones
lsdDaRDC       = -22.43                                       # Da-value when using fixed values
lsdRhRDC       = 0.2174                                       # Rhombicity when using fixed values

######################## end of lsd mod ########################

xplor.parseArguments()  # check for typos on the command-line

simWorld.setRandomSeed(seed)

#
# Create the PSF and initial PDB files as an extended structure

import protocol
protocol.initParams("lsd_old_nucleic")
protocol.initTopology("lsd_old_nucleic")
protocol.initStruct(lsdInitPSF)


#
# starting coords
#
protocol.initCoords(lsdInitCoord)

# list of potential terms used in refinement
from potList import PotList
potList = PotList()
crossTerms=PotList('cross terms') # can add some pot terms which are not
                                  # refined against- but included in analysis


# parameters to ramp up during the simulated annealing protocol
#
from simulationTools import MultRamp, StaticRamp, InitialParams
rampedParams=[]
highTempParams=[]

from varTensorTools import create_VarTensor, calcTensor
media={}
for medium in ['pf1']:
    media[medium] = create_VarTensor(medium)
    pass
```

```
from xplorPot import XplorPot

#planarity restraints
xplor.command("@%s" % lsdPlan)
potList.append(XplorPot("plan",xplor.simulation))

#NOE potentials
from noePotTools import create_NOEPot
noePots = PotList("noe")
noe = create_NOEPot("noeAll",lsdNOEexp)
noe.setPotType("hard")
noe.setThreshold(lsdNOEthresh)
noePots.append(noe)

# need to be satisfied by all structures
noeHB  = create_NOEPot("noeNH",lsdHbond)
noeHB.setPotType("hard")
noeHB.setScale(1000)
noeHB.setThreshold( 0.1 )
noePots.append(noeHB)
potList.append(noePots)
rampedParams.append( StaticRamp("noePots.setScale( 50 )") )

protocol.initDihedrals(lsdDihe)
potList.append(XplorPot("CDIH"))
highTempParams.append( StaticRamp("potList['CDIH'].setScale(200)") )
rampedParams.append( StaticRamp("potList['CDIH'].setScale(200)") )
#rampedParams.append( MultRamp(10,200,"potList['CDIH'].setScale(VALUE)") )

from rdcPotTools import Da_prefactor, create_RDCPot, scale_toCH

rdcPots = PotList('rdcs')
# weight is the relative weighting of expts, as determined by expt. error
for (name,medium,weight,files) in [
    ('JCH' ,'pf1',lsdRDCscale,lsdDipoInp),('methyl' ,'pf1',lsdMeRDCscale,lsdDipoInpMe)
    ]:
    term = create_RDCPot(name,oTensor=media[medium],defThreshold=lsdRDCthresh)
    if type(files)==type('string'):
        files=(files,)
        pass
    for file in files:
        term.addRestraints( open(file).read() )
        pass
    term.setShowAllRestraints(1)
    term.setScale(weight)
    #term.setAveType("average")
    term.setAveType("sum")
    print name
    scale_toCH(term) #also sets useDistance
    print term.info()
    print term.gyroA()
    rdcPots.append(term)
    pass
```

```
potList.append(rdcPots)
rampedParams.append( MultRamp(0.01,1,"rdcPots.setScale( VALUE )") )

from rdcPotTools import Da_prefactor

print "factor:", Da_prefactor['CH'] / Da_prefactor["NH"]
for medium in media.values():
    calcTensor(medium)
    print "medium: ", medium.instanceName(), \
          "Da: ",medium.Da(), "Rh: ",medium.Rh()
    pass

#let's try fixing Da, Rh:
print medium
for (medium,Da,Rh) in (('pf1',lsdDaRDC,lsdRhRDC),):
    medium = media[medium]
    medium.setDa(Da)
    medium.setRh(Rh)
    pass


potList.append( XplorPot("VDW") )
potList.append( XplorPot("elec") )
rampedParams.append( StaticRamp("""xplor.command('''param nbonds
 atom
 repel=0
 wmin=0.01
 nbxmod=5
 cutnb=58.5
 ctonnb=56.5
 ctofnb=57.5
 tolerance=0.5
 rdie
 vswitch
 switch
   end end''')""") )

for name in ("bond","angl","impr"):
    potList.append( XplorPot(name) )
    pass
rampedParams.append( MultRamp(0.4,1.0,"potList['ANGL'].setScale(VALUE)"))
rampedParams.append( MultRamp(0.1,1.0,"potList['IMPR'].setScale(VALUE)"))


from ivm import IVM
import varTensorTools
mini = IVM()             #initial alignment of orientation tensor axes

for medium in (('pf1'),): media[medium].setFreedom("fixDa, fixRh")

varTensorTools.topologySetup(mini,media.values())

protocol.initMinimize(mini,
                      numSteps=20)
mini.fix("not resname ANI")
mini.run()               #this initial minimization is not strictly necessary
```

```
#uncomment to allow Da, Rh to vary
if lsdFixedRDC==0:
  pass
else:
  for medium in (('pf1'),): media[medium].setFreedom("varyDa, varyRh")


dyn = IVM()
protocol.initDynamics(dyn,potList=potList)
varTensorTools.topologySetup(dyn,media.values())
protocol.torsionTopology(dyn)

# Give atoms uniform weights, except for the anisotropy axis
from atomAction import SetProperty
AtomSel("not resname ANI").apply( SetProperty("mass",100.) )
varTensorTools.massSetup(media.values(),300)
AtomSel("all            ").apply( SetProperty("fric",10.) )


##
## minc used for final cartesian minimization
##
from selectTools import IVM_groupRigidSidechain
minc = IVM()
protocol.initMinimize(minc,potList=potList)
IVM_groupRigidSidechain(minc)
protocol.cartesianTopology(minc,"not resname ANI")
varTensorTools.topologySetup(minc,media.values())

init_t1 = 200000
init_t2 = 20000
init_t3 = 3000

from simulationTools import AnnealIVM
anneal2= AnnealIVM(initTemp =init_t2,
                  finalTemp=init_t3,
                  tempStep =500,
                  ivm=dyn,
                  rampedParams = rampedParams)
anneal3= AnnealIVM(initTemp =init_t3,
                  finalTemp=25,
                  tempStep =25,
                  ivm=dyn,
                  rampedParams = rampedParams)


# initialize parameters for initial minimization.
InitialParams( rampedParams )
# high-temp dynamics setup - only need to specify parameters which
#   differfrom initial values in rampedParams
InitialParams( highTempParams )

# initial minimization
protocol.initMinimize(dyn,
                      potList=[potList['CDIH'],potList['IMPR']],
```

```
                      numSteps=50)
dyn.run()

# initial minimization
protocol.initMinimize(dyn,
                      potList=potList,
                      numSteps=1000)
minc.run()

def calcOneStructure(loopInfo):

# mod by lsd: second annealing loop, actual annealing
    # initialize parameters for high temp dynamics.
    InitialParams( rampedParams )
    # high-temp dynamics setup - only need to specify parameters which
    #   differfrom initial values in rampedParams
    InitialParams( highTempParams )

    protocol.initDynamics(dyn,
                          initVelocities=1,
                          bathTemp=init_t2,
                          potList=potList,
                          finalTime=50)
    dyn.setETolerance( init_t2/100 )  #used to det. stepsize. default: t/1000
    dyn.run()

    # initialize parameters for cooling loop
    InitialParams( rampedParams )

    # perform simulated annealing
    #
    protocol.initDynamics(dyn,
                          finalTime=0.5,  #time to integrate at a given temp.
                          numSteps=0,     # take as many steps as necessary
                          #eTol_minimum=0.001 # cutoff for auto-TS det.
                          )
    anneal2.run()
    anneal3.run()

    #
    # torsion angle minimization
    #
    protocol.initMinimize(dyn,numSteps=5000)
    dyn.run()

    ##
    ##all atom minimization
    ##
    protocol.initMinimize(minc,potList=potList,numSteps=3000)
    minc.run()

    #
    # perform analysis and write structure
    loopInfo.writeStructure(potList,crossTerms)
    pass
```

```
from simulationTools import StructureLoop
StructureLoop(numStructures=numberOfStructures,
              startStructure=startStructure,
              structLoopAction=calcOneStructure,
              pdbTemplate=outFilename,
              genViolationStats=1,
              averageFilename="average_min.pdb",
              averageFitSel="not resname ANI and not (name H71 or name H72 or name H73)",
              averageRefineSteps=15,
              averageTopFraction=0.1,
              averagePotList=potList).run()
```

2.4 Parameter file used by Xplor-NIH

```
!RNA PARAMETER FILE 'FRAMEOWRK' FROM PARALLHDG.DNA AND ATOM NAMES
! AND HEAVY ATOM PARAMETERS FROM DNA-RNA.PARAM
!INCLUDES ALL NONEXCHANGEABLE HYDROGEN TERMS FOR BOND, ANGLE, AND
!IMPROPERS WITH ENERGY CONSTANT VARIABLES: $kchbond, $kchangle, AND $kchimpr.
!BOND, ANGLE, AND IMPROPERS WERE ESTIMATED FROM VALUES FROM THE STANDARD
!NUCLEOTIDES OF INSIGHTII 95.0 (BIOSYM/MOLECULAR SIMULATIONS).
!CREATED 2/24/96-- JASON P. RIFE AND PETER B. MOORE
! DNA-RNA-ALLATOM.PARAM

set echo=off message=off end


! checkversion 1.0


evaluate ($kchbond = 2000)
evaluate ($kchangle  = 1000)
evaluate ($kchimpr = 1000)



!****************** change by lsd - DAP *********************

 { Note: edit if necessary }
 BOND OY1  CY2    1000.0  1.426 ! Nobs =    1
 BOND OY1  HY29   1000.0  0.963 ! Nobs =    1
 BOND CY2  HY3    1000.0  1.096 ! Nobs =    1
 BOND CY2  HY4    1000.0  1.097 ! Nobs =    1
 BOND CY2  CY5    1000.0  1.517 ! Nobs =    1
 BOND CY5  HY6    1000.0  1.094 ! Nobs =    1
 BOND CY5  OY7    1000.0  1.450 ! Nobs =    1
 BOND CY5  CY18   1000.0  1.543 ! Nobs =    1
 BOND OY7  CY8    1000.0  1.441 ! Nobs =    1
 BOND CY8  HY9    1000.0  1.098 ! Nobs =    1
 BOND CY8  CY20   1000.0  1.527 ! Nobs =    1
 BOND C1D  CY25   1000.0  1.509 ! modded by lsd
 BOND CY10 HY11   1000.0  1.079 ! Nobs =    1
 BOND CY10 NY24   1000.0  1.374 ! Nobs =    1
```

```
BOND CY10 CY25   1000.0  1.351 ! Nobs =    1
BOND CY12 NY13   1000.0  1.339 ! Nobs =    1
BOND CY12 NY16   1000.0  1.314 ! Nobs =    1
BOND CY12 NY24   1000.0  1.366 ! Nobs =    1
BOND NY13 HY14   1000.0  1.008 ! Nobs =    1
BOND NY13 HY15   1000.0  1.006 ! Nobs =    1
BOND NY16 CY17   1000.0  1.342 ! Nobs =    1
BOND CY17 CY25   1000.0  1.450 ! Nobs =    1
BOND CY17 NY26   1000.0  1.326 ! Nobs =    1
BOND CY18 HY19   1000.0  1.092 ! Nobs =    1
BOND CY18 CY20   1000.0  1.527 ! Nobs =    1
BOND CY18 OY23   1000.0  1.425 ! Nobs =    1
BOND CY20 HY21   1000.0  1.089 ! Nobs =    1
BOND CY20 HY22   1000.0  1.089 ! Nobs =    1
BOND OY23 HY30   1000.0  0.963 ! Nobs =    1
BOND NY24 HY31   1000.0  1.010 ! Nobs =    1
BOND NY26 HY27   1000.0  1.014 ! Nobs =    1
BOND NY26 HY28   1000.0  1.008 ! Nobs =    1

{ Note: edit if necessary }
ANGLe CY2  OY1  HY29    500.0   108.93 ! Nobs =    1
ANGLe OY1  CY2  HY3     500.0   111.24 ! Nobs =    1
ANGLe OY1  CY2  HY4     500.0   111.10 ! Nobs =    1
ANGLe OY1  CY2  CY5     500.0   109.30 ! Nobs =    1
ANGLe HY3  CY2  HY4     500.0   108.20 ! Nobs =    1
ANGLe HY3  CY2  CY5     500.0   108.77 ! Nobs =    1
ANGLe HY4  CY2  CY5     500.0   108.14 ! Nobs =    1
ANGLe CY2  CY5  HY6     500.0   108.20 ! Nobs =    1
ANGLe CY2  CY5  OY7     500.0   110.08 ! Nobs =    1
ANGLe CY2  CY5  CY18    500.0   115.10 ! Nobs =    1
ANGLe HY6  CY5  OY7     500.0   107.82 ! Nobs =    1
ANGLe HY6  CY5  CY18    500.0   109.84 ! Nobs =    1
ANGLe OY7  CY5  CY18    500.0   105.57 ! Nobs =    1
ANGLe CY5  OY7  CY8     500.0   110.21 ! Nobs =    1
ANGLe OY7  CY8  HY9     500.0   108.09 ! Nobs =    1
ANGLe OY7  CY8  CY20    500.0   105.09 ! Nobs =    1
ANGLe O4D  C1D  CY25    500.0   109.05 ! modded by lsd
ANGLe HY9  CY8  CY20    500.0   109.22 ! Nobs =    1
ANGLe H    C1D  CY25    500.0   107.80 ! modded by lsd
ANGLe C2D  C1D  CY25    500.0   117.27 ! modded by lsd
ANGLe HY11 CY10 NY24    500.0   115.49 ! Nobs =    1
ANGLe HY11 CY10 CY25    500.0   123.38 ! Nobs =    1
ANGLe NY24 CY10 CY25    500.0   121.13 ! Nobs =    1
ANGLe NY13 CY12 NY16    500.0   119.62 ! Nobs =    1
ANGLe NY13 CY12 NY24    500.0   119.02 ! Nobs =    1
ANGLe NY16 CY12 NY24    500.0   121.36 ! Nobs =    1
ANGLe CY12 NY13 HY14    500.0   117.80 ! Nobs =    1
ANGLe CY12 NY13 HY15    500.0   123.59 ! Nobs =    1
ANGLe HY14 NY13 HY15    500.0   118.59 ! Nobs =    1
ANGLe CY12 NY16 CY17    500.0   119.65 ! Nobs =    1
ANGLe NY16 CY17 CY25    500.0   121.96 ! Nobs =    1
ANGLe NY16 CY17 NY26    500.0   117.38 ! Nobs =    1
ANGLe CY25 CY17 NY26    500.0   120.65 ! Nobs =    1
ANGLe CY5  CY18 HY19    500.0   111.68 ! Nobs =    1
ANGLe CY5  CY18 CY20    500.0   102.81 ! Nobs =    1
ANGLe CY5  CY18 OY23    500.0   111.93 ! Nobs =    1
```

```
ANGLe HY19 CY18 CY20    500.0   111.73 ! Nobs =    1
ANGLe HY19 CY18 OY23    500.0   111.35 ! Nobs =    1
ANGLe CY20 CY18 OY23    500.0   106.94 ! Nobs =    1
ANGLe CY8  CY20 CY18    500.0   101.56 ! Nobs =    1
ANGLe CY8  CY20 HY21    500.0   110.76 ! Nobs =    1
ANGLe CY8  CY20 HY22    500.0   113.50 ! Nobs =    1
ANGLe CY18 CY20 HY21    500.0   108.80 ! Nobs =    1
ANGLe CY18 CY20 HY22    500.0   110.87 ! Nobs =    1
ANGLe HY21 CY20 HY22    500.0   110.92 ! Nobs =    1
ANGLe CY18 OY23 HY30    500.0   109.89 ! Nobs =    1
ANGLe CY10 NY24 CY12    500.0   120.27 ! Nobs =    1
ANGLe CY10 NY24 HY31    500.0   119.01 ! Nobs =    1
ANGLe CY12 NY24 HY31    500.0   120.70 ! Nobs =    1
ANGLe C1D  CY25 CY10    500.0   122.92 ! modded by lsd
ANGLe C1D  CY25 CY17    500.0   121.43 ! modded by lsd
ANGLe CY10 CY25 CY17    500.0   115.50 ! Nobs =    1
ANGLe CY17 NY26 HY27    500.0   119.25 ! Nobs =    1
ANGLe CY17 NY26 HY28    500.0   118.79 ! Nobs =    1
ANGLe HY27 NY26 HY28    500.0   120.67 ! Nobs =    1

{ Note: edit if necessary }
DIHEdral HY29 OY1  CY2  HY3     750.0 0    60.00 ! Nobs =    1 ... Value =    60.14
DIHEdral HY29 OY1  CY2  HY4     750.0 0   -60.00 ! Nobs =    1 ... Value =   -60.46
DIHEdral HY29 OY1  CY2  CY5     750.0 0   180.00 ! Nobs =    1 ... Value =  -179.72
DIHEdral OY1  CY2  CY5  HY6     750.0 0   180.00 ! Nobs =    1 ... Value =   172.91
DIHEdral OY1  CY2  CY5  OY7     750.0 0   -60.00 ! Nobs =    1 ... Value =   -69.50
DIHEdral HY3  CY2  CY5  HY6     750.0 0   -60.00 ! Nobs =    1 ... Value =   -65.45
DIHEdral HY3  CY2  CY5  OY7     750.0 0    60.00 ! Nobs =    1 ... Value =    52.14
DIHEdral HY3  CY2  CY5  CY18    750.0 0   180.00 ! Nobs =    1 ... Value =   171.28
DIHEdral HY4  CY2  CY5  HY6     750.0 0    60.00 ! Nobs =    1 ... Value =    51.84
DIHEdral CY2  CY5  OY7  CY8     750.0 0   120.00 ! Nobs =    1 ... Value =   122.24
DIHEdral HY6  CY5  OY7  CY8     750.0 0  -120.00 ! Nobs =    1 ... Value =  -119.94
DIHEdral CY18 CY5  OY7  CY8     750.0 0     0.00 ! Nobs =    1 ... Value =    -2.57
DIHEdral CY2  CY5  CY18 CY20    750.0 0   -90.00 ! Nobs =    1 ... Value =   -96.57
DIHEdral HY6  CY5  CY18 HY19    750.0 0   -90.00 ! Nobs =    1 ... Value =   -99.03
DIHEdral OY7  CY5  CY18 OY23    750.0 0   -90.00 ! Nobs =    1 ... Value =   -89.40
DIHEdral CY5  OY7  CY8  HY9     750.0 0    90.00 ! Nobs =    1 ... Value =    95.37
DIHEdral CY25 CY8  CY20 HY22    750.0 0   -90.00 ! Nobs =    1 ... Value =   -83.70
DIHEdral HY11 CY10 NY24 CY12    750.0 0   180.00 ! Nobs =    1 ... Value =   177.39
DIHEdral HY11 CY10 NY24 HY31    750.0 0     0.00 ! Nobs =    1 ... Value =    -1.13
DIHEdral CY25 CY10 NY24 CY12    750.0 0     0.00 ! Nobs =    1 ... Value =    -1.67
DIHEdral CY25 CY10 NY24 HY31    750.0 0   180.00 ! Nobs =    1 ... Value =   179.81
DIHEdral HY11 CY10 CY25 CY8     750.0 0     0.00 ! Nobs =    1 ... Value =    -4.80
DIHEdral HY11 CY10 CY25 CY17    750.0 0   180.00 ! Nobs =    1 ... Value =   179.53
DIHEdral NY24 CY10 CY25 CY8     750.0 0   180.00 ! Nobs =    1 ... Value =   174.19
DIHEdral NY24 CY10 CY25 CY17    750.0 0     0.00 ! Nobs =    1 ... Value =    -1.49
DIHEdral NY16 CY12 NY13 HY14    750.0 0     0.00 ! Nobs =    1 ... Value =    -0.22
DIHEdral NY16 CY12 NY13 HY15    750.0 0   180.00 ! Nobs =    1 ... Value =  -178.44
DIHEdral NY24 CY12 NY13 HY14    750.0 0   180.00 ! Nobs =    1 ... Value =  -179.63
DIHEdral NY24 CY12 NY13 HY15    750.0 0     0.00 ! Nobs =    1 ... Value =     2.16
DIHEdral NY13 CY12 NY16 CY17    750.0 0   180.00 ! Nobs =    1 ... Value =   179.93
DIHEdral NY24 CY12 NY16 CY17    750.0 0     0.00 ! Nobs =    1 ... Value =    -0.68
DIHEdral NY13 CY12 NY24 CY10    750.0 0   180.00 ! Nobs =    1 ... Value =  -177.69
DIHEdral NY13 CY12 NY24 HY31    750.0 0     0.00 ! Nobs =    1 ... Value =     0.81
DIHEdral NY16 CY12 NY24 CY10    750.0 0     0.00 ! Nobs =    1 ... Value =     2.91
DIHEdral NY16 CY12 NY24 HY31    750.0 0   180.00 ! Nobs =    1 ... Value =  -178.58
```

```
 DIHEdral CY12 NY16 CY17 CY25    750.0 0     0.00 ! Nobs =    1 ... Value =    -2.69
 DIHEdral CY12 NY16 CY17 NY26    750.0 0   180.00 ! Nobs =    1 ... Value =   178.27
 DIHEdral NY16 CY17 CY25 CY8     750.0 0   180.00 ! Nobs =    1 ... Value =  -172.01
 DIHEdral NY16 CY17 CY25 CY10    750.0 0     0.00 ! Nobs =    1 ... Value =     3.74
 DIHEdral NY26 CY17 CY25 CY8     750.0 0     0.00 ! Nobs =    1 ... Value =     7.00
 DIHEdral NY26 CY17 CY25 CY10    750.0 0   180.00 ! Nobs =    1 ... Value =  -177.26
 DIHEdral NY16 CY17 NY26 HY28    750.0 0     0.00 ! Nobs =    1 ... Value =    -1.69
 DIHEdral CY25 CY17 NY26 HY28    750.0 0   180.00 ! Nobs =    1 ... Value =   179.27
 DIHEdral CY5  CY18 CY20 HY21    750.0 0    90.00 ! Nobs =    1 ... Value =    80.22
 DIHEdral HY19 CY18 CY20 HY22    750.0 0    90.00 ! Nobs =    1 ... Value =    82.56
 DIHEdral OY23 CY18 CY20 CY8     750.0 0    90.00 ! Nobs =    1 ... Value =    81.38
 DIHEdral CY20 CY18 OY23 HY30    750.0 0   180.00 ! Nobs =    1 ... Value =   170.64

 { Note: edit if necessary }
 IMPRoper CY2  OY1  HY3  HY4     750.0 0   35.000 ! Nobs =    1 ... Value =   33.228
 IMPRoper CY5  CY2  HY6  OY7     750.0 0  -35.000 ! Nobs =    1 ... Value =  -38.195
! >>> NOTE - unusual value for following improper :    41.37 reset to +35.0
 IMPRoper CY8  OY7  HY9  CY20    750.0 0   35.000 ! Nobs =    1 ... Value =   41.369
 IMPRoper CY10 HY11 NY24 CY25    750.0 0    0.000 ! Nobs =    1 ... Value =   -0.545
 IMPRoper CY12 NY13 NY16 NY24    750.0 0    0.000 ! Nobs =    1 ... Value =    0.351
 IMPRoper NY13 CY12 HY14 HY15    750.0 0    0.000 ! Nobs =    1 ... Value =    0.940
 IMPRoper CY17 NY16 CY25 NY26    750.0 0    0.000 ! Nobs =    1 ... Value =    0.570
! >>> NOTE - unusual value for following improper :    40.17 reset to +35.0
 IMPRoper CY18 CY5  HY19 CY20    750.0 0   35.000 ! Nobs =    1 ... Value =   40.170
 IMPRoper CY20 CY8  CY18 HY21    750.0 0   35.000 ! Nobs =    1 ... Value =   30.715
 IMPRoper NY24 CY10 CY12 HY31    750.0 0    0.000 ! Nobs =    1 ... Value =    0.769
 IMPRoper CY25 C1D  CY10 CY17    750.0 0    0.000 ! Nobs =    1 ... Value = -2.663 mod by lsd
! >>> NOTE - unusual value for following improper :    -7.15 reset to 0.0
 IMPRoper NY26 CY17 HY27 HY28    750.0 0    0.000 ! Nobs =    1 ... Value =   -7.146
!add IMPRoper for chirality around C1'
 IMPRoper H    C2D  O4D  CY25    $kchimpr 0  -65.280

 { Note: edit if necessary }
 NONBonded OY1  0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
 NONBonded CY2  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HY3  0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HY4  0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded CY5  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HY6  0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded OY7  0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
 NONBonded CY8  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HY9  0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded CY10 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HY11 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded CY12 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded NY13 0.2384  2.8509    0.2384  2.8509 ! assuming Nitrogen
 NONBonded HY14 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HY15 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded NY16 0.2384  2.8509    0.2384  2.8509 ! assuming Nitrogen
 NONBonded CY17 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CY18 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HY19 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded CY20 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HY21 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HY22 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded OY23 0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
```

```
 NONBonded NY24 0.2384  2.8509    0.2384  2.8509 ! assuming Nitrogen
 NONBonded CY25 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded NY26 0.2384  2.8509    0.2384  2.8509 ! assuming Nitrogen
 NONBonded HY27 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HY28 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HY29 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HY30 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HY31 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen

!***************end of change by lsd - DAP *********************


!******************* change by lsd - 4AP **********************

 { Note: edit if necessary }
 BOND OX1  CX2    1000.0  1.428 ! Nobs =    1
 BOND OX1  HX21   1000.0  0.962 ! Nobs =    1
 BOND CX2  HX3    1000.0  1.096 ! Nobs =    1
 BOND CX2  HX4    1000.0  1.099 ! Nobs =    1
 BOND CX2  CX5    1000.0  1.518 ! Nobs =    1
 BOND CX5  HX6    1000.0  1.098 ! Nobs =    1
 BOND CX5  OX7    1000.0  1.430 ! Nobs =    1
 BOND CX5  CX14   1000.0  1.551 ! Nobs =    1
 BOND OX7  CX8    1000.0  1.435 ! Nobs =    1
 BOND CX8  HX9    1000.0  1.094 ! Nobs =    1
 BOND CX8  CX16   1000.0  1.536 ! Nobs =    1
 BOND C1D  CX20   1000.0  1.509 ! modded by lsd
 BOND CX10 HX11   1000.0  1.082 ! Nobs =    1
 BOND CX10 CX20   1000.0  1.391 ! Nobs =    1
 BOND CX10 CX23   1000.0  1.406 ! Nobs =    1
 BOND CX12 CX23   1000.0  1.407 ! Nobs =    1
 BOND CX12 CX24   1000.0  1.374 ! Nobs =    1
 BOND CX12 HX34   1000.0  1.083 ! Nobs =    1
 BOND CX13 CX20   1000.0  1.398 ! Nobs =    1
 BOND CX13 CX24   1000.0  1.396 ! Nobs =    1
 BOND CX13 CX30   1000.0  1.482 ! Nobs =    1
 BOND CX14 HX15   1000.0  1.093 ! Nobs =    1
 BOND CX14 CX16   1000.0  1.524 ! Nobs =    1
 BOND CX14 OX19   1000.0  1.430 ! Nobs =    1
 BOND CX16 HX17   1000.0  1.090 ! Nobs =    1
 BOND CX16 HX18   1000.0  1.088 ! Nobs =    1
 BOND OX19 HX22   1000.0  0.963 ! Nobs =    1
 BOND CX23 NX25   1000.0  1.381 ! Nobs =    1
 BOND CX24 CX29   1000.0  1.497 ! Nobs =    1
 BOND NX25 HX26   1000.0  1.007 ! Nobs =    1
 BOND NX25 HX27   1000.0  1.006 ! Nobs =    1
 BOND NX28 CX29   1000.0  1.393 ! Nobs =    1
 BOND NX28 CX30   1000.0  1.408 ! Nobs =    1
 BOND NX28 HX33   1000.0  1.009 ! Nobs =    1
 BOND CX29 OX32   1000.0  1.208 ! Nobs =    1
 BOND CX30 OX31   1000.0  1.210 ! Nobs =    1

 { Note: edit if necessary }
 ANGLe CX2  OX1  HX21    500.0   108.71 ! Nobs =    1
 ANGLe OX1  CX2  HX3     500.0   111.02 ! Nobs =    1
 ANGLe OX1  CX2  HX4     500.0   110.85 ! Nobs =    1
```

```
ANGLe OX1  CX2  CX5     500.0   109.55 ! Nobs =    1
ANGLe HX3  CX2  HX4     500.0   108.36 ! Nobs =    1
ANGLe HX3  CX2  CX5     500.0   108.57 ! Nobs =    1
ANGLe HX4  CX2  CX5     500.0   108.42 ! Nobs =    1
ANGLe CX2  CX5  HX6     500.0   107.54 ! Nobs =    1
ANGLe CX2  CX5  OX7     500.0   109.63 ! Nobs =    1
ANGLe CX2  CX5  CX14    500.0   114.88 ! Nobs =    1
ANGLe HX6  CX5  OX7     500.0   108.82 ! Nobs =    1
ANGLe HX6  CX5  CX14    500.0   108.72 ! Nobs =    1
ANGLe OX7  CX5  CX14    500.0   107.12 ! Nobs =    1
ANGLe CX5  OX7  CX8     500.0   109.83 ! Nobs =    1
ANGLe OX7  CX8  HX9     500.0   109.62 ! Nobs =    1
ANGLe OX7  CX8  CX16    500.0   104.20 ! Nobs =    1
ANGLe O4D  C1D  CX20    500.0   109.88 ! modded by lsd
ANGLe HX9  CX8  CX16    500.0   109.16 ! Nobs =    1
ANGLe H    C1D  CX20    500.0   109.27 ! modded by lsd
ANGLe C2D  C1D  CX20    500.0   114.55 ! modded by lsd
ANGLe HX11 CX10 CX20    500.0   118.07 ! Nobs =    1
ANGLe HX11 CX10 CX23    500.0   119.45 ! Nobs =    1
ANGLe CX20 CX10 CX23    500.0   122.48 ! Nobs =    1
ANGLe CX23 CX12 CX24    500.0   117.51 ! Nobs =    1
ANGLe CX23 CX12 HX34    500.0   121.57 ! Nobs =    1
ANGLe CX24 CX12 HX34    500.0   120.92 ! Nobs =    1
ANGLe CX20 CX13 CX24    500.0   120.33 ! Nobs =    1
ANGLe CX20 CX13 CX30    500.0   131.38 ! Nobs =    1
ANGLe CX24 CX13 CX30    500.0   108.28 ! Nobs =    1
ANGLe CX5  CX14 HX15    500.0   111.38 ! Nobs =    1
ANGLe CX5  CX14 CX16    500.0   103.02 ! Nobs =    1
ANGLe CX5  CX14 OX19    500.0   112.03 ! Nobs =    1
ANGLe HX15 CX14 CX16    500.0   111.74 ! Nobs =    1
ANGLe HX15 CX14 OX19    500.0   110.45 ! Nobs =    1
ANGLe CX16 CX14 OX19    500.0   107.98 ! Nobs =    1
ANGLe CX8  CX16 CX14    500.0   102.58 ! Nobs =    1
ANGLe CX8  CX16 HX17    500.0   109.82 ! Nobs =    1
ANGLe CX8  CX16 HX18    500.0   112.71 ! Nobs =    1
ANGLe CX14 CX16 HX17    500.0   109.59 ! Nobs =    1
ANGLe CX14 CX16 HX18    500.0   111.72 ! Nobs =    1
ANGLe HX17 CX16 HX18    500.0   110.18 ! Nobs =    1
ANGLe CX14 OX19 HX22    500.0   108.93 ! Nobs =    1
ANGLe C1D  CX20 CX10    500.0   120.99 ! modded by lsd
ANGLe C1D  CX20 CX13    500.0   121.86 ! modded by lsd
ANGLe CX10 CX20 CX13    500.0   117.13 ! Nobs =    1
ANGLe CX10 CX23 CX12    500.0   119.62 ! Nobs =    1
ANGLe CX10 CX23 NX25    500.0   120.09 ! Nobs =    1
ANGLe CX12 CX23 NX25    500.0   120.24 ! Nobs =    1
ANGLe CX12 CX24 CX13    500.0   122.92 ! Nobs =    1
ANGLe CX12 CX24 CX29    500.0   128.41 ! Nobs =    1
ANGLe CX13 CX24 CX29    500.0   108.67 ! Nobs =    1
ANGLe CX23 NX25 HX26    500.0   117.56 ! Nobs =    1
ANGLe CX23 NX25 HX27    500.0   117.77 ! Nobs =    1
ANGLe HX26 NX25 HX27    500.0   114.53 ! Nobs =    1
ANGLe CX29 NX28 CX30    500.0   113.29 ! Nobs =    1
ANGLe CX29 NX28 HX33    500.0   123.64 ! Nobs =    1
ANGLe CX30 NX28 HX33    500.0   123.04 ! Nobs =    1
ANGLe CX24 CX29 NX28    500.0   104.62 ! Nobs =    1
ANGLe CX24 CX29 OX32    500.0   129.17 ! Nobs =    1
```

```
ANGLe NX28 CX29 OX32    500.0   126.21 ! Nobs =    1
ANGLe CX13 CX30 NX28    500.0   105.13 ! Nobs =    1
ANGLe CX13 CX30 OX31    500.0   130.41 ! Nobs =    1
ANGLe NX28 CX30 OX31    500.0   124.46 ! Nobs =    1

{ Note: edit if necessary }
DIHEdral HX21 OX1  CX2  HX3     750.0 0    60.00 ! Nobs =    1 ... Value =    59.00
DIHEdral HX21 OX1  CX2  HX4     750.0 0   -60.00 ! Nobs =    1 ... Value =   -61.50
DIHEdral HX21 OX1  CX2  CX5     750.0 0   180.00 ! Nobs =    1 ... Value =   178.90
DIHEdral OX1  CX2  CX5  HX6     750.0 0   180.00 ! Nobs =    1 ... Value =   171.53
DIHEdral OX1  CX2  CX5  CX14    750.0 0    60.00 ! Nobs =    1 ... Value =    50.36
DIHEdral HX3  CX2  CX5  HX6     750.0 0   -60.00 ! Nobs =    1 ... Value =   -67.09
DIHEdral HX3  CX2  CX5  OX7     750.0 0    60.00 ! Nobs =    1 ... Value =    51.08
DIHEdral HX3  CX2  CX5  CX14    750.0 0   180.00 ! Nobs =    1 ... Value =   171.74
DIHEdral HX4  CX2  CX5  HX6     750.0 0    60.00 ! Nobs =    1 ... Value =    50.44
DIHEdral CX14 CX5  OX7  CX8     750.0 0     0.00 ! Nobs =    1 ... Value =     9.07
DIHEdral CX5  OX7  CX8  HX9     750.0 0    90.00 ! Nobs =    1 ... Value =    87.96
DIHEdral HX9  CX8  CX16 CX14    750.0 0   -90.00 ! Nobs =    1 ... Value =   -80.38
DIHEdral CX20 CX8  CX16 HX18    750.0 0   -90.00 ! Nobs =    1 ... Value =   -82.97
DIHEdral CX16 CX8  CX20 CX10    750.0 0   -90.00 ! Nobs =    1 ... Value =   -96.73
DIHEdral CX16 CX8  CX20 CX13    750.0 0    90.00 ! Nobs =    1 ... Value =    81.66
DIHEdral HX11 CX10 CX20 CX8     750.0 0     0.00 ! Nobs =    1 ... Value =    -2.07
DIHEdral HX11 CX10 CX20 CX13    750.0 0   180.00 ! Nobs =    1 ... Value =   179.46
DIHEdral CX23 CX10 CX20 CX8     750.0 0   180.00 ! Nobs =    1 ... Value =   178.46
DIHEdral CX23 CX10 CX20 CX13    750.0 0     0.00 ! Nobs =    1 ... Value =     0.00
DIHEdral HX11 CX10 CX23 CX12    750.0 0   180.00 ! Nobs =    1 ... Value =  -179.04
DIHEdral HX11 CX10 CX23 NX25    750.0 0     0.00 ! Nobs =    1 ... Value =    -1.73
DIHEdral CX20 CX10 CX23 CX12    750.0 0     0.00 ! Nobs =    1 ... Value =     0.42
DIHEdral CX20 CX10 CX23 NX25    750.0 0   180.00 ! Nobs =    1 ... Value =   177.73
DIHEdral CX24 CX12 CX23 CX10    750.0 0     0.00 ! Nobs =    1 ... Value =    -0.40
DIHEdral CX24 CX12 CX23 NX25    750.0 0   180.00 ! Nobs =    1 ... Value =  -177.71
DIHEdral HX34 CX12 CX23 CX10    750.0 0   180.00 ! Nobs =    1 ... Value =   179.18
DIHEdral HX34 CX12 CX23 NX25    750.0 0     0.00 ! Nobs =    1 ... Value =     1.86
DIHEdral CX23 CX12 CX24 CX13    750.0 0     0.00 ! Nobs =    1 ... Value =    -0.02
DIHEdral CX23 CX12 CX24 CX29    750.0 0   180.00 ! Nobs =    1 ... Value =   179.67
DIHEdral HX34 CX12 CX24 CX13    750.0 0   180.00 ! Nobs =    1 ... Value =  -179.59
DIHEdral HX34 CX12 CX24 CX29    750.0 0     0.00 ! Nobs =    1 ... Value =     0.10
DIHEdral CX24 CX13 CX20 CX8     750.0 0   180.00 ! Nobs =    1 ... Value =  -178.86
DIHEdral CX24 CX13 CX20 CX10    750.0 0     0.00 ! Nobs =    1 ... Value =    -0.41
DIHEdral CX30 CX13 CX20 CX8     750.0 0     0.00 ! Nobs =    1 ... Value =     2.05
DIHEdral CX30 CX13 CX20 CX10    750.0 0   180.00 ! Nobs =    1 ... Value =  -179.49
DIHEdral CX20 CX13 CX24 CX12    750.0 0     0.00 ! Nobs =    1 ... Value =     0.43
DIHEdral CX20 CX13 CX24 CX29    750.0 0   180.00 ! Nobs =    1 ... Value =  -179.31
DIHEdral CX30 CX13 CX24 CX12    750.0 0   180.00 ! Nobs =    1 ... Value =   179.71
DIHEdral CX30 CX13 CX24 CX29    750.0 0     0.00 ! Nobs =    1 ... Value =    -0.03
DIHEdral CX20 CX13 CX30 NX28    750.0 0   180.00 ! Nobs =    1 ... Value =   178.75
DIHEdral CX20 CX13 CX30 OX31    750.0 0     0.00 ! Nobs =    1 ... Value =    -1.46
DIHEdral CX24 CX13 CX30 NX28    750.0 0     0.00 ! Nobs =    1 ... Value =    -0.42
DIHEdral CX24 CX13 CX30 OX31    750.0 0   180.00 ! Nobs =    1 ... Value =   179.37
DIHEdral CX5  CX14 CX16 HX17    750.0 0    90.00 ! Nobs =    1 ... Value =    86.21
DIHEdral HX15 CX14 CX16 HX18    750.0 0    90.00 ! Nobs =    1 ... Value =    88.94
DIHEdral OX19 CX14 CX16 CX8     750.0 0    90.00 ! Nobs =    1 ... Value =    88.25
DIHEdral CX5  CX14 OX19 HX22    750.0 0   -60.00 ! Nobs =    1 ... Value =   -67.94
DIHEdral HX15 CX14 OX19 HX22    750.0 0    60.00 ! Nobs =    1 ... Value =    56.85
DIHEdral CX16 CX14 OX19 HX22    750.0 0   180.00 ! Nobs =    1 ... Value =   179.31
DIHEdral CX12 CX24 CX29 NX28    750.0 0   180.00 ! Nobs =    1 ... Value =  -179.25
```

```
 DIHEdral CX12 CX24 CX29 OX32    750.0 0     0.00 ! Nobs =    1 ... Value =     0.74
 DIHEdral CX13 CX24 CX29 NX28    750.0 0     0.00 ! Nobs =    1 ... Value =     0.48
 DIHEdral CX13 CX24 CX29 OX32    750.0 0   180.00 ! Nobs =    1 ... Value =  -179.54
 DIHEdral CX30 NX28 CX29 CX24    750.0 0     0.00 ! Nobs =    1 ... Value =    -0.78
 DIHEdral CX30 NX28 CX29 OX32    750.0 0   180.00 ! Nobs =    1 ... Value =   179.24
 DIHEdral HX33 NX28 CX29 CX24    750.0 0   180.00 ! Nobs =    1 ... Value =  -179.19
 DIHEdral HX33 NX28 CX29 OX32    750.0 0     0.00 ! Nobs =    1 ... Value =     0.82
 DIHEdral CX29 NX28 CX30 CX13    750.0 0     0.00 ! Nobs =    1 ... Value =     0.77
 DIHEdral CX29 NX28 CX30 OX31    750.0 0   180.00 ! Nobs =    1 ... Value =  -179.04
 DIHEdral HX33 NX28 CX30 CX13    750.0 0   180.00 ! Nobs =    1 ... Value =   179.19
 DIHEdral HX33 NX28 CX30 OX31    750.0 0     0.00 ! Nobs =    1 ... Value =    -0.62

 { Note: edit if necessary }
 IMPRoper CX2  OX1  HX3  HX4     750.0 0   35.000 ! Nobs =    1 ... Value =   33.335
 IMPRoper CX5  CX2  HX6  OX7     750.0 0  -35.000 ! Nobs =    1 ... Value =  -37.677
! >>> NOTE - unusual value for following improper :    41.94 reset to +35.0
 IMPRoper CX8  OX7  HX9  CX16    750.0 0   35.000 ! Nobs =    1 ... Value =   41.939
 IMPRoper CX10 HX11 CX20 CX23    750.0 0    0.000 ! Nobs =    1 ... Value =   -0.313
 IMPRoper CX12 CX23 CX24 HX34    750.0 0    0.000 ! Nobs =    1 ... Value =   -0.218
 IMPRoper CX13 CX20 CX24 CX30    750.0 0    0.000 ! Nobs =    1 ... Value =    0.475
! >>> NOTE - unusual value for following improper :    40.01 reset to +35.0
 IMPRoper CX14 CX5  HX15 CX16    750.0 0   35.000 ! Nobs =    1 ... Value =   40.005
 IMPRoper CX16 CX8  CX14 HX17    750.0 0   35.000 ! Nobs =    1 ... Value =   30.720
 IMPRoper CX20 CX8  CX10 CX13    750.0 0    0.000 ! Nobs =    1 ... Value =   -0.905
 IMPRoper CX23 CX10 CX12 NX25    750.0 0    0.000 ! Nobs =    1 ... Value =   -1.535
 IMPRoper CX24 CX12 CX13 CX29    750.0 0    0.000 ! Nobs =    1 ... Value =   -0.171
! >>> NOTE - unusual value for following improper :   -20.06 reset to -35.0
 IMPRoper NX25 CX23 HX26 HX27    750.0 0    0.000 ! Nobs =    1 ... Value =  -20.059
 IMPRoper NX28 CX29 CX30 HX33    750.0 0    0.000 ! Nobs =    1 ... Value =    0.749
 IMPRoper CX29 CX24 NX28 OX32    750.0 0    0.000 ! Nobs =    1 ... Value =   -0.008
 IMPRoper CX30 CX13 NX28 OX31    750.0 0    0.000 ! Nobs =    1 ... Value =   -0.091
!Improper to keep both rings parallel
 IMPRoper NX28 CX24 CX13 CX10    750.0 0    0.000
 IMPRoper NX28 CX24 CX13 CX23    750.0 0    0.000
 IMPRoper CX29 CX24 CX13 CX20    750.0 0    0.000
 IMPRoper CX12 CX24 CX13 CX30    750.0 0    0.000
!secure chirality
 IMPRoper H    C2D  O4D  CX20    $kchimpr 0  -65.280
 IMPRoper CX10 NX25 CX23 HX26    750.0 0    0.000

 { Note: edit if necessary }
 NONBonded OX1  0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
 NONBonded CX2  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HX3  0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HX4  0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded CX5  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HX6  0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded OX7  0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
 NONBonded CX8  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HX9  0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded CX10 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HX11 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded CX12 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CX13 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CX14 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HX15 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
```

```
NONBonded CX16 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
NONBonded HX17 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
NONBonded HX18 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
NONBonded OX19 0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
NONBonded CX20 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
NONBonded HX21 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
NONBonded HX22 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
NONBonded CX23 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
NONBonded CX24 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
NONBonded NX25 0.2384  2.8509    0.2384  2.8509 ! assuming Nitrogen
NONBonded HX26 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
NONBonded HX27 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
NONBonded NX28 0.2384  2.8509    0.2384  2.8509 ! assuming Nitrogen
NONBonded CX29 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
NONBonded CX30 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
NONBonded OX31 0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
NONBonded OX32 0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
NONBonded HX33 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
NONBonded HX34 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen

set echo=true end


!***************** end of change by lsd - 4AP *****************


!******************* change by lsd - 6HQ **********************

{ Note: edit if necessary }
BOND P    OQ21     3350.720    1.593      ! added bond by lsd
BOND P    OQ19     2326.889    1.607      ! added bond by lsd

BOND CQ1  NQ2    1000.0  1.495 ! Nobs =    1
BOND CQ1  CQ20   1000.0  1.522 ! Nobs =    1
BOND CQ1  HQ26   1000.0  1.086 ! Nobs =    1
BOND CQ1  HQ27   1000.0  1.091 ! Nobs =    1
BOND NQ2  CQ3    1000.0  1.389 ! Nobs =    1
BOND NQ2  CQ7    1000.0  1.337 ! Nobs =    1
BOND CQ3  CQ4    1000.0  1.428 ! Nobs =    1
BOND CQ3  CQ9    1000.0  1.411 ! Nobs =    1
BOND CQ4  CQ5    1000.0  1.410 ! Nobs =    1
BOND CQ4  CQ12   1000.0  1.411 ! Nobs =    1
BOND CQ5  CQ6    1000.0  1.371 ! Nobs =    1
BOND CQ5  HQ14   1000.0  1.083 ! Nobs =    1
BOND CQ6  CQ7    1000.0  1.393 ! Nobs =    1
BOND CQ6  HQ15   1000.0  1.081 ! Nobs =    1
BOND CQ7  HQ8    1000.0  1.080 ! Nobs =    1
BOND CQ9  CQ10   1000.0  1.370 ! Nobs =    1
BOND CQ9  HQ16   1000.0  1.077 ! Nobs =    1
BOND CQ10 CQ11   1000.0  1.416 ! Nobs =    1
BOND CQ10 HQ17   1000.0  1.082 ! Nobs =    1
BOND CQ11 CQ12   1000.0  1.376 ! Nobs =    1
BOND CQ11 OQ13   1000.0  1.346 ! Nobs =    1
BOND CQ12 HQ18   1000.0  1.083 ! Nobs =    1
BOND OQ13 HQ30   1000.0  0.966 ! Nobs =    1
BOND OQ19 CQ20   1000.0  1.424 ! Nobs =    1
```

```
BOND OQ19 HQ29   1000.0  0.966 ! Nobs =    1
BOND CQ20 CQ22   1000.0  1.532 ! Nobs =    1
BOND CQ20 HQ25   1000.0  1.097 ! Nobs =    1
BOND OQ21 CQ22   1000.0  1.420 ! Nobs =    1
BOND OQ21 HQ28   1000.0  0.964 ! Nobs =    1
BOND CQ22 HQ23   1000.0  1.096 ! Nobs =    1
BOND CQ22 HQ24   1000.0  1.096 ! Nobs =    1

{ Note: edit if necessary }
ANGLe CQ22 OQ21 P       1175.163  120.900 ! data taken from normal dna by lsd
ANGLe O1P  P    OQ21    357.719  108.100 !again modded by lsd
ANGLe O2P  P    OQ21    412.677  108.300 !again modded by lsd
ANGLe O3R  P    OQ21    833.356  104.000 !again modded by lsd

ANGLe O1P  P    OQ19    357.719  108.100 !again modded by lsd
ANGLe O2P  P    OQ19    412.677  108.300 !again modded by lsd
ANGLe O5R  P    OQ19    833.356  104.000 !again modded by lsd

ANGLe NQ2  CQ1  CQ20    500.0   116.04 ! Nobs =    1
ANGLe NQ2  CQ1  HQ26    500.0   107.23 ! Nobs =    1
ANGLe NQ2  CQ1  HQ27    500.0   107.76 ! Nobs =    1
ANGLe CQ20 CQ1  HQ26    500.0   106.40 ! Nobs =    1
ANGLe CQ20 CQ1  HQ27    500.0   109.68 ! Nobs =    1
ANGLe HQ26 CQ1  HQ27    500.0   109.59 ! Nobs =    1
ANGLe CQ1  NQ2  CQ3     500.0   118.99 ! Nobs =    1
ANGLe CQ1  NQ2  CQ7     500.0   119.88 ! Nobs =    1
ANGLe CQ3  NQ2  CQ7     500.0   121.11 ! Nobs =    1
ANGLe NQ2  CQ3  CQ4     500.0   118.64 ! Nobs =    1
ANGLe NQ2  CQ3  CQ9     500.0   122.50 ! Nobs =    1
ANGLe CQ4  CQ3  CQ9     500.0   118.86 ! Nobs =    1
ANGLe CQ3  CQ4  CQ5     500.0   118.72 ! Nobs =    1
ANGLe CQ3  CQ4  CQ12    500.0   119.66 ! Nobs =    1
ANGLe CQ5  CQ4  CQ12    500.0   121.62 ! Nobs =    1
ANGLe CQ4  CQ5  CQ6     500.0   120.23 ! Nobs =    1
ANGLe CQ4  CQ5  HQ14    500.0   119.26 ! Nobs =    1
ANGLe CQ6  CQ5  HQ14    500.0   120.50 ! Nobs =    1
ANGLe CQ5  CQ6  CQ7     500.0   119.49 ! Nobs =    1
ANGLe CQ5  CQ6  HQ15    500.0   121.80 ! Nobs =    1
ANGLe CQ7  CQ6  HQ15    500.0   118.71 ! Nobs =    1
ANGLe NQ2  CQ7  CQ6     500.0   121.77 ! Nobs =    1
ANGLe NQ2  CQ7  HQ8     500.0   116.11 ! Nobs =    1
ANGLe CQ6  CQ7  HQ8     500.0   122.11 ! Nobs =    1
ANGLe CQ3  CQ9  CQ10    500.0   120.12 ! Nobs =    1
ANGLe CQ3  CQ9  HQ16    500.0   121.44 ! Nobs =    1
ANGLe CQ10 CQ9  HQ16    500.0   118.43 ! Nobs =    1
ANGLe CQ9  CQ10 CQ11    500.0   121.41 ! Nobs =    1
ANGLe CQ9  CQ10 HQ17    500.0   120.46 ! Nobs =    1
ANGLe CQ11 CQ10 HQ17    500.0   118.14 ! Nobs =    1
ANGLe CQ10 CQ11 CQ12    500.0   119.52 ! Nobs =    1
ANGLe CQ10 CQ11 OQ13    500.0   115.70 ! Nobs =    1
ANGLe CQ12 CQ11 OQ13    500.0   124.78 ! Nobs =    1
ANGLe CQ4  CQ12 CQ11    500.0   120.42 ! Nobs =    1
ANGLe CQ4  CQ12 HQ18    500.0   118.69 ! Nobs =    1
ANGLe CQ11 CQ12 HQ18    500.0   120.90 ! Nobs =    1
ANGLe CQ11 OQ13 HQ30    500.0   111.90 ! Nobs =    1
ANGLe CQ20 OQ19 HQ29    500.0   109.90 ! Nobs =    1
```

```
 ANGLe CQ1  CQ20 OQ19    500.0   108.46 ! Nobs =    1
 ANGLe CQ1  CQ20 CQ22    500.0   108.93 ! Nobs =    1
 ANGLe CQ1  CQ20 HQ25    500.0   109.41 ! Nobs =    1
 ANGLe OQ19 CQ20 CQ22    500.0   110.93 ! Nobs =    1
 ANGLe OQ19 CQ20 HQ25    500.0   111.16 ! Nobs =    1
 ANGLe CQ22 CQ20 HQ25    500.0   107.91 ! Nobs =    1
 ANGLe CQ22 OQ21 HQ28    500.0   109.93 ! Nobs =    1
 ANGLe CQ20 CQ22 OQ21    500.0   107.25 ! Nobs =    1
 ANGLe CQ20 CQ22 HQ23    500.0   109.58 ! Nobs =    1
 ANGLe CQ20 CQ22 HQ24    500.0   108.15 ! Nobs =    1
 ANGLe OQ21 CQ22 HQ23    500.0   112.04 ! Nobs =    1
 ANGLe OQ21 CQ22 HQ24    500.0   111.52 ! Nobs =    1
 ANGLe HQ23 CQ22 HQ24    500.0   108.22 ! Nobs =    1

{ Note: edit if necessary }
 DIHEdral HQ27 CQ1  NQ2  CQ7     750.0 0    90.00 ! Nobs =    1 ... Value =    99.41
 DIHEdral HQ27 CQ1  CQ20 HQ25    750.0 0   180.00 ! Nobs =    1 ... Value =  -170.45
 DIHEdral CQ1  NQ2  CQ3  CQ4     750.0 0   180.00 ! Nobs =    1 ... Value =  -179.29
 DIHEdral CQ1  NQ2  CQ3  CQ9     750.0 0     0.00 ! Nobs =    1 ... Value =     1.29
 DIHEdral CQ7  NQ2  CQ3  CQ4     750.0 0     0.00 ! Nobs =    1 ... Value =     2.40
 DIHEdral CQ7  NQ2  CQ3  CQ9     750.0 0   180.00 ! Nobs =    1 ... Value =  -177.02
 DIHEdral CQ1  NQ2  CQ7  CQ6     750.0 0   180.00 ! Nobs =    1 ... Value =   179.55
 DIHEdral CQ1  NQ2  CQ7  HQ8     750.0 0     0.00 ! Nobs =    1 ... Value =    -1.94
 DIHEdral CQ3  NQ2  CQ7  CQ6     750.0 0     0.00 ! Nobs =    1 ... Value =    -2.15
 DIHEdral CQ3  NQ2  CQ7  HQ8     750.0 0   180.00 ! Nobs =    1 ... Value =   176.35
 DIHEdral NQ2  CQ3  CQ4  CQ5     750.0 0     0.00 ! Nobs =    1 ... Value =    -1.07
 DIHEdral NQ2  CQ3  CQ4  CQ12    750.0 0   180.00 ! Nobs =    1 ... Value =   179.45
 DIHEdral CQ9  CQ3  CQ4  CQ5     750.0 0   180.00 ! Nobs =    1 ... Value =   178.37
 DIHEdral CQ9  CQ3  CQ4  CQ12    750.0 0     0.00 ! Nobs =    1 ... Value =    -1.11
 DIHEdral NQ2  CQ3  CQ9  CQ10    750.0 0   180.00 ! Nobs =    1 ... Value =  -179.89
 DIHEdral NQ2  CQ3  CQ9  HQ16    750.0 0     0.00 ! Nobs =    1 ... Value =     1.05
 DIHEdral CQ4  CQ3  CQ9  CQ10    750.0 0     0.00 ! Nobs =    1 ... Value =     0.70
 DIHEdral CQ4  CQ3  CQ9  HQ16    750.0 0   180.00 ! Nobs =    1 ... Value =  -178.36
 DIHEdral CQ3  CQ4  CQ5  CQ6     750.0 0     0.00 ! Nobs =    1 ... Value =    -0.51
 DIHEdral CQ3  CQ4  CQ5  HQ14    750.0 0   180.00 ! Nobs =    1 ... Value =  -179.88
 DIHEdral CQ12 CQ4  CQ5  CQ6     750.0 0   180.00 ! Nobs =    1 ... Value =   178.97
 DIHEdral CQ12 CQ4  CQ5  HQ14    750.0 0     0.00 ! Nobs =    1 ... Value =    -0.41
 DIHEdral CQ3  CQ4  CQ12 CQ11    750.0 0     0.00 ! Nobs =    1 ... Value =     0.57
 DIHEdral CQ3  CQ4  CQ12 HQ18    750.0 0   180.00 ! Nobs =    1 ... Value =  -179.72
 DIHEdral CQ5  CQ4  CQ12 CQ11    750.0 0   180.00 ! Nobs =    1 ... Value =  -178.90
 DIHEdral CQ5  CQ4  CQ12 HQ18    750.0 0     0.00 ! Nobs =    1 ... Value =     0.81
 DIHEdral CQ4  CQ5  CQ6  CQ7     750.0 0     0.00 ! Nobs =    1 ... Value =     0.82
 DIHEdral CQ4  CQ5  CQ6  HQ15    750.0 0   180.00 ! Nobs =    1 ... Value =  -178.93
 DIHEdral HQ14 CQ5  CQ6  CQ7     750.0 0   180.00 ! Nobs =    1 ... Value =  -179.81
 DIHEdral HQ14 CQ5  CQ6  HQ15    750.0 0     0.00 ! Nobs =    1 ... Value =     0.44
 DIHEdral CQ5  CQ6  CQ7  NQ2     750.0 0     0.00 ! Nobs =    1 ... Value =     0.50
 DIHEdral CQ5  CQ6  CQ7  HQ8     750.0 0   180.00 ! Nobs =    1 ... Value =  -177.91
 DIHEdral HQ15 CQ6  CQ7  NQ2     750.0 0   180.00 ! Nobs =    1 ... Value =  -179.74
 DIHEdral HQ15 CQ6  CQ7  HQ8     750.0 0     0.00 ! Nobs =    1 ... Value =     1.84
 DIHEdral CQ3  CQ9  CQ10 CQ11    750.0 0     0.00 ! Nobs =    1 ... Value =     0.26
 DIHEdral CQ3  CQ9  CQ10 HQ17    750.0 0   180.00 ! Nobs =    1 ... Value =  -179.76
 DIHEdral HQ16 CQ9  CQ10 CQ11    750.0 0   180.00 ! Nobs =    1 ... Value =   179.35
 DIHEdral HQ16 CQ9  CQ10 HQ17    750.0 0     0.00 ! Nobs =    1 ... Value =    -0.67
 DIHEdral CQ9  CQ10 CQ11 CQ12    750.0 0     0.00 ! Nobs =    1 ... Value =    -0.82
 DIHEdral CQ9  CQ10 CQ11 OQ13    750.0 0   180.00 ! Nobs =    1 ... Value =   179.56
 DIHEdral HQ17 CQ10 CQ11 CQ12    750.0 0   180.00 ! Nobs =    1 ... Value =   179.20
```

```
DIHEdral HQ17 CQ10 CQ11 OQ13    750.0 0     0.00 ! Nobs =    1 ... Value =    -0.42
DIHEdral CQ10 CQ11 CQ12 CQ4     750.0 0     0.00 ! Nobs =    1 ... Value =     0.38
DIHEdral CQ10 CQ11 CQ12 HQ18    750.0 0   180.00 ! Nobs =    1 ... Value =  -179.31
DIHEdral OQ13 CQ11 CQ12 CQ4     750.0 0   180.00 ! Nobs =    1 ... Value =   179.97
DIHEdral OQ13 CQ11 CQ12 HQ18    750.0 0     0.00 ! Nobs =    1 ... Value =     0.27
DIHEdral CQ10 CQ11 OQ13 HQ30    750.0 0   180.00 ! Nobs =    1 ... Value =   179.93
DIHEdral CQ12 CQ11 OQ13 HQ30    750.0 0     0.00 ! Nobs =    1 ... Value =     0.33
DIHEdral HQ29 OQ19 CQ20 CQ1     750.0 0   180.00 ! Nobs =    1 ... Value =   175.20
DIHEdral HQ29 OQ19 CQ20 CQ22    750.0 0    60.00 ! Nobs =    1 ... Value =    55.60
DIHEdral HQ29 OQ19 CQ20 HQ25    750.0 0   -60.00 ! Nobs =    1 ... Value =   -64.48
DIHEdral CQ1  CQ20 CQ22 OQ21    750.0 0    60.00 ! Nobs =    1 ... Value =    60.68
DIHEdral CQ1  CQ20 CQ22 HQ23    750.0 0   -60.00 ! Nobs =    1 ... Value =   -61.16
DIHEdral CQ1  CQ20 CQ22 HQ24    750.0 0   180.00 ! Nobs =    1 ... Value =  -178.92
DIHEdral OQ19 CQ20 CQ22 OQ21    750.0 0   180.00 ! Nobs =    1 ... Value =   180.00
DIHEdral OQ19 CQ20 CQ22 HQ23    750.0 0    60.00 ! Nobs =    1 ... Value =    58.16
DIHEdral OQ19 CQ20 CQ22 HQ24    750.0 0   -60.00 ! Nobs =    1 ... Value =   -59.60
DIHEdral HQ25 CQ20 CQ22 OQ21    750.0 0   -60.00 ! Nobs =    1 ... Value =   -58.01
DIHEdral HQ25 CQ20 CQ22 HQ23    750.0 0   180.00 ! Nobs =    1 ... Value =  -179.85
DIHEdral HQ25 CQ20 CQ22 HQ24    750.0 0    60.00 ! Nobs =    1 ... Value =    62.39
DIHEdral HQ28 OQ21 CQ22 CQ20    750.0 0   180.00 ! Nobs =    1 ... Value =   172.83
DIHEdral HQ28 OQ21 CQ22 HQ23    750.0 0   -60.00 ! Nobs =    1 ... Value =   -66.89
DIHEdral HQ28 OQ21 CQ22 HQ24    750.0 0    60.00 ! Nobs =    1 ... Value =    54.59

{ Note: edit if necessary }
IMPRoper CQ1  NQ2  CQ20 HQ26    750.0 0   35.000 ! Nobs =    1 ... Value =   33.167
IMPRoper NQ2  CQ1  CQ3  CQ7     750.0 0    0.000 ! Nobs =    1 ... Value =   -0.935
IMPRoper CQ3  NQ2  CQ4  CQ9     750.0 0    0.000 ! Nobs =    1 ... Value =    0.328
IMPRoper CQ4  CQ3  CQ5  CQ12    750.0 0    0.000 ! Nobs =    1 ... Value =    0.298
IMPRoper CQ5  CQ4  CQ6  HQ14    750.0 0    0.000 ! Nobs =    1 ... Value =    0.330
IMPRoper CQ6  CQ5  CQ7  HQ15    750.0 0    0.000 ! Nobs =    1 ... Value =    0.130
IMPRoper CQ7  NQ2  CQ6  HQ8     750.0 0    0.000 ! Nobs =    1 ... Value =   -0.832
IMPRoper CQ9  CQ3  CQ10 HQ16    750.0 0    0.000 ! Nobs =    1 ... Value =    0.489
IMPRoper CQ10 CQ9  CQ11 HQ17    750.0 0    0.000 ! Nobs =    1 ... Value =   -0.011
IMPRoper CQ11 CQ10 CQ12 OQ13    750.0 0    0.000 ! Nobs =    1 ... Value =    0.224
IMPRoper CQ12 CQ4  CQ11 HQ18    750.0 0    0.000 ! Nobs =    1 ... Value =   -0.158
IMPRoper CQ20 CQ1  OQ19 CQ22    750.0 0  -35.000 ! Nobs =    1 ... Value =  -35.348
! >>> NOTE - unusual value for following improper :    29.83 reset to +35.0
IMPRoper CQ22 CQ20 OQ21 HQ23    750.0 0   35.000 ! Nobs =    1 ... Value =   29.826

{ Note: edit if necessary }
NONBonded CQ1  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
NONBonded NQ2  0.2384  2.8509    0.2384  2.8509 ! assuming Nitrogen
NONBonded CQ3  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
NONBonded CQ4  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
NONBonded CQ5  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
NONBonded CQ6  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
NONBonded CQ7  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
NONBonded HQ8  0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
NONBonded CQ9  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
NONBonded CQ10 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
NONBonded CQ11 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
NONBonded CQ12 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
NONBonded OQ13 0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
NONBonded HQ14 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
NONBonded HQ15 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
NONBonded HQ16 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
```

```
 NONBonded HQ17 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HQ18 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded OQ19 0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
 NONBonded CQ20 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded OQ21 0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
 NONBonded CQ22 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HQ23 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HQ24 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HQ25 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HQ26 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HQ27 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HQ28 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HQ29 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HQ30 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen

 set echo=true end

!***************end of change by lsd - 6HQ *******************

!****************** change by lsd - HCF *********************

 { Note: edit if necessary }
 BOND C1D  HZ29   1000.0  1.092 ! Nobs =    1 !added bond C1'-H1'' by lsd
 BOND C1D  OZ14   1000.0  1.435 ! Nobs =    1 !added bond C1'-O2  by lsd

 BOND CZ1  CZ2    1500.0  1.402 ! Nobs =    1 !mod. to 1500 from 1000 by lsd
 BOND CZ1  CZ6    1500.0  1.402 ! Nobs =    1 !mod. to 1500 from 1000 by lsd
 BOND CZ1  OZ14   1000.0  1.362 ! Nobs =    1
 BOND CZ2  CZ3    1500.0  1.393 ! Nobs =    1 !mod. to 1500 from 1000 by lsd
 BOND CZ2  HZ25   1000.0  1.080 ! Nobs =    1
 BOND CZ3  CZ4    1500.0  1.388 ! Nobs =    1 !mod. to 1500 from 1000 by lsd
 BOND CZ3  HZ18   1000.0  1.084 ! Nobs =    1
 BOND CZ4  CZ5    1500.0  1.410 ! Nobs =    1 !mod. to 1500 from 1000 by lsd
 BOND CZ4  CZ7    1500.0  1.462 ! Nobs =    1 !mod. to 1500 from 1000 by lsd
 BOND CZ5  CZ6    1500.0  1.379 ! Nobs =    1 !mod. to 1500 from 1000 by lsd
 BOND CZ5  CZ9    1500.0  1.512 ! Nobs =    1 !mod. to 1500 from 1000 by lsd
 BOND CZ6  HZ19   1000.0  1.083 ! Nobs =    1
 BOND CZ7  CZ8    1500.0  1.409 ! Nobs =    1 !mod. to 1500 from 1000 by lsd
 BOND CZ7  CZ10   1500.0  1.394 ! Nobs =    1 !mod. to 1500 from 1000 by lsd
 BOND CZ8  CZ9    1500.0  1.513 ! Nobs =    1 !mod. to 1500 from 1000 by lsd
 BOND CZ8  CZ13   1500.0  1.380 ! Nobs =    1 !mod. to 1500 from 1000 by lsd
 BOND CZ9  HZ21   1000.0  1.095 ! Nobs =    1
 BOND CZ9  HZ22   1000.0  1.095 ! Nobs =    1
 BOND CZ10 CZ11   1500.0  1.389 ! Nobs =    1 !mod. to 1500 from 1000 by lsd
 BOND CZ10 HZ23   1000.0  1.083 ! Nobs =    1
 BOND CZ11 CZ12   1500.0  1.401 ! Nobs =    1 !mod. to 1500 from 1000 by lsd
 BOND CZ11 HZ24   1000.0  1.081 ! Nobs =    1
 BOND CZ12 CZ13   1500.0  1.403 ! Nobs =    1 !mod. to 1500 from 1000 by lsd
 BOND CZ12 CZ15   1000.0  1.480 ! Nobs =    1
 BOND CZ13 HZ20   1000.0  1.083 ! Nobs =    1
 BOND CZ15 OZ16   1000.0  1.211 ! Nobs =    1
 BOND CZ15 OZ17   1000.0  1.361 ! Nobs =    1
 BOND OZ17 HZ43   1000.0  0.969 ! Nobs =    1
 BOND CZ26 OZ27   1000.0  1.396 ! Nobs =    1
 BOND CZ26 CZ28   1000.0  1.543 ! Nobs =    1
 BOND OZ27 CZ30   1000.0  1.436 ! Nobs =    1
```

```
BOND CZ28 CZ31   1000.0  1.535 ! Nobs =    1
BOND CZ28 HZ32   1000.0  1.090 ! Nobs =    1
BOND CZ28 HZ33   1000.0  1.090 ! Nobs =    1
BOND CZ30 CZ31   1000.0  1.539 ! Nobs =    1
BOND CZ30 HZ34   1000.0  1.096 ! Nobs =    1
BOND CZ30 CZ36   1000.0  1.509 ! Nobs =    1
BOND CZ31 HZ35   1000.0  1.098 ! Nobs =    1
BOND CZ31 OZ39   1000.0  1.422 ! Nobs =    1
BOND CZ36 HZ37   1000.0  1.101 ! Nobs =    1
BOND CZ36 HZ38   1000.0  1.097 ! Nobs =    1
BOND CZ36 OZ41   1000.0  1.421 ! Nobs =    1

{ Note: edit if necessary }
ANGLe CZ2  CZ1  CZ6     500.0   120.21 ! Nobs =    1
ANGLe CZ2  CZ1  OZ14    500.0   124.48 ! Nobs =    1
ANGLe CZ6  CZ1  OZ14    500.0   115.30 ! Nobs =    1
ANGLe CZ1  CZ2  CZ3     500.0   120.19 ! Nobs =    1
ANGLe CZ1  CZ2  HZ25    500.0   120.77 ! Nobs =    1
ANGLe CZ3  CZ2  HZ25    500.0   119.04 ! Nobs =    1
ANGLe CZ2  CZ3  CZ4     500.0   119.74 ! Nobs =    1
ANGLe CZ2  CZ3  HZ18    500.0   119.21 ! Nobs =    1
ANGLe CZ4  CZ3  HZ18    500.0   121.04 ! Nobs =    1
ANGLe CZ3  CZ4  CZ5     500.0   119.77 ! Nobs =    1
ANGLe CZ3  CZ4  CZ7     500.0   131.57 ! Nobs =    1
ANGLe CZ5  CZ4  CZ7     500.0   108.67 ! Nobs =    1
ANGLe CZ4  CZ5  CZ6     500.0   120.93 ! Nobs =    1
ANGLe CZ4  CZ5  CZ9     500.0   109.97 ! Nobs =    1
ANGLe CZ6  CZ5  CZ9     500.0   129.10 ! Nobs =    1
ANGLe CZ1  CZ6  CZ5     500.0   119.15 ! Nobs =    1
ANGLe CZ1  CZ6  HZ19    500.0   118.53 ! Nobs =    1
ANGLe CZ5  CZ6  HZ19    500.0   122.32 ! Nobs =    1
ANGLe CZ4  CZ7  CZ8     500.0   108.63 ! Nobs =    1
ANGLe CZ4  CZ7  CZ10    500.0   131.15 ! Nobs =    1
ANGLe CZ8  CZ7  CZ10    500.0   120.22 ! Nobs =    1
ANGLe CZ7  CZ8  CZ9     500.0   109.98 ! Nobs =    1
ANGLe CZ7  CZ8  CZ13    500.0   120.64 ! Nobs =    1
ANGLe CZ9  CZ8  CZ13    500.0   129.38 ! Nobs =    1
ANGLe CZ5  CZ9  CZ8     500.0   102.76 ! Nobs =    1
ANGLe CZ5  CZ9  HZ21    500.0   111.76 ! Nobs =    1
ANGLe CZ5  CZ9  HZ22    500.0   111.84 ! Nobs =    1
ANGLe CZ8  CZ9  HZ21    500.0   111.92 ! Nobs =    1
ANGLe CZ8  CZ9  HZ22    500.0   111.91 ! Nobs =    1
ANGLe HZ21 CZ9  HZ22    500.0   106.76 ! Nobs =    1
ANGLe CZ7  CZ10 CZ11    500.0   119.12 ! Nobs =    1
ANGLe CZ7  CZ10 HZ23    500.0   120.97 ! Nobs =    1
ANGLe CZ11 CZ10 HZ23    500.0   119.91 ! Nobs =    1
ANGLe CZ10 CZ11 CZ12    500.0   120.68 ! Nobs =    1
ANGLe CZ10 CZ11 HZ24    500.0   120.08 ! Nobs =    1
ANGLe CZ12 CZ11 HZ24    500.0   119.24 ! Nobs =    1
ANGLe CZ11 CZ12 CZ13    500.0   120.14 ! Nobs =    1
ANGLe CZ11 CZ12 CZ15    500.0   122.01 ! Nobs =    1
ANGLe CZ13 CZ12 CZ15    500.0   117.85 ! Nobs =    1
ANGLe CZ8  CZ13 CZ12    500.0   119.19 ! Nobs =    1
ANGLe CZ8  CZ13 HZ20    500.0   121.95 ! Nobs =    1
ANGLe CZ12 CZ13 HZ20    500.0   118.86 ! Nobs =    1
ANGLe CZ1  OZ14 C1D     500.0   120.09 ! Nobs =    1 !modded by lsd
```

```
 ANGLe CZ12 CZ15 OZ16    500.0   125.17 ! Nobs =    1
 ANGLe CZ12 CZ15 OZ17    500.0   113.31 ! Nobs =    1
 ANGLe OZ16 CZ15 OZ17    500.0   121.52 ! Nobs =    1
 ANGLe CZ15 OZ17 HZ43    500.0   106.35 ! Nobs =    1
 ANGLe OZ14 C1D  O4D     500.0   107.07 ! Nobs =    1 !modded by lsd
 ANGLe OZ14 C1D  C2D     500.0   111.77 ! Nobs =    1 !modded by lsd
 ANGLe OZ14 C1D  HZ29    500.0   108.68 ! Nobs =    1 !modded by lsd
 ANGLe OZ27 CZ26 CZ28    500.0   106.71 ! Nobs =    1
 ANGLe O4D  C1D  HZ29    500.0   107.49 ! Nobs =    1 !modded by lsd
 ANGLe C2D  C1D  HZ29    500.0   114.73 ! Nobs =    1 !modded by lsd
 ANGLe CZ26 OZ27 CZ30    500.0   108.21 ! Nobs =    1
 ANGLe CZ26 CZ28 CZ31    500.0   104.34 ! Nobs =    1
 ANGLe CZ26 CZ28 HZ32    500.0   111.89 ! Nobs =    1
 ANGLe CZ26 CZ28 HZ33    500.0   110.79 ! Nobs =    1
 ANGLe CZ31 CZ28 HZ32    500.0   112.22 ! Nobs =    1
 ANGLe CZ31 CZ28 HZ33    500.0   109.17 ! Nobs =    1
 ANGLe HZ32 CZ28 HZ33    500.0   108.39 ! Nobs =    1
 ANGLe OZ27 CZ30 CZ31    500.0   103.43 ! Nobs =    1
 ANGLe OZ27 CZ30 HZ34    500.0   110.34 ! Nobs =    1
 ANGLe OZ27 CZ30 CZ36    500.0   109.80 ! Nobs =    1
 ANGLe CZ31 CZ30 HZ34    500.0   109.13 ! Nobs =    1
 ANGLe CZ31 CZ30 CZ36    500.0   114.51 ! Nobs =    1
 ANGLe HZ34 CZ30 CZ36    500.0   109.47 ! Nobs =    1
 ANGLe CZ28 CZ31 CZ30    500.0   102.79 ! Nobs =    1
 ANGLe CZ28 CZ31 HZ35    500.0   110.28 ! Nobs =    1
 ANGLe CZ28 CZ31 OZ39    500.0   109.69 ! Nobs =    1
 ANGLe CZ30 CZ31 HZ35    500.0   109.23 ! Nobs =    1
 ANGLe CZ30 CZ31 OZ39    500.0   114.42 ! Nobs =    1
 ANGLe HZ35 CZ31 OZ39    500.0   110.18 ! Nobs =    1
 ANGLe CZ30 CZ36 HZ37    500.0   108.31 ! Nobs =    1
 ANGLe CZ30 CZ36 HZ38    500.0   108.62 ! Nobs =    1
 ANGLe CZ30 CZ36 OZ41    500.0   109.34 ! Nobs =    1
 ANGLe HZ37 CZ36 HZ38    500.0   108.31 ! Nobs =    1
 ANGLe HZ37 CZ36 OZ41    500.0   110.48 ! Nobs =    1
 ANGLe HZ38 CZ36 OZ41    500.0   111.71 ! Nobs =    1

 { Note: edit if necessary }
 IMPRoper CZ1  CZ2  CZ6  OZ14    750.0 0    0.000 ! Nobs =    1 ... Value =   -0.175
 IMPRoper CZ2  CZ1  CZ3  HZ25    750.0 0    0.000 ! Nobs =    1 ... Value =    0.334
 IMPRoper CZ3  CZ2  CZ4  HZ18    750.0 0    0.000 ! Nobs =    1 ... Value =    0.234
 IMPRoper CZ4  CZ3  CZ5  CZ7     750.0 0    0.000 ! Nobs =    1 ... Value =    0.070
 IMPRoper CZ5  CZ4  CZ6  CZ9     750.0 0    0.000 ! Nobs =    1 ... Value =   -0.069
 IMPRoper CZ6  CZ1  CZ5  HZ19    750.0 0    0.000 ! Nobs =    1 ... Value =    0.230
 IMPRoper CZ7  CZ4  CZ8  CZ10    750.0 0    0.000 ! Nobs =    1 ... Value =   -0.008
 IMPRoper CZ8  CZ7  CZ9  CZ13    750.0 0    0.000 ! Nobs =    1 ... Value =   -0.011
! >>> NOTE - unusual value for following improper :   -28.83 reset to -35.0
IMPRoper CZ9  CZ5  CZ8  HZ21    750.0 0  28.835 ! Nobs =    1 ... Value =  -28.835 !default par-file value
! IMPRoper CZ9  CZ5  CZ8  HZ21     94.5 0   28.808 ! mod by lsd, copied from HNF
 IMPRoper CZ10 CZ7  CZ11 HZ23    750.0 0    0.000 ! Nobs =    1 ... Value =    0.013
 IMPRoper CZ11 CZ10 CZ12 HZ24    750.0 0    0.000 ! Nobs =    1 ... Value =   -0.040
 IMPRoper CZ12 CZ11 CZ13 CZ15    750.0 0    0.000 ! Nobs =    1 ... Value =   -0.027
 IMPRoper CZ13 CZ8  CZ12 HZ20    750.0 0    0.000 ! Nobs =    1 ... Value =   -0.016
 IMPRoper CZ15 CZ12 OZ16 OZ17    750.0 0    0.000 ! Nobs =    1 ... Value =   -0.002
 IMPRoper CZ26 OZ14 OZ27 CZ28    750.0 0  -35.000 ! Nobs =    1 ... Value =  -37.104
! >>> NOTE - unusual value for following improper :    28.14 reset to +35.0
 IMPRoper CZ28 CZ26 CZ31 HZ32    750.0 0   35.000 ! Nobs =    1 ... Value =   28.137
```

```
 IMPRoper CZ30 OZ27 CZ31 HZ34    750.0 0   35.000 ! Nobs =    1 ... Value =   31.191
 IMPRoper CZ31 CZ28 CZ30 HZ35    750.0 0   35.000 ! Nobs =    1 ... Value =   30.698
 IMPRoper CZ36 CZ30 HZ37 HZ38    750.0 0  -35.000 ! Nobs =    1 ... Value =  -34.749
 IMPRoper OZ14 C2D  O4D  HZ29    $kchimpr    0    -65.280 !modded by lsd
 { Note: edit if necessary }
 NONBonded CZ1  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CZ2  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CZ3  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CZ4  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CZ5  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CZ6  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CZ7  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CZ8  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CZ9  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CZ10 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CZ11 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CZ12 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CZ13 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded OZ14 0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
 NONBonded CZ15 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded OZ16 0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
 NONBonded OZ17 0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
 NONBonded HZ18 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HZ19 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HZ20 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HZ21 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HZ22 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HZ23 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HZ24 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HZ25 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded CZ26 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded OZ27 0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
 NONBonded CZ28 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HZ29 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded CZ30 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CZ31 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HZ32 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HZ33 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HZ34 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HZ35 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded CZ36 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HZ37 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HZ38 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded OZ39 0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
 NONBonded OZ41 0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
 NONBonded HZ43 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen

!************** end of change by lsd - HCF *******************


!the generic bonds were taken from param11.dna with 3*kq
BOND  C5R  OH       876.000    1.4300     ! 5’ end
BOND  C5D  OH       876.000    1.4300     ! 5’ end
BOND  C3R  OH       876.000    1.4300     ! 3’ end
BOND  C3D  OH       876.000    1.4300     ! 3’ end
BOND  O2R  HO      1350.000    0.9572
```

```
!Phos.  - combined RNA/DNA statistics used
!                    kq       x_eq       sigma
BOND  P    O1P     1489.209    1.485      ! 0.015  Phos
BOND  P    O2P     1489.209    1.485      ! 0.015  P
BOND  P    O5R     3350.720    1.593      ! 0.010  P
BOND  P    OH      3350.720    1.593      ! 0.010  P ! For 5pho patch
BOND  P    O3R     2326.889    1.607      ! 0.012  P
BOND  P    OX17    2326.889    1.607      ! 0.012  P    ! mod by anda
BOND  PX1  O3R     2326.889    1.607      ! 0.012  P    ! mod by anda


!Sugars
!RNA statistics
BOND  O5R  C5R     1709.551    1.425      ! 0.014  Sugar
BOND  C5R  C4R     1982.674    1.510      ! 0.013  S
BOND  C4R  C3R     2769.190    1.524      ! 0.011  S
BOND  C3R  C2R     2769.190    1.525      ! 0.011  S
BOND  C2R  C1R     3350.720    1.528      ! 0.010  S
BOND  O4R  C1R     2326.888    1.414      ! 0.012  S
BOND  O4R  C4R     2326.888    1.453      ! 0.012  S
BOND  O3R  C3R     1982.674    1.423      ! 0.013  S
BOND  C2R  O2R     1982.674    1.413      ! 0.013  S

!DNA statistics
BOND  O5R  C5D     1709.551    1.427      ! 0.014  Sugar
BOND  C5D  C4D     5235.500    1.511      ! 0.008  S
BOND  C4D  C3D     3350.720    1.528      ! 0.010  S
BOND  C3D  C2D     3350.720    1.518      ! 0.010  S
BOND  C2D  C1D     1709.551    1.521      ! 0.014  S
BOND  O4D  C1D     1982.674    1.420      ! 0.013  S
BOND  O4D  C4D     2769.190    1.446      ! 0.011  S
BOND  O3R  C3D     1982.674    1.431      ! 0.013  S

!hydrogen/carbon
BOND  C4R  H       $kchbond   1.09
BOND  C3R  H       $kchbond   1.09
BOND  C2R  H       $kchbond   1.09
BOND  C1R  H       $kchbond   1.09
BOND  C5R  H       $kchbond   1.09

BOND  C4D  H       $kchbond   1.09
BOND  C3D  H       $kchbond   1.09
BOND  C2D  H       $kchbond   1.09
BOND  C1D  H       $kchbond   1.09
BOND  C5D  H       $kchbond   1.09

!Bases
!base specific bonds taken from param11.dna , 3*kq
BOND  O2U  HO      1350.000    0.957      ! UR
BOND  HN   NNA     1416.000    1.010      ! URA
BOND  HN   N1T     1416.000    1.010      ! Infer.
BOND  HN   N1C     1416.000    1.010
BOND  HN   N9G     1416.000    1.010
BOND  HN   N9A     1416.000    1.010
BOND  HN   N9P     1416.000    1.010
```

```
BOND  HN   N3U     1416.000    1.010
BOND  HN   N3T     1416.000    1.010
BOND  H2   N2      1416.000    1.010
BOND  H2   N4C     1416.000    1.010
BOND  H2   N2G     1416.000    1.010
BOND  H2   N6A     1416.000    1.010

BOND  HO   OH      1350.000    0.960      ! PARAM7 (IR stretch 3400 cm-1)

!Base sugar joint bonds (scale from sugar)
!                    kq       x_eq       sigma
BOND  C1R  N1T     1709.551    1.473      ! 0.014  Base
BOND  C1R  N1U     4136.691    1.469      ! 0.009  B
BOND  C1R  N1C     2326.889    1.470      ! 0.012  B
BOND  C1R  N9G     4136.691    1.459      ! 0.009  B
BOND  C1R  N9A     3350.720    1.462      ! 0.010  B
BOND  C1R  N9P     3350.720    1.462      ! 0.010  B

BOND  C1D  N1T     1709.551    1.473      ! 0.014  B   !DNA
BOND  C1D  N1U     4136.691    1.469      ! 0.009  B
BOND  C1D  N1C     2326.889    1.470      ! 0.012  B
BOND  C1D  N9G     4136.691    1.459      ! 0.009  B
BOND  C1D  N9A     3350.720    1.462      ! 0.010  B
BOND  C1D  N9P     3350.720    1.462      ! 0.010  B

!cytosine                     kq          x_eq  sigma
BOND       C2C    ON         1370.370    1.240 !0.009  B
BOND       C4C    N4C        1370.370    1.335 !0.009  B
BOND       N1C    C2C        1110.000    1.397 !0.010  B
BOND       N1C    C6C        3083.333    1.367 !0.006  B
BOND       C2C    NC         1734.375    1.353 !0.008  B
BOND       NC     C4C        2265.306    1.335 !0.007  B
BOND       C4C    C5C        1734.375    1.425 !0.008  B
BOND       C5C    C6C        1734.375    1.339 !0.008  B
BOND       C5C    H          $kchbond    1.09
BOND       C6C    H          $kchbond    1.09

!thymine
BOND       N1T    C2T        1734.375    1.376 !0.008  B
BOND       C2T    N3T        1734.375    1.373 !0.008  B
BOND       N3T    C4T        1734.375    1.382 !0.008  B
BOND       C4T    C5T        1370.370    1.445 !0.009  B
BOND       C5T    C6T        2265.306    1.339 !0.007  B
BOND       C6T    N1T        2265.306    1.378 !0.007  B
BOND       C2T    ON         1734.375    1.220 !0.008  B
BOND       C4T    ON         1370.370    1.228 !0.009  B
BOND       C5T    CC3E       3083.333    1.496 !0.006  B
BOND       C6T    H          $kchbond    1.09
BOND       CC3E   H          $kchbond    1.09

!adenine
BOND       NC     C2A        1370.370    1.339 !0.009  B
BOND       C2A    N3A        1370.370    1.331 !0.009  B
BOND       N3A    C4A        3083.333    1.344 !0.006  B
BOND       C4A    C5A        2265.306    1.383 !0.007  B
BOND       C5A    C6A        1370.370    1.406 !0.009  B
```

```
BOND      C6A   NC      2265.306   1.351 !0.007  B
BOND      C5A   N7A     3083.333   1.388 !0.006  B
BOND      N7A   C8A     2265.306   1.311 !0.007  B
BOND      C8A   N9A     1734.375   1.373 !0.008  B
BOND      N9A   C4A     3083.333   1.374 !0.006  B
BOND      C6A   N6A     1734.375   1.335 !0.008  B
BOND      C8A   H       $kchbond   1.08
BOND      C2A   H       $kchbond   1.09

!purine
BOND      NC    C2P     1370.370   1.339 !0.009  B
BOND      C2P   N3P     1370.370   1.331 !0.009  B
BOND      N3P   C4P     3083.333   1.344 !0.006  B
BOND      C4P   C5P     2265.306   1.383 !0.007  B
BOND      C5P   C6P     1370.370   1.406 !0.009  B
BOND      C6P   NC      2265.306   1.351 !0.007  B
BOND      C5P   N7P     3083.333   1.388 !0.006  B
BOND      N7P   C8P     2265.306   1.311 !0.007  B
BOND      C8P   N9P     1734.375   1.373 !0.008  B
BOND      N9P   C4P     3083.333   1.374 !0.006  B
BOND      C6P   H       $kchbond   1.09  !0.008  B
BOND      C8P   H       $kchbond   1.08
BOND      C2P   H       $kchbond   1.09

!guanine
BOND      NNA   C2G     1734.375   1.373 !0.008  B
BOND      C2G   N3G     1734.375   1.323 !0.008  B
BOND      N3G   C4G     2265.306   1.350 !0.007  B
BOND      C4G   C5G     2265.306   1.379 !0.007  B
BOND      C5G   C6G     1110.000   1.419 !0.010  B
BOND      C6G   NNA     2265.306   1.391 !0.007  B
BOND      C5G   N7G     3083.333   1.388 !0.006  B
BOND      N7G   C8G     3083.333   1.305 !0.006  B
BOND      C8G   N9G     2265.306   1.374 !0.007  B
BOND      N9G   C4G     1734.375   1.375 !0.008  B
BOND      C2G   N2G     1110.000   1.341 !0.010  B
BOND      C6G   O6G     1370.370   1.237 !0.009  B
BOND      C8G   H       $kchbond   1.08

!uracil
BOND      C2U   ON      1370.370   1.219 !0.009  B
BOND      C4U   ON      1734.375   1.232 !0.008  B
BOND      N1U   C2U     1370.370   1.381 !0.009  B
BOND      N1U   C6U     1370.370   1.375 !0.009  B
BOND      C2U   N3U     2265.306   1.373 !0.007  B
BOND      N3U   C4U     1370.370   1.380 !0.009  B
BOND      C4U   C5U     1370.370   1.431 !0.009  B
BOND      C5U   C6U     1370.370   1.337 !0.009  B
BOND      C5U   H       $kchbond   1.09
BOND      C6U   H       $kchbond   1.09
BOND      C2D   NX29    2265.306   1.479 !check param, added for pyr


!Phos.
!the ANGLe s were taken from param11.dna with 3*kq
ANGLe      HO    OH    C5R       139.500  107.300
```

```
ANGLe       HO     O5R    C5R        139.500  107.300
ANGLe       HO     OH     C5D        139.500  107.300
ANGLe       HO     O5R    C5D        139.500  107.300
ANGLe       HO     O3R    P          139.500  107.300
ANGLe       HO     OH   P          139.500  107.300 ! For 5pho patch
ANGLe       HO     O2R    C2R        139.500  107.300
ANGLe       OH     P      O3R        144.300  102.600 !
ANGLe       OH     P      O5R        144.300  102.600 !
ANGLe       OH     P      O1P        296.700  108.230 !
ANGLe       OH     P      O2P        296.700  108.230 !
ANGLe       OH     C5R    C4R        210.000  112.000 !
ANGLe       OH     C5D    C4D        210.000  112.000 !
ANGLe       C4D    C3D    OH         139.500  111.000 !
ANGLe       C4R    C3R    OH         139.500  111.000 !
ANGLe       C2D    C3D    OH         139.500  111.000 !
ANGLe       C2R    C3R    OH         139.500  111.000 !
ANGLe       C3R    OH     HO         139.500  107.300 !
ANGLe       C3D    OH     HO         139.500  107.300 !

!Phos.  - combined RNA/DNA statistics used
!                                     kq      x_eq  sigma
ANGLe       O1P    P      O2P       1337.074  119.600 !1.5 P
ANGLe       O5R    P      O1P        357.719  108.100 !2.9 P
ANGLe       O5R    P      O2P        412.677  108.300 !2.7 P
ANGLe       O3R    P      O5R        833.356  104.000 !1.9 P
ANGLe       OX17   P      O5R        833.356  104.000 !1.9 P !mod by anda
ANGLe       O3R    PX1    OX4        833.356  104.000 !1.9 P !mod by anda

ANGLe       O2P    P      O3R        293.791  108.300 !3.2 P
ANGLe       O1P    P      O3R        293.791  107.400 !3.2 P
ANGLe       OX3    PX1    O3R        293.791  108.300 !3.2 P !mod by anda
ANGLe       OX2    PX1    O3R        293.791  107.400 !3.2 P !mod by anda
ANGLe       O2P    P      OX17       293.791  108.300 !3.2 P !mod by anda
ANGLe       O1P    P      OX17       293.791  107.400 !3.2 P !mod by anda

ANGLe       O5R    C5R    C4R       1534.906  110.200 !1.4 P
ANGLe       P      O5R    C5R       1175.163  120.900 !1.6 P
ANGLe       P      O3R    C3R       2089.178  119.700 !1.2 P

ANGLe       O5R    C5D    C4D       1534.906  110.200 !1.4 P !DNA
ANGLe       P      O5R    C5D       1175.163  120.900 !1.6 P
ANGLe       P      O3R    C3D       2089.178  119.700 !1.2 P
ANGLe       PX1    O3R    C3D       2089.178  119.700 !1.2 P !mod by anda
ANGLe       P      OX17   CX12      2089.178  119.700 !1.2 P !mod by anda


!Sugars
!RNA statistics
!                                     kq      x_eq  sigma
ANGLe       O4R    C4R    C3R        561.212  105.500 !1.4 S
ANGLe       C5R    C4R    C3R        488.878  115.500 !1.5 S
ANGLe       C5R    C4R    O4R        561.212  109.200 !1.4 S
ANGLe       C1R    O4R    C4R       1357.996  109.600 !0.9 S
ANGLe       C4R    C3R    C2R       1099.976  102.700 !1.0 S
ANGLe       C3R    C2R    C1R       1357.996  101.500 !0.9 S
ANGLe       O4R    C1R    C2R        561.212  106.400 !1.4 S
```

```
ANGLe      N1T   C1R   C2R       429.678  113.400 !1.6 S
ANGLe      N1C   C1R   C2R       429.678  113.400 !1.6 S
ANGLe      N1U   C1R   C2R       429.678  113.400 !1.6 S
ANGLe      N9G   C1R   C2R       429.678  113.400 !1.6 S
ANGLe      N9A   C1R   C2R       429.678  113.400 !1.6 S
ANGLe      N9P   C1R   C2R       429.678  113.400 !1.6 S
ANGLe      O4R   C1R   N1T      1099.976  108.200 !1.0 S
ANGLe      O4R   C1R   N1C      1099.976  108.200 !1.0 S
ANGLe      O4R   C1R   N1U      1099.976  108.200 !1.0 S
ANGLe      O4R   C1R   N9A      1099.976  108.200 !1.0 S
ANGLe      O4R   C1R   N9P      1099.976  108.200 !1.0 S
ANGLe      O4R   C1R   N9G      1099.976  108.200 !1.0 S
ANGLe      C1R   C2R   O2R       357.719  110.600 !2.9 S scale from phos.
ANGLe      C3R   C2R   O2R       357.719  113.300 !2.9 S scale from phos.
ANGLe      C4R   C3R   O3R       445.032  110.500 !2.6 S scale from phos.
ANGLe      C2R   C3R   O3R       383.726  111.000 !2.8 S scale from phos.

!DNA statistics
ANGLe      O4D   C4D   C3D      1099.976  105.600 !1.0 S
ANGLe      C5D   C4D   C3D       488.878  114.700 !1.5 S
ANGLe      C5D   C4D   O4D       429.678  109.400 !1.6 S
ANGLe      C1D   O4D   C4D       650.874  109.700 !1.3 S
ANGLe      C4D   C3D   C2D      1099.976  103.200 !1.0 S
ANGLe      C3D   C2D   C1D       650.874  102.700 !1.3 S
ANGLe      O4D   C1D   C2D       909.071  106.100 !1.1 S
ANGLe      N1T   C1D   C2D       488.878  114.200 !1.5 S
ANGLe      N1C   C1D   C2D       488.878  114.200 !1.5 S
ANGLe      N1U   C1D   C2D       488.878  114.200 !1.5 S
ANGLe      N9G   C1D   C2D       488.878  114.200 !1.5 S
ANGLe      N9A   C1D   C2D       488.878  114.200 !1.5 S
ANGLe      N9P   C1D   C2D       488.878  114.200 !1.5 S
ANGLe      O4D   C1D   N1T      1357.996  107.800 !0.9 S
ANGLe      O4D   C1D   N1C      1357.996  107.800 !0.9 S
ANGLe      O4D   C1D   N1U      1357.996  107.800 !0.9 S
ANGLe      O4D   C1D   N9A      1357.996  107.800 !0.9 S
ANGLe      O4D   C1D   N9P      1357.996  107.800 !0.9 S
ANGLe      O4D   C1D   N9G      1357.996  107.800 !0.9 S
ANGLe      C4D   C3D   O3R       621.574  110.300 !2.2 S scale from phos.
ANGLe      C2D   C3D   O3R       412.677  110.600 !2.7 S scale from phos.

!Ribose terms involving non-exchageables
ANGLe      OH    C5R   H        $kchangle     109.83
ANGLe      O5R   C5R   H        $kchangle     109.83
ANGLe      H     C5R   H        $kchangle     109.11
ANGLe      C4R   C5R   H        $kchangle     109.11
ANGLe      C5R   C4R   H        $kchangle     107.93
ANGLe      H     C4R   C3R      $kchangle     107.13
ANGLe      H     C4R   O4R      $kchangle     113.74
ANGLe      H     C3R   C4R      $kchangle     111.35
ANGLe      H     C3R   O3R      $kchangle     105.87
ANGLe      H     C3R   OH       $kchangle     105.87
ANGLe      H     C3R   C2R      $kchangle     112.27
ANGLe      H     C2R   C3R      $kchangle     111.41
ANGLe      H     C2R   O2R      $kchangle     113.07
ANGLe      H     C2R   C1R      $kchangle     112.38
ANGLe      H     C1R   C2R      $kchangle     111.95
```

```
ANGLe      H      C1R    N1C       $kchangle     107.70
ANGLe      H      C1R    N1U       $kchangle     107.70
ANGLe      H      C1R    N1T       $kchangle     107.70
ANGLe      H      C1R    N9A       $kchangle     107.70
ANGLe      H      C1R    N9P       $kchangle     107.70
ANGLe      H      C1R    N9G       $kchangle     107.70
ANGLe      H      C1R    O4R       $kchangle     106.86

!Deoxyribose terms involving non-exchageables
!
ANGLe      OH     C5D    H         $kchangle     109.70
ANGLe      O5R    C5D    H         $kchangle     109.70
ANGLe      H      C5D    H         $kchangle     109.17
ANGLe      C4D    C5D    H         $kchangle     109.17
ANGLe      C5D    C4D    H         $kchangle     107.78
ANGLe      H      C4D    C3D       $kchangle     106.91
ANGLe      H      C4D    O4D       $kchangle     112.98
ANGLe      H      C3D    C4D       $kchangle     111.16
ANGLe      H      C3D    O3R       $kchangle     109.34
ANGLe      H      C3D    OH        $kchangle     109.34
ANGLe      H      C3D    C2D       $kchangle     111.98
ANGLe      H      C2D    C3D       $kchangle     111.36
ANGLe      H      C2D    H         $kchangle     107.52
ANGLe      H      C1D    H         $kchangle     107.52 !mod by anda
ANGLe      H      C2D    C1D       $kchangle     112.29
ANGLe      H      C1D    C2D       $kchangle     110.94
ANGLe      H      C1D    N1C       $kchangle     108.25
ANGLe      H      C1D    N1U       $kchangle     108.25
ANGLe      H      C1D    N1T       $kchangle     108.25
ANGLe      H      C1D    N9A       $kchangle     108.25
ANGLe      H      C1D    N9P       $kchangle     108.25
ANGLe      H      C1D    N9G       $kchangle     108.25
ANGLe      H      C1D    O4D       $kchangle     107.95

!Bases
!cytosine                           kq       x_eq   sigma
ANGLe      C6C    N1C    C2C       2277.447  120.300 !0.40 B
ANGLe      N1C    C2C    NC         743.656  119.200 !0.70 B
ANGLe      C2C    NC     C4C       1457.566  119.900 !0.50 B
ANGLe      NC     C4C    C5C       2277.447  121.900 !0.40 B
ANGLe      C4C    C5C    C6C       1457.566  117.400 !0.50 B
ANGLe      C5C    C6C    N1C       1457.566  121.000 !0.50 B
ANGLe      N1C    C2C    ON        1012.199  118.900 !0.60 B
ANGLe      NC     C2C    ON         743.656  121.900 !0.70 B
ANGLe      NC     C4C    N4C        743.656  118.000 !0.70 B
ANGLe      C5C    C4C    N4C        743.656  120.200 !0.70 B
ANGLe      C6C    N1C    C1R        763.873  120.800 !1.20 B scale from sugar
ANGLe      C2C    N1C    C1R        909.071  118.800 !1.10 B scale from sugar
ANGLe      C6C    N1C    C1D        763.873  120.800 !1.20 B !DNA
ANGLe      C2C    N1C    C1D        909.071  118.800 !1.10 B
ANGLe      C4C    N4C    H2         105.000  120.000 !from param11.dna, 3*keq
ANGLe      H2     N4C    H2         105.000  120.000
ANGLe      N1C    C6C    H         $kchangle  119.63
ANGLe      C5C    C6C    H         $kchangle  119.36
ANGLe      C4C    C5C    H         $kchangle  121.54
ANGLe      C6C    C5C    H         $kchangle  121.54
```

```
!thymine                        kq        x_eq    sigma
ANGLe       C6T    N1T    C2T       1457.566  121.300 !0.50 B
ANGLe       N1T    C2T    N3T       1012.199  114.600 !0.60 B
ANGLe       C2T    N3T    C4T       1012.199  127.200 !0.60 B
ANGLe       N3T    C4T    C5T       1012.199  115.200 !0.60 B
ANGLe       C4T    C5T    C6T       1012.199  118.000 !0.60 B
ANGLe       C5T    C6T    N1T       1012.199  123.700 !0.60 B
ANGLe       N1T    C2T    ON         569.362  123.100 !0.80 B
ANGLe       N3T    C2T    ON        1012.199  122.300 !0.60 B
ANGLe       N3T    C4T    ON        1012.199  119.900 !0.60 B
ANGLe       C5T    C4T    ON         743.656  124.900 !0.70 B
ANGLe       C4T    C5T    CC3E      1012.199  119.000 !0.60 B
ANGLe       C6T    C5T    CC3E      1012.199  122.900 !0.60 B
ANGLe       C6T    N1T    C1R        488.878  120.400 !1.50 B scale from sugar
ANGLe       C2T    N1T    C1R        429.678  118.200 !1.60 B scale from sugar
ANGLe       C6T    N1T    C1D        488.878  120.400 !1.50 B !DNA
ANGLe       C2T    N1T    C1D        429.678  118.200 !1.60 B
ANGLe       C2T    N3T    HN         105.000  116.500  !from param11.dna, 3*keq
ANGLe       C4T    N3T    HN         105.000  116.500
ANGLe      C5T    CC3E   H         $kchangle  109.50
ANGLe       H      CC3E   H         $kchangle  109.44
ANGLe       N1T    C6T    H         $kchangle  119.52
ANGLe       C5T    C6T    H         $kchangle  119.52

!adenine                        kq        x_eq    sigma
ANGLe       C6A    NC     C2A       1012.199  118.600 !0.60 B
ANGLe       NC     C2A    N3A       1457.566  129.300 !0.50 B
ANGLe       C2A    N3A    C4A       1457.566  110.600 !0.50 B
ANGLe       N3A    C4A    C5A        743.656  126.800 !0.70 B
ANGLe       C4A    C5A    C6A       1457.566  117.000 !0.50 B
ANGLe       C5A    C6A    NC        1457.566  117.700 !0.50 B
ANGLe       C4A    C5A    N7A       1457.566  110.700 !0.50 B
ANGLe       C5A    N7A    C8A       1457.566  103.900 !0.50 B
ANGLe       N7A    C8A    N9A       1457.566  113.800 !0.50 B
ANGLe       C8A    N9A    C4A       2277.447  105.800 !0.40 B
ANGLe       N9A    C4A    C5A       2277.447  105.800 !0.40 B
ANGLe       N3A    C4A    N9A        569.362  127.400 !0.80 B
ANGLe       C6A    C5A    N7A        743.656  132.300 !0.70 B
ANGLe       NC     C6A    N6A       1012.199  118.600 !0.60 B
ANGLe       C5A    C6A    N6A        569.362  123.700 !0.80 B
ANGLe       C8A    N9A    C1R        339.499  127.700 !1.80 B scale from sugar
ANGLe       C4A    N9A    C1R        339.499  126.300 !1.80 B scale from sugar
ANGLe       C8A    N9A    C1D        339.499  127.700 !1.80 B !DNA
ANGLe       C4A    N9A    C1D        339.499  126.300 !1.80 B
ANGLe       C6A    N6A    H2         105.000  120.000  !from param11.dna, 3*keq
ANGLe       H2     N6A    H2         105.000  120.000
ANGLe       N7A    C8A    H         $kchangle  123.16
ANGLe       N9A    C8A    H         $kchangle  123.16
ANGLe       NC     C2A    H         $kchangle  115.54
ANGLe       N3A    C2A    H         $kchangle  115.54

!purine                         kq        x_eq    sigma
ANGLe       C6P    NC     C2P       1012.199  118.600 !0.60 B
ANGLe       NC     C2P    N3P       1457.566  129.300 !0.50 B
ANGLe       C2P    N3P    C4P       1457.566  110.600 !0.50 B
```

```
ANGLe     N3P   C4P   C5P      743.656  126.800 !0.70 B
ANGLe     C4P   C5P   C6P     1457.566  117.000 !0.50 B
ANGLe     C5P   C6P   NC      1457.566  117.700 !0.50 B
ANGLe     C4P   C5P   N7P     1457.566  110.700 !0.50 B
ANGLe     C5P   N7P   C8P     1457.566  103.900 !0.50 B
ANGLe     N7P   C8P   N9P     1457.566  113.800 !0.50 B
ANGLe     C8P   N9P   C4P     2277.447  105.800 !0.40 B
ANGLe     N9P   C4P   C5P     2277.447  105.800 !0.40 B
ANGLe     N3P   C4P   N9P      569.362  127.400 !0.80 B
ANGLe     C6P   C5P   N7P      743.656  132.300 !0.70 B
ANGLe     NC    C6P   H       $kchangle 120.164 !0.60 B !modified by anda
ANGLe     C5P   C6P   H       $kchangle 120.164 !0.80 B !modified by anda
ANGLe     C8P   N9P   C1R      339.499  127.700 !1.80 B scale from sugar
ANGLe     C4P   N9P   C1R      339.499  126.300 !1.80 B scale from sugar
ANGLe     C8P   N9P   C1D      339.499  127.700 !1.80 B !DNA
ANGLe     C4P   N9P   C1D      339.499  126.300 !1.80 B
ANGLe     N7P   C8P   H       $kchangle  123.16
ANGLe     N9P   C8P   H       $kchangle  123.16
ANGLe     NC    C2P   H       $kchangle  115.54
ANGLe     N3P   C2P   H       $kchangle  115.54

!guanine                         kq        x_eq    sigma
ANGLe     C6G   NNA   C2G     1012.199  125.100 !0.60 B
ANGLe     NNA   C2G   N3G     1012.199  123.900 !0.60 B
ANGLe     C2G   N3G   C4G     1457.566  111.900 !0.50 B
ANGLe     N3G   C4G   C5G     1457.566  128.600 !0.50 B
ANGLe     C4G   C5G   C6G     1012.199  118.800 !0.60 B
ANGLe     C5G   C6G   NNA     1457.566  111.500 !0.50 B
ANGLe     C4G   C5G   N7G     2277.447  110.800 !0.40 B
ANGLe     C5G   N7G   C8G     1457.566  104.300 !0.50 B
ANGLe     N7G   C8G   N9G     1457.566  113.100 !0.50 B
ANGLe     C8G   N9G   C4G     2277.447  106.400 !0.40 B
ANGLe     N9G   C4G   C5G     2277.447  105.400 !0.40 B
ANGLe     N3G   C4G   N9G     1012.199  126.000 !0.60 B
ANGLe     C6G   C5G   N7G     1012.199  130.400 !0.60 B
ANGLe     NNA   C2G   N2G      449.866  116.20  !0.90 B
ANGLe     N3G   C2G   N2G      743.656  119.900 !0.70 B
ANGLe     NNA   C6G   O6G     1012.199  119.900 !0.60 B
ANGLe     C5G   C6G   O6G     1012.199  128.600 !0.60 B
ANGLe     C8G   N9G   C1R      650.874  127.000 !1.30 B scale from sugar
ANGLe     C4G   N9G   C1R      650.874  126.500 !1.30 B scale from sugar
ANGLe     C8G   N9G   C1D      650.874  127.000 !1.30 B !DNA
ANGLe     C4G   N9G   C1D      650.874  126.500 !1.30 B
ANGLe     C2G   N2G   H2       105.000  120.000  !from param11.dna, 3*keq
ANGLe     H2    N2G   H2       105.000  120.000
ANGLe     C2G   NNA   HN       105.000  119.300
ANGLe     C6G   NNA   HN       105.000  119.300
ANGLe     N7G   C8G   H       $kchangle  122.91
ANGLe     N9G   C8G   H       $kchangle  122.91

!uracile                         kq        x_eq    sigma
ANGLe     C6U   N1U   C2U     1012.199  121.000 !0.60 B
ANGLe     N1U   C2U   N3U     1012.199  114.900 !0.60 B
ANGLe     C2U   N3U   C4U     1012.199  127.000 !0.60 B
ANGLe     N3U   C4U   C5U     1012.199  114.600 !0.60 B
ANGLe     C4U   C5U   C6U     1012.199  119.700 !0.60 B
```

```
ANGLe      C5U   C6U   N1U      1457.566  122.700 !0.50 B
ANGLe      N1U   C2U   ON        743.656  122.800 !0.70 B
ANGLe      N3U   C2U   ON        743.656  122.200 !0.70 B
ANGLe      N3U   C4U   ON        743.656  119.400 !0.70 B
ANGLe      C5U   C4U   ON       1012.199  125.900 !0.60 B
ANGLe      C6U   N1U   C1R       561.212  121.200 !1.40 B
ANGLe      C2U   N1U   C1R       763.872  117.700 !1.20 B
ANGLe      C6U   N1U   C1D       561.212  121.200 !1.40 B !DNA
ANGLe      C2U   N1U   C1D       763.872  117.700 !1.20 B
ANGLe      C4U   ON    HO        105.000  120.000 !from param11.dna, 3*keq
ANGLe      C2U   N3U   HN        105.000  116.500
ANGLe      C4U   N3U   HN        105.000  116.500
ANGLe      N1U   C6U   H        $kchangle  119.38
ANGLe      C5U   C6U   H        $kchangle  119.38
ANGLe      C4U   C5U   H        $kchangle  119.56
ANGLe      C6U   C5U   H        $kchangle  119.56
ANGLe      C3D   C2D   NX29     1457.566   110.00 !check param, added fpr pyr
ANGLe      H     C2D   NX29      500.00    109.51 !check param, added fpr pyr



{
!Dihedrals from param11.dna (included for terminal residues)

!DIHEdral    X    C2R  C3R  X      4.50 3    0.000
!DIHEdral    X    C4R  C3R  X      4.50 3    0.000
!DIHEdral    X    C2R  C1R  X      4.50 3    0.000
!DIHEdral    X    C5R  O5R  X      1.50 3    0.000
!DIHEdral    X    C3R  O3R  X      1.50 3    0.000
DIHEdral    X    C3R  OH   X      1.50 3    0.000
DIHEdral    X    C5R  OH   X      1.50 3    0.000
!DIHEdral    X    C2R  O2R  X      1.50 3    0.000
!DIHEdral    X    O5R  P    X      2.25 3    0.000
DIHEdral    X    OH   P    X      2.25 3    0.000
!DIHEdral    OH   C5R  C4R  O4R    4.50 3    0.000
!DIHEdral    OH   C5R  C4R  C3R    4.50 3    0.000 ! gamma
!DIHEdral    C3R  O3R  P    OH     2.25 3    0.000 ! added by infer
!DIHEdral    C3R  O3R  P    OH     2.25 2    0.000 ! ATB, 7-SEP-84
DIHEdral    C5R  O5R  P    OH     2.25 3    0.000 ! added by infer
!DIHEdral    C5R  O5R  P    OH     2.25 2    0.000 ! ATB, 7-SEP-84

!DIHEdral    X    C2D  C3D  X      4.50 3    0.000
!DIHEdral    X    C4D  C3D  X      4.50 3    0.000  !DNA
!DIHEdral    X    C2D  C1D  X      4.50 3    0.000
!DIHEdral    X    C5D  O5R  X      1.50 3    0.000
!DIHEdral    X    C3D  O3R  X      1.50 3    0.000
DIHEdral    X    C3D  OH   X      1.50 3    0.000
DIHEdral    X    C5D  OH   X      1.50 3    0.000
!DIHEdral    OH   C5D  C4D  O4D    4.50 3    0.000
!DIHEdral    OH   C5D  C4D  C3D    4.50 3    0.000
!DIHEdral    C3D  O3R  P    OH     2.25 3    0.000
!DIHEdral    C3D  O3R  P    OH     2.25 2    0.000
DIHEdral    C5D  O5R  P    OH     2.25 3    0.000
!DIHEdral    C5D  O5R  P    OH     2.25 2    0.000
}{
!Base hydrogen DIHEdrals taken from param11.dna
```

```
DIHEdral    X    C2G  N2G  X     18.0  2  180.000
DIHEdral    X    C6A  N6A  X     18.0  2  180.000
!DIHEdral    X    C6A  N4C  X     18.00 2  180.000
DIHEdral    X    C4C  N4C  X     18.00 2  180.000
}


!IMPRopers to keep the two purine rings parallel:
!guanine
IMPRoper    C8G  C4G  C5G  NNA  250.0  2  180.000
IMPRoper    C8G  C5G  C4G  C2G  250.0  2  180.000
IMPRoper    N3G  C4G  C5G  N7G  250.0  2  180.000
IMPRoper    C6G  C5G  C4G  N9G  250.0  2  180.000
!adenine
IMPRoper    C8A  C4A  C5A  N9A  250.0  2  180.000 ! WYE AND PATCHED RESIDUES
IMPRoper    C8A  C5A  C4A  C2A  250.0  2  180.000
IMPRoper    C8A  C4A  C5A  NC   250.0  2  180.000
IMPRoper    N3A  C4A  C5A  N7A  250.0  2  180.000
IMPRoper    C6A  C5A  C4A  N9A  250.0  2  180.000
!purine
IMPRoper    C8P  C4P  C5P  N9P  250.0  2  180.000 ! WYE AND PATCHED RESIDUES
IMPRoper    C8P  C5P  C4P  C2P  250.0  2  180.000
IMPRoper    C8P  C4P  C5P  NC   250.0  2  180.000
IMPRoper    N3P  C4P  C5P  N7P  250.0  2  180.000
IMPRoper    C6P  C5P  C4P  N9P  250.0  2  180.000


!other base specific non-exch hydrogen IMPRopers
IMPRoper    H    C4C  C6C  C5C  $kchimpr    0    0.000
IMPRoper    H    N1C  C5C  C6C  $kchimpr    0    0.000
IMPRoper    H    C4U  C6U  C5U  $kchimpr    0    0.000
IMPRoper    H    N1U  C5U  C6U  $kchimpr    0    0.000
IMPRoper    H    N1T  C5T  C6T  $kchimpr    0    0.000
IMPRoper    H    N7A  N9A  C8A  $kchimpr    0    0.000
IMPRoper    H    NC   N3A  C2A  $kchimpr    0    0.000
IMPRoper    H    N7P  N9P  C8P  $kchimpr    0    0.000
IMPRoper    H    NC   N3P  C2P  $kchimpr    0    0.000
IMPRoper    H    NC   C5P  C6P  $kchimpr    0    0.000
IMPRoper    H    N7G  N9G  C8G  $kchimpr    0    0.000

!Impropers for ribose chirality
IMPRoper    H    C2R  O4R  N9A    $kchimpr    0    -65.000!C1R
IMPRoper    H    C2R  O4R  N9P    $kchimpr    0    -65.000!C1R
IMPRoper    H    C2R  O4R  N9G    $kchimpr    0    -65.000!C1R
IMPRoper    H    C2R  O4R  N1C    $kchimpr    0    -65.000!C1R
IMPRoper    H    C2R  O4R  N1U    $kchimpr    0    -65.000!C1R
IMPRoper    H    C2R  O4R  N1T    $kchimpr    0    -65.000!C1R

IMPRoper    H    C3R  C1R  O2R    $kchimpr    0    65.000!C2R
IMPRoper    H    C4R  C2R  O3R    $kchimpr    0    60.300!C3R
IMPRoper    H    C4R  C2R  OH     $kchimpr    0    60.300!C3R; TERMINAL RES
IMPRoper    H    C5R  C3R  O4R    $kchimpr    0    70.300!C4R
IMPRoper    H    O5R  H    C4R    $kchimpr    0    72.000!C5R;
IMPRoper    H    OH   H    C4R    $kchimpr    0    72.000!C5R; TERMINAL RES
!Impropers for deoxyribose chirality
IMPRoper    H    C2D  O4D  N9A    $kchimpr    0    -65.280!C1D
```

```
IMPRoper    H    C2D  O4D  N9P    $kchimpr    0    -65.280!C1D
IMPRoper    H    C2D  O4D  N9G    $kchimpr    0    -65.280!C1D
IMPRoper    H    C2D  O4D  N1C    $kchimpr    0    -65.280!C1D
IMPRoper    H    C2D  O4D  N1T    $kchimpr    0    -65.280!C1D
IMPRoper    H    C2D  O4D  N1U    $kchimpr    0    -65.280!C1D
IMPRoper    H    C2D  O4D  H      $kchimpr    0    -65.280!C1D !mod by anda (ABA)

IMPRoper    H    C3D  H    C1D    $kchimpr    0    -73.500!C2D
IMPRoper    H    C4D  C2D  O3R    $kchimpr    0    62.660!C3D
IMPRoper    H    C4D  C2D  OH     $kchimpr    0    62.660!C3D; TERMINAL RES
IMPRoper    H    C5D  C3D  O4D    $kchimpr    0    70.220!C4D
IMPRoper    H    O5R  H    C4D    $kchimpr    0    71.430!C5D;
IMPRoper    H    OH   H    C4D    $kchimpr    0    71.430!C5D; TERMINAL RES

{
!Phos.  - periodical potentials from combined RNA/DNA statistics
!                                 kq       x_eq       (sigma in parenthesis)
DIHEdral    O3R  P    O5R  C5R    1.41 3   24     ! alpha !P (20.3)
DIHEdral    P    O5R  C5R  C4R    3.45 0  178     ! beta  !P (13.0)
DIHEdral    O5R  C5R  C4R  C3R   12.24 3   18     ! gamma !S (6.9)
DIHEdral    O5R  C5R  C4R  O4R   24.28 3   14.1   !       !S (4.9)
DIHEdral    C4R  C3R  O3R  P      7.88 0 -153     ! eps   !P (8.6)
DIHEdral    C3R  O3R  P    O5R    1.75 3   33     ! zeta  !P (18.3)

DIHEdral    O3R  P    O5R  C5D    1.41 3    6.0 !DNA
DIHEdral    P    O5R  C5D  C4D    3.45 0  183.5
DIHEdral    O5R  C5D  C4D  C3D   12.42 3   18.3
DIHEdral    O5R  C5D  C4D  O4D   24.28 3   14.1
DIHEdral    C4D  C3D  O3R  P      7.88 0  214.0
DIHEdral    C3D  O3R  P    O5R    1.75 3    0.3

!Phos.  - discrete values from combined RNA/DNA statistics
!                                 kq       x_eq       (sigma in parenthesis)
!DIHEdral    O3R  P    O5R  C5R    6.07 0  285.3   ! (9.8)  alpha1 !P
!DIHEdral    O3R  P    O5R  C5R    3.98 0   81.1   ! (12.1) alpha2;alpha3=180.
!DIHEdral    P    O5R  C5R  C4R    3.44 0  183.5   ! (13.0) beta   !P
!DIHEdral    O5R  C5R  C4R  C3R   17.94 0   52.5   ! (5.7)  gamma1 !S
!DIHEdral    O5R  C5R  C4R  C3R   14.23 0  179.4   ! (6.4)  gamma2 !S
!DIHEdral    O5R  C5R  C4R  C3R    3.85 0  292.9   ! (12.3) gamma3 !S
!DIHEdral    C4R  C3R  O3R  P      7.88 0  214.0   ! (8.6)  eps    !P
!DIHEdral    C3R  O3R  P    O5R   25.30 0  289.2   ! (4.8)  zeta1  !P
!DIHEdral    C3R  O3R  P    O5R    2.85 0   80.7   ! (14.3) zeta2;zeta3=180.
}{
!Sugars
! c3'-endo conformation as the default for for RNA, c2'-endo for DNA,
!RNA statistics, C3'-endo
DIHEdral    C5R  C4R  C3R  O3R  30.12 0   81.1   ! delta ! c3'-endo S (4.4)
DIHEdral    O4R  C4R  C3R  O3R  33.10 0  201.8   !  4.2  ! c3'-endo S
DIHEdral    O4R  C1R  C2R  C3R  24.28 0  335.4   !  4.9  ! c3'-endo S
DIHEdral    C1R  C2R  C3R  C4R  74.36 0   35.9   !  2.8  ! c3'-endo S
DIHEdral    C2R  C3R  C4R  O4R  60.67 0  324.7   !  3.1  ! c3'-endo S
DIHEdral    C3R  C4R  O4R  C1R  22.42 0   20.5   !  5.1  ! c3'-endo S
DIHEdral    C4R  O4R  C1R  C2R  15.67 0    2.8   !  6.1  ! c3'-endo S
DIHEdral    C5R  C4R  C3R  C2R  60.67 0  204.0   !  3.1  ! c3'-endo S
DIHEdral    O3R  C3R  C2R  O2R  28.79 0   44.3   !  4.5  ! c3'-endo S
DIHEdral    C4R  O4R  C1R  N1T  13.80 0  241.4   !  6.5  ! c3'-endo S
```

```
DIHEdral    C4R  O4R  C1R  N1C  13.80 0  241.4   !  6.5  ! c3'-endo S
DIHEdral    C4R  O4R  C1R  N1U  13.80 0  241.4   !  6.5  ! c3'-endo S
DIHEdral    C4R  O4R  C1R  N9G  13.80 0  241.4   !  6.5  ! c3'-endo S
DIHEdral    C4R  O4R  C1R  N9A  13.80 0  241.4   !  6.5  ! c3'-endo S
DIHEdral    C4R  O4R  C1R  N9P  13.80 0  241.4   !  6.5  ! c3'-endo S

!RNA c3'-endo sugar base joint torsions  (combined RNA/DNA statistics used)
DIHEdral    O4R  C1R  N1T  C2T  13.38 0  195.7   !  6.6  ! c3'-endo S
DIHEdral    O4R  C1R  N1C  C2C  13.38 0  195.7   !  6.6  ! c3'-endo S
DIHEdral    O4R  C1R  N1U  C2U  13.38 0  195.7   !  6.6  ! c3'-endo S
DIHEdral    O4R  C1R  N9A  C4A   2.97 0  193.3   ! 14.0  ! c3'-endo S
DIHEdral    O4R  C1R  N9P  C4P   2.97 0  193.3   ! 14.0  ! c3'-endo S
DIHEdral    O4R  C1R  N9G  C4G   2.97 0  193.3   ! 14.0  ! c3'-endo S

!DNA statistics (c2'-endo)
DIHEdral    C5D  C4D  C3D  O3R   36.44 0  145.2   ! delta ! c2'-endo S (4.0)
DIHEdral    O4D  C1D  C2D  C3D   24.28 0   32.8   !  4.9  ! c2'-endo S
DIHEdral    O4D  C4D  C3D  O3R   31.53 0  265.8   !  4.3  ! c2'-endo S
DIHEdral    C1D  C2D  C3D  C4D   44.99 0  326.9   !  3.6  ! c2'-endo S
DIHEdral    C2D  C3D  C4D  O4D   28.79 0   22.6   !  4.5  ! c2'-endo S
DIHEdral    C3D  C4D  O4D  C1D   15.67 0  357.7   !  6.1  ! c2'-endo S
DIHEdral    C4D  O4D  C1D  C2D   14.69 0  340.7   !  6.3  ! c2'-endo S
DIHEdral    C5D  C4D  C3D  C2D   34.68 0  262.0   !  4.1  ! c2'-endo S
DIHEdral    C4D  O4D  C1D  N1T   12.99 0  217.7   !  6.7  ! c2'-endo S
DIHEdral    C4D  O4D  C1D  N1C   12.99 0  217.7   !  6.7  ! c2'-endo S
DIHEdral    C4D  O4D  C1D  N1U   12.99 0  217.7   !  6.7  ! c2'-endo S
DIHEdral    C4D  O4D  C1D  N9G   12.99 0  217.7   !  6.7  ! c2'-endo S
DIHEdral    C4D  O4D  C1D  N9A   12.99 0  217.7   !  6.7  ! c2'-endo S
DIHEdral    C4D  O4D  C1D  N9P   12.99 0  217.7   !  6.7  ! c2'-endo S

!DNA c2'-endo sugar base joint torsions  (combined RNA/DNA statistics used)
DIHEdral    O4D  C1D  N1T  C2T    1.72 0  229.8   ! 18.4  ! c2'-endo S
DIHEdral    O4D  C1D  N1C  C2C    1.72 0  229.8   ! 18.4  ! c2'-endo S
DIHEdral    O4D  C1D  N1U  C2U    1.72 0  229.8   ! 18.4  ! c2'-endo S
DIHEdral    O4D  C1D  N9A  C4A    1.00 0  237.0   ! 24.3  ! c2'-endo S
DIHEdral    O4D  C1D  N9P  C4P    1.00 0  237.0   ! 24.3  ! c2'-endo S
DIHEdral    O4D  C1D  N9G  C4G    1.00 0  237.0   ! 24.3  ! c2'-endo S

!--------------------------------------------------------------------------
!In the case of c3'-endo conformation, the following DIHEdrals are provided
!to overwrite the c2'-endo DIHEdrals

!RNA statistics (c2'-endo)
!DIHEdral    C5R  C4R  C3R  O3R   24.28 0  147.3   ! delta ! c2'-endo S (4.9)
!DIHEdral    O4R  C1R  C2R  C3R   50.43 0   35.2   !  3.4  ! c2'-endo S
!DIHEdral    O4R  C4R  C3R  O3R   20.75 0  268.1   !  5.3  ! c2'-endo S
!DIHEdral    C1R  C2R  C3R  C4R   74.36 0  324.6   !  2.8  ! c2'-endo S
!DIHEdral    C2R  C3R  C4R  O4R   31.53 0   24.2   !  4.3  ! c2'-endo S
!DIHEdral    C3R  C4R  O4R  C1R   17.94 0  357.7   !  5.7  ! c2'-endo S
!DIHEdral    C4R  O4R  C1R  C2R   21.56 0  339.2   !  5.2  ! c2'-endo S
!DIHEdral    C5R  C4R  C3R  C2R   34.68 0  263.4   !  4.1  ! c2'-endo S
!DIHEdral    O3R  C3R  C2R  O2R   33.05 0  319.7   !  4.2  ! c2'-endo S
!DIHEdral    C4R  O4R  C1R  N1T   19.27 0  216.6   !  5.5  ! c2'-endo S
!DIHEdral    C4R  O4R  C1R  N1C   19.27 0  216.6   !  5.5  ! c2'-endo S
!DIHEdral    C4R  O4R  C1R  N1U   19.27 0  216.6   !  5.5  ! c2'-endo S
!DIHEdral    C4R  O4R  C1R  N9G   19.27 0  216.6   !  5.5  ! c2'-endo S
```

```
!DIHEdral    C4R  O4R  C1R  N9A   19.27 0  216.6   !  5.5  ! c2'-endo S

!RNA c2'-endo sugar base joint torsions  (combined RNA/DNA statistics used)
!DIHEdral    O4R  C1R  N1T  C2T    1.72 0  229.8   ! 18.4  ! c2'-endo S
!DIHEdral    O4R  C1R  N1C  C2C    1.72 0  229.8   ! 18.4  ! c2'-endo S
!DIHEdral    O4R  C1R  N1U  C2U    1.72 0  229.8   ! 18.4  ! c2'-endo S
!DIHEdral    O4R  C1R  N9A  C4A    1.00 0  237.0   ! 24.3  ! c2'-endo S
!DIHEdral    O4R  C1R  N9G  C4G    1.00 0  237.0   ! 24.3  ! c2'-endo S

!DNA statistics, c3'-endo (insuficient data, RNA values used)
!DIHEdral    C5D  C4D  C3D  O3R  30.12 0   81.1   ! delta ! c3'-endo S (4.4)
!DIHEdral    O4D  C4D  C3D  O3R  33.10 0  201.8   !  4.2  ! c3'-endo S
!DIHEdral    O4D  C1D  C2D  C3D  24.28 0  335.4   !  4.9  ! c3'-endo S
!DIHEdral    C1D  C2D  C3D  C4D  74.36 0   35.9   !  2.8  ! c3'-endo S
!DIHEdral    C2D  C3D  C4D  O4D  60.67 0  324.7   !  3.1  ! c3'-endo S
!DIHEdral    C3D  C4D  O4D  C1D  22.42 0   20.5   !  5.1  ! c3'-endo S
!DIHEdral    C4D  O4D  C1D  C2D  15.67 0    2.8   !  6.1  ! c3'-endo S
!DIHEdral    C5D  C4D  C3D  C2D  60.67 0  204.0   !  3.1  ! c3'-endo S
!DIHEdral    C4D  O4D  C1D  N1T  13.80 0  241.4   !  6.5  ! c3'-endo S
!DIHEdral    C4D  O4D  C1D  N1C  13.80 0  241.4   !  6.5  ! c3'-endo S
!DIHEdral    C4D  O4D  C1D  N1U  13.80 0  241.4   !  6.5  ! c3'-endo S
!DIHEdral    C4D  O4D  C1D  N9G  13.80 0  241.4   !  6.5  ! c3'-endo S
!DIHEdral    C4D  O4D  C1D  N9A  13.80 0  241.4   !  6.5  ! c3'-endo S

!DNA c3'-endo sugar base joint torsions  (combined RNA/DNA statistics used)
!DIHEdral    O4D  C1D  N1T  C2T  13.38 0  195.7   !  6.6  ! c3'-endo S
!DIHEdral    O4D  C1D  N1C  C2C  13.38 0  195.7   !  6.6  ! c3'-endo S
!DIHEdral    O4D  C1D  N1U  C2U  13.38 0  195.7   !  6.6  ! c3'-endo S
!DIHEdral    O4D  C1D  N9A  C4A   2.97 0  193.3   ! 14.0  ! c3'-endo S
!DIHEdral    O4D  C1D  N9G  C4G   2.97 0  193.3   ! 14.0  ! c3'-endo S
!----------------------------------------------------------------------------
!----------------------------------------------------------------------------
}
```

```
!Impropers taken from param11.dna , 3*kq
IMPRoper    C5R  X    X    C2R   94.5  0   35.260
IMPRoper    C5R  X    X    C1R   94.5  0   35.260
IMPRoper    OH   X    X    C3R   94.5  0   35.260
IMPRoper    OH   X    X    C4R   94.5  0   35.260
IMPRoper    OH   X    X    C1R   94.5  0   35.260
IMPRoper    O3R  X    X    C3R   94.5  0   35.260
IMPRoper    O5R  X    X    C1R   94.5  0   35.260
IMPRoper    O2R  X    X    C2R   94.5  0   35.260
IMPRoper    C4R  O5R  C1R  N1T   94.5  0   35.260
IMPRoper    C4R  O5R  C1R  N1C   94.5  0   35.260
IMPRoper    C4R  O5R  C1R  N9G   94.5  0   35.260
IMPRoper    C4R  O5R  C1R  N9A   94.5  0   35.260
IMPRoper    C4R  O5R  C1R  N9P   94.5  0   35.260
IMPRoper    C5R  O4R  C3R  C4R   94.5  0   35.260
IMPRoper    N1T  C2R  O4R  C1R   94.5  0   35.260
IMPRoper    N1C  C2R  O4R  C1R   94.5  0   35.260
IMPRoper    N9A  C2R  O4R  C1R   94.5  0   35.260
IMPRoper    N9P  C2R  O4R  C1R   94.5  0   35.260
IMPRoper    N9G  C2R  O4R  C1R   94.5  0   35.260
```

```
IMPRoper    C4R  O5R  C1R  N1U   94.5  0   35.260
IMPRoper    N1U  C2R  O4R  C1R   94.5  0   35.260

IMPRoper    C5D  X    X    C2D   94.5  0   35.260  !DNA
IMPRoper    C5D  X    X    C1D   94.5  0   35.260
IMPRoper    OH   X    X    C3D   94.5  0   35.260
IMPRoper    OH   X    X    C4D   94.5  0   35.260
IMPRoper    OH   X    X    C1D   94.5  0   35.260
IMPRoper    O3R  X    X    C3D   94.5  0   35.260
IMPRoper    O5R  X    X    C1D   94.5  0   35.260
IMPRoper    C4D  O5R  C1D  N1T   94.5  0   35.260
IMPRoper    C4D  O5R  C1D  N1C   94.5  0   35.260
IMPRoper    C4D  O5R  C1D  N9G   94.5  0   35.260
IMPRoper    C4D  O5R  C1D  N9A   94.5  0   35.260
IMPRoper    C4D  O5R  C1D  N9P   94.5  0   35.260
IMPRoper    C5D  O4D  C3D  C4D   94.5  0   35.260
IMPRoper    N1T  C2D  O4D  C1D   94.5  0   35.260
IMPRoper    N1C  C2D  O4D  C1D   94.5  0   35.260
IMPRoper    N9A  C2D  O4D  C1D   94.5  0   35.260
IMPRoper    N9P  C2D  O4D  C1D   94.5  0   35.260
IMPRoper    N9G  C2D  O4D  C1D   94.5  0   35.260
IMPRoper    C4D  O5R  C1D  N1U   94.5  0   35.260
IMPRoper    N1U  C2D  O4D  C1D   94.5  0   35.260

!the following impropers were taken from param11x.dna
!the higher kq was used to enforce the ring planarity
!cytosine
IMPRoper    C4C  X    X    ON  2400.0  0    0.000
IMPRoper    C4C  X    X    N1C  250.0  0    0.000
IMPRoper    C6C  X    X    NC   250.0  0    0.000
IMPRoper    C4C  X    X    N2  2400.0  0    0.000
IMPRoper    C2C  X    X    ON  2400.0  0    0.000
!infer
IMPRoper    C1R  C2C  C6C  N1C 2400.0  0    0.000
IMPRoper    C1D  C2C  C6C  N1C 2400.0  0    0.000
IMPRoper    N4C  NC   C5C  C4C 2400.0  0    0.000
IMPRoper    C2C  NC   C4C  C5C  250.0  0    0.000
IMPRoper    C5C  C6C  N1C  C2C  250.0  0    0.000
IMPRoper    H2   C4C  H2   N4C  250.0  0    0.000
IMPRoper    C5C  C4C  N4C  H2  2000.0  0    0.000

!uracil
IMPRoper    C4U  X    X    ON  2400.0  0    0.000
IMPRoper    C4U  X    X    N1U  250.0  0    0.000
IMPRoper    C6U  X    X    N3U  250.0  0    0.000
IMPRoper    C4U  X    X    N2  2400.0  0    0.000
IMPRoper    C2U  X    X    ON  2400.0  0    0.000
IMPRoper    C1R  C2U  C6U  N1U 2400.0  0    0.000
IMPRoper    C1D  C2U  C6U  N1U 2400.0  0    0.000
IMPRoper    ON   N3U  C5U  C4U  250.0  0    0.000
IMPRoper    C2U  N3U  C4U  C5U  250.0  0    0.000
IMPRoper    C5U  C6U  N1U  C2U  250.0  0    0.000
IMPRoper    H2   C4U  H2   ON   250.0  0    0.000
IMPRoper    HN   C2U  C4U  N3U  250.0  0    0.000
IMPRoper    H    C3D  NX29 C1D  500.0  0     -73.5 !added for pyr, jt
```

```
!thymidine
IMPRoper    C4T  X    X    ON  2400.0  0    0.000
IMPRoper    C4T  X    X    N1T  250.0  0    0.000
IMPRoper    C6T  X    X    N3T  250.0  0    0.000
IMPRoper    C4T  X    X    N2  2400.0  0    0.000
IMPRoper    C2T  X    X    ON  2400.0  0    0.000
IMPRoper    C1R  C2T  C6T  N1T 2400.0  0    0.000
IMPRoper    C1D  C2T  C6T  N1T 2400.0  0    0.000
IMPRoper    ON   N3T  C5T  C4T  250.0  0    0.000
IMPRoper    C2T  N3T  C4T  C5T  250.0  0    0.000
IMPRoper    C5T  C6T  N1T  C2T  250.0  0    0.000
IMPRoper    H2   C4T  H2   ON   250.0  0    0.000
IMPRoper    CC3E C4T  C6T  C5T 2400.0  0    0.000

!infer
IMPRoper    HN   C2T  C4T  N3T  250.0  0    0.000

! The ring-spanning impropers have been left out.
!adenine
IMPRoper    N2A  N3A  NC   C2A  250.0  0    0.000
IMPRoper    H2   C2A  H2   N2A  250.0  0    0.000
IMPRoper    C4A  C5A  N7A  C8A  250.0  0    0.000
IMPRoper    C5A  C4A  N9A  C8A  250.0  0    0.000
IMPRoper    C4A  X    X    NC   250.0  0    0.000
IMPRoper    C2A  X    X    N9A  250.0  0    0.000
IMPRoper    C2A  X    X    C5A  250.0  0    0.000
IMPRoper    C6A  C5A  C4A  N3A  250.0  0    0.000
IMPRoper    C5A  X    X    N9A  250.0  0    0.000
IMPRoper    C6A  X    X    N6A 2400.0  0    0.000
IMPRoper    H2   X    X    N6A  250.0  0    0.000

!infer
IMPRoper    C1R  C4A  C8A  N9A 2400.0  0    0.000
IMPRoper    C1D  C4A  C8A  N9A 2400.0  0    0.000
IMPRoper    N9A  C4A  C5A  N7A  250.0  0    0.000
IMPRoper    N7A  C8A  N9A  C4A  250.0  0    0.000
IMPRoper    N3A  C2A  NC   C6A  250.0  0    0.000
IMPRoper    C5A  C6A  N6A  H2  2000.0  0    0.000

! The ring-spanning impropers have been left out.
!purine
IMPRoper    N2P  N3P  NC   C2P  250.0  0    0.000
IMPRoper    C4P  C5P  N7P  C8P  250.0  0    0.000
IMPRoper    C5P  C4P  N9P  C8P  250.0  0    0.000
IMPRoper    C4P  X    X    NC   250.0  0    0.000
IMPRoper    C2P  X    X    N9P  250.0  0    0.000
IMPRoper    C2P  X    X    C5P  250.0  0    0.000
IMPRoper    C6P  C5P  C4P  N3P  250.0  0    0.000
IMPRoper    C5P  X    X    N9P  250.0  0    0.000

!infer
IMPRoper    C1R  C4P  C8P  N9P 2400.0  0    0.000
IMPRoper    C1D  C4P  C8P  N9P 2400.0  0    0.000
IMPRoper    N9P  C4P  C5P  N7P  250.0  0    0.000
IMPRoper    N7P  C8P  N9P  C4P  250.0  0    0.000
IMPRoper    N3P  C2P  NC   C6P  250.0  0    0.000
```

```
! The ring-spanning impropers have been left out.
!guanine
IMPRoper    C4G  C5G  N7G  C8G  250.0  0    0.000
IMPRoper    C5G  C4G  N9G  C8G  250.0  0    0.000
IMPRoper    C4G  X    X    NNA  250.0  0    0.000
IMPRoper    C2G  X    X    N9G  250.0  0    0.000
IMPRoper    C2G  X    X    C5G  250.0  0    0.000
IMPRoper    C6G  C5G  C4G  N3G  250.0  0    0.000
IMPRoper    C5G  X    X    N9G  250.0  0    0.000
IMPRoper    C6G  X    X    O6G 2400.0  0    0.000
IMPRoper    C2G  X    X    N2G 2400.0  0    0.000

!infer
IMPRoper    C1R  C4G  C8G  N9G 2400.0  0    0.000
IMPRoper    C1D  C4G  C8G  N9G 2400.0  0    0.000
IMPRoper    N9G  C4G  C5G  N7G  250.0  0    0.000
IMPRoper    N7G  C8G  N9G  C4G  250.0  0    0.000
IMPRoper    N3G  C2G  NNA  C6G  250.0  0    0.000
IMPRoper    H2   H2   C2G  N2G  250.0  0    0.000
IMPRoper    HN   C2G  C6G  NNA 2000.0  0    0.000
IMPRoper    N3G  C2G  N2G  H2  2000.0  0    0.000

!                   Lennard-Jones parameters
!                                       ------1-4-------
!                  epsilon   sigma       epsilon   sigma
!                 (Kcal/mol)   (A)      (Kcal/mol) (A)
! Taken from Rossky Karplus and Rahman BIOPOLY (1979)
! 0.05 ADDED TO RADII TO IMPRoperOVE ON NUCL.ACID STACKING/LN
!
!                  eps     sigma       eps(1:4) sigma(1:4)


NONBonded  C5R     0.0900   3.2970      0.0900   3.2970
NONBonded  C1R     0.0900   3.2970      0.0900   3.2970
NONBonded  C2R     0.0900   3.2970      0.0900   3.2970
NONBonded  C3R     0.0900   3.2970      0.0900   3.2970
NONBonded  C4R     0.0900   3.2970      0.0900   3.2970

NONBonded  C5D     0.0900   3.2970      0.0900   3.2970  !DNA
NONBonded  C1D     0.0900   3.2970      0.0900   3.2970
NONBonded  C2D     0.0900   3.2970      0.0900   3.2970
NONBonded  C3D     0.0900   3.2970      0.0900   3.2970
NONBonded  C4D     0.0900   3.2970      0.0900   3.2970

NONBonded  HN      0.0045   2.6160      0.0045   2.6160
NONBonded  H2      0.0045   1.6040      0.0045   1.6040
NONBonded  H       0.0045   2.6160      0.0045   2.6160
!
! give it the same as th Hn from RKR
NONBonded  HO      0.0045   1.6040      0.0045   1.6040

!
! THIS STILL IS AN EXTENDED ATOM
NONBonded  O3R      0.2304   2.7290      0.2304   2.7290
NONBonded  O4R      0.2304   2.7290      0.2304   2.7290
```

```
NONBonded  O4D      0.2304   2.7290      0.2304   2.7290
NONBonded  O5R      0.2304   2.7290      0.2304   2.7290
NONBonded  O1P      0.2304   2.7290      0.2304   2.7290
NONBonded  O2P      0.2304   2.7290      0.2304   2.7290
NONBonded  P        0.5849   3.3854      0.5849   3.3854

!bases
NONBonded  C2      0.0900   3.2970      0.0900   3.2970
NONBonded  C3      0.0900   3.2970      0.0900   3.2970
NONBonded  CB      0.0900   3.2970      0.0900   3.2970
NONBonded  CE      0.0900   3.2970      0.0900   3.2970
NONBonded  CH      0.0900   3.2970      0.0900   3.2970

NONBonded  N2      0.1600   2.8591      0.1600   2.8591
NONBonded  N3U     0.1600   2.8591      0.1600   2.8591
NONBonded  N3T     0.1600   2.8591      0.1600   2.8591
NONBonded  NNA     0.1600   2.8591      0.1600   2.8591
NONBonded  NB      0.1600   2.8591      0.1600   2.8591
NONBonded  NC      0.1600   2.8591      0.1600   2.8591

NONBonded  NH2E    0.1600   3.0291      0.1600   3.0291
NONBonded  NS      0.1600   2.8591      0.1600   2.8591
NONBonded  N1T     0.1600   2.8591      0.1600   2.8591
NONBonded  N1C     0.1600   2.8591      0.1600   2.8591
NONBonded  N9A     0.1600   2.8591      0.1600   2.8591
NONBonded  N9P     0.1600   2.8591      0.1600   2.8591
NONBonded  N9G     0.1600   2.8591      0.1600   2.8591
NONBonded  N1U     0.1600   2.8591      0.1600   2.8591
NONBonded  ON      0.2304   2.7290      0.2304   2.7290
NONBonded  O2R     0.2304   2.7290      0.2304   2.7290
NONBonded  OH      0.2304   2.5508      0.2304   2.5508
NONBonded  SD      0.3515   2.6727      0.3515   2.6727  ! G U E S S
NONBonded  O2      0.2304   2.7290      0.2304   2.7290

! NEW
NONBonded  C6C      0.0900   3.2970      0.0900   3.2970
NONBonded  C5C      0.0900   3.2970      0.0900   3.2970
NONBonded  C4C      0.0900   3.2970      0.0900   3.2970
NONBonded  C2C      0.0900   3.2970      0.0900   3.2970
NONBonded  C6U      0.0900   3.2970      0.0900   3.2970
NONBonded  C5U      0.0900   3.2970      0.0900   3.2970
NONBonded  C4U      0.0900   3.2970      0.0900   3.2970
NONBonded  C2U      0.0900   3.2970      0.0900   3.2970
NONBonded  C8A      0.0900   3.2970      0.0900   3.2970
NONBonded  C6A      0.0900   3.2970      0.0900   3.2970
NONBonded  C5A      0.0900   3.2970      0.0900   3.2970
NONBonded  C4A      0.0900   3.2970      0.0900   3.2970
NONBonded  C2A      0.0900   3.2970      0.0900   3.2970
NONBonded  C8P      0.0900   3.2970      0.0900   3.2970
NONBonded  C6P      0.0900   3.2970      0.0900   3.2970
NONBonded  C5P      0.0900   3.2970      0.0900   3.2970
NONBonded  C4P      0.0900   3.2970      0.0900   3.2970
NONBonded  C2P      0.0900   3.2970      0.0900   3.2970
NONBonded  C8G      0.0900   3.2970      0.0900   3.2970
NONBonded  C6G      0.0900   3.2970      0.0900   3.2970
NONBonded  C5G      0.0900   3.2970      0.0900   3.2970
```

```
NONBonded  C4G      0.0900   3.2970      0.0900   3.2970
NONBonded  C2G      0.0900   3.2970      0.0900   3.2970
NONBonded  C6T      0.0900   3.2970      0.0900   3.2970
NONBonded  C5T      0.0900   3.2970      0.0900   3.2970
NONBonded  C4T      0.0900   3.2970      0.0900   3.2970
NONBonded  C2T      0.0900   3.2970      0.0900   3.2970
NONBonded  N4C      0.1600   2.8591      0.1600   2.8591
NONBonded  O4U      0.2304   2.7290      0.2304   2.7290
NONBonded  N7G      0.1600   2.8591      0.1600   2.8591
NONBonded  N3G      0.1600   2.8591      0.1600   2.8591
NONBonded  N2G      0.1600   2.8591      0.1600   2.8591
NONBonded  N3A      0.1600   2.8591      0.1600   2.8591
NONBonded  N7A      0.1600   2.8591      0.1600   2.8591
NONBonded  N6A      0.1600   2.8591      0.1600   2.8591
NONBonded  O6G      0.2304   2.7290      0.2304   2.7290
NONBonded  CC3E     0.0900   3.2970      0.0900   3.2970
NONBonded  N2A      0.1600   2.8591      0.1600   2.8591
NONBonded  N2P      0.1600   2.8591      0.1600   2.8591
NONBonded  N3P      0.1600   2.8591      0.1600   2.8591
NONBonded  N7P      0.1600   2.8591      0.1600   2.8591

! special solute-solute hydrogen bonding potential parameters
!AEXP 4
!REXP 6

!HAEX 4

!AAEX 2
! "all" possible combinations of HB-pairs in nucleic acids:
!  WELL DEPTHS DEEPENED BY 0.5 KCAL TO IMPROVE BASEPAIR ENERGIES /LN
!  AND DISTANCES INCREASED BY 0.05
!                  Emin      Rmin
!               (Kcal/mol)   (A)
!hbond  N*   O*      -14.0       2.95
!hbond  N*   N*      -14.5       3.05
!hbond  O*   O*      -15.75      2.80
!hbond  O*   N*      -15.50      2.90

! the following NBFIXes are for DNA-DNA hydrogen bonding

! terms
!                                          ------1-4------
!                  A           B           A          B
!            [Kcal/(mol A^12)] [Kcal/(mol A^6)]
!
nbfix  HO   ON      0.05       0.1      0.05       0.1
nbfix  HO   O3R     0.05       0.1      0.05       0.1
nbfix  HO   O5R     0.05       0.1      0.05       0.1
nbfix  HO   OH      0.05       0.1      0.05       0.1
nbfix  HO   O2R     0.05       0.1      0.05       0.1
nbfix  HO   NC      0.05       0.1      0.05       0.1


nbfix  H    ON       0.05       0.1      0.05       0.1
nbfix  H    O2       0.05       0.1      0.05       0.1
nbfix  H    O5R      0.05       0.1      0.05       0.1
```

```
nbfix  H    O4R      0.05      0.1      0.05      0.1
nbfix  H    O4D      0.05      0.1      0.05      0.1
nbfix  H    O3R      0.05      0.1      0.05      0.1
nbfix  H    O2R      0.05      0.1      0.05      0.1
nbfix  H    OH       0.05      0.1      0.05      0.1
nbfix  H    N7A      0.05      0.1      0.05      0.1
nbfix  H    N7P      0.05      0.1      0.05      0.1
nbfix  H    N7G      0.05      0.1      0.05      0.1
nbfix  H    N3A      0.05      0.1      0.05      0.1
nbfix  H    N3P      0.05      0.1      0.05      0.1
nbfix  H    N3G      0.05      0.1      0.05      0.1


nbfix  HN   ON      0.05      0.1      0.05      0.1
nbfix  HN   O2R     0.05      0.1      0.05      0.1
nbfix  HN   OH      0.05      0.1      0.05      0.1
nbfix  HN   NC      0.05      0.1      0.05      0.1

nbfix  H2   ON      0.05      0.1      0.05      0.1
nbfix  H2   O2R     0.05      0.1      0.05      0.1
nbfix  H2   OH      0.05      0.1      0.05      0.1

nbfix  H2   NC      0.05      0.1      0.05      0.1

set echo=on message=on end
```

2.5 Topology file used by Xplor-NIH

```
  !RNA TOPOLOGY FILE 'FRAMEWORK' FROM TOPALLHDG.DNA AND ATOM NAMES
    !  FROM DNA-RNA.PARAM
    !INCLUDES ALL NONEXCHANGEABLE HYDROGENS AND TERMS FOR BOND, ANGLE, AND
    !IMPROPERS. NONEXCHANGEABLE HYDROGEN CHARGES WERE ASSIGNED 0.035.
    !CARBON CHARGES WERE REDUCED 0.035 FOR EACH ATTACHED HYDROGEN.
    !CREATED 2/24/96-- JASON P. RIFE AND PETER B. MOORE
    ! DNA-RNA-ALLATOM.TOP

    set echo=false end


    ! checkversion 1.0


    AUTOGENERATE  ANGLES=TRUE  END
    {*==========================*}


!************************* change by lsd - DAP ************************

 { Note: edit masses if necessary }
 MASS OY1    15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS CY2    12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS HY3     1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
```

```
 MASS HY4     1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS CY5    12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS HY6     1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS OY7    15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS CY8    12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS HY9     1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS CY10   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS HY11    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS CY12   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS NY13   14.00700 ! assuming N -> 14.00700 + 1.008 * 0 (Hs)
 MASS HY14    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HY15    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS NY16   15.01500 ! assuming N -> 14.00700 + 1.008 * 1 (Hs)
 MASS CY17   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CY18   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS HY19    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS CY20   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS HY21    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HY22    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS OY23   15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS NY24   15.01500 ! assuming N -> 14.00700 + 1.008 * 1 (Hs)
 MASS CY25   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS NY26   14.00700 ! assuming N -> 14.00700 + 1.008 * 0 (Hs)
 MASS HY27    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HY28    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HY29    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HY30    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HY31    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)

!********************** end of change by lsd - DAP *******************

!************************* change by lsd - 4AP ***********************

 { Note: edit masses if necessary }
 MASS OX1    15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS CX2    12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS HX3     1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HX4     1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS CX5    12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS HX6     1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS OX7    15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS CX8    12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS HX9     1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS CX10   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS HX11    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS CX12   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CX13   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CX14   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS HX15    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS CX16   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS HX17    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HX18    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS OX19   15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS CX20   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS HX21    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HX22    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
```

```
 MASS CX23   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CX24   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS NX25   14.00700 ! assuming N -> 14.00700 + 1.008 * 0 (Hs)
 MASS HX26    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HX27    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS NX28   14.00700 ! assuming N -> 14.00700 + 1.008 * 0 (Hs)
 MASS CX29   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CX30   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS OX31   15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS OX32   15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS HX33    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HX34    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)

 autogenerate angles=true end

!****************** end of change by lsd - 4AP ***********************



!************************* change by lsd - 6HQ ***********************

{ Note: edit masses if necessary }
 MASS CQ1    12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS NQ2    14.00700 ! assuming N -> 14.00700 + 1.008 * 0 (Hs)
 MASS CQ3    12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CQ4    12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CQ5    12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CQ6    12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CQ7    12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS HQ8     1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS CQ9    12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CQ10   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CQ11   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CQ12   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS OQ13   15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS HQ14    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HQ15    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HQ16    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HQ17    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HQ18    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS OQ19   15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS CQ20   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS OQ21   15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS CQ22   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS HQ23    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HQ24    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HQ25    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HQ26    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HQ27    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HQ28    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HQ29    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HQ30    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
!******************** end of change by lsd - 6HQ *********************

!************************* change by lsd - HCF ***********************
```

```
 { Note: edit masses if necessary }
 MASS PZ1    30.97400 ! assuming P -> 30.97400 + 1.008 * 0 (Hs)
 MASS OZ2    15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS OZ3    15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS CZ1    12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CZ2    12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CZ3    12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CZ4    12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CZ5    12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CZ6    12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CZ7    12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CZ8    12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CZ9    12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CZ10   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CZ11   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CZ12   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CZ13   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS OZ14   15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS CZ15   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS OZ16   15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS OZ17   15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS HZ18    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HZ19    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HZ20    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HZ21    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HZ22    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HZ23    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HZ24    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HZ25    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS CZ26   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS OZ27   15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS CZ28   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS HZ29    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS CZ30   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CZ31   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS HZ32    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HZ33    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HZ34    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HZ35    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS CZ36   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS HZ37    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS HZ38    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS OZ39   15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS OZ41   15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS HZ43    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)

!********************* end of change by lsd - HCF ********************


    {* DNA/RNA default masses *}

    MASS   P     30.97400! phosphorus
    MASS   O1P   15.99940! O in phosphate
    MASS   O2P   15.99940! O in phosphate
    MASS   O5R   15.99940! ester -P-O-C-
    MASS   C5R   12.011! corresp. to CH2E
```

```
MASS   C4R   12.011! corresp. to CH1E
MASS   C3R   12.011! corresp. to CH1E
MASS   C2R   12.011! corresp. to CH1E
MASS   C1R   12.011! corresp. to CH1E
MASS   O4R   15.99940! ester -P-O-C-
MASS   O3R   15.99940! ester -P-O-C-
MASS   O2R   15.99940! ester -P-O-C-
MASS   OH    15.99940! corresp. to OH1

!DEOXY SUGAR
MASS   C5D   14.02700! corresp. to CH2E
MASS   C4D   13.01900! corresp. to CH1E
MASS   C3D   13.01900! corresp. to CH1E
MASS   C2D   13.01900! corresp. to CH1E
MASS   C1D   13.01900! corresp. to CH1E
MASS   O4D   15.99940! ester -P-O-C-
MASS   O5D   15.99940!
MASS   O3D   15.99940!

! Insert Bases
! Generic

MASS   N2    14.00670! nitrogen in -NH2
MASS   NNA   14.00670! corresp. to NH1
MASS   ON    15.99940! corresp. to O
MASS   NC    14.00670! corresp. to NR
MASS   NS    14.00670! nitrogen in ring >N-

! Insert 4 Bases
!  GUA
MASS   N9G    14.00670! nitrogen in ring >N-
MASS   C2G    12.011! (prev CE)
MASS   N3G    14.00670! (prev NC)
MASS   C4G    12.01100! (prev CB)
MASS   C5G    12.01100! (prev CB)
MASS   C6G    12.01100! (prev CN)
MASS   N7G    14.00670! (prev NB)
MASS   C8G    12.011! (prev CE)
MASS   O6G    15.99940! (prev CE)
MASS   N2G    14.00670! nitrogen in -NH2
!  ADE
MASS   N9A    14.00670! nitrogen in ring >N-
MASS   C2A    12.011! (prev CE)
MASS   N3A    14.00670! (prev NC)
MASS   C4A    12.01100! (prev CB)
MASS   C5A    12.01100! (prev CB)
MASS   C6A    12.01100! (prev CA)
MASS   N7A    14.00670! (prev NB)
MASS   C8A    12.011! (prev CE)
MASS   N6A    14.00670! nitrogen in -NH2

! PUR
MASS   N9P    14.00670! nitrogen in ring >N-
MASS   C2P    12.011! (prev CE)
MASS   N3P    14.00670! (prev NC)
MASS   C4P    12.01100! (prev CB)
```

```
    MASS   C5P    12.01100! (prev CB)
    MASS   C6P    12.01100! (prev CA)
    MASS   N7P    14.00670! (prev NB)
    MASS   C8P    12.011! (prev CE)

    !  CYT
    MASS   N1C    14.00670! nitrogen in ring >N-
    MASS   C2C    12.01100! (prev CN)
    MASS   C4C    12.01100! (prev CA)
    MASS   C5C    12.011! (prev CF)
    MASS   C6C    12.011! (prev CF)
    MASS   N4C    14.00670! nitrogen in -NH2

    !  THY
    MASS   N1T    14.00670! nitrogen in ring >N-
    MASS   N3T    14.00670! nitrogen in ring >N-
    MASS   C2T    12.01100! (prev CN)
    MASS   C4T    12.01100! (prev CN)
    MASS   C5T    12.011! (prev CS)
    MASS   C6T    12.011! (prev CF)
    MASS   CC3E   12.01100! (prev CF)

    ! END

    MASS   H      1.00800! non-exchangeable Hydrogens
    MASS   HN     1.00800! corresp. to H
    MASS   H2     1.00800! hydrogen in -NH2
    MASS   HO     1.00800! hydroxy hydrogen

    !  URI
    MASS   N1U    14.00670! nitrogen in ring >N-
    MASS   C2U    12.01100! (prev CN)
    MASS   C4U    12.01100! (prev CA)
    MASS   C5U    12.011! (prev CF)
    MASS   C6U    12.011! (prev CF)
    MASS   N3U    14.00670!


    !-------------------------------------------------------------------
    !-------------------------------------------------------------------

RESIdue DAP

 { Note: electrostatics should normally not be used in }
 { crystallographic refinement since it can produce }
 { artefacts. For this reason, all charges are set to }
 { zero by default. Edit them if necessary }
GROUp
 ATOM P    TYPE=P     CHARGE=1.20    END
 ATOM O1P  TYPE=O1P   CHARGE=-0.40   END
 ATOM O2P  TYPE=O2P   CHARGE=-0.40   END
 ATOM O5'  TYPE=O5R   CHARGE=-0.36   END
!Charge of the group: 0.04
GROUp
 ATOM C5'  TYPE=C5R   CHARGE=-0.070  END
 ATOM H5'  TYPE=H     CHARGE=0.035   END
```

```
 ATOM H5'' TYPE=H     CHARGE=0.035   END
!Charge of the group: 0.00
GROUp
 ATOM C4'  TYPE=C4R   CHARGE=0.065   END
 ATOM H4'  TYPE=H     CHARGE=0.035   END
 ATOM O4'  TYPE=O4R   CHARGE=-0.30   END
 ATOM C1'  TYPE=C1R   CHARGE=0.412   END !increased from 0.386 by lsd
 ATOM H1'  TYPE=H     CHARGE=0.035   END
!Charge of the group: 0.247 sums up to +1 with DAP
GROUP
 ATOM C2'  TYPE=C2R   CHARGE=0.115   END
 ATOM H2'  TYPE=H     CHARGE=0.035   END
 ATOM O2'  TYPE=O2R   CHARGE=-0.40   END
 ATOM HO2' TYPE=HO    CHARGE=0.25    END
!Charge of the group: 0.00
GROUp
 ATOM  C3'  TYPE=C3R  CHARGE=-0.035 END
 ATOM  H3'  TYPE=H    CHARGE=0.035  END
!Charge of the group: 0.00
GROUp
 ATOM  O3'  TYPE=O3R  CHARGE=-0.36  END
!Charge of the group: -0.36

!DAP-base
GROUp
 ATOM  C6   TYPE CY10  CHARge  0.215  END ! Nr of Hs =  0
 ATOM  H6   TYPE HY11  CHARge  0.123  END ! Nr of Hs =  0
 ATOM  C2   TYPE CY12  CHARge  1.034  END ! Nr of Hs =  0
 ATOM  N2   TYPE NY13  CHARge -1.028  END ! Nr of Hs =  0
 ATOM  H22  TYPE HY14  CHARge  0.478  END ! Nr of Hs =  0
 ATOM  H21  TYPE HY15  CHARge  0.457  END ! Nr of Hs =  0
 ATOM  N3   TYPE NY16  CHARge -0.799  END ! Nr of Hs =  1
 ATOM  C4   TYPE CY17  CHARge  0.907  END ! Nr of Hs =  0
 ATOM  N1   TYPE NY24  CHARge -0.610  END ! Nr of Hs =  1
 ATOM  C5   TYPE CY25  CHARge -0.403  END ! Nr of Hs =  0
 ATOM  N4   TYPE NY26  CHARge -0.863  END ! Nr of Hs =  0
 ATOM  H41  TYPE HY27  CHARge  0.403  END ! Nr of Hs =  0
 ATOM  H42  TYPE HY28  CHARge  0.440  END ! Nr of Hs =  0
 ATOM  H1   TYPE HY31  CHARge  0.399  END ! Nr of Hs =  0
!Charge of the group: 0.753
 BOND  C6   H6       BOND  C6   N1       BOND  C6   C5       BOND  C2   N2
 BOND  C2   N3       BOND  C2   N1       BOND  N2   H22      BOND  N2   H21
 BOND  N3   C4       BOND  C4   C5       BOND  C4   N4       BOND  N1   H1
 BOND  N4   H41      BOND  C1'  C5       BOND  N4   H42


!Ribose
    BOND P    O1P              BOND P    O2P              BOND P    O5'
    BOND O5'  C5'              BOND C5'  C4'              BOND C4'  O4'
    BOND C4'  C3'              BOND O4'  C1'
    BOND C1'  C2'              BOND C2'  C3'         BOND C3'  O3'
    BOND C2'  O2' BOND O2'  HO2'     BOND C5'  H5'
    BOND C5'  H5''   BOND C3'  H3'     BOND C2'  H2'
    BOND C1'  H1'   BOND C4'  H4'

 { Note: edit these DIHEdrals if necessary }
```

```
  DIHEdral  N3   C2   N2   H22 ! flat ? (0 degrees = cis)    -0.22
  DIHEdral  N3   C2   N2   H21 ! flat ? (180 degrees = trans)   181.56
  DIHEdral  N1   C2   N2   H22 ! flat ? (180 degrees = trans)   180.37
  DIHEdral  N1   C2   N2   H21 ! flat ? (0 degrees = cis)     2.16
  DIHEdral  N3   C4   N4   H42 ! flat ? (0 degrees = cis)    -1.69
  DIHEdral  C5   C4   N4   H42 ! flat ? (180 degrees = trans)   179.27

 { Note: edit these IMPRopers if necessary }
 IMPRoper  C6   H6   N1   C5  ! chirality or flatness improper    -0.54
 IMPRoper  C2   N2   N3   N1  ! chirality or flatness improper     0.35
 IMPRoper  N2   C2   H22  H21 ! chirality or flatness improper     0.94
 IMPRoper  C4   N3   C5   N4  ! chirality or flatness improper     0.57
 IMPRoper  N1   C6   C2   H1  ! chirality or flatness improper     0.77
 IMPRoper  C5   C1'  C6   C4  ! chirality or flatness improper    -2.66
 IMPRoper  N4   C4   H41  H42 ! chirality or flatness improper    -7.15
!Ribose
 IMPRoper     H1'  C2'  O4'  C5  !C1'
 IMPRoper     H2'  C3'  C1'  O2' !C2'
 IMPRoper     H3'  C4'  C2'  O3' !C3'
 IMPRoper     H4'  C5'  C3'  O4' !C4'
 IMPRoper     H5'  O5'  H5'' C4' !C5'

END { RESIdue DAP  }

RESIdue 4AP

 { Note: electrostatics should normally not be used in }
 { crystallographic refinement since it can produce }
 { artefacts. For this reason, all charges are set to }
 { zero by default. Edit them if necessary }
GROUp
 ATOM P    TYPE=P     CHARGE=1.20    END
 ATOM O1P  TYPE=O1P   CHARGE=-0.40   END
 ATOM O2P  TYPE=O2P   CHARGE=-0.40   END
 ATOM O5'  TYPE=O5R   CHARGE=-0.36   END
!Charge of the group: 0.04
GROUp
 ATOM C5'  TYPE=C5R   CHARGE=-0.070  END
 ATOM H5'  TYPE=H     CHARGE=0.035   END
 ATOM H5'' TYPE=H     CHARGE=0.035   END
!Charge of the group: 0.00
GROUp
 ATOM C4'  TYPE=C4R   CHARGE=0.065   END
 ATOM H4'  TYPE=H     CHARGE=0.035   END
 ATOM O4'  TYPE=O4R   CHARGE=-0.30   END
 ATOM C1'  TYPE=C1R   CHARGE=0.238   END !increased from 0.165 by lsd
 ATOM H1'  TYPE=H     CHARGE=0.035   END
!Charge of the group: 0.073 is complmentary to -0.073 of 4AP
GROUP
 ATOM C2'  TYPE=C2R   CHARGE=0.115   END
 ATOM H2'  TYPE=H     CHARGE=0.035   END
 ATOM O2'  TYPE=O2R   CHARGE=-0.40   END
 ATOM HO2' TYPE=HO    CHARGE=0.25    END
!Charge of the group: 0.00
GROUp
 ATOM  C3'  TYPE=C3R  CHARGE=-0.035 END
```

```
 ATOM  H3'  TYPE=H    CHARGE=0.035  END
!Charge of the group: 0.00
GROUp
 ATOM  O3'  TYPE=O3R  CHARGE=-0.36  END
!Charge of the group: -0.36

!4AP-base
GROUp
 ATOM  C5   TYPE CX10  CHARge -0.267  END ! Nr of Hs =  0
 ATOM  H5   TYPE HX11  CHARge  0.161  END ! Nr of Hs =  0
 ATOM  C3   TYPE CX12  CHARge -0.269  END ! Nr of Hs =  0
 ATOM  C6a  TYPE CX13  CHARge -0.239  END ! Nr of Hs =  0
 ATOM  C6   TYPE CX20  CHARge  0.095  END ! Nr of Hs =  0
 ATOM  C4   TYPE CX23  CHARge  0.434  END ! Nr of Hs =  0
 ATOM  C2a  TYPE CX24  CHARge -0.038  END ! Nr of Hs =  0
 ATOM  N4   TYPE NX25  CHARge -0.824  END ! Nr of Hs =  0
 ATOM  H41  TYPE HX26  CHARge  0.348  END ! Nr of Hs =  0
 ATOM  H42  TYPE HX27  CHARge  0.362  END ! Nr of Hs =  0
 ATOM  N1   TYPE NX28  CHARge -0.642  END ! Nr of Hs =  0
 ATOM  C2   TYPE CX29  CHARge  0.702  END ! Nr of Hs =  0
 ATOM  C7   TYPE CX30  CHARge  0.663  END ! Nr of Hs =  0
 ATOM  O7   TYPE OX31  CHARge -0.523  END ! Nr of Hs =  0
 ATOM  O2   TYPE OX32  CHARge -0.551  END ! Nr of Hs =  0
 ATOM  H1   TYPE HX33  CHARge  0.367  END ! Nr of Hs =  0
 ATOM  H3   TYPE HX34  CHARge  0.148  END ! Nr of Hs =  0


 BOND  C1'  C6
 BOND  C5   H5      BOND  C5   C6      BOND  C5   C4      BOND  C3   C4
 BOND  C3   C2a     BOND  C3   H3      BOND  C6a  C6      BOND  C6a  C2a
 BOND  C6a  C7      BOND  C4   N4      BOND  C2a  C2      BOND  N4   H41
 BOND  N4   H42     BOND  N1   C2      BOND  N1   C7      BOND  N1   H1
 BOND  C2   O2     BOND  C7   O7
!Ribose
     BOND P    O1P              BOND P    O2P              BOND P    O5'
     BOND O5'  C5'              BOND C5'  C4'              BOND C4'  O4'
     BOND C4'  C3'              BOND O4'  C1'
     BOND C1'  C2'              BOND C2'  C3'         BOND C3'  O3'
     BOND C2'  O2' BOND O2'  HO2'     BOND C5'  H5'
     BOND C5'  H5''   BOND C3'  H3'     BOND C2'  H2'
     BOND C1'  H1'   BOND C4'  H4'


 { Note: edit these IMPRopers if necessary }
 IMPRoper  C5   H5   C6   C4  ! chirality or flatness improper    -0.31
 IMPRoper  C3   C4   C2a  H3  ! chirality or flatness improper    -0.22
 IMPRoper  C6a  C6   C2a  C7  ! chirality or flatness improper     0.47
 IMPRoper  C4   C5   C3   N4  ! chirality or flatness improper    -1.53
 IMPRoper  C2a  C3   C6a  C2  ! chirality or flatness improper    -0.17
 IMPRoper  N4   C4   H41  H42 ! chirality or flatness improper   -20.06
 IMPRoper  C5   N4   C4   H41
 IMPRoper  N1   C2   C7   H1  ! chirality or flatness improper     0.75
 IMPRoper  C2   C2a  N1   O2  ! chirality or flatness improper    -0.01
 IMPRoper  C7   C6a  N1   O7  ! chirality or flatness improper    -0.09
!Improper to keep both rings parallel
! IMPRoper  N1   C2a  C6a  C5
```

```
! IMPRoper  N1   C2a  C6a  C4
! IMPRoper  C2   C2a  C6a  C6
! IMPRoper  C3   C2a  C6a  C7


!Ribose
 IMPRoper     H1’  C2’  O4’  C6  !C1’
 IMPRoper     H2’  C3’  C1’  O2’ !C2’
 IMPRoper     H3’  C4’  C2’  O3’ !C3’
 IMPRoper     H4’  C5’  C3’  O4’ !C4’
 IMPRoper     H5’  O5’  H5’’ C4’ !C5’

END { RESIdue 4AP  }


RESIdue 6HQ

 { Note: electrostatics should normally not be used in }
 { crystallographic refinement since it can produce }
 { artefacts. For this reason, all charges are set to }
 { zero by default. Edit them if necessary }
GROUp
 ATOM P    TYPE=P     CHARGE=1.20    END
 ATOM O1P  TYPE=O1P   CHARGE=-0.40   END
 ATOM O2P  TYPE=O2P   CHARGE=-0.40   END
 ATOM O3g  TYPE OQ21  CHARGE=0.40    END
GROUp
 ATOM  C1g  TYPE CQ1   CHARGE= 0.365  END ! inc. from 0.095 by lsd
 ATOM  N1   TYPE NQ2   CHARGE=-0.05   END ! inc. from -0.102 by lsd
 ATOM  C8a  TYPE CQ3   CHARGE= 0.167  END
 ATOM  C4a  TYPE CQ4   CHARGE= 0.162  END
 ATOM  C4   TYPE CQ5   CHARGE=-0.015  END
 ATOM  C3   TYPE CQ6   CHARGE=-0.207  END
 ATOM  C2   TYPE CQ7   CHARGE= 0.151  END
 ATOM  H2   TYPE HQ8   CHARGE= 0.134  END
 ATOM  C8   TYPE CQ9   CHARGE=-0.195  END
 ATOM  C7   TYPE CQ10  CHARGE=-0.120  END
 ATOM  C6   TYPE CQ11  CHARGE= 0.472  END
 ATOM  C5   TYPE CQ12  CHARGE=-0.421  END
 ATOM  O6   TYPE OQ13  CHARGE=-0.586  END
 ATOM  H4   TYPE HQ14  CHARGE= 0.142  END
 ATOM  H3   TYPE HQ15  CHARGE= 0.161  END
 ATOM  H8   TYPE HQ16  CHARGE= 0.139  END
 ATOM  H7   TYPE HQ17  CHARGE= 0.176  END
 ATOM  H5   TYPE HQ18  CHARGE= 0.175  END

 ATOM  C2g  TYPE CQ20  CHARGE= 0.271  END
 ATOM  C3g  TYPE CQ22  CHARGE= 0.213  END
 ATOM  H3’’ TYPE HQ23  CHARGE= 0.008  END
 ATOM  H3’  TYPE HQ24  CHARGE= 0.028  END
 ATOM  H2’  TYPE HQ25  CHARGE= 0.010  END
 ATOM  H1’’ TYPE HQ26  CHARGE= 0.077  END
 ATOM  H1’  TYPE HQ27  CHARGE= 0.085  END
 ATOM  H6   TYPE HQ30  CHARGE= 0.457  END

ATOM  O2g  TYPE OQ19  CHARGE=-0.400  END ! inc. from -0.674 by lsd
```

```
BOND  C1g  N1       BOND  C1g  C2g      BOND  C1g  H1''     BOND  C1g  H1'
 BOND  N1   C8a      BOND  N1   C2       BOND  C8a  C4a      BOND  C8a  C8
BOND  C4a  C4       BOND  C4a  C5       BOND  C4   C3       BOND  C4   H4
BOND  C3   C2       BOND  C3   H3       BOND  C2   H2       BOND  C8   C7
BOND  C8   H8       BOND  C7   C6       BOND  C7   H7       BOND  C6   C5
BOND  C6   O6       BOND  C5   H5       BOND  O6   H6       BOND  O2g  C2g
BOND  C2g  C3g      BOND  C2g  H2'      BOND  O3g  C3g
BOND  C3g  H3''     BOND  C3g  H3'
BOND  P    O1P      BOND  P    O2P      BOND  P    O3g

{ Note: edit these DIHEdrals if necessary }
! DIHEdral  H1'  C1g   N1   C2  ! flexible dihedral ???    99.41
! DIHEdral  H1'  C1g   C2g  H2' ! flat ? (180 degrees = trans)   189.55
! DIHEdral  C1g   N1   C8a  C4a ! flat ? (180 degrees = trans)   180.71
! DIHEdral  C1g   N1   C8a  C8  ! flat ? (0 degrees = cis)     1.29
  DIHEdral  C2    N1   C8a  C4a ! flat ? (0 degrees = cis)     2.40
  DIHEdral  C2    N1   C8a  C8  ! flat ? (180 degrees = trans)   182.98
! DIHEdral  C1g   N1   C2   C3  ! flat ? (180 degrees = trans)   179.55
! DIHEdral  C1g   N1   C2   H2  ! flat ? (0 degrees = cis)    -1.94
  DIHEdral  C8a   N1   C2   C3  ! flat ? (0 degrees = cis)    -2.15
  DIHEdral  C8a   N1   C2   H2  ! flat ? (180 degrees = trans)   176.35
  DIHEdral  N1    C8a  C4a  C4  ! flat ? (0 degrees = cis)    -1.07
  DIHEdral  N1    C8a  C4a  C5  ! flat ? (180 degrees = trans)   179.45
  DIHEdral  C8    C8a  C4a  C4  ! flat ? (180 degrees = trans)   178.37
  DIHEdral  C8    C8a  C4a  C5  ! flat ? (0 degrees = cis)    -1.11
  DIHEdral  N1    C8a  C8   C7 ! flat ? (180 degrees = trans)   180.11
  DIHEdral  N1    C8a  C8   H8  ! flat ? (0 degrees = cis)     1.05
  DIHEdral  C4a   C8a  C8   C7 ! flat ? (0 degrees = cis)     0.70
  DIHEdral  C4a   C8a  C8   H8  ! flat ? (180 degrees = trans)   181.64
  DIHEdral  C8a   C4a  C4   C3  ! flat ? (0 degrees = cis)    -0.51
  DIHEdral  C8a   C4a  C4   H4  ! flat ? (180 degrees = trans)   180.12
  DIHEdral  C5    C4a  C4   C3  ! flat ? (180 degrees = trans)   178.97
  DIHEdral  C5    C4a  C4   H4  ! flat ? (0 degrees = cis)    -0.41
  DIHEdral  C8a   C4a  C5   C6  ! flat ? (0 degrees = cis)     0.57
  DIHEdral  C8a   C4a  C5   H5  ! flat ? (180 degrees = trans)   180.28
  DIHEdral  C4    C4a  C5   C6  ! flat ? (180 degrees = trans)   181.10
  DIHEdral  C4    C4a  C5   H5  ! flat ? (0 degrees = cis)     0.81
  DIHEdral  C4a   C4   C3   C2  ! flat ? (0 degrees = cis)     0.82
  DIHEdral  C4a   C4   C3   H3  ! flat ? (180 degrees = trans)   181.07
  DIHEdral  H4    C4   C3   C2  ! flat ? (180 degrees = trans)   180.19
  DIHEdral  H4    C4   C3   H3  ! flat ? (0 degrees = cis)     0.44
  DIHEdral  C4    C3   C2   N1  ! flat ? (0 degrees = cis)     0.50
  DIHEdral  C4    C3   C2   H2  ! flat ? (180 degrees = trans)   182.09
  DIHEdral  H3    C3   C2   N1  ! flat ? (180 degrees = trans)   180.26
  DIHEdral  H3    C3   C2   H2  ! flat ? (0 degrees = cis)     1.84
  DIHEdral  C8a   C8   C7   C6  ! flat ? (0 degrees = cis)     0.26
  DIHEdral  C8a   C8   C7   H7  ! flat ? (180 degrees = trans)   180.24
  DIHEdral  H8    C8   C7   C6  ! flat ? (180 degrees = trans)   179.35
  DIHEdral  H8    C8   C7   H7  ! flat ? (0 degrees = cis)    -0.67
  DIHEdral  C8    C7   C6   C5  ! flat ? (0 degrees = cis)    -0.82
  DIHEdral  C8    C7   C6   O6  ! flat ? (180 degrees = trans)   179.56
  DIHEdral  H7    C7   C6   C5  ! flat ? (180 degrees = trans)   179.20
  DIHEdral  H7    C7   C6   O6  ! flat ? (0 degrees = cis)    -0.42
  DIHEdral  C7    C6   C5   C4a ! flat ? (0 degrees = cis)     0.38
  DIHEdral  C7    C6   C5   H5  ! flat ? (180 degrees = trans)   180.69
```

```
  DIHEdral  O6    C6   C5   C4a ! flat ? (180 degrees = trans)   179.97
  DIHEdral  O6    C6   C5   H5  ! flat ? (0 degrees = cis)     0.27
  DIHEdral  C7    C6   O6   H6 ! flat ? (180 degrees = trans)   179.93
  DIHEdral  C5    C6   O6   H6 ! flat ? (0 degrees = cis)     0.33
! DIHEdral  C1g   C2g  C3g  O3g ! flexible dihedral ???    60.68
! DIHEdral  C1g   C2g  C3g  H3'' ! flexible dihedral ???   -61.16
  DIHEdral  C1g   C2g  C3g  H3' ! flat ? (180 degrees = trans)   181.08
  DIHEdral  O2g   C2g  C3g  O3g ! flat ? (180 degrees = trans)   180.00
! DIHEdral  O2g   C2g  C3g  H3'' ! flexible dihedral ???    58.16
! DIHEdral  O2g   C2g  C3g  H3' ! flexible dihedral ???   -59.60
! DIHEdral  H2'   C2g  C3g  O3g ! flexible dihedral ???   -58.01
  DIHEdral  H2'   C2g  C3g  H3'' ! flat ? (180 degrees = trans)   180.15
! DIHEdral  H2'   C2g  C3g  H3' ! flexible dihedral ???    62.39


 { Note: edit these IMPRopers if necessary }
! IMPRoper  C1g  N1   C2g  H1'' ! chirality or flatness improper    33.17
! IMPRoper  N1   C1g  C8a  C2  ! chirality or flatness improper    -0.94
 IMPRoper  C8a  N1   C4a  C8  ! chirality or flatness improper     0.33
 IMPRoper  C4a  C8a  C4   C5 ! chirality or flatness improper     0.30
 IMPRoper  C4   C4a  C3   H4 ! chirality or flatness improper     0.33
 IMPRoper  C3   C4   C2   H3 ! chirality or flatness improper     0.13
 IMPRoper  C2   N1   C3   H2  ! chirality or flatness improper    -0.83
 IMPRoper  C8   C8a  C7   H8 ! chirality or flatness improper     0.49
 IMPRoper  C7   C8   C6   H7 ! chirality or flatness improper    -0.01
 IMPRoper  C6   C7   C5   O6 ! chirality or flatness improper     0.22
 IMPRoper  C5   C4a  C6   H5 ! chirality or flatness improper    -0.16
 IMPRoper  C2g  C1g  O2g  C3g ! chirality or flatness improper   -35.35
 IMPRoper  C3g  C2g  O3g  H3'' ! chirality or flatness improper    29.83

END { RESIdue 6HQ   }




RESIdue HCF

 { Note: electrostatics should normally not be used in }
 { crystallographic refinement since it can produce }
 { artefacts. For this reason, all charges are set to }
 { zero by default. Edit them if necessary }
GROUp
 ATOM P    TYPE=P     CHARGE=1.20    END
 ATOM O1P  TYPE=O1P   CHARGE=-0.40   END
 ATOM O2P  TYPE=O2P   CHARGE=-0.40   END
 ATOM O5'  TYPE=O5R   CHARGE=-0.36   END
GROUp
 ATOM C5'  TYPE=C5R   CHARGE=-0.070  END
 ATOM H5'  TYPE=H     CHARGE=0.035   END
 ATOM H5'' TYPE=H     CHARGE=0.035   END
GROUp
 ATOM C4'  TYPE=C4R   CHARGE=0.065   END
 ATOM H4'  TYPE=H     CHARGE=0.035   END
 ATOM O4'  TYPE=O4R   CHARGE=-0.30   END
 ATOM C1'  TYPE=C1R   CHARGE=0.390   END !increased from 0.165 by lsd
```

```
 ATOM H1'' TYPE=HZ29  CHARGE=0.022   END !changed from 0.035 by lsd
!Charge of the group: 0.212
GROUP
 ATOM C2'  TYPE=C2R   CHARGE=0.115   END
 ATOM H2'  TYPE=H     CHARGE=0.035   END
 ATOM O2'  TYPE=O2R   CHARGE=-0.40   END
 ATOM HO2' TYPE=HO    CHARGE=0.25    END
!Charge of the group: 0.00
GROUp
 ATOM  C3'  TYPE=C3R  CHARGE=-0.035 END
 ATOM  H3'  TYPE=H    CHARGE=0.035  END
!Charge of the group: 0.00
GROUp
 ATOM  O3'  TYPE=O3R  CHARGE=-0.36  END
!Charge of the group: -0.36
!HCF-Base
GROUp
 ATOM  C2   TYPE=CZ1   CHARge= 0.390   END ! Nr of Hs =  0
 ATOM  C3   TYPE=CZ2   CHARge= -0.262  END ! Nr of Hs =  0
 ATOM  C4   TYPE=CZ3   CHARge= -0.101  END ! Nr of Hs =  0
 ATOM  C10  TYPE=CZ4   CHARge= -0.065  END ! Nr of Hs =  0
 ATOM  C11  TYPE=CZ5   CHARge= 0.066   END ! Nr of Hs =  0
 ATOM  C1   TYPE=CZ6   CHARge= -0.308  END ! Nr of Hs =  0
 ATOM  C13  TYPE=CZ7   CHARge= 0.096   END ! Nr of Hs =  0
 ATOM  C12  TYPE=CZ8   CHARge= -0.025  END ! Nr of Hs =  0
 ATOM  C9   TYPE=CZ9   CHARge= 0.122   END ! Nr of Hs =  0
 ATOM  C5   TYPE=CZ10  CHARge= -0.154  END ! Nr of Hs =  0
 ATOM  C6   TYPE=CZ11  CHARge= -0.108  END ! Nr of Hs =  0
 ATOM  C7   TYPE=CZ12  CHARge= -0.032  END ! Nr of Hs =  0
 ATOM  C8   TYPE=CZ13  CHARge= -0.157  END ! Nr of Hs =  0
 ATOM  O2   TYPE=OZ14  CHARge= -0.343  END ! reduced charge from -0.361 to zero charge of HCF-base + C1'-C4'-group
 ATOM  C14  TYPE=CZ15  CHARge= 0.707   END ! Nr of Hs =  0
 ATOM  O142 TYPE=OZ16  CHARge= -0.575  END ! Nr of Hs =  0
 ATOM  O141 TYPE=OZ17  CHARge= -0.625  END ! Nr of Hs =  0
 ATOM  H4   TYPE=HZ18  CHARge= 0.117   END ! Nr of Hs =  0
 ATOM  H1   TYPE=HZ19  CHARge= 0.141   END ! Nr of Hs =  0
 ATOM  H8   TYPE=HZ20  CHARge= 0.124   END ! Nr of Hs =  0
 ATOM  H91  TYPE=HZ21  CHARge= 0.009   END ! Nr of Hs =  0
 ATOM  H92  TYPE=HZ22  CHARge= 0.010   END ! Nr of Hs =  0
 ATOM  H5   TYPE=HZ23  CHARge= 0.103   END ! Nr of Hs =  0
 ATOM  H6   TYPE=HZ24  CHARge= 0.119   END ! Nr of Hs =  0
 ATOM  H3   TYPE=HZ25  CHARge= 0.110   END ! Nr of Hs =  0
 ATOM  H141 TYPE=HZ43  CHARge= 0.429   END ! Nr of Hs =  0


!HCF-Base
 BOND  C2   C3      BOND  C2   C1      BOND  C2   O2      BOND  C3   C4
 BOND  C3   H3      BOND  C4   C10     BOND  C4   H4      BOND  C10  C11
 BOND  C10  C13     BOND  C11  C1      BOND  C11  C9      BOND  C1   H1
 BOND  C13  C12     BOND  C13  C5      BOND  C12  C9      BOND  C12  C8
 BOND  C9   H91     BOND  C9   H92     BOND  C5   C6      BOND  C5   H5
 BOND  C6   C7      BOND  C6   H6      BOND  C7   C8      BOND  C7   C14
 BOND  C8   H8      BOND  C14  O142    BOND  C14  O141    BOND  O141 H141
!Ribose
    BOND P    O1P              BOND P    O2P              BOND P    O5'
    BOND O5'  C5'              BOND C5'  C4'              BOND C4'  O4'
```

```
    BOND C4'  C3'               BOND O4'  C1'               BOND C1'  O2
    BOND C1'  C2'               BOND C2'  C3'          BOND C3'  O3'
    BOND C2'  O2' BOND O2'  HO2'     BOND C5'  H5'
    BOND C5'  H5''   BOND C3'  H3'     BOND C2'  H2'
    BOND C1'  H1''   BOND C4'  H4'

 { Note: edit these IMPRopers if necessary }
!HCF-Base
 IMPRoper  C2   C3   C1   O2   ! chirality or flatness improper    -0.17
 IMPRoper  C3   C2   C4   H3   ! chirality or flatness improper     0.33
 IMPRoper  C4   C3   C10  H4   ! chirality or flatness improper     0.23
 IMPRoper  C10  C4   C11  C13  ! chirality or flatness improper     0.07
 IMPRoper  C11  C10  C1   C9   ! chirality or flatness improper    -0.07
 IMPRoper  C1   C2   C11  H1   ! chirality or flatness improper     0.23
 IMPRoper  C13  C10  C12  C5   ! chirality or flatness improper    -0.01
 IMPRoper  C12  C13  C9   C8   ! chirality or flatness improper    -0.01
 IMPRoper  C9   C11  C12  H91  ! chirality or flatness improper   -28.83
 IMPRoper  C5   C13  C6   H5   ! chirality or flatness improper     0.01
 IMPRoper  C6   C5   C7   H6   ! chirality or flatness improper    -0.04
 IMPRoper  C7   C6   C8   C14  ! chirality or flatness improper    -0.03
 IMPRoper  C8   C12  C7   H8   ! chirality or flatness improper    -0.02
 IMPRoper  C14  C7   O142 O141 ! chirality or flatness improper     0.00
!Ribose
 IMPRoper     O2   C2'  O4'  H1''!C1'
 IMPRoper     H2'  C3'  C1'  O2' !C2'
 IMPRoper     H3'  C4'  C2'  O3' !C3'
 IMPRoper     H4'  C5'  C3'  O4' !C4'
 IMPRoper     H5'  O5'  H5'' C4' !C5'


END { RESIdue HCF }

!********************* end of change by lsd - HCF *********************


   RESIdue GUA
     GROUp
     ATOM P    TYPE=P     CHARGE=1.20    END
     ATOM O1P  TYPE=O1P   CHARGE=-0.40   END
     ATOM O2P  TYPE=O2P   CHARGE=-0.40   END
     ATOM O5'  TYPE=O5R   CHARGE=-0.36   END
    GROUp
     ATOM C5'  TYPE=C5R   CHARGE=-0.070  END
     ATOM H5'  TYPE=H     CHARGE=0.035   END !JPR
     ATOM H5'' TYPE=H     CHARGE=0.035   END !JPR
    GROUp
     ATOM C4'  TYPE=C4R   CHARGE=0.065  END
     ATOM H4'  TYPE=H     CHARGE=0.035   END !JPR
     ATOM O4'  TYPE=O4R   CHARGE=-0.30   END
     ATOM C1'  TYPE=C1R   CHARGE=0.165   END !JPR
     ATOM H1'  TYPE=H     CHARGE=0.035   END !JPR

   ! Insert Base
    GROUp
     ATOM N9   TYPE=N9G   CHARGE=-0.19   END
     ATOM C4   TYPE=C4G   CHARGE=0.19    EXCLusion=( N1 )   END
```

```
 GROUp
  ATOM N3   TYPE=N3G   CHARGE=-0.35   EXCLusion=( C6 )   END
  ATOM C2   TYPE=C2G   CHARGE=0.35    EXCLusion=( C5 )   END
 GROUp
  ATOM N2   TYPE=N2G   CHARGE=-0.42   END
  ATOM H21  TYPE=H2    CHARGE=0.21    END
  ATOM H22  TYPE=H2    CHARGE=0.21    END
 GROUp
  ATOM N1   TYPE=NNA   CHARGE=-0.26   END
  ATOM H1   TYPE=HN    CHARGE=0.26    END
 GROUp
  ATOM C6   TYPE=C6G   CHARGE=0.30    END
  ATOM O6   TYPE=O6G   CHARGE=-0.30   END
 GROUp
  ATOM C5   TYPE=C5G   CHARGE=0.02    END
  ATOM N7   TYPE=N7G   CHARGE=-0.25   END
  ATOM C8   TYPE=C8G   CHARGE=0.145   END
  ATOM H8   TYPE=H     CHARGE=0.035   END

!

 GROUP
  ATOM C2'  TYPE=C2R   CHARGE=0.115   END
  ATOM H2'  TYPE=H     CHARGE=0.035   END
  ATOM O2'  TYPE=O2R   CHARGE=-0.40   END
  ATOM HO2' TYPE=HO    CHARGE=0.25    END
 GROUP
  ATOM C3'  TYPE=C3R    CHARGE=-0.035 END
  ATOM H3'  TYPE=H      CHARGE=0.035  END
 GROUP
  ATOM O3'  TYPE=O3R   CHARGE=-0.36   END

 BOND P    O1P         BOND P    O2P           BOND P    O5'

 BOND O5'  C5'         BOND C5'  C4'           BOND C4'  O4'
 BOND C4'  C3'         BOND O4'  C1'           BOND C1'  N9
 BOND C1'  C2'         BOND N9   C4            BOND N9   C8
 BOND C4   N3          BOND C4   C5            BOND N3   C2
 BOND C2   N2          BOND C2   N1            BOND N2   H21

 BOND N2   H22         BOND N1   H1            BOND N1   C6
 BOND C6   O6          BOND C6   C5            BOND C5   N7
 BOND N7   C8          BOND C2'  C3'           BOND C3'  O3'
 BOND C2'  O2'         BOND C8   H8

 BOND O2'  HO2'
 BOND C5'  H5'         BOND C5'  H5''          BOND C4'  H4'
 BOND C3'  H3'         BOND C2'  H2'           BOND C1'  H1'
{
 DIHEdral  P    O5'  C5'  C4'        DIHEdral  O5'  C5'  C4'  O4'
 DIHEdral  O5'  C5'  C4'  C3'
}{
 DIHEdral  C3'  C4'  O4'  C1'
 DIHEdral  C4'  O4'  C1'  C2'        DIHEdral  O4'  C1'  C2'  C3'

 DIHEdral  C1'  C2'  C3'  C4'        DIHEdral  O4'  C4'  C3'  O3'
```

```
 DIHEdral  C5’  C4’  C3’  C2’        DIHEdral  O3’  C3’  C2’  O2’
 DIHEdral  O4’  C1’  N9   C4
 DIHEdral  C3’  C2’  O2’  H2’
}

 !
 IMPRoper N3   C2   N2   H21         IMPRoper C1’  C4   C8   N9
 IMPRoper N9   C4   C5   N7          IMPRoper C4   C5   N7   C8
 IMPRoper C5   N7   C8   N9          IMPRoper N7   C8   N9   C4
 IMPRoper C8   N9   C4   C5          IMPRoper N2   N3   N1   C2

 IMPRoper H1   C2   C6   N1          IMPRoper O6   N1   C5   C6
 IMPRoper C4   N3   C2   N1          IMPRoper N3   C2   N1   C6
 IMPRoper C2   N1   C6   C5          IMPRoper N1   C6   C5   C4

 IMPRoper C6   C5   C4   N3          IMPRoper C5   C4   N3   C2
 IMPRoper H22  H21  C2   N2
 IMPRoper H8   N7   N9   C8

!IMPRoper to keep the two purine rings parallel:
 IMPRoper C8   C4   C5   N1          IMPRoper C8   C5   C4   C2
 IMPRoper N3   C4   C5   N7          IMPRoper C6   C5   C4   N9

!RIBOSE IMPROPERS
IMPRoper     H1’  C2’  O4’  N9  !C1’
IMPRoper     H2’  C3’  C1’  O2’ !C2’
IMPRoper     H3’  C4’  C2’  O3’ !C3’
IMPRoper     H4’  C5’  C3’  O4’ !C4’
IMPRoper     H5’  O5’  H5’’ C4’ !C5’



 END {GUA}

! ----------------------------------------------------------------------

RESIdue ADE
GROUp
  ATOM P    TYPE=P     CHARGE=1.20    END
  ATOM O1P  TYPE=O1P   CHARGE=-0.40   END
  ATOM O2P  TYPE=O2P   CHARGE=-0.40   END
  ATOM O5’  TYPE=O5R   CHARGE=-0.36   END
 GROUp
  ATOM C5’  TYPE=C5R   CHARGE=-0.070  END
  ATOM H5’  TYPE=H     CHARGE=0.035   END
  ATOM H5’’ TYPE=H     CHARGE=0.035   END
 GROUp
  ATOM C4’  TYPE=C4R   CHARGE=0.065   END
  ATOM H4’  TYPE=H     CHARGE=0.035   END
  ATOM O4’  TYPE=O4R   CHARGE=-0.30   END
  ATOM C1’  TYPE=C1R   CHARGE=0.165   END
  ATOM H1’  TYPE=H     CHARGE=0.035   END

! Insert Base
 GROUp
  ATOM N9    TYPE=N9A   CHARGE=-0.19   END
```

```
  ATOM C4    TYPE=C4A   CHARGE=0.19    EXCLusion=( N1 )  END
 GROUp
  ATOM N3    TYPE=N3A   CHARGE=-0.26   EXCLusion=( C6 )  END
  ATOM C2    TYPE=C2A   CHARGE=0.225   EXCLusion=( C5 )  END
  ATOM H2    TYPE=H     CHARGE=0.035   END
 GROUp
  ATOM N1    TYPE=NC    CHARGE=-0.28   END
  ATOM C6    TYPE=C6A   CHARGE=0.28    END
 GROUp
  ATOM N6    TYPE=N6A   CHARGE=-0.42   END
  ATOM H61   TYPE=H2    CHARGE=0.21    END
  ATOM H62   TYPE=H2    CHARGE=0.21    END
 GROUp
  ATOM C5    TYPE=C5A   CHARGE=0.02    END
  ATOM N7    TYPE=N7A   CHARGE=-0.25   END
  ATOM C8    TYPE=C8A   CHARGE=0.195   END
  ATOM H8    TYPE=H     CHARGE=0.035   END
! END

 GROUP
  ATOM C2'  TYPE=C2R   CHARGE=0.115   END
  ATOM H2'  TYPE=H     CHARGE=0.035   END
  ATOM O2'  TYPE=O2R   CHARGE=-0.40   END
  ATOM HO2' TYPE=HO    CHARGE=0.25    END
 GROUP
  ATOM C3'  TYPE=C3R   CHARGE=-0.035  END
  ATOM H3'  TYPE=H     CHARGE=0.035   END
 GROUP
  ATOM O3'  TYPE=O3R   CHARGE=-0.36   END

 BOND P    O1P             BOND P    O2P             BOND P    O5'
 BOND O5'  C5'             BOND C5'  C4'             BOND C4'  O4'
 BOND C4'  C3'             BOND O4'  C1'             BOND C1'  N9
 BOND C1'  C2'             BOND N9   C4              BOND N9   C8
 BOND C4   N3              BOND C4   C5              BOND N3   C2
 BOND C2   N1              BOND N1   C6              BOND C6   N6

 BOND N6   H61             BOND N6   H62             BOND C6   C5

 BOND C5   N7              BOND N7   C8              BOND C2'  C3'
 BOND C2'  O2'             BOND C3'  O3'
 BOND C8   H8              BOND C2   H2
 BOND O2'  HO2'
 BOND C5'  H5'        BOND C5'  H5''         BOND C4'  H4'
 BOND C3'  H3'        BOND C2'  H2'          BOND C1'  H1'
{
 DIHEdral P   O5'  C5'  C4'             DIHEdral O5'  C5'  C4'  O4'
 DIHEdral O5'  C5'  C4'  C3'
}{
 DIHEdral C3'  C4'  O4'  C1'
 DIHEdral C4'  O4'  C1'  C2'            DIHEdral O4'  C1'  C2'  C3'

 DIHEdral C1'  C2'  C3'  C4'            DIHEdral O4'  C4'  C3'  O3'
 DIHEdral C5'  C4'  C3'  C2'            DIHEdral O2'  C2'  C3'  O3'
 DIHEdral O4'  C1'  N9   C4
 DIHEdral C3'  C2'  O2'  H2'
```

```
 }

 !
 IMPRoper C5   C6   N6   H61            IMPRoper C1'  C4   C8   N9
 IMPRoper N9   C4   C5   N7             IMPRoper C4   C5   N7   C8
 IMPRoper C5   N7   C8   N9             IMPRoper N7   C8   N9   C4
 IMPRoper C8   N9   C4   C5             IMPRoper N6   N1   C5   C6
 IMPRoper H62  C6   H61  N6             IMPRoper C4   N3   C2   N1
 IMPRoper N3   C2   N1   C6             IMPRoper C2   N1   C6   C5
 IMPRoper N1   C6   C5   C4             IMPRoper C6   C5   C4   N3
 IMPRoper C5   C4   N3   C2
 IMPRoper H2   N1   N3   C2             IMPRoper H8   N7   N9   C8
 ! IMPRoper to keep the two purine rings parallel:
 IMPRoper C8   C4   C5   N1             IMPRoper C8   C5   C4   C2
 IMPRoper N3   C4   C5   N7             IMPRoper C6   C5   C4   N9


!RIBOSE IMPROPERS
IMPRoper     H1'  C2'  O4'  N9  !C1'
IMPRoper     H2'  C3'  C1'  O2' !C2'
IMPRoper     H3'  C4'  C2'  O3' !C3'
IMPRoper     H4'  C5'  C3'  O4' !C4'
IMPRoper     H5'  O5'  H5'' C4' !C5'



 END {ADE}

! ----------------------------------------------------------------------

RESIdue PUR
GROUp
  ATOM P    TYPE=P     CHARGE=1.20    END
  ATOM O1P  TYPE=O1P   CHARGE=-0.40   END
  ATOM O2P  TYPE=O2P   CHARGE=-0.40   END
  ATOM O5'  TYPE=O5R   CHARGE=-0.36   END
 GROUp
  ATOM C5'  TYPE=C5R   CHARGE=-0.070  END
  ATOM H5'  TYPE=H     CHARGE=0.035   END
  ATOM H5'' TYPE=H     CHARGE=0.035   END
 GROUp
  ATOM C4'  TYPE=C4R   CHARGE=0.065   END
  ATOM H4'  TYPE=H     CHARGE=0.035   END
  ATOM O4'  TYPE=O4R   CHARGE=-0.30   END
  ATOM C1'  TYPE=C1R   CHARGE=0.165   END
  ATOM H1'  TYPE=H     CHARGE=0.035   END

! Insert Base
 GROUp
  ATOM N9    TYPE=N9P   CHARGE=-0.19   END
  ATOM C4    TYPE=C4P   CHARGE=0.19    EXCLusion=( N1 )  END
 GROUp
  ATOM N3    TYPE=N3P   CHARGE=-0.26   EXCLusion=( C6 )  END
  ATOM C2    TYPE=C2P   CHARGE=0.225   EXCLusion=( C5 )  END
  ATOM H2    TYPE=H     CHARGE=0.035   END
 GROUp
```

```
  ATOM N1    TYPE=NC    CHARGE=-0.28   END
  ATOM C6    TYPE=C6P   CHARGE=0.28    END
  ATOM H6    TYPE=H     CHARge= 0.035  END
 GROUp
  ATOM C5    TYPE=C5P   CHARGE=0.02    END
  ATOM N7    TYPE=N7P   CHARGE=-0.25   END
  ATOM C8    TYPE=C8P   CHARGE=0.195   END
  ATOM H8    TYPE=H     CHARGE=0.035   END
! END

 GROUP
  ATOM C2'  TYPE=C2R   CHARGE=0.115   END
  ATOM H2'  TYPE=H     CHARGE=0.035   END
  ATOM O2'  TYPE=O2R   CHARGE=-0.40   END
  ATOM HO2' TYPE=HO    CHARGE=0.25    END
 GROUP
  ATOM C3'  TYPE=C3R   CHARGE=-0.035  END
  ATOM H3'  TYPE=H     CHARGE=0.035   END
 GROUP
  ATOM O3'  TYPE=O3R   CHARGE=-0.36   END

 BOND P    O1P             BOND P    O2P             BOND P    O5'
 BOND O5'  C5'             BOND C5'  C4'             BOND C4'  O4'
 BOND C4'  C3'             BOND O4'  C1'             BOND C1'  N9
 BOND C1'  C2'             BOND N9   C4              BOND N9   C8
 BOND C4   N3              BOND C4   C5              BOND N3   C2
 BOND C2   N1              BOND N1   C6              BOND C6   H6

 BOND C6   C5

 BOND C5   N7              BOND N7   C8              BOND C2'  C3'
 BOND C2'  O2'             BOND C3'  O3'
 BOND C8   H8              BOND C2   H2
 BOND O2'  HO2'
 BOND C5'  H5'        BOND C5'  H5''         BOND C4'  H4'
 BOND C3'  H3'        BOND C2'  H2'          BOND C1'  H1'
{
 DIHEdral P   O5'  C5'  C4'             DIHEdral O5'  C5'  C4'  O4'
 DIHEdral O5'  C5'  C4'  C3'
}{
 DIHEdral C3'  C4'  O4'  C1'
 DIHEdral C4'  O4'  C1'  C2'            DIHEdral O4'  C1'  C2'  C3'

 DIHEdral C1'  C2'  C3'  C4'            DIHEdral O4'  C4'  C3'  O3'
 DIHEdral C5'  C4'  C3'  C2'            DIHEdral O2'  C2'  C3'  O3'
 DIHEdral O4'  C1'  N9   C4
 DIHEdral C3'  C2'  O2'  H2'
 }

 !
 IMPRoper H6   N1   C5   C6             IMPRoper C1'  C4   C8   N9
 IMPRoper N9   C4   C5   N7             IMPRoper C4   C5   N7   C8
 IMPRoper C5   N7   C8   N9             IMPRoper N7   C8   N9   C4
 IMPRoper C8   N9   C4   C5             IMPRoper N6   N1   C5   C6
 IMPRoper C4   N3   C2   N1
 IMPRoper N3   C2   N1   C6             IMPRoper C2   N1   C6   C5
```

```
 IMPRoper N1   C6   C5   C4              IMPRoper C6   C5   C4   N3
 IMPRoper C5   C4   N3   C2
 IMPRoper H2   N1   N3   C2              IMPRoper H8   N7   N9   C8
 ! IMPRoper to keep the two purine rings parallel:
 IMPRoper C8   C4   C5   N1              IMPRoper C8   C5   C4   C2
 IMPRoper N3   C4   C5   N7              IMPRoper C6   C5   C4   N9


!RIBOSE IMPROPERS
IMPRoper     C2'  C3'  C1'  O2'
IMPRoper     H1'  C2'  O4'  N9  !C1'
IMPRoper     H2'  C3'  C1'  O2' !C2'
IMPRoper     H3'  C4'  C2'  O3' !C3'
IMPRoper     H4'  C5'  C3'  O4' !C4'
IMPRoper     H5'  O5'  H5'' C4' !C5'



 END {PUR}

! ---------------------------------------------------------------------

RESIdue ABA
  GROUp
  ATOM P    TYPE=P     CHARGE=1.20    END
  ATOM O1P  TYPE=O1P   CHARGE=-0.40   END
  ATOM O2P  TYPE=O2P   CHARGE=-0.40   END
  ATOM O5'  TYPE=O5R   CHARGE=-0.36   END
 GROUp
  ATOM C5'  TYPE=C5R   CHARGE=-0.070  END
  ATOM H5'  TYPE=H     CHARGE=0.035   END
  ATOM H5'' TYPE=H     CHARGE=0.035   END
 GROUp
  ATOM C4'  TYPE=C4R   CHARGE=0.065   END
  ATOM H4'  TYPE=H     CHARGE=0.035   END
  ATOM O4'  TYPE=O4R   CHARGE=-0.30   END
  ATOM C1'  TYPE=C1R   CHARGE=0.165   END
  ATOM H1'  TYPE=H     CHARGE=0.018   END
  ATOM H1'' TYPE=H     CHARGE=0.017   END

GROUP
  ATOM C2'  TYPE=C2R   CHARGE=0.115   END
  ATOM H2'  TYPE=H     CHARGE=0.035   END
  ATOM O2'  TYPE=O2R   CHARGE=-0.40   END
  ATOM HO2' TYPE=HO    CHARGE=0.25    END
 GROUP
  ATOM C3'  TYPE=C3R   CHARGE=-0.035  END
  ATOM H3'  TYPE=H     CHARGE=0.035   END
 GROUP
  ATOM O3'  TYPE=O3R   CHARGE=-0.36   END

 BOND P    O1P             BOND P    O2P             BOND P    O5'
 BOND O5'  C5'             BOND C5'  C4'             BOND C4'  O4'
 BOND C4'  C3'             BOND O4'  C1'
 BOND C1'  C2'             BOND C2'  C3'
 BOND C3'  O3'             BOND C2'  O2'
```

```
 BOND O2'  HO2'       BOND C1'  H1''
 BOND C5'  H5'        BOND C5'  H5''        BOND C4'  H4'
 BOND C3'  H3'        BOND C2'  H2'         BOND C1'  H1'

{
 DIHEdral P    O5'  C5'  C4'             DIHEdral O5'  C5'  C4'  O4'
 DIHEdral O5'  C5'  C4'  C3'
}{
 DIHEdral C3'  C4'  O4'  C1'
 DIHEdral C4'  O4'  C1'  C2'             DIHEdral O4'  C1'  C2'  C3'

 DIHEdral C1'  C2'  C3'  C4'             DIHEdral O4'  C4'  C3'  O3'
 DIHEdral C5'  C4'  C3'  C2'             DIHEdral O2'  C2'  C3'  O3'
 DIHEdral O4'  C1'  H1'' C2
 DIHEdral C3'  C2'  O2'  H2'

 ! New dihedrals
 DIHEdral  C5'  C4'  C3'  O3'            DIHEdral  C4'  O4'  C1'  H1''
}

    !RIBOSE IMPROPERS
!IMPRoper     H1'  C2'  O4'  H1''  !C1' !mod by anda
IMPRoper     H2'  C3'  C1'  O2' !C2'
IMPRoper     H3'  C4'  C2'  O3' !C3'
IMPRoper     H4'  C5'  C3'  O4' !C4'
IMPRoper     H5'  O5'  H5'' C4' !C5'



 END {ABA}


! ---------------------------------------------------------------------


RESIdue CYT
  GROUp
  ATOM P    TYPE=P     CHARGE=1.20    END
  ATOM O1P  TYPE=O1P   CHARGE=-0.40   END
  ATOM O2P  TYPE=O2P   CHARGE=-0.40   END
  ATOM O5'  TYPE=O5R   CHARGE=-0.36   END
 GROUp
  ATOM C5'  TYPE=C5R   CHARGE=-0.070  END
  ATOM H5'  TYPE=H     CHARGE=0.035   END
  ATOM H5'' TYPE=H     CHARGE=0.035   END
 GROUp
  ATOM C4'  TYPE=C4R   CHARGE=0.065   END
  ATOM H4'  TYPE=H     CHARGE=0.035   END
  ATOM O4'  TYPE=O4R   CHARGE=-0.30   END
  ATOM C1'  TYPE=C1R   CHARGE=0.165   END
  ATOM H1'  TYPE=H     CHARGE=0.035   END

! Insert Base

 GROUp
  ATOM N1    TYPE=N1C   CHARGE=-0.19  EXCLUSION=( C4 )  END
  ATOM C6    TYPE=C6C   CHARGE=0.155  EXCLUSION=( N3 )  END
```

```
  ATOM H6    TYPE=H     CHARGE=0.035  END
 GROUp
  ATOM C2    TYPE=C2C   CHARGE=0.30    EXCLUSION=( C5 )  END
  ATOM O2    TYPE=ON    CHARGE=-0.30   END
 GROUp
  ATOM N3    TYPE=NC    CHARGE=-0.28   END
  ATOM C4    TYPE=C4C   CHARGE=0.28    END
 GROUp
  ATOM N4    TYPE=N4C   CHARGE=-0.42   END
  ATOM H41   TYPE=H2    CHARGE=0.21    END
  ATOM H42   TYPE=H2    CHARGE=0.21    END
 GROUp
  ATOM C5    TYPE=C5C   CHARGE=-0.035  END !CHRG
  ATOM H5    TYPE=H     CHARGE=0.035   END
 GROUp

! END



GROUP
  ATOM C2'  TYPE=C2R   CHARGE=0.115   END
  ATOM H2'  TYPE=H     CHARGE=0.035   END
  ATOM O2'  TYPE=O2R   CHARGE=-0.40   END
  ATOM HO2' TYPE=HO    CHARGE=0.25    END
 GROUP
  ATOM C3'  TYPE=C3R   CHARGE=-0.035  END
  ATOM H3'  TYPE=H     CHARGE=0.035   END
 GROUP
  ATOM O3'  TYPE=O3R   CHARGE=-0.36   END

 BOND P    O1P            BOND P    O2P            BOND P    O5'
 BOND O5'  C5'            BOND C5'  C4'            BOND C4'  O4'
 BOND C4'  C3'            BOND O4'  C1'            BOND C1'  N1
 BOND C1'  C2'            BOND N1   C2             BOND N1   C6
                          BOND C2   N3             BOND N3   C4
 BOND C4   N4             BOND N4   H41            BOND N4   H42
 BOND C2   O2
 BOND C4   C5             BOND C5   C6             BOND C2'  C3'
 BOND C3'  O3'            BOND C2'  O2'
 BOND C6   H6             BOND C5   H5
 BOND O2'  HO2'
 BOND C5'  H5'        BOND C5'  H5''         BOND C4'  H4'
 BOND C3'  H3'        BOND C2'  H2'          BOND C1'  H1'

{
 DIHEdral P    O5'  C5'  C4'             DIHEdral O5'  C5'  C4'  O4'
 DIHEdral O5'  C5'  C4'  C3'
}{
 DIHEdral C3'  C4'  O4'  C1'
 DIHEdral C4'  O4'  C1'  C2'             DIHEdral O4'  C1'  C2'  C3'

 DIHEdral C1'  C2'  C3'  C4'             DIHEdral O4'  C4'  C3'  O3'
 DIHEdral C5'  C4'  C3'  C2'             DIHEdral O2'  C2'  C3'  O3'
 DIHEdral O4'  C1'  N1   C2
 DIHEdral C3'  C2'  O2'  H2'
```

```
 ! New dihedrals
 DIHEdral  C5'  C4'  C3'  O3'            DIHEdral  C4'  O4'  C1'  N1
}

 IMPRoper C5   C4   N4   H41              IMPRoper C1'  C2   C6   N1
 IMPRoper O2   N1   N3   C2               IMPRoper N4   N3   C5   C4
 IMPRoper N1   C2   N3   C4               IMPRoper C2   N3   C4   C5
 IMPRoper N3   C4   C5   C6               IMPRoper C4   C5   C6   N1
 IMPRoper C5   C6   N1   C2               IMPRoper C6   N1   C2   N3
 IMPRoper H42  C4   H41  N4
 IMPRoper H5  C4  C6   C5                 IMPRoper  H6   N1   C5   C6

!RIBOSE IMPROPERS
IMPRoper     H1'  C2'  O4'  N1  !C1'
IMPRoper     H2'  C3'  C1'  O2' !C2'
IMPRoper     H3'  C4'  C2'  O3' !C3'
IMPRoper     H4'  C5'  C3'  O4' !C4'
IMPRoper     H5'  O5'  H5'' C4' !C5'



 END {CYT}

! ---------------------------------------------------------------------

RESIdue THY
 GROUp
  ATOM P    TYPE=P     CHARGE=1.20    END
  ATOM O1P  TYPE=O1P   CHARGE=-0.40   END
  ATOM O2P  TYPE=O2P   CHARGE=-0.40   END
  ATOM O5'  TYPE=O5R   CHARGE=-0.36   END
 GROUp
  ATOM C5'  TYPE=C5R   CHARGE=-0.070  END
  ATOM H5'  TYPE=H     CHARGE=0.035   END
  ATOM H5''  TYPE=H     CHARGE=0.035  END
 GROUp
  ATOM C4'  TYPE=C4R   CHARGE=0.065   END
  ATOM H4'  TYPE=H     CHARGE=0.035   END
  ATOM O4'  TYPE=O4R   CHARGE=-0.30   END
  ATOM C1'  TYPE=C1R   CHARGE=0.20    END
  ATOM H1'  TYPE=H     CHARGE=0.165   END

! Insert Base
 GROUp
  ATOM N1    TYPE=N1T   CHARGE=-0.19  EXCLUSION=( C4 )  END
  ATOM C6    TYPE=C6T   CHARGE=0.155  EXCLUSION=( N3 )  END
  ATOM H6    TYPE=H     CHARGE=0.035  END
 GROUp
  ATOM C2    TYPE=C2T   CHARGE=0.35   EXCLUSION=( C5 )  END
  ATOM O2    TYPE=ON    CHARGE=-0.35  END
 GROUp
  ATOM N3    TYPE=N3T   CHARGE=-0.26  END
  ATOM H3    TYPE=HN    CHARGE=0.26   END
 GROUp
  ATOM C4    TYPE=C4T   CHARGE=0.30   END
  ATOM O4    TYPE=ON    CHARGE=-0.30  END
```

```
 GROUp
  ATOM C5    TYPE=C5T   CHARGE=-0.035 END
  ATOM C7    TYPE=CC3E  CHARGE=-0.070 END ! name per IUPAC-IUB recomm.
  ATOM H71   TYPE=H     CHARGE=0.035  END ! name per IUPAC-IUB recomm.
  ATOM H72   TYPE=H     CHARGE=0.035  END ! name per IUPAC-IUB recomm.
  ATOM H73   TYPE=H     CHARGE=0.035  END ! name per IUPAC-IUB recomm.

 GROUp

! END



 GROUP
  ATOM C2'  TYPE=C2R   CHARGE=0.115    END
  ATOM H2'  TYPE=H     CHARGE=0.035    END
  ATOM O2'  TYPE=O2R   CHARGE=-0.40    END
  ATOM HO2' TYPE=HO    CHARGE=0.25     END
 GROUP
  ATOM C3'  TYPE=C3R   CHARGE=-0.035   END
  ATOM H3'  TYPE=H     CHARGE=0.035    END

 GROUP
  ATOM O3'  TYPE=O3R   CHARGE=-0.36   END


 BOND P    O1P              BOND P    O2P              BOND P    O5'
 BOND O5'  C5'              BOND C5'  C4'              BOND C4'  O4'
 BOND C4'  C3'              BOND O4'  C1'              BOND C1'  N1
 BOND C1'  C2'              BOND N1   C2               BOND N1   C6
 BOND C2   O2               BOND C2   N3               BOND N3   H3
 BOND N3   C4               BOND C4   O4               BOND C4   C5
 BOND C5   C7               BOND C5   C6               BOND C2'  C3'
 BOND C3'  O3'              BOND C2'  O2'
 BOND O2'  HO2'
 BOND C5'  H5'              BOND C5'  H5''
 BOND C3'  H3' BOND C2'  H2'              BOND C1'  H1'
 BOND C4'  H4'              BOND C7   H71              BOND C7   H72
 BOND C7   H73              BOND C6   H6
{
 DIHEdral P   O5'  C5'  C4'             DIHEdral O5'  C5'  C4'  O4'
 DIHEdral O5'  C5'  C4'  C3'
}{
 DIHEdral C3'  C4'  O4'  C1'
 DIHEdral C4'  O4'  C1'  C2'            DIHEdral O4'  C1'  C2'  C3'

 DIHEdral C1'  C2'  C3'  C4'            DIHEdral O4'  C4'  C3'  O3'
 DIHEdral C5'  C4'  C3'  C2'            DIHEdral O2'  C2'  C3'  O3'
 DIHEdral O4'  C1'  N1   C2
 DIHEdral C3'  C2'  O2'  H2'


 ! New dihedrals
 DIHEdral  C5'  C4'  C3'  O3'           DIHEdral  C4'  O4'  C1'  N1
}

 IMPRoper O4   N3   C5   C4               IMPRoper C1'  C2   C6   N1
```

```
 IMPRoper O2   N1   N3   C2              IMPRoper C4   C5   C6   N1
 IMPRoper N1   C2   N3   C4              IMPRoper C2   N3   C4   C5
                                         IMPRoper N3   C4   C5   C6
 IMPRoper C5   C6   N1   C2              IMPRoper C6   N1   C2   N3
 IMPRoper H3   C2   C4   N3
 IMPRoper C7   C4   C6   C5              IMPRoper  H6   N1   C5   C6

!RIBOSE IMPROPERS
IMPRoper     H1'  C2'  O4'  N1 ! C1'
IMPRoper     H2'  C3'  C1'  O2' !C2'
IMPRoper     H3'  C4'  C2'  O3' !C3'
IMPRoper     H4'  C5'  C3'  O4' !C4'
IMPRoper     H5'  O5'  H5'' C4' !C5'



 END {THY}


!------------------------------------------------------------------


RESIdue URI
GROUp
  ATOM P    TYPE=P     CHARGE=1.20    END
  ATOM O1P  TYPE=O1P   CHARGE=-0.40   END
  ATOM O2P  TYPE=O2P   CHARGE=-0.40   END
  ATOM O5'  TYPE=O5R   CHARGE=-0.36   END
 GROUp
  ATOM C5'  TYPE=C5R   CHARGE=-0.070  END
  ATOM H5'  TYPE=H     CHARGE=0.035   END
  ATOM H5'' TYPE=H     CHARGE=0.035   END
 GROUp
  ATOM C4'  TYPE=C4R   CHARGE=0.065   END
  ATOM H4'  TYPE=H     CHARGE=0.035   END
  ATOM O4'  TYPE=O4R   CHARGE=-0.30   END
  ATOM C1'  TYPE=C1R   CHARGE=0.165   END
  ATOM H1'  TYPE=H     CHARGE=0.035   END



 GROUp
  ATOM N1    TYPE=N1U   CHARGE=-0.19  EXCLUSION=( C4 )  END
  ATOM C6    TYPE=C6U   CHARGE=0.155  EXCLUSION=( N3 )  END
  ATOM H6    TYPE=H     CHARGE=0.035  END
 GROUp
  ATOM C2    TYPE=C2U   CHARGE=0.30   EXCLUSION=( C5 )  END
  ATOM O2    TYPE=ON    CHARGE=-0.30  END
 GROUp
  ATOM N3    TYPE=N3U   CHARGE=-0.28  END
  ATOM H3    TYPE=HN    CHARGE=0.26   END
 GROUp
  ATOM C4    TYPE=C4U   CHARGE=0.28   END
  ATOM O4    TYPE=ON    CHARGE=-0.30  END
 GROUp
  ATOM C5    TYPE=C5U   CHARGE=-0.035 END !JPR
  ATOM H5    TYPE=H     CHARGE=0.035  END !JPR
```

```
 GROUP
  ATOM C2'  TYPE=C2R   CHARGE=0.115   END
  ATOM H2'  TYPE=H     CHARGE=0.035   END !
  ATOM O2'  TYPE=O2R   CHARGE=-0.40   END
  ATOM HO2'   TYPE=HO    CHARGE=0.25  END
 GROUP
  ATOM C3'  TYPE=C3R    CHARGE=-0.035 END
  ATOM H3'  TYPE=H     CHARGE=0.035   END
 GROUP
  ATOM O3'  TYPE=O3R   CHARGE=-0.36   END

 BOND P    O1P              BOND P    O2P             BOND P    O5'
 BOND O5'  C5'              BOND C5'  C4'             BOND C4'  O4'
 BOND C4'  C3'              BOND O4'  C1'             BOND C1'  N1
 BOND C1'  C2'              BOND N1   C2              BOND N1   C6
 BOND C2   O2               BOND C2   N3              BOND N3   H3
 BOND N3   C4               BOND C4   O4              BOND C4   C5

 BOND C5   C6               BOND C2'  C3'             BOND C3'  O3'
 BOND C2'  O2'
 BOND C5  H5                BOND C6   H6

 BOND O2'  HO2'
 BOND C5'  H5'        BOND C5'  H5''         BOND C4'  H4'
 BOND C3'  H3'        BOND C2'  H2'          BOND C1'  H1'
{
 DIHEdral P   O5'  C5'  C4'             DIHEdral O5'  C5'  C4'  O4'
 DIHEdral O5'  C5'  C4'  C3'
}{
 DIHEdral C3'  C4'  O4'  C1'
 DIHEdral C4'  O4'  C1'  C2'            DIHEdral O4'  C1'  C2'  C3'

 DIHEdral C1'  C2'  C3'  C4'            DIHEdral O4'  C4'  C3'  O3'
 DIHEdral C5'  C4'  C3'  C2'            DIHEdral O2'  C2'  C3'  O3'
 DIHEdral O4'  C1'  N1   C2
 DIHEdral C3'  C2'  O2'  H2'

 DIHEdral P    O3'  C3'  C2'            DIHEdral P    O3'  C3'  C4'
 ! New dihedrals
 DIHEdral  C5'  C4'  C3'  O3'           DIHEdral  C4'  O4'  C1'  N1
}
                                           IMPRoper C1'  C2   C6   N1
 IMPRoper O2   N1   N3   C2                IMPRoper H3   C2   C4   N3
 IMPRoper O4   N3   C5   C4                IMPRoper N1   C2   N3   C4
 IMPRoper C2   N3   C4   C5                IMPRoper N3   C4   C5   C6
 IMPRoper C4   C5   C6   N1                IMPRoper C5   C6   N1   C2
 IMPRoper C6   N1   C2   N3
 IMPRoper H5   C4   C6   C5                IMPRoper H6   N1   C5   C6

 !GENERAL RIBOSE IMPROPERS
IMPRoper     H1'  C2'  O4'  N1  !C1'
IMPRoper     H2'  C3'  C1'  O2' !C2'
IMPRoper     H3'  C4'  C2'  O3' !C3'
IMPRoper     H4'  C5'  C3'  O4' !C4'
IMPRoper     H5'  O5'  H5'' C4' !C5'
```

```
END {URI}


!-------------------------------------------------------------------
PRESidue DEOX    ! Patch to make DEOXYribose of the ribose
 DELETE ATOM O2'   END
 DELETE ATOM HO2'   END
GROUP
 MODIFY  ATOM C2'  TYPE=C2D   CHARGE=-0.07    END
 MODIFY  ATOM C5'  TYPE=C5D   CHARGE=-0.07    END
 MODIFY  ATOM C4'  TYPE=C4D   CHARGE=0.065    END
 MODIFY  ATOM O4'  TYPE=O4D   CHARGE=-0.30    END
 MODIFY  ATOM C1'  TYPE=C1D   CHARGE=0.165    END
 MODIFY  ATOM C3'  TYPE=C3D   CHARGE=-0.035   END
 ADD     ATOM H2'' TYPE=H     CHARGE=0.035    END

 ADD BOND   C2'  H2''
 ADD ANGLE  C1'  C2'  H2'
 ADD ANGLE  C3'  C2'  H2''
 ADD ANGLE  H2'  C2'  H2''
 ADD IMPRoper   H2'  C3'  H2'' C1'! C2' chirality term
END {DEOX}

!-------------------------------------------------------------------
PRESidue HCFDEOX    ! Patch to make DEOXYribose of the ribose
  MODIFY  ATOM C1'  TYPE=C1D   CHARGE=0.390    END
END {DEOX}

!-------------------------------------------------------------------


PRESidue 3TER              ! 3-terminus (without phosphate)
                           ! should be used as "LAST 3TER  HEAD - * END"
 GROUp                     ! i.e. to be patched to the last RNA residue
  MODIFY ATOM -C3'    TYPE=C3R    CHARGE=0.15    END
  MODIFY ATOM -O3'    TYPE=OH     CHARGE=-0.40   END
  ADD ATOM -H3T  TYPE=HO     CHARGE=0.25        END
 !
 ADD BOND -O3'  -H3T
 ADD ANGLe -C3'  -O3'  -H3T
! ADD DIHEdral -C4'  -C3'  -O3'  -H3T
END {3TER}



! -------------------------------------------------------------------
PRESidue 5TER        ! 5-terminus (without phosphate)
 !                   ! should be used as "FIRST  5TER  TAIL + * END"
 GROUp               ! i.e. to be patched to the first RNA residue
  ADD ATOM +H5T  TYPE=HO    CHARGE=0.25    END
  MODIFY ATOM +O5'  TYPE=OH  CHARGE=-0.40  END
  MODIFY ATOM +C5'  TYPE=C5R  CHARGE=0.15  END
```

```
  DELETE ATOM +P   END
  DELETE ATOM +O1P END
  DELETE ATOM +O2P END
 !
 ADD BOND +H5T   +O5'
 ADD ANGLe +H5T   +O5'  +C5'
! ADD DIHEdral +H5T   +O5'  +C5'  +C4'
END {5TER}

!------------------------------------------------------------------

PRESidue NUC         ! patch for nucleic acid backbone
                     ! should be used as "LINK NUC  HEAD - *  TAIL + *  END"
                     ! i.e. it links the previous RNA residue (-) with
                     ! the current one (+)
 GROUp

  MODIFY ATOM -O3' END   !
  MODIFY ATOM +P   END   !
  MODIFY ATOM +O1P END   ! this should correctly define the electrostatic

  MODIFY ATOM +O2P END   ! group boundary

  MODIFY ATOM +O5' END   !
 ADD BOND -O3' +P
 ADD ANGLE -C3' -O3' +P
 ADD ANGLE -O3' +P  +O1P
 ADD ANGLE -O3' +P  +O2P
 ADD ANGLE -O3' +P  +O5'
!ADD DIHEdral -O3' +P  +O5'  +C5'

! ADD DIHEdral -C4' -C3' -O3'  +P
! ADD DIHEdral -C3' -O3' +P   +O5'

END {NUC}

!------------------------------------------------------------------

!-----------------------6HQ mod by lsd----------------------------

PRESidue N6HQ        ! patch for nucleic acid backbone of Glycerol-6HQ
                     ! should be used as "LINK NUC  HEAD - *  TAIL + *  END"
                     ! i.e. it links the previous RNA residue (-) with
                     ! the current one (+)
 GROUp

  MODIFY ATOM -O3' END   !
  MODIFY ATOM +P   END   !
  MODIFY ATOM +O1P END   ! this should correctly define the electrostatic

  MODIFY ATOM +O2P END   ! group boundary

  MODIFY ATOM +O3g END   !
 ADD BOND -O3' +P
 ADD ANGLE -C3' -O3' +P
 ADD ANGLE -O3' +P  +O1P
```

```
 ADD ANGLE -O3' +P  +O2P
 ADD ANGLE -O3' +P  +O3g

END {N6HQ}

PRESidue S6HQ        ! patch for nucleic acid backbone of Glycerol-6HQ
                     ! should be used as "LINK NUC  HEAD - *  TAIL + *  END"
                     ! i.e. it links the previous RNA residue (-) with
                     ! the current one (+)
 GROUp

 MODIFY ATOM -O2g END   !
  MODIFY ATOM +P   END   !
  MODIFY ATOM +O1P END   ! this should correctly define the electrostatic

  MODIFY ATOM +O2P END   ! group boundary

  MODIFY ATOM +O5' END   !
 ADD BOND -O2g +P
 ADD ANGLE -C2g -O3' +P
 ADD ANGLE -O2g +P  +O1P
 ADD ANGLE -O2g +P  +O2P
 ADD ANGLE -O2g +P  +O5'

END {S6HQ}
!-------------------end of mod by lsd-----------------------------

set echo=true end
```

3 Script Code

3.1 Script to export distances from Cara to XPLOR-NIH

```
-- First part: Script to output all chosen and integrated peaks from one project
-- and combine them in one peaklist.

-- Second part: Choose the best integrated peak among same ones or average over
-- equivalently rated peaks

-- Third part: Convert peak volumes to distances. Tricky is here the
-- differentiation of d2o and h2o and methyl peaks (all have different
-- reference peaks)!

-- Fourth part: An XPLOR-inputfile is generated where the distance information
-- and some predefined upper and lower limits (deduced from the maximum
-- deviation of the standard peaks) are used

-- written by Andre Dallmann April-05-2007


--------------------------------------------------------------------------------
--FIRST PART   --
```

```
--------------------------------------------------------------------------------


----------------------------- PREPARATIONS ------------------------------------

t = {} -- table for all the variables used in the script

-- choosing one project
local ProjectNames = {}
i = 0
for a,b in pairs(cara:getProjects()) do
i = i + 1
ProjectNames[ i ] = b:getName()
end
t.ProjectName=dlg.getSymbol("Select Project","", unpack( ProjectNames ) )
t.project = cara:getProject( t.ProjectName )

-- Get Output Filename
t.Filename = dlg.getText("Enter the output filename", "", t.ProjectName)

-- open outfile
outfile = io.output( t.Filename.."_all.peaks" )

-- Write header to peaklist
label = string.format ("%25.25s", "Peaklabel")
id = string.format ("%9.9s", "PeakID")
assx = string.format ("%9.9s", "ID(X)")
assy = string.format ("%9.9s", "ID(Y)")
posx = string.format ("%9.9s", "PPM(X)")
posy = string.format ("%9.9s", "PPM(Y)")
ampl = string.format ("%7.7s", "Ampl")
grade = string.format ("%9.9s", "Grade")
vol = string.format ("%15.15s", "VolumeInt")
outfile:write("IDnew"..id..label..assx..assy..posx..posy..ampl..vol..grade.."\n")


-- generate tables for information
count = 0
i = 0
t.label = {}
t.id_old = {}
t.assx = {}
t.assy = {}
t.posx = {}
t.posy = {}
t.ampl = {}
t.grade = {}
t.vol = {}

--------------------------------------------------------------------------------
----------------------------- Main Body ----------------------------------
--------------------------------------------------------------------------------

-- generate list of all peaks graded abc of all peaklists in specified project
for peaklistid, peaklist in pairs(t.project:getPeakLists()) do
--cycle through all peaklists
```

```
t.peaklist = t.project:getPeakList(peaklistid)
for peakid,peak in pairs(t.peaklist:getPeaks()) do --cycle through all peaks
t.peak = t.peaklist:getPeak(peakid)
if ((t.peak:getAttr("grade")== "a") or (t.peak:getAttr("grade")== "b") or
(t.peak:getAttr("grade")== "c")) then
-- choose only peaks with grade abc
i = i + 1 -- this is the index for all the tables, corresponds to new peakid
t.label[i] = string.format ("%25.25s", t.peak:getLabel())
t.id_old[i] = string.format ("%9.0f", t.peak:getId())
t.ass = {t.peak:getAssig()}
t.assx[i] = string.format ("%9.0f", t.ass[1])
t.assy[i] = string.format ("%9.0f", t.ass[2])
t.pos = {t.peak:getPos()}
t.posx[i] = string.format ("%9.3f", t.pos[1])
t.posy[i] = string.format ("%9.3f", t.pos[2])
t.ampl[i] = string.format ("%7.0f", t.peak:getAmp())
t.grade[i] = string.format ("%7.7s", t.peak:getAttr("grade"))
t.vol[i] = string.format ("%15.3f", t.peak:getVol())
outfile:write(i.."    "..t.id_old[i]..t.label[i]..t.assx[i]..t.assy[i]..t.posx[i]
..t.posy[i]..t.ampl[i]..t.vol[i]..t.grade[i].."\n")
end --of if loop
end -- of second for loop
end -- of first for loop

--------------------------- End of Main Body ------------------------------
----------------------------------------------------------------------------------

-- close outfile
outfile:close()



----------------------------------------------------------------------------------
-------------------------- End of FIRST PART  ------------------------------
----------------------------------------------------------------------------------




----------------------------------------------------------------------------------
----------------------------------------------------------------------------------
--  SECOND PART   --
----------------------------------------------------------------------------------
----------------------------------------------------------------------------------


----------------------------- PREPARATIONS -----------------------------

-- open outfile
outfile = io.output( t.Filename.."_combo.peaks" )

-- initialize variables
x = 0
counter = 1
a = string.format ("%7.7s","a")
b = string.format ("%7.7s","b")
c = string.format ("%7.7s","c")
```

```
--------------------------------------------------------------------------------
------------------------------ Main Body ----------------------------------
--------------------------------------------------------------------------------


------------------------ Preparing combination -----------------------
for i,assx in pairs (t.assx) do
for j,assy in pairs (t.assy) do
if ((((t.assx[i] == t.assx[j]) and (t.assy[i] == t.assy[j])) or ((t.assx[i] == t.assy[j]) and
(t.assy[i] == t.assx[j]))) and (j~=i)) then
-- select all peaks that have the same assignment (including cross-diagonal peaks)

counter = counter + 1
if (t.grade[i]==t.grade[j]) then
t.vol[i] = string.format ("%15.3f",(t.vol[i] + t.vol[j])) -- average volumes, rest stays
t.label[j] = nil -- set jth peak to nil
t.assx[j] = nil
t.assy[j] = nil
t.posx[j] = nil
t.posy[j] = nil
t.ampl[j] = nil
t.grade[j] = nil
t.vol[j] = nil
t.id_old[j] = nil
end
if (((a==t.grade[i]) and ((t.grade[j]==b) or (t.grade[j]==b))) or ((b==t.grade[i]) and
(t.grade[j]==c))) then
counter = 1
t.label[j] = nil -- set jth peak to nil
t.assx[j] = nil
t.assy[j] = nil
t.posx[j] = nil
t.posy[j] = nil
t.ampl[j] = nil
t.grade[j] = nil
t.vol[j] = nil
t.id_old[j] = nil
end
if (((a==t.grade[j]) and ((t.grade[i]==b) or (t.grade[i]==b))) or ((b==t.grade[j]) and
(t.grade[i]==c))) then
counter = 1
t.vol[i] = t.vol[j] --transfer volume and grade of better integrated peak (j)
t.grade[i] = t.grade[j]
t.label[j] = nil -- set jth peak to nil
t.assx[j] = nil
t.assy[j] = nil
t.posx[j] = nil
t.posy[j] = nil
t.ampl[j] = nil
t.grade[j] = nil
t.vol[j] = nil
t.id_old[j] = nil
end
end -- if loop
```

```
end -- second for loop
if (counter > 1) then -- only valid if grades are the same and averaging is needed
t.vol[i] = string.format ("%15.3f",(t.vol[i]/counter))
end
counter = 1
end -- first for loop


--------------------------------------------------------------------------------
---------------------- Generating new combined peaklist ------------------------

-- initiliaize new tables for the combined peaklist
t.labelnew = {}
t.assxnew = {}
t.assynew = {}
t.assxlabel = {}
t.assylabel = {}
t.assxresid = {}
t.assyresid = {}
t.gradenew= {}
t.volnew = {}

for i,assx in pairs (t.assx) do -- generate new table with combined peaks
x = x + 1
t.labelnew[x] = t.label[i]
t.assxnew[x] = t.assx[i]
t.assynew[x] = t.assy[i]
t.assxlabel[x] = string.format ("%7.7s", t.project:getSpin(t.assx[i]):getLabel())
-- get Peaklabel
t.assylabel[x] = string.format ("%7.7s", t.project:getSpin(t.assy[i]):getLabel())
-- get Peaklabel
t.assxresid[x] = string.format ("%5.5s", t.project:getSpin(t.assx[i]):getSystem():getId())
-- get residue id, works only if SpinsystemId equal to residue number !!
t.assyresid[x] = string.format ("%5.5s", t.project:getSpin(t.assy[i]):getSystem():getId())
-- get residue id, works only if SpinsystemId equal to residue number !!
t.gradenew[x] = t.grade[i]
t.volnew[x] = t.vol[i]
outfile:write (x.."  "..t.labelnew[x]..t.assxlabel[x]..t.assylabel[x]..t.gradenew[x]..t.volnew[x].."\n")
end

-- loop to correct for base rectangle sum method error
-- for peaks with grade c or b divide volume by 2 or 1.5 respectively
-- this is a very rough approximation!!!
for i,vol in pairs (t.volnew) do
if (t.gradenew[i]==c) then
t.volnew[i]=string.format("%15.3f",vol/2)
elseif (t.gradenew[i]==b) then
t.volnew[i]=string.format("%15.3f",vol/1.5)
end
end

--------------------------- End of Main Body -----------------------------
--------------------------------------------------------------------------------

-- close outfile
outfile:close()
```

```
i = 0

--------------------------------------------------------------------------------
-------------------------- End of SECOND PART  ------------------------
--------------------------------------------------------------------------------


--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--  THIRD PART   --
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------


----------------------------- PREPARATIONS -------------------------------

-- open outfile
outfile = io.output( t.Filename.."_dist.peaks" )

-- initialize variables
sumcyt = 0
summet = 0
sumcytamino = 0
sumcyth42h5 = 0
countcyt = 0
countmet = 0
countcytamino = 0
countcyth42h5 = 0

--------------------------------------------------------------------------------
------------------------------ Main Body ----------------------------------
--------------------------------------------------------------------------------


----------------- Setting up Reference Volumes and Distances ------

-- sum up reference peaks
for j,assx in pairs (t.assxnew) do
for y in string.gfind (t.labelnew[j],"H[56]/H[56] [0-9]:C[0-9]+") do
-- establish reference for d2o peaks
countcyt = countcyt + 1
sumcyt = sumcyt + t.volnew[j]
end
for y in string.gfind (t.labelnew[j],"H[67]/H[67] [0-9]:T[0-9]+") do
-- establish reference for methyl peaks
countmet = countmet + 1
summet = summet + t.volnew[j]
end
for y in string.gfind (t.labelnew[j],"H4[12]/H4[12] [0-9]:C[0-9]+") do
-- establish reference for h2o exchangeable peaks
countcytamino = countcytamino + 1
sumcytamino = sumcytamino + t.volnew[j]
end
for y in string.gfind (t.labelnew[j],"H42/H5 [0-9]:C[0-9]+") do
-- establish reference for h2o exchangeable-non-exchangeable peaks
```

```
countcyth42h5 = countcyth42h5 + 1
sumcyth42h5 = sumcyth42h5 + t.volnew[j]
end
for y in string.gfind (t.labelnew[j],"H5/H42 [0-9]:C[0-9]+") do
-- establish reference for h2o exchangeable-non-exchangeable peaks
countcyth42h5 = countcyth42h5 + 1
sumcyth42h5 = sumcyth42h5 + t.volnew[j]
end
end

refvolcyt = string.format ("%13.3f", sumcyt / countcyt)
-- average volume of CYT H5-H6
refdistcyt = 2.48 -- distance of CYT H5-H6
refvolmet = string.format ("%13.3f", summet / countmet)
-- average volume of THY H6-H7
refdistmet = 3.09 -- distance of THY H6-H7
refvolcytamino = string.format ("%13.3f", sumcytamino / countcytamino)
-- average volume of CYT H41-H42
refdistcytamino = 1.70 -- distance of CYT H41-H42
refvolcyth42h5 = string.format ("%13.3f", sumcyth42h5 / countcyth42h5)
-- average volume of CYT H42-H5
refdistcyth42h5 = 2.40 -- distance of CYT H42-H5


----------------- Prepare standard deviations for references -------------------

-- initialize variables
stddevsumcyt = 0
stddevsummet = 0
stddevsumcytamino = 0
stddevsumcyth42h5 = 0
maxdev1 = 0
maxdev2 = 0
maxdev3 = 0
maxdev4 = 0

for j,assx in pairs (t.assxnew) do
for y in string.gfind (t.labelnew[j],"H[56]/H[56] [0-9]:C[0-9]+") do
stddevsumcyt = stddevsumcyt + (t.volnew[j]-refvolcyt)^2 -- standard deviation
dummy1 = math.abs(t.volnew[j]-refvolcyt) -- dummy for maximum deviation
if (dummy1 > maxdev1)then
maxdev1 = dummy1
end
end
for y in string.gfind (t.labelnew[j],"H[67]/H[67] [0-9]:T[0-9]+") do
-- establish reference for methyl peaks
stddevsummet = stddevsummet + (t.volnew[j]-refvolmet)^2 -- standard deviation
dummy2 = math.abs(t.volnew[j]-refvolmet) -- dummy for maximum deviation
if (dummy2 > maxdev2) then
maxdev2 = dummy2
end
end
for y in string.gfind (t.labelnew[j],"H4[12]/H4[12] [0-9]:C[0-9]+") do
-- establish reference for h2o exchangeable peaks
stddevsumcytamino = stddevsumcytamino + (t.volnew[j]-refvolcytamino)^2
-- standard deviation
```

```
dummy3 = math.abs(t.volnew[j]-refvolcytamino) -- dummy for maximum deviation
if (dummy3 > maxdev3) then
maxdev3 = dummy3
end
end
for y in string.gfind (t.labelnew[j],"H42/H5 [0-9]:C[0-9]+") do
-- establish reference for h2o exchangeable-non-exchangeable peaks
--(appears twice because of selection reasons)
stddevsumcyth42h5 = stddevsumcyth42h5 + (t.volnew[j]-refvolcyth42h5)^2
-- standard deviation
dummy4 = math.abs(t.volnew[j]-refvolcyth42h5) -- dummy for maximum deviation
if (dummy4 > maxdev4) then
maxdev4 = dummy4
end
end
for y in string.gfind (t.labelnew[j],"H5/H42 [0-9]:C[0-9]+") do
-- establish reference for h2o exchangeable-non-exchangeable peaks
stddevsumcyth42h5 = stddevsumcyth42h5 + (t.volnew[j]-refvolcyth42h5)^2
-- standard deviation
dummy4 = math.abs(t.volnew[j]-refvolcyth42h5) -- dummy for maximum deviation
if (dummy4 > maxdev4) then
maxdev4 = dummy4
end
end
end

---------------- Calculate standard deviations for references ------------------

stddevcyt = string.format ("%13.3f", (stddevsumcyt / countcyt)^(1/2))
stddevmet = string.format ("%13.3f", (stddevsummet / countmet)^(1/2))
stddevcytamino = string.format ("%13.3f", (stddevsumcytamino / countcytamino)^(1/2))
stddevcyth42h5 = string.format ("%13.3f", (stddevsumcyth42h5 / countcyth42h5)^(1/2))

------------ Prepare maximum deviations in percent for references --------------

if (maxdev1==nil) then
maxdevcyt = string.format ("%13.3f", dummy1/refvolcyt)
else
maxdevcyt = string.format ("%13.3f", maxdev1/refvolcyt)
end
if (maxdev2==nil) then
maxdevmet = string.format ("%13.3f", dummy2/refvolmet)
else
maxdevmet = string.format ("%13.3f", maxdev2/refvolmet)
end
if (maxdev3==nil) then
maxdevcytamino = string.format ("%13.3f", dummy3/refvolcytamino)
else
maxdevcytamino = string.format ("%13.3f", maxdev3/refvolcytamino)
end
if (maxdev4==nil) then
maxdevcyth42h5 = string.format ("%13.3f", dummy4/refvolcyth42h5)
else
maxdevcyth42h5 = string.format ("%13.3f", maxdev4/refvolcyth42h5)
end
```

```
--------------------- Prepare for distance calculation ------------------------

function f ( String ) -- function to format the atomlabels
FormattedString = string.format( "%7.7s", String )
return FormattedString
end
function f2 ( String ) -- function to format the atomlabels
FormattedString = string.format( "%9.9s", String )
return FormattedString
end

-- initialize new tables for distance and the lower and upper limit (same)
t.distance = {}
t.limit = {}
i = nil
assx = nil

---------------------- Distance and Limit calculation -------------------------

-- select atom pairs corresponding to references and calculate distances and
-- limits and write out new peaklist to file

-- limits are calculated by taking the maximum deviation of the corresponding
--reference peak times the distance

for i,assx in pairs (t.assxlabel) do
print(t.label[i])
if (((assx==f("H5")) or (assx==f("H6")) or (assx==f("H8")) or (assx==f("H1'")) or
(assx==f("H2")) or (assx==f("H2'")) or
(assx==f("H2''")) or (assx==f("H3'")) or (assx==f("H4'")) or (assx==f("H5'")) or
(assx==f("H5''"))) and
((t.assylabel[i]==f("H5")) or (t.assylabel[i]==f("H6")) or (t.assylabel[i]==f("H8"))
or (t.assylabel[i]==f("H1'")) or
(t.assylabel[i]==f("H2")) or (t.assylabel[i]==f("H2'")) or (t.assylabel[i]==f("H2''"))
or (t.assylabel[i]==f("H3'")) or
(t.assylabel[i]==f("H4'")) or (t.assylabel[i]==f("H5'")) or (t.assylabel[i]==f("H5''"))))
then
t.distance[i] = string.format ("%7.2f", refdistcyt*(refvolcyt/t.volnew[i])^(1/6))
print(t.label[i])
print(t.distance[i])
if (t.gradenew[i]==a) then -- error bounds scaled by grading of integration
t.limit[i] = string.format ("%7.1f", t.distance[i]*maxdevcyt)
else
if (t.gradenew[i]==b) then
t.limit[i] = string.format ("%7.1f", t.distance[i]*maxdevcyt*1.2)
else
t.limit[i] = string.format ("%7.1f", t.distance[i]*maxdevcyt*1.4)
end
end
outfile:write (f2("d2o:  ")..f(i)..t.labelnew[i]..assx..t.assylabel[i]..t.gradenew[i]
..t.volnew[i]..t.distance[i]..t.limit[i].."\n")
end
if ((assx==f("H7")) or (t.assylabel[i]==f("H7"))) then
t.distance[i] = string.format ("%7.2f", refdistmet*(refvolmet/t.volnew[i])^(1/6))
if (t.gradenew[i]==a) then -- error bounds scaled by grading of integration
```

```
t.limit[i] = string.format ("%7.1f", t.distance[i]*maxdevmet)
else
if (t.gradenew[i]==b) then
t.limit[i] = string.format ("%7.1f", t.distance[i]*maxdevmet*1.2)
else
t.limit[i] = string.format ("%7.1f", t.distance[i]*maxdevmet*1.4)
end
end
outfile:write (f2("methyl:  ")..f(i)..t.labelnew[i]..assx..t.assylabel[i]..t.gradenew[i]
..t.volnew[i]..t.distance[i]..t.limit[i].."\n")
end
if (((assx==f("H1")) or (assx==f("H3")) or (assx==f("H41")) or (assx==f("H42")))
and ((t.assylabel[i]==f("H1")) or
(t.assylabel[i]==f("H3")) or (t.assylabel[i]==f("H41")) or (t.assylabel[i]==f("H42"))))
then
t.distance[i] = string.format ("%7.2f", refdistcytamino*(refvolcytamino/t.volnew[i])^(1/6))
if (t.gradenew[i]==a) then -- error bounds scaled by grading of integration
t.limit[i] = string.format ("%7.1f", t.distance[i]*maxdevcytamino)
else
if (t.gradenew[i]==b) then
t.limit[i] = string.format ("%7.1f", t.distance[i]*maxdevcytamino*1.2)
else
t.limit[i] = string.format ("%7.1f", t.distance[i]*maxdevcytamino*1.4)
end
end
outfile:write (f2("h2o:  ")..f(i)..t.labelnew[i]..assx..t.assylabel[i]..t.gradenew[i]
..t.volnew[i]..t.distance[i]..t.limit[i].."\n")
end
if ((((assx==f("H5")) or (assx==f("H6")) or (assx==f("H8")) or (assx==f("H1'")) or
(assx==f("H2")) or (assx==f("H2'")) or (assx==f("H2''")) or (assx==f("H3'")) or
(assx==f("H4'")) or (assx==f("H5'")) or (assx==f("H5''")))) and ((t.assylabel[i]==f("H1"))
or (t.assylabel[i]==f("H3")) or (t.assylabel[i]==f("H41")) or (t.assylabel[i]==f("H42")))) then
t.distance[i] = string.format ("%7.2f", refdistcyth42h5*(refvolcyth42h5/t.volnew[i])^(1/6))
if (t.gradenew[i]==a) then -- error bounds scaled by grading of integration
t.limit[i] = string.format ("%7.1f", t.distance[i]*maxdevcyth42h5)
else
if (t.gradenew[i]==b) then
t.limit[i] = string.format ("%7.1f", t.distance[i]*maxdevcyth42h5*1.2)
else
t.limit[i] = string.format ("%7.1f", t.distance[i]*maxdevcyth42h5*1.4)
end
end
outfile:write (f2("d2o_h2o:  ")..f(i)..t.labelnew[i]..assx..t.assylabel[i]
..t.gradenew[i]..t.volnew[i]..t.distance[i]..t.limit[i].."\n")
end
if (((assx==f("H1")) or (assx==f("H3")) or (assx==f("H41")) or (assx==f("H42")))
and ((t.assylabel[i]==f("H5")) or  (t.assylabel[i]==f("H6")) or (t.assylabel[i]==f("H8"))
or (t.assylabel[i]==f("H1'")) or (t.assylabel[i]==f("H2")) or  (t.assylabel[i]==f("H2'"))
or (t.assylabel[i]==f("H2''")) or (t.assylabel[i]==f("H3'")) or (t.assylabel[i]==f("H4'")) or
(t.assylabel[i]==f("H5'")) or (t.assylabel[i]==f("H5''")))) then
t.distance[i] = string.format ("%7.2f", refdistcyth42h5*(refvolcyth42h5/t.volnew[i])^(1/6))
if (t.gradenew[i]==a) then -- error bounds scaled by grading of integration
t.limit[i] = string.format ("%7.1f", t.distance[i]*maxdevcyth42h5)
else
if (t.gradenew[i]==b) then
t.limit[i] = string.format ("%7.1f", t.distance[i]*maxdevcyth42h5*1.2)
```

```
else
t.limit[i] = string.format ("%7.1f", t.distance[i]*maxdevcyth42h5*1.4)
end
end
outfile:write (f2("d2o_h2o:  ")..f(i)..t.labelnew[i]..assx..t.assylabel[i]
..t.gradenew[i]..t.volnew[i]..t.distance[i]..t.limit[i].."\n")
end
end

-------------------------- End of Main Body ----------------------------
--------------------------------------------------------------------------------

-- close outfile
outfile:close()

i = 0

--------------------------------------------------------------------------------
-------------------------- End of THIRD PART  ------------------------
--------------------------------------------------------------------------------




--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--  FOURTH PART   --
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------



---------------------------- PREPARATIONS -----------------------------

-- open outfile
outfile = io.output( t.Filename.."_reference.peaks" )

function f ( String ) -- function to format the atomlabels
FormattedString = string.format( "%7.2f", String )
return FormattedString
end


--------------------------------------------------------------------------------
----------------------------- Main Body ----------------------------------
--------------------------------------------------------------------------------

outfile:write ("\n-----------------------------------------------------------------------------------\n\nReference
for non-exchangeable proton cross-peaks: CYT H5-H6\n\nreference_vol  ref_dist  standard_dev
maximum_dev(%)\n"..refvolcyt..f(refdistcyt)..stddevcyt..maxdevcyt.."\n\n
Peaklabel    Volume     Dist   Dev\n")
for j,assx in pairs (t.assxnew) do
for y in string.gfind (t.labelnew[j],"H[56]/H[56] [0-9]:C[0-9]+") do -- reference for d2o peaks
outfile:write (t.labelnew[j]..t.volnew[j]..t.distance[j]..f(t.distance[j]-refdistcyt).."\n")
end
end
outfile:write ("\n-----------------------------------------------------------------------------------\n\nReference
for methyl proton cross-peaks: MET H6-H7\n\nreference_vol  ref_dist  standard_dev
```

```
maximum_dev(%)\n"..refvolmet..f(refdistmet)..stddevmet..maxdevmet.."\n\n
Peaklabel    Volume     Dist   Dev\n")
for j,assx in pairs (t.assxnew) do
for y in string.gfind (t.labelnew[j],"H[67]/H[67] [0-9]:T[0-9]+") do -- reference for methyl peaks
outfile:write (t.labelnew[j]..t.volnew[j]..t.distance[j]..f(t.distance[j]-refdistmet).."\n")
end
end
outfile:write ("\n----------------------------------------------------------------------------------\n\nReference
for exchangeable proton cross-peaks: CYT H41-H42\n\nreference_vol  ref_dist  standard_dev
maximum_dev(%)\n"..refvolcytamino..f(refdistcytamino)..stddevcytamino..maxdevcytamino.."\n\n
Peaklabel    Volume     Dist   Dev\n")
for j,assx in pairs (t.assxnew) do
for y in string.gfind (t.labelnew[j],"H4[12]/H4[12] [0-9]:C[0-9]+") do
-- reference for h2o exchangeable peaks
outfile:write (t.labelnew[j]..t.volnew[j]..t.distance[j]..f(t.distance[j]-refdistcytamino).."\n")
end
end
outfile:write ("\n----------------------------------------------------------------------------------\n\nReference
for non-exchangeable/exchangeable proton cross-peaks: CYT H42-H5\n\nreference_vol
ref_dist  standard_dev
maximum_dev(%)\n"..refvolcyth42h5..f(refdistcyth42h5)..stddevcyth42h5..maxdevcyth42h5.."\n\n
Peaklabel    Volume     Dist   Dev\n")
for j,assx in pairs (t.assxnew) do
for y in string.gfind (t.labelnew[j],"H42/H5 [0-9]:C[0-9]+") do
-- reference for h2o exchangeable-non-exchangeable peaks
--(appears twice because of selection reasons)
outfile:write (t.labelnew[j]..t.volnew[j]..t.distance[j]..f(t.distance[j]-refdistcyth42h5).."\n")
end
end
for j,assx in pairs (t.assxnew) do
for y in string.gfind (t.labelnew[j],"H5/H42 [0-9]:C[0-9]+") do
-- reference for h2o exchangeable-non-exchangeable peaks
outfile:write (t.labelnew[j]..t.volnew[j]..t.distance[j]..f(t.distance[j]-refdistcyth42h5).."\n")
end
end
--------------------------- End of Main Body -----------------------------------
--------------------------------------------------------------------------------

-- close outfile
outfile:close()

i = 0
--------------------------------------------------------------------------------
-------------------------- End of FOURTH PART  ------------------------
--------------------------------------------------------------------------------



--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--  FIFTH PART   --
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------

----------------------------- PREPARATIONS ------------------------------
```

```
outfile = io.output( t.Filename.."_xplor.list" )
outfile2 = io.output( t.Filename.."_xplor_all.list" )
outfile3 = io.output( "picktbl_"..t.Filename)
outfile4 = io.output( t.Filename.."_xplor.noe" )
outfile5 = io.output( t.Filename.."_xplor_all.noe" )

function find (index) -- function to format the atomlabels
local Boolean = false
local Booleanx = false
local Booleany = false
for x in string.gfind(t.assxlabel[index],"H[2345]'[']*") do
Booleanx=true
end
for y in string.gfind(t.assylabel[index],"H[2345]'[']*") do
Booleany=true
end
if (Booleanx==true) and (Booleany==true) then
Boolean=true
else
Boolean=false
end
return Boolean
end

function find_h1 (index) -- function to format the atomlabels
local Boolean2 = false
local Boolean2x = false
local Boolean2y = false
for x in string.gfind(t.assxlabel[index],"H[12345]'[']*") do
Boolean2x=true
end
for y in string.gfind(t.assylabel[index],"H[12345]'[']*") do
Boolean2y=true
end
if (Boolean2x==true) and (Boolean2y==true) then
Boolean2=false
else
return Boolean2
end
end

for i,assx in pairs (t.assxnew) do -- iterate over all peaks
if (t.distance[i]) and  (find(i)==false) then -- filter out negative volume peaks
if (find_h1(i)==false) and ((t.distance[i]/1) < 4.5) then
outfile:write ("assign (resid"..t.assxresid[i].." and name"..t.assxlabel[i]..")
(resid"..t.assyresid[i].." and name"..t.assylabel[i]..")"..t.distance[i]..t.limit[i]..t.limit[i].."\n")
outfile2:write ("assign (resid"..t.assxresid[i].." and name"..t.assxlabel[i]..")
(resid"..t.assyresid[i].." and name"..t.assylabel[i]..")"..t.distance[i]..t.limit[i]..t.limit[i].."\n")
outfile3:write ("pick bond (resid"..t.assxresid[i].." and name"..t.assxlabel[i]..")
(resid"..t.assyresid[i].." and name"..t.assylabel[i]..")".." geometry\ndisplay \$result".."\n")
outfile4:write (t.assxresid[i].. t.assxlabel[i]..t.assyresid[i]..t.assylabel[i]
..t.labelnew[i]..t.gradenew[i]..t.distance[i].."\n")
outfile5:write (t.assxresid[i].. t.assxlabel[i]..t.assyresid[i]..t.assylabel[i]
..t.labelnew[i]..t.gradenew[i]..t.distance[i].."\n")
else
outfile2:write ("assign (resid"..t.assxresid[i].." and name"..t.assxlabel[i]..")
```

```
(resid"..t.assyresid[i].." and name"..t.assylabel[i]..")"..t.distance[i]
..t.limit[i]..t.limit[i].." !added!\n")
outfile5:write (t.assxresid[i].. t.assxlabel[i]..t.assyresid[i]..t.assylabel[i]
..t.labelnew[i]..t.gradenew[i]..t.distance[i].."\n")
end
end
end

-------------------------- End of Main Body ----------------------------
--------------------------------------------------------------------------------

-- close outfile
outfile:close()
outfile2:close()
outfile3:close()
outfile4:close()
outfile5:close()

i = 0
t = nil

--------------------------------------------------------------------------------
-------------------------- End of FIFTH PART  ---------------------------
--------------------------------------------------------------------------------


print ( "\ngenerateinput_byanda is done." )
print ( "Have a nice day!" )


------------------------- End generateinput  ---------------------------
```

3.2 Mathematica script to visualize NOESY back-calculation

```
NOESY - spectra back-calculation

Read data

SetDirectory["D:\\Lars\\Promotion\\13merHCF\\Fertige Strukturen\\HC-_diss\\198_face-down\\gifa"]

ppm - file containing chemical shifts

ppmfile=Import["13merHCF_gifa.ppm","Table"];
Dimensions[ppmfile]
TableForm[Sort[ppmfile,#1[[3]]<#2[[3]]&]]; (* Can be used for inspection of data *)
ppmscanlist=Table[{ToString[ppmfile[[m,3]]+ppmfile[[m,4]]],ppmfile[[m,5]]},{m,1,Length[ppmfile]}];
Dimensions[ppmscanlist]

Read xplor-intensities calculated from Full Matrix Relaxation Approach
Choose between GIFA or XPLOR file, the latter is recommended.
GIFA

(*Intfile=Import["13merHCF_gifa.spect","Table"];*)
(*Intlist=Table[{ToString[Intfile[[m,3]]+Intfile[[m,4]]],ToString[Intfile[[m,6]]+Intfile[[m,7]]],
```

```
Intfile[[m,8]]},{m,1,Length[Intfile]}];*)
```

XPLOR

```
rawIntfile=Import["13merHCF.spect","Table"];
Intfile=Table[rawIntfile[[2*n]],{n,2,Length[rawIntfile]/2}];
Intlist=Table[{ToString[Intfile[[m,1]]+Intfile[[m,3]]],ToString[Intfile[[m,4]]+Intfile[[m,6]]],
Intfile[[m,7]]},{m,1,Length[Intfile]}];
Dimensions[Intlist]
TableForm[Sort[Intlist,#1[[1]]<#2[[1]]&]]; (* Can be used for inspection of data *)
```

Build a peaklist that uses the data of both input files

```
peaklist={}; (* Enthält dann (x,y,Int) *)
Do[
xvar=false;
yvar=false;
Do[
If[Intlist[[mInt,1]]==ppmscanlist[[mppm,1]],
xcoord=ppmscanlist[[mppm,2]];xvar=true];
If[Intlist[[mInt,2]]==ppmscanlist[[mppm,1]],ycoord=ppmscanlist[[mppm,2]];yvar=true];,
{mppm,1,Length[ppmscanlist]}];
tmp={xcoord,ycoord,Intlist[[mInt,3]]};
If[xvar==true&&yvar==true,peaklist=Append[peaklist,tmp]];
(*Second round to add missing symmetry around diagonal axis to peaklist *)
xvar=false;
yvar=false;
Do[
If[Intlist[[mInt,1]]==ppmscanlist[[mppm,1]],ycoord=ppmscanlist[[mppm,2]];yvar=true];
If[Intlist[[mInt,2]]==ppmscanlist[[mppm,1]],xcoord=ppmscanlist[[mppm,2]];xvar=true];
If[Intlist[[mInt,2]]==Intlist[[mInt,1]],xvar=false;yvar=false;],{mppm,1,Length[ppmscanlist]}];
tmp={xcoord,ycoord,Intlist[[mInt,3]]};
If[xvar==true&&yvar==true,peaklist=Append[peaklist,tmp]];
,{mInt,1,Length[Intlist]}]
Dimensions[peaklist]
TableForm[Sort[peaklist,#1[[1]]<#2[[1]]&]]; (* Can be used for inspection of data *)
```

Use peaklist to build NOESY spectrum
Define peaks with different lineshapes. First is Lorenz-type, second Gaussian lineshape.

```
scale=1; (* Parameters that effect lineshape *)
w=0.015;
alpha=0.9;
peakL[peaknummer_,x_,y_]:=
1/(1+((x-peaklist[[peaknummer,1]])/(0.5w))^2)*1/(1+((y-peaklist[[peaknummer,2]])/(0.5w))^2);

peakG[peaknummer_,x_,y_]:=
Exp[-Log[2]((x-peaklist[[peaknummer,1]])/(0.5w))^2]*Exp[-Log[2]((y-peaklist[[peaknummer,2]])/(0.5w))^2];
peakfun[peaknummer_,x_,y_]:=
10^6*peaklist[[peaknummer,3]]*scale*(alpha*peakG[peaknummer,x,y]+(1-alpha)peakL[peaknummer,x,y])
```

Define intensity for a single set of coordinates in the spectrum

```
Intensity[x_,y_]:=Sum[peakfun[peaknummer,x,y],{peaknummer,1,Length[peaklist]}]
Intensity[5,6]
Plot[peakfun[5,x,6],{x,6,10},PlotRange->All]
```

Build a selected region of the spectrum using a targetfunction that is automatically composed
Choose spectral region

```
xrange={6.925,8.14}; (* This is H6/H8 region *)
yrange={5.1,6.2};(* This is H1' region *)
Targetfunction=0;
Do[
xvar=false;
yvar=false;
If[peaklist[[mInt,1]]>=xrange[[1]]-0.1&&peaklist[[mInt,1]]<=xrange[[2]]+0.1,xvar=true];
If[peaklist[[mInt,2]]>=yrange[[1]]-0.1&&peaklist[[mInt,2]]<=yrange[[2]]+0.1,yvar=true];
If[xvar==true&&yvar==true,Targetfunction=Targetfunction+peakfun[mInt,x,y]];
,{mInt,1,Length[peaklist]}]
```

Build spectrum

```
ContourPlot[Targetfunction,{x,xrange[[1]],xrange[[2]]},{y,yrange[[1]],yrange[[2]]},
PlotPoints->20,Contours->Table[10^k,{k,4,8,0.5}],PlotRange->All,
FrameLabel->{ppm,ppm},ImageSize->600,ColorFunction->"Aquamarine",
AspectRatio->1/2]
```

Build for export

```
ContourPlot[Targetfunction,{x,xrange[[1]],xrange[[2]]},{y,yrange[[1]],yrange[[2]]},
PlotPoints->20,Contours->Table[10^k,{k,4,8,0.5}],PlotRange->All,
FrameLabel->{ppm,ppm},ImageSize->900,ContourShading->None,
ContourStyle->Hue[0],AspectRatio->1/2,Frame->False]
```

HCF H1'' - H1/H3 region
Choose spectral region for H1/H3 to H1'' of HCF

```
xrange={5.5,5.8};
yrange={4.9,5.1};
Targetfunction=0;
Do[
xvar=false;
yvar=false;
If[peaklist[[mInt,1]]>=xrange[[1]]-0.1&&peaklist[[mInt,1]]<=xrange[[2]]+0.1,xvar=true];
If[peaklist[[mInt,2]]>=yrange[[1]]-0.1&&peaklist[[mInt,2]]<=yrange[[2]]+0.1,yvar=true];
If[xvar==true&&yvar==true,Targetfunction=Targetfunction+peakfun[mInt,x,y]];
,{mInt,1,Length[peaklist]}]
```

Build spectrum

```
ContourPlot[Targetfunction,{x,xrange[[1]],xrange[[2]]},{y,yrange[[1]],yrange[[2]]},PlotPoints->20,
Contours->Table[10^k,{k,5,9,0.5}],PlotRange->All,FrameLabel->{ppm,ppm},ImageSize->600,
ColorFunction->"Aquamarine",AspectRatio->1/2]
```

Bibliography

[1] J. Watson and F. Crick. Molecular structure of nucleic acids. *Nature*, 177:737 – 738, 1953.

[2] Richard Wing, Horace Drew, Tsunehiro Takano, Chris Broka, Shoji Tanaka, Keiichi Itakura, and Richard E. Dickerson. Crystal structure analysis of a complete turn of b-dna. *Nature*, 287(5784):755–758, 1980.

[3] H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. E. Dickerson. Structure of a b-dna dodecamer: conformation and dynamics. *Proceedings of the National Academy of Sciences*, 78(4):2179–2183, 1981.

[4] Richard E. Dickerson and Horace R. Drew. Structure of a b-dna dodecamer: Ii. influence of base sequence on helix structure. *Journal of Molecular Biology*, 149(4):761 – 786, 1981.

[5] Horace R. Drew and Richard E. Dickerson. Structure of a b-dna dodecamer: Iii. geometry of hydration. *Journal of Molecular Biology*, 151(3):535 – 556, 1981.

[6] H. R. Drew, S. Samson, and R. E. Dickerson. Structure of a b-dna dodecamer at 16 k. *Proceedings of the National Academy of Sciences of the United States of America-biological Sciences*, 79(13):4040–4044, 1982.

[7] A. V. Fratini, M. L. Kopka, H. R. Drew, and R. E. Dickerson. Reversible bending and helix geometry in a b-dna dodecamer - cgcgaattbrcgcg. *Journal of Biological Chemistry*, 257(24):4686–4707, 1982.

[8] R. E. Dickerson, M. L. Kopka, and P. Pjura. A random-walk model for helix bending in b-dna. *Proceedings of the National Academy of Sciences of the United States of America-biological Sciences*, 80(23):7099–7103, 1983.

[9] U. Heinemann and C. Alings. Crystallographic study of one turn of g-c-rich b-dna. *Journal of Molecular Biology*, 210(2):369–381, November 1989.

[10] R. E. Dickerson, H. R. Drew, B. N. Conner, R. M. Wing, A. V. Fratini, and M. L. Kopka. The anatomy of a-dna, b-dna, and z-dna. *Science*, 216(4545):475–485, 1982.

[11] Z. Shakked, D. Rabinovich, W. B. T. Cruse, E. Egert, O. Kennard, G. Sala, S. A. Salisbury, and M. A. Viswamitra. Crystalline a-dna - the x-ray-analysis of the fragment d(g-g-t-a-t-a-c-c). *Proceedings of the Royal Society Series B-biological Sciences*, 213(1193):479–487, 1981.

[12] Benjamin N. Conner, Tsunehiro Takano, Shoji Tanaka, Keiichi Itakura, and Richard E. Dickerson. The molecular structure of d(icpcpgpg), a fragment of right-handed double helical a-dna. *Nature*, 295(5847):294 – 299, January 1982.

[13] A. H. J. Wang, S. Fujii, J. H. Vanboom, G. A. Vandermarel, S. A. A. Vanboeckel, and A. Rich. Molecular-structure of r(gcg)d(tatacgc) - a dna rna hybrid helix joined to double helical dna. *Nature*, 299(5884):601–604, 1982.

[14] A. H. J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. Vanboom, G. Vandermarel, and A. Rich. Molecular-structure of a left-handed double helical dna fragment at atomic resolution. *Nature*, 282(5740):680–686, 1979.

[15] H. Drew, T. Takano, S. Tanaka, K. Itakura, and R. E. Dickerson. High-salt d(cpgpcpg), a left-handed z' dna double helix. *Nature*, 286(5773):567–573, 1980.

[16] A. H. J. Wang, G. J. Quigley, F. J. Kolpak, G. Vandermarel, J. H. Vanboom, and A. Rich. Left-handed double helical dna - variations in the backbone conformation. *Science*, 211(4478):171–176, 1981.

[17] A. Rich, A. Nordheim, and A. H. J. Wang. The chemistry and biology of left-handed z-dna. *Annual Review of Biochemistry*, 53:791–846, 1984.

[18] Stephen Neidle. *Principles of Nucleic Acid Structure*. Academic Press Inc., 2007.

[19] Victor A. Bloomfield, Donald M. Crothers, and Jr. Ignacio Tinoco. *Nucleic Acids: structures, properties and functions*. University Science Books, 2000.

[20] G. R. Clark, D. G. Brown, M. R. Sanderson, T. Chwalinski, S. Neidle, J. M. Veal, R. L. Jones, W. D. Wilson, G. Zon, E. Garman, and D. I. Stuart. Crystal and solution structures of the oligonucleotide d(atgcgcat)2 - a combined x-ray and nmr-study. *Nucleic Acids Research*, 18(18):5521–5528, September 1990.

[21] N. Verdaguer, J. Aymami, D. Fernandezforner, I. Fita, M. Coll, T. Huynhdinh, J. Igolen, and J. A. Subirana. Molecular-structure of a complete turn of a-dna. *Journal of Molecular Biology*, 221(2):623–635, September 1991.

[22] S. Jain and M. Sundaralingam. Effect of crystal packing environment on conformation of the dna duplex - molecular-structure of the a-dna octamer d(g-t-g-t-a-c-a-c) in 2 crystal forms. *Journal of Biological Chemistry*, 264(22):12780–12784, August 1989.

[23] Z. Shakked, G. Guersteinguzikevich, M. Eisenstein, F. Frolow, and D. Rabinovich. The conformation of the dna double helix in the crystal is dependent on its environment. *Nature*, 342(6248):456–460, November 1989.

[24] R. E. Dickerson. Definitions and nomenclature of nucleic-acid structure parameters. *Journal of Biomolecular Structure & Dynamics*, 6(4):627–634, February 1989.

[25] J. Jeener, B. H. Meier, P. Bachmann, and R. R. Ernst. Investigation of exchange processes by 2-dimensional nmr-spectroscopy. *Journal of Chemical Physics*, 71(11):4546–4553, 1979.

[26] G. Bodenhausen, R. Freeman, and D. L. Turner. 2-dimensional j-spectroscopy - proton-coupled c-13 nmr. *Journal of Chemical Physics*, 65(2):839–840, 1976.

[27] G. Bodenhausen and R. Freeman. Correlation of proton and c-13 nmr-spectra by heteronuclear 2-dimensional spectroscopy. *Journal of Magnetic Resonance*, 28(3):471–476, 1977.

[28] A. Bax, R. Freeman, and G. Morris. Correlation of proton chemical-shifts by two-dimensional fourier-transform nmr. *Journal of Magnetic Resonance*, 42(1):164–168, 1981.

[29] A. Bax and R. Freeman. Investigation of complex networks of spin-spin coupling by two-dimensional nmr. *Journal of Magnetic Resonance*, 44(3):542–561, 1981.

[30] L. Müller, A. Kumar, and R. R. Ernst. Two-dimensional carbon-13 nmr-spectroscopy. *Journal of Chemical Physics*, 63(12):5490–5491, 1975.

[31] Anil Kumar, R. R. Ernst, and K. Wüthrich. A two-dimensional nuclear overhauser enhancement (2d noe) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochemical and Biophysical Research Communications*, 95(1):1 – 6, 1980.

[32] Anil Kumar, G. Wagner, R. R. Ernst, and K. Wüthrich. Studies of j-connectivities and selective 1h-1h overhauser effects in h_2o solutions of biological macromolecules by two-dimensional nmr experiments. *Biochemical and Biophysical Research Communications*, 96(3):1156 – 1163, 1980.

[33] A. Kumar, G. Wagner, R. R. Ernst, and K. Wüthrich. Buildup rates of the nuclear overhauser effect measured by two-dimensional proton magnetic-resonance spectroscopy - implications for studies of protein conformation. *Journal of the American Chemical Society*, 103(13):3654–3658, 1981.

[34] G. Wagner, A. Kumar, and K. Wüthrich. Systematic application of two-dimensional h-1-nmr techniques for studies of proteins .2. combined use of correlated spectroscopy and nuclear overhauser spectroscopy for sequential assignments of backbone resonances and elucidation of polypeptide secondary structures. *European Journal of Biochemistry*, 114(2):375–384, 1981.

[35] G. Wagner and K. Wüthrich. Sequential resonance assignments in protein h-1 nuclear magnetic-resonance spectra - basic pancreatic trypsin-inhibitor. *Journal of Molecular Biology*, 155(3):347–366, 1982.

[36] G. Wagner and K. Wüthrich. Amide proton-exchange and surface conformation of the basic pancreatic trypsin-inhibitor in solution - studies with two-dimensional nuclear magnetic-resonance. *Journal of Molecular Biology*, 160(2):343–361, 1982.

[37] E. R. P. Zuiderweg, R. Kaptein, and K. Wüthrich. Secondary structure of the lac repressor dna-binding domain by two-dimensional h-1 nuclear magnetic-resonance in solution. *Proceedings of the National Academy of Sciences of the United States of America-biological Sciences*, 80(19):5837–5841, 1983.

[38] M. Rance, O. W. Sorensen, G. Bodenhausen, G. Wagner, R. R. Ernst, and K. Wüthrich. Improved spectral resolution in cosy h-1-nmr spectra of proteins via double quantum filtering. *Biochemical and Biophysical Research Communications*, 117(2):479–485, 1983.

[39] A. Pardi, G. Wagner, and K. Wüthrich. Protein conformation and proton nmr chemical-shifts. *European Journal of Biochemistry*, 137(3):445–454, 1983.

[40] K. Wüthrich, M. Billeter, and W. Braun. Pseudo-structures for the 20 common amino-acids for use in studies of protein conformations by measurements of intramolecular proton proton distance constraints with nuclear magnetic-resonance. *Journal of Molecular Biology*, 169(4):949–961, 1983.

[41] D. Marion and K. Wüthrich. Application of phase sensitive two-dimensional correlated spectroscopy (cosy) for measurements of h-1-h-1 spin-spin coupling-constants in proteins. *Biochemical and Biophysical Research Communications*, 113(3):967–974, 1983.

[42] A. Pardi, M. Billeter, and K. Wüthrich. Calibration of the angular-dependence of the amide proton-c-alpha proton coupling-constants, 3jhn-alpha, in a globular

protein - use of 3jhn-alpha for identification of helical secondary structure. *Journal of Molecular Biology*, 180(3):741–751, 1984.

[43] G. M. Clore and A. M. Gronenborn. Sequence-dependent structural variations in 2 right-handed alternating pyrimidine-purine dna oligomers in solution determined by nuclear overhauser enhancement measurements. *Embo Journal*, 2(12):2109–2115, 1983.

[44] Dennis R. Hare, David E. Wemmer, Shan-Ho Chou, Gary Drobny, and Brian R. Reid. Assignment of the non-exchangeable proton resonances of d(c-g-c-g-a-a-t-t-c-g-c-g) using two-dimensional nuclear magnetic resonance methods. *Journal of Molecular Biology*, 171(3):319 – 336, 1983.

[45] J. Feigon, W. Leupin, W. A. Denny, and D. R. Kearns. Two-dimensional proton nuclear magnetic-resonance investigation of the synthetic deoxyribonucleic-acid decamer d(atatcgatat)2. *Biochemistry*, 22(25):5943–5951, 1983.

[46] D. G. Reid, S. A. Salisbury, S. Bellard, Z. Shakked, and D. H. Williams. Proton nuclear overhauser effect study of the structure of a deoxyoligonucleotide duplex in aqueous-solution. *Biochemistry*, 22(8):2019–2025, 1983.

[47] R. M. Scheek, N. Russo, R. Boelens, and R. Kaptein. Sequential resonance assignments in dna h-1-nmr spectra by two-dimensional noe spectroscopy. *Journal of the American Chemical Society*, 105(9):2914–2916, 1983.

[48] Timothy Havel and Kurt Wüthrich. A distance geometry program for determining the structures of small proteins and other macromolecules from nuclear magnetic resonance measurements of intramolecular ^{1}h-^{1}h proximities in solution. *Bulletin of Mathematical Biology*, 46(4):673–698, 1984.

[49] A. T. Brünger, G. M. Clore, A. M. Gronenborn, and M. Karplus. 3-dimensional structure of proteins determined by molecular-dynamics with interproton distance restraints - application to crambin. *Proceedings of the National Academy of Sciences of the United States of America*, 83(11):3801–3805, June 1986.

[50] M. M. Bluhm, G. Bodo, H. M. Dintzis, and J. C. Kendrew. The crystal structure of myoglobin .4. a fourier projection of sperm-whale myoglobin by the method of isomorphous replacement. *Proceedings of the Royal Society of London Series A-mathematical and Physical Sciences*, 246(1246):369–389, 1958.

[51] M. P. Williamson, T. F. Havel, and K. Wuthrich. Solution conformation of proteinase inhibitor-iia from bull seminal plasma by h-1 nuclear magnetic-resonance and distance geometry. *Journal of Molecular Biology*, 182(2):295–315, 1985.

[52] R. Kaptein, E. R. P. Zuiderweg, R. M. Scheek, R. Boelens, and W. F. Vangunsteren. A protein-structure from nuclear magnetic-resonance data - lac repressor headpiece. *Journal of Molecular Biology*, 182(1):179–182, 1985.

[53] G.Marius Clore and Angela M. Gronenborn. Probing the three-dimensional structures of dna and rna oligonucleotides in solution by nuclear overhauser enhancement measurements. *FEBS Letters*, 179(2):187 – 198, 1985.

[54] S. L. Beaucage and M. H. Caruthers. Deoxynucleoside phosphoramidites - a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Letters*, 22(20):1859 – 1862, 1981.

[55] N. D. Sinha, J. Biernat, J. Mcmanus, and H. Koster. Polymer support oligonucleotide synthesis .18. use of beta-cyanoethyl-n,n-dialkylamino-/n-morpholino phosphoramidite of deoxynucleosides for the synthesis of dna fragments simplifying deprotection and isolation of the final product. *Nucleic Acids Research*, 12(11):4539–4557, 1984.

[56] J. C. Schulhof, D. Molko, and R. Teoule. The final deprotection step in oligonucleotide synthesis is reduced to a mild and rapid ammonia treatment by using labile base-protecting groups. *Nucleic Acids Research*, 15(2):397–416, January 1987.

[57] M. H. Caruthers, A. D. Barone, S. L. Beaucage, D. R. Dodds, E. F. Fisher, L. J. Mcbride, M. Matteucci, Z. Stabinsky, and J. Y. Tang. Chemical synthesis of deoxy-

oligonucleotides by the phosphoramidite method. *Methods In Enzymology*, 154:287–313, 1987.

[58] B. F. Li, P. F. Swann, M. Kalnik, and D. J. Patel. Synthesis and structural studies by nuclear magnetic resonance of dodecadeoxynucleotides containing o6-methylguanine, o6-ethylguanine and o4-methylthymine. *IARC scientific publications*, (84):44–8, 1987.

[59] M. W. Kalnik, M. Kouchakdjian, B. F. L. Li, P. F. Swann, and D. J. Patel. Base pair mismatches and carcinogen-modified bases in dna - an nmr-study of a.c and a.o4met pairing in dodecanucleotide duplexes. *Biochemistry*, 27(1):100–108, January 1988.

[60] Frederick E. Evans and Robert A. Levine. Nmr study of stacking interactions and conformational adjustments in the dinucleotide-carcinogen adduct 2'-deoxycytidylyl-(3 .fwdarw. 5)-2'-deoxy-8-(n-fluoren-2-ylacetamido)guanosine. *Biochemistry*, 27(8):3046–3055, 1988.

[61] J. G. Pelton and D. E. Wemmer. Structural modeling of the distamycin-a-d(cgcgaattcgcg)$_2$ complex using 2d nmr and molecular mechanics. *Biochemistry*, 27(21):8088–8096, October 1988.

[62] J. M. L. Pieters, E. Devroom, G. A. Vandermarel, J. H. Vanboom, and C. Altona. Conformational consequences of the incorporation of arabinofuranosylcytidine in dna - an nmr-study of the dna fragments d(cgctagcg) and d(cgactagcg) in solution. *European Journal of Biochemistry*, 184(2):415–425, September 1989.

[63] Peter M. M. Rae and Robert E. Steele. Modified bases in the dnas of unicellular eukaryotes: an examination of distributions and possible roles, with emphasis on hydroxymethyluracil in dinoflagellates. *Biosystems*, 10(1-2):37 – 53, 1978.

[64] R. A. J. Warren. Modified bases in bacteriophage dnas. *Annual Review of Microbiology*, 34:137–158, 1980.

[65] Mark Lukin and Carlos de los Santos. Nmr structures of damaged dna. *Chemical Reviews*, 106(2):607–686, 2006. PMID: 16464019.

[66] P. E. Nielsen, M. Egholm, R. H. Berg, and O. Buchardt. Sequence-selective recognition of dna by strand displacement with a thymine-substituted polyamide. *Science*, 254(5037):1497–1500, December 1991.

[67] J. C. Hanvey, N. J. Peffer, J. E. Bisi, S. A. Thomson, R. Cadilla, J. A. Josey, D. J. Ricca, C. F. Hassman, M. A. Bonham, K. G. Au, S. G. Carter, D. A. Bruckenstein, A. L. Boyd, S. A. Noble, and L. E. Babiss. Antisense and antigene properties of peptide nucleic-acids. *Science*, 258(5087):1481–1485, November 1992.

[68] A. A. Koshkin, S. K. Singh, P. Nielsen, V. K. Rajwanshi, R. Kumar, M. Meldgaard, C. E. Olsen, and J. Wengel. Lna (locked nucleic acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition. *Tetrahedron*, 54(14):3607–3630, April 1998.

[69] Jakob T. Nielsen, Khalil Arar, and Michael Petersen. Solution structure of a locked nucleic acid modified quadruplex: Introducing the v4 folding topology. *Angewandte Chemie*, 121(17):3145–3149, 2009.

[70] K.-U. Schöning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy, and A. Eschenmoser. Chemical etiology of nucleic acid structure: The alpha-threofuranosyl-(3'->2') oligonucleotide system. *Science*, 290(5495):1347–1351, 2000.

[71] N. Ueda, T. Kawabata, and K. Takemoto. Synthesis of n-(2,3-dihydroxypropyl) derivatives of nucleic bases. *Journal of Heterocyclic Chemistry*, 8(5):827–&, 1971.

[72] O. L. Acevedo and R. S. Andrews. Synthesis of propane-2,3-diol combinatorial monomers. *Tetrahedron Letters*, 37(23):3931–3934, June 1996.

[73] L. L. Zhang, A. Peritz, and E. Meggers. A simple glycol nucleic acid. *Journal of the American Chemical Society*, 127(12):4174–4175, March 2005.

[74] Allen T. Horhota, Jack W. Szostak, and Larry W. McLaughlin. Glycerol nucleoside triphosphates: Synthesis and polymerase substrate activities. *Organic Letters*, 8(23):5345–5347, 2006. PMID: 17078714.

[75] D. Pinkel, T. Straume, and J. W. Gray. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proceedings of the National Academy of Sciences of the United States of America*, 83(9):2934–2938, May 1986.

[76] T. R. Krugh, D. E. Graves, and M. P. Stone. 2-dimensional nmr-studies on the anthramycin-d(atgcat)2 adduct. *Biochemistry*, 28(26):9988–9994, December 1989.

[77] A. C. Pease, D. Solas, E. J. Sullivan, M. T. Cronin, C. P. Holmes, and S. P. A. Fodor. Light-generated oligonucleotide arrays for rapid dna-sequence analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 91(11):5022–5026, May 1994.

[78] J. L. Schwartz, J. S. Rice, B. A. Luxon, J. M. Sayer, G. Xie, H. J. C. Yeh, X. Liu, D. M. Jerina, and D. G. Gorenstein. Solution structure of the minor conformer of a dna duplex containing a dg mismatch opposite a benzo[a]pyrene diol epoxide/da adduct: Glycosidic rotation from syn to anti at the modified deoxyadenosine. *Biochemistry*, 36(37):11069–11076, 1997. PMID: 9333324.

[79] R. T. Ranasinghe and T. Brown. Fluorescence based strategies for genetic analysis. *Chemical Communications*, (44):5487–5502, November 2005.

[80] J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. Dewinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Ko-

rlach, and S. Turner. Real-time dna sequencing from single polymerase molecules. *Science*, 323(5910):133–138, January 2009.

[81] F. Sanger, S. Nicklen, and A. R. Coulson. Dna sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12):5463–5467, 1977.

[82] L. M. Smith, J. Z. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connell, C. Heiner, S. B. H. Kent, and L. E. Hood. Fluorescence detection in automated dna-sequence analysis. *Nature*, 321(6071):674–679, June 1986.

[83] L. S. Lerman. Structural considerations in interaction of dna and acridines. *Journal of Molecular Biology*, 3(1):18–30, 1961.

[84] A. Fede, M. Billeter, W. Leupin, and K. Wüthrich. Determination of the nmr solution structure of the hoechst 33258-d(gtggaattccac)$_2$ complex and comparison with the x-ray crystal-structure. *Structure*, 1(3):177–186, November 1993.

[85] H. P. Spielmann, D. E. Wemmer, and J. P. Jacobsen. Solution structure of a dna complex with the fluorescent bis-intercalator toto determined by nmr-spectroscopy. *Biochemistry*, 34(27):8542–8553, July 1995.

[86] A. N. Glazer and R. A. Mathies. Energy-transfer fluorescent reagents for dna analyses. *Current Opinion In Biotechnology*, 8(1):94–102, February 1997.

[87] K. E. Erkkila, D. T. Odom, and J. K. Barton. Recognition and reaction of metallointercalators with dna. *Chemical Reviews*, 99(9):2777–2795, September 1999.

[88] G. Subramaniam, M. M. Paz, G. S. Kumar, A. Das, Y. Palom, C. C. Clement, D. J. Patel, and M. Tomasz. Solution structure of a guanine-n7-linked complex of the mitomycin c metabolite 2,7-diaminomitosene and dna. basis of sequence selectivity. *Biochemistry*, 40(35):10473–10484, September 2001.

[89] Adolf Wacker, Sigrid Kirschfeld, and Lothar Träger. Über den einbau purin-analoger verbindungen in die bakterien-nukleinsäure. *Journal of Molecular Biology*, 2(4):241 – 242, 1960.

[90] Adolf Wacker, Sigrid Kirschfeld, Dieter Hartmann, and Dieter Weinblum. Über den einbau von 5-nitrouracil, 5-aminouracil und 2-thiothymin in die bakterien-desoxyribonukleinsäure. *Journal of Molecular Biology*, 2(1):69 – 71, 1960.

[91] Jorge R. Barrio, John A. Secrist III, and Nelson J. Leonard. Fluorescent adenosine and cytidine derivatives. *Biochemical and Biophysical Research Communications*, 46(2):597 – 604, 1972.

[92] T. M. Nordlund, S. Andersson, L. Nilsson, R. Rigler, A. Graeslund, and L. W. McLaughlin. Structure and dynamics of a fluorescent dna oligomer containing the ecori recognition sequence: fluorescence, molecular dynamics, and nmr studies. *Biochemistry*, 28(23):9095–9103, 1989. PMID: 2605243.

[93] K. M. Guckian, T. R. Krugh, and E. T. Kool. Solution structure of a dna duplex containing a replicable difluorotoluene-adenine pair. *Nature Structural Biology*, 5(11):954–959, November 1998.

[94] James N. Wilson and Eric T. Kool. Fluorescent dna base replacements: reporters and sensors for biological systems. *Organic & Biomolecular Chemistry*, 4:4265–4274, 2006.

[95] Vyacheslav V. Filichev and Erik B. Pedersen. *DNA-Conjugated Organic Chromophores in DNA Stacking Interactions*. John Wiley & Sons, Inc., 2008.

[96] R. W. Sinkeldam, N. J. Greco, and Y. Tor. Fluorescent analogs of biomolecular building blocks: Design, properties, and applications. *Chemical Reviews*, 110(5):2579–2619, May 2010.

[97] Tracy J. Matray and Eric T. Kool. Selective and stable dna base pairing without hydrogen bonds. *Journal of the American Chemical Society*, 120(24):6191–6192, 1998.

[98] K. M. Guckian, T. R. Krugh, and E. T. Kool. Solution structure of a nonpolar, non-

hydrogen-bonded base pair surrogate in dna. *Journal of the American Chemical Society*, 122(29):6841–6847, July 2000.

[99] D. Loakes. The applications of universal dna base analogues. *Nucleic Acids Research*, 29(12):2437–2447, June 2001.

[100] S. Tyagi and F. R. Kramer. Molecular beacons: Probes that fluoresce upon hybridization. *Nature Biotechnology*, 14(3):303–308, March 1996.

[101] S. Tyagi, D. P. Bratu, and F. R. Kramer. Multicolor molecular beacons for allele discrimination. *Nature Biotechnology*, 16(1):49–53, January 1998.

[102] B. Dubertret, M. Calame, and A. J. Libchaber. Single-mismatch detection using gold-quenched fluorescent oligonucleotides. *Nature Biotechnology*, 19(4):365–370, April 2001.

[103] K. M. Wang, Z. W. Tang, C. Y. J. Yang, Y. M. Kim, X. H. Fang, W. Li, Y. R. Wu, C. D. Medley, Z. H. Cao, J. Li, P. Colon, H. Lin, and W. H. Tan. Molecular engineering of dna: Molecular beacons. *Angewandte Chemie-international Edition*, 48(5):856–870, 2009.

[104] Frédéric Godde, Jean-Jacques Toulmé, and Serge Moreau. Benzoquinazoline derivatives as substitutes for thymine in nucleic acid complexes. use of fluorescence emission of benzo[g]quinazoline-2,4-(1h,3h)-dione in probing duplex and triplex formation. *Biochemistry*, 37(39):13765–13775, 1998.

[105] Akimitsu Okamoto, Yoshio Saito, and Isao Saito. Design of base-discriminating fluorescent nucleosides. *Journal of Photochemistry and Photobiology C: Photochemistry Reviews*, 6(2-3):108 – 122, 2005.

[106] H.-A. Wagenknecht. Fluorescent dna base modifications and substitutes: Multiple fluorophore labeling and the deteq concept. *Fluorescence Methods and Applications: Spectroscopy, Imaging, and Probes*, 1130:122–130, 2008.

[107] Olaf Köhler, Dilip Venkatrao Jarikote, and Oliver Seitz. Forced intercalation probes (fit probes): Thiazole orange as a fluorescent base in peptide nucleic acids for homogeneous single-nucleotide-polymorphism detection. *ChemBioChem*, 6(1):69–77, 2005.

[108] C. A. Heid, J. Stevens, K. J. Livak, and P. M. Williams. Real time quantitative pcr. *Genome Research*, 6(10):986–994, October 1996.

[109] S. A. Bustin, V. Benes, J. A. Garson, J. Hellemans, J. Huggett, M. Kubista, R. Mueller, T. Nolan, M. W. Pfaffl, G. L. Shipley, J. Vandesompele, and C. T. Wittwer. The miqe guidelines: Minimum information for publication of quantitative real-time pcr experiments. *Clinical Chemistry*, 55(4):611–622, April 2009.

[110] K. J. Livak, S. J. A. Flood, J. Marmaro, W. Giusti, and K. Deetz. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting pcr product and nucleic-acid hybridization. *Pcr-methods and Applications*, 4(6):357–362, June 1995.

[111] Elke Socher, Dilip V. Jarikote, Andrea Knoll, Lars Röglin, Jens Burmeister, and Oliver Seitz. Fit probes: Peptide nucleic acid probes with a fluorescent base surrogate enable real-time dna quantification and single nucleotide polymorphism discovery. *Analytical Biochemistry*, 375(2):318 – 330, 2008.

[112] Felix Hövelmann, Lucas Bethge, and Oliver Seitz. Single labeled dna fit probes for avoiding false-positive signaling in the detection of dna/rna in qpcr or cell media. *ChemBioChem*, 13(14):2072–2081, 2012.

[113] S. Dhanasekaran, T. M. Doherty, and J. Kenneth. Comparison of different standards for real-time pcr-based absolute quantification. *Journal of Immunological Methods*, 354(1-2):34–39, March 2010.

[114] L. Stryer. Fluorescence energy-transfer as a spectroscopic ruler. *Annual Review of Biochemistry*, 47:819–846, 1978.

[115] J.Y. Ju, I. Kheterpal, J.R. Scherer, C.C. Ruan, C.W. Fuller, A.N. Glazer, and R.A. Mathies. Design and synthesis of fluorescence energy transfer dye-labeled primers and their application for dna sequencing and analysis. *Analytical Biochemistry*, 231(1):131 – 140, 1995.

[116] D. M. J. Lilley and T. J. Wilson. Fluorescence resonance energy transfer as a structural tool for nucleic acids. *Current Opinion In Chemical Biology*, 4(5):507–517, October 2000.

[117] O. Schiemann, N. Piton, Y. G. Mu, G. Stock, J. W. Engels, and T. F. Prisner. A peldor-based nanometer distance ruler for oligonucleotides. *Journal of the American Chemical Society*, 126(18):5722–5729, May 2004.

[118] O. Schiemann and T. F. Prisner. Long-range distance determinations in biomacromolecules by epr spectroscopy. *Quarterly Reviews of Biophysics*, 40(1):1–53, February 2007.

[119] J. M. Obliosca, C. Liu, and H. C. Yeh. Fluorescent silver nanoclusters as dna probes. *Nanoscale*, 5(18):8443–8461, 2013.

[120] C. Z. Wan, T. Fiebig, O. Schiemann, J. K. Barton, and A. H. Zewail. Femtosecond direct observation of charge transfer between bases in dna. *Proceedings of the National Academy of Sciences of the United States of America*, 97(26):14052–14055, December 2000.

[121] A. H. Zewail. Femtochemistry: Atomic-scale dynamics of the chemical bond. *Journal of Physical Chemistry A*, 104(24):5660–5694, June 2000.

[122] J. L. N. Lustres, S. A. Kovalenko, M. Mosquera, T. Senyushkina, W. Flasche, and N. P. Ernsting. Ultrafast solvation of n-methyl-6-quinolone probes local ir spectrum. *Angewandte Chemie-international Edition*, 44(35):5635–5639, 2005.

[123] Mohsen Sajadi, Falko Berndt, Celin Richter, Mario Gerecke, Rainer Mahrwald, and Nikolaus P. Ernsting. Observing the hydration layer of trehalose with a linked

molecular terahertz probe. *The Journal of Physical Chemistry Letters*, 5(11):1845–1849, 2014.

[124] Anders Holmén, Bengt Nordén, and Bo Albinsson. Electronic transition moments of 2-aminopurine. *J. Am. Chem. Soc.*, 119:3114–3121, 1997.

[125] Lawrence C. Sowers, Yves Boulard, and G. Victor Fazakerley. Multiple structures for the 2-aminopurine-cytosine mispair. *Biochemistry*, 39:7613–7620, 2000.

[126] Edward L. Rachofsky, Roman Osman, and J. B. Alexander Ross. Probing structure and dynamics of dna with 2-aminopurine: Effects of local environment on fluorescence. *Biochemistry*, 40:946–956, 2001.

[127] S. K. Pal, L. Zhao, T. B. Xia, and A. H. Zewail. Site- and sequence-selective ultrafast hydration of dna. *Proceedings of the National Academy of Sciences of the United States of America*, 100(24):13746–13751, November 2003.

[128] D. C. Ward, E. Reich, and L. Stryer. Fluorescence studies of nucleotides and polynucleotides: I. formycin, 2,6-diaminopurine riboside, and 2-aminopurine riboside, polynucleotides: I. formycin,. *J. Biol. Chem.*, 244:1228–1237, 1969.

[129] André Dallmann, Lars Dehmel, Torben Peters, Clemens Mügge, Christian Griesinger, Jennifer Tuma, and Nikolaus P. Ernsting. 2-aminopurine incorporation perturbs the dynamics and structure of dna. *Angewandte Chemie International Edition*, 49(34):5989–5992, 2010.

[130] S. Hess, W. B. Davis, A. A. Voityuk, N. Rosch, M. E. Michel-Beyerle, N. P. Ernsting, S. A. Kovalenko, and J. L. P. Lustres. Excited-state photophysics of an acridine derivative selectively intercalated in duplex dna. *Chemphyschem*, 3(5):452, May 2002.

[131] Mary E. Hawkins, Wolfgang Pfleiderer, Oliver Jungmann, and Frank M. Balis. Synthesis and fluorescence characterization of pteridine adenosine nucleoside analogs for dna incorporation. *Analytical Biochemistry*, 298(2):231 – 240, 2001.

[132] Mary E. Hawkins, Wolfgang Pfleiderer, Frank M. Balis, Denise Porter, and Jay R. Knutson. Fluorescence properties of pteridine nucleoside analogs as monomers and incorporated into oligonucleotides. *Analytical Biochemistry*, 244(1):86 – 95, 1997.

[133] Mary E. Hawkins, Wolfgang Pfleiderer, Abhijit Mazumder, Yves G. Pommier, and Frank M. Balis. Incorporation of a fluorescent guanosine analog into oligonucleotides and its application to a real time assay for the hiv-1 integrase 3'-processing reaction. *Nucleic Acids Research*, 23(15):2872–2880, 1995.

[134] MaryE. Hawkins. Fluorescent pteridine nucleoside analogs. *Cell Biochemistry and Biophysics*, 34(2):257–281, 2001.

[135] K. Y. Lin, R. J. Jones, and M. Matteucci. Tricyclic 2'-deoxycytidine analogs - syntheses and incorporation into oligodeoxynucleotides which have enhanced binding to complementary rna. *Journal of the American Chemical Society*, 117(13):3873–3874, April 1995.

[136] Peter Sandin, L. Marcus Wilhelmsson, Per Lincoln, Vicki E. C. Powers, Tom Brown, and Bo Albinsson. Fluorescent properties of dna base analogue tc upon incorporation into dna — negligible influence of neighbouring bases on fluorescence quantum yield. *Nucleic Acids Research*, 33(16):5019–5025, 2005.

[137] P. Sandin, P. Lincoln, T. Brown, and L. M. Wilhelmsson. Synthesis and oligonucleotide incorporation of fluorescent cytosine analogue tc: a promising nucleic acid probe. *Nature Protocols*, 2(3):615–623, 2007.

[138] K. Cecilia Engman, Peter Sandin, Sadie Osborne, Tom Brown, Martin Billeter, Per Lincoln, Bengt Nordén, Bo Albinsson, and L. Marcus Wilhelmsson. Dna adopts normal b-form upon incorporation of highly fluorescent dna base analogue tc: Nmr structure and uv-vis spectroscopy characterization. *Nucleic Acids Research*, 32(17):5087–5095, 2004.

[139] Peter Sandin, Karl Börjesson, Hong Li, Jerker Mårtensson, Tom Brown, L. Marcus Wilhelmsson, and Bo Albinsson. Characterization and use of an unprecedentedly

bright and structurally non-perturbing fluorescent dna base analogue. *Nucleic Acids Research*, 36(1):157–167, 2008.

[140] K. Borjesson, S. Preus, A. H. El-Sagheer, T. Brown, B. Albinsson, and L. M. Wilhelmsson. Nucleic acid base analog fret-pair facilitating detailed structural measurements in nucleic acid containing systems. *Journal of the American Chemical Society*, 131(12):4288–4293, April 2009.

[141] B. J. Rodgers, N. A. Elsharif, N. Vashisht, M. M. Mingus, M. A. Mulvahill, G. Stengel, R. D. Kuchta, and B. W. Purse. Functionalized tricyclic cytosine analogues provide nucleoside fluorophores with improved photophysical properties and a range of solvent sensitivities. *Chemistry-a European Journal*, 20(7):2010–2015, February 2014.

[142] J. S. Woo, R. B. Meyer, and H. B. Gamper. G/c-modified oligodeoxynucleotides with selective complementarity: Synthesis and hybridization properties. *Nucleic Acids Research*, 24(13):2470–2475, July 1996.

[143] D. A. Berry, K. Y. Jung, D. S. Wise, A. D. Sercel, W. H. Pearson, H. Mackie, J. B. Randolph, and R. L. Somers. Pyrrolo-dc and pyrrolo-c: fluorescent analogs of cytidine and 2 '-deoxycytidine for the study of oligonucleotides. *Tetrahedron Letters*, 45(11):2457–2461, March 2004.

[144] C. Dash, J. W. Rausch, and S. F. J. Le Grice. Using pyrrolo-deoxycytosine to probe rna/dna hybrids containing the human immunodeficiency virus type-1 3 ' polypurine tract. *Nucleic Acids Research*, 32(4):1539–1547, February 2004.

[145] L. M. Wilhelmsson. Fluorescent nucleic acid base analogues. *Quarterly Reviews of Biophysics*, 43(2):159–183, May 2010.

[146] C. H. Liu and C. T. Martin. Fluorescence characterization of the transcription bubble in elongation complexes of t7 rna polymerase. *Journal of Molecular Biology*, 308(3):465–475, May 2001.

[147] Hong Zang, Qingming Fang, Anthony E. Pegg, and F. Peter Guengerich. Kinetic analysis of steps in the repair of damaged dna by human o6-alkylguanine-dna alkyltransferase. *The Journal of biological chemistry*, 280(35):30873–81, September 2005.

[148] A. A. Marti, S. Jockusch, Z. M. Li, J. Y. Ju, and N. J. Turro. Molecular beacons with intrinsically fluorescent nucleotides. *Nucleic Acids Research*, 34(6):e50, 2006.

[149] X. Ming and F. Seela. A nucleobase-discriminating pyrrolo-dc click adduct designed for dna fluorescence mismatch sensing. *Chemistry-a European Journal*, 18(31):9590–9600, July 2012.

[150] M. S. Noe, A. C. Rios, and Y. Tor. Design, synthesis, and spectroscopic properties of extended and fused pyrrolo-dc and pyrrolo-c analogs. *Organic Letters*, 14(12):3150–3153, June 2012.

[151] R. S. Coleman and M. L. Madaras. Synthesis of a novel coumarin c-riboside as a photophysical probe of oligonucleotide dynamics. *J. Org. Chem.*, 63(2):5700–5703, 1998.

[152] E. B. Brauns, M. L. Madaras, R. S. Coleman, C. J. Murphy, and M. A. Berg. Measurement of local dna reorganization on the picosecond and nanosecond time scales. *Journal of the American Chemical Society*, 121(50):11644–11649, December 1999.

[153] M. M. Somoza, D. Andreatta, C. J. Murphy, R. S. Coleman, and M. A. Berg. Effect of lesions on the dynamics of dna on the picosecond and nanosecond timescales using a polarity sensitive probe. *Nucleic Acids Research*, 32(8):2494–2507, April 2004.

[154] D. Andreatta, J. L. P. Lustres, S. A. Kovalenko, N. P. Ernsting, C. J. Murphy, R. S. Coleman, and M. A. Berg. Power-law solvation dynamics in dna over six decades in time. *Journal of the American Chemical Society*, 127(20):7270–7271, May 2005.

[155] D. Andreatta, S. Sen, J. L. P. Lustres, S. A. Kovalenko, N. P. Ernsting, C. J. Murphy,

R. S. Coleman, and M. A. Berg. Ultrafast dynamics in dna: "fraying" at the end of the helix. *Journal of the American Chemical Society*, 128(21):6885–6892, May 2006.

[156] R. S. Coleman, M. A. Berg, and C. J. Murphy. Coumarin base-pair replacement as a fluorescent probe of ultrafast dna dynamics. *Tetrahedron*, 63(17):3450–3456, April 2007.

[157] M. A. Berg, R. S. Coleman, and C. J. Murphy. Nanoscale structure and dynamics of dna. *Physical Chemistry Chemical Physics*, 10(9):1229–1242, 2008.

[158] E. B. Brauns, M. L. Madaras, R. S. Coleman, C. J. Murphy, and M. A. Berg. Complex local dynamics in dna on the picosecond and nanosecond time scales. *Physical Review Letters*, 88(15):158101, April 2002.

[159] Akimitsu Okamoto, Kazuki Tainaka, and Yoshimasa Fujiwara. Nile red nucleoside: Design of a solvatofluorochromic nucleoside as an indicator of micropolarity around dna. *The Journal of Organic Chemistry*, 71(9):3592–3598, 2006. PMID: 16626146.

[160] Tracy J. Matray and Eric T. Kool. A specific partner for abasic damage in dna. *Nature*, 399:704–708, 1999.

[161] Christine Brotschi, Adrian Häberli, and Christian J. Leumann. A stable dna duplex containing a non-hydrogen-bonding and non-shape-complementary base couple: Interstrand stacking as the stability determining factor. *Angewandte Chemie International Edition*, 40(16):3012–3014, 2001.

[162] Hugo Morales-Rojas and Eric T. Kool. A porphyrin c-nucleoside incorporated into dna. *Organic Letters*, 4(25):4377–4380, 2002.

[163] Simon M. Langenegger and Robert Häner. A simple, non-nucleosidic base surrogate increases the duplex stability of dna containing an abasic site. *Chemistry & Biodiversity*, 1(2):259–264, 2004.

[164] Simon M. Langenegger and Robert Häner. Remarkable stabilization of duplex dna containing an abasic site by non-nucleosidic phenanthroline and pyrene building blocks. *ChemBioChem*, 6(5):848–851, 2005.

[165] Natalia N. Dioubankova, Andrey D. Malakhov, Dmitry A. Stetsenko, Vladimir A. Korshun, and Michael J. Gait. (r)-2,4-dihydroxybutyramide seco-pseudonucleosides: New versatile homochiral synthons for synthesis of modified oligonucleotides. *Organic Letters*, 4(26):4607–4610, 2002. PMID: 12489941.

[166] Robert Huber, Nicole Amann, and Hans-Achim Wagenknecht. Synthesis of dna with phenanthridinium as an artificial dna base. *J. Org. Chem.*, 69:744–751, 2004.

[167] Jinsong Ren and Jonathan B. Chaires. Sequence and structural selectivity of nucleic acid binding ligands. *Biochemistry*, 38(49):16067–16075, 1999.

[168] John Olmsted and David R. Kearns. Mechanism of ethidium bromide fluorescence enhancement on binding to nucleic acids. *Biochemistry*, 16(16):3647–3654, 1977.

[169] Shana O. Kelley and Jacqueline K. Barton. Dna-mediated electron transfer from a modified base to ethidium: Pi-stacking as a modulator of reactivity. *Chemistry & Biology*, 5(8):413 – 425, 1998.

[170] Linda Valis, Nicole Amann, and Hans-Achim Wagenknecht. Detection of single base mismatches and abasic sites using phenanthridinium as an artificial dna base and charge donor. *Org. Biomol. Chem.*, 3:36–38, 2005.

[171] Clemens Wagner and Hans-Achim Wagenknecht. Perylene-3,4:9,10-tetracarboxylic acid bisimide dye as an artificial dna base surrogate. *Organic Letters*, 8(19):4191–4194, 2006.

[172] Andrei D. Malakhov, Mikhail V. Skorobogatyi, Igor A. Prokhorenko, Sergei V. Gontarev, Dmitry T. Kozhich, Dmitry A. Stetsenko, Irina A. Stepanova, Zakhar O. Shenkarev, Yuri A. Berlin, and Vladimir A. Korshun. 1-(phenylethynyl)pyrene and

9,10-bis(phenylethynyl)anthracene, useful fluorescent dyes for dna labeling: Excimer formation and energy transfer. *European Journal of Organic Chemistry*, 2004(6):1298–1307, 2004.

[173] Hajime Maeda, Tomohiro Maeda, Kazuhiko Mizuno, Kazuhisa Fujimoto, Hisao Shimizu, and Masahiko Inouye. Alkynylpyrenes as improved pyrene-based biomolecular probes with the advantages of high fluorescence quantum yields and long absorption/emission wavelengths. *Chemistry A European Journal*, 12(3):824–831, 2006.

[174] Igor A. Prokhorenko, Andrei D. Malakhov, Anna A. Kozlova, Kuvat Momynaliev, Vadim M. Govorun, and Vladimir A. Korshun. Phenylethynylpyrene-labeled oligonucleotide probes for excimer fluorescence snp analysis of 23s rrna gene in clarithromycin-resistant helicobacter pylori strains. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 599(1â€“2):144 – 151, 2006.

[175] M. Kimoto, T. Mitsui, S. Yokoyama, and I. Hirao. A unique fluorescent base analogue for the expansion of the genetic alphabet. *Journal of the American Chemical Society*, 132(14):4988–+, April 2010.

[176] A. A. Tanpure and S. G. Srivatsan. Synthesis and photophysical characterisation of a fluorescent nucleoside analogue that signals the presence of an abasic site in rna. *Chembiochem*, 13(16):2392–2399, November 2012.

[177] M. G. Pawar and S. G. Srivatsan. Environment-responsive fluorescent nucleoside analogue probe for studying oligonucleotide dynamics in a model cell-like compartment. *Journal of Physical Chemistry B*, 117(46):14273–14282, November 2013.

[178] P. M. Sabale, A. Nuthanakanti, and S. G. Srivatsan. Synthesis and fluorescence properties of a full set of extended rna base analogues. *Indian Journal of Chemistry Section A-inorganic Bio-inorganic Physical Theoretical & Analytical Chemistry*, 52(8-9):1004–1013, August 2013.

[179] P. M. Sabale, J. T. George, and S. G. Srivatsan. A base-modified pna-graphene

oxide platform as a turn-on fluorescence sensor for the detection of human telomeric repeats. *Nanoscale*, 6(18):10460–10469, 2014.

[180] A. A. Tanpure and S. G. Srivatsan. Synthesis, photophysical properties and incorporation of a highly emissive and environment-sensitive uridine analogue based on the lucifer chromophore. *Chembiochem*, 15(9):1309–1316, June 2014.

[181] Carolin Holzhauser and Hans-Achim Wagenknecht. "dna traffic lights": Concept of wavelength-shifting dna probes and application in an aptasensor. *ChemBioChem*, 13(8):1136–1138, 2012.

[182] C. Holzhauser and H.-A. Wagenknecht. Dna and rna "traffic lights": Synthetic wavelength-shifting fluorescent probes based on nucleic acid base substitutes for molecular imaging. *Journal of Organic Chemistry*, 78(15):7373–7379, August 2013.

[183] S. Barrois, S. Worner, and H.-A. Wagenknecht. The role of duplex stability for wavelength-shifting fluorescent dna probes: energy transfer vs. exciton interactions in dna "traffic lights". *Photochemical & Photobiological Sciences*, 13(8):1126–1129, 2014.

[184] André Nadler, Julian Strohmeier, and Ulf Diederichsen. 8-vinyl-2'-deoxyguanosine as a fluorescent 2'-deoxyguanosine mimic for investigating dna hybridization and topology. *Angewandte Chemie International Edition*, 50(23):5392–5396, 2011.

[185] Bastian Holzberger, Julian Strohmeier, Vanessa Siegmund, Ulf Diederichsen, and Andreas Marx. Enzymatic synthesis of 8-vinyl- and 8-styryl-2'-deoxyguanosine modified dna - novel fluorescent molecular probes. *Bioorganic & Medicinal Chemistry Letters*, 22(9):3136 – 3139, 2012.

[186] V. Singh, S. L. Wang, and E. T. Kool. Genetically encoded multispectral labeling of proteins with polyfluorophores on a dna backbone. *Journal of the American Chemical Society*, 135(16):6184–6191, April 2013.

[187] L. H. Yuen, R. M. Franzini, S. L. Wang, P. Crisalli, V. Singh, W. Jiang, and E. T. Kool. Pattern-based detection of toxic metals in surface water with dna polyfluorophores. *Angewandte Chemie-international Edition*, 53(21):5361–5365, May 2014.

[188] Hiroyuki Asanuma, Taiga Fujii, Tomohiro Kato, and Hiromu Kashida. Coherent interactions of dyes assembled on dna. *Journal of Photochemistry and Photobiology C: Photochemistry Reviews*, 13(2):124 – 135, 2012. Photochemical Methodology Utilized in Chemical Biology.

[189] Y. N. Teo and E. T. Kool. Dna-multichromophore systems. *Chemical Reviews*, 112(7):4221–4245, July 2012.

[190] M. Winnacker and E. T. Kool. Artificial genetic sets composed of size-expanded base pairs. *Angewandte Chemie-international Edition*, 52(48):12498–12508, November 2013.

[191] Andrew T. Krueger, Haige Lu, Alex H. F. Lee, and Eric T. Kool. Synthesis and properties of size-expanded dnas: Toward designed, functional genetic systems. *Acc Chem Res.*, 40(2):141–150, 2007.

[192] F. Wojciechowski, J. Lietard, and C. J. Leumann. 2-pyrenyl-dna: Synthesis, pairing, and fluorescence properties. *Organic Letters*, 14(20):5176–5179, October 2012.

[193] Peter Kaden, Elke Mayer-Enthart, Anton Trifonov, Torsten Fiebig, and Hans-Achim Wagenknecht. Real-time spectroscopic and chemical probing of reductive electron transfer in dna. *Angewandte Chemie International Edition*, 44(11):1636–1639, 2005.

[194] Elke Mayer-Enthart and Hans-Achim Wagenknecht. Structure-sensitive and self-assembled helical pyrene array based on dna architecture. *Angewandte Chemie International Edition*, 45(20):3372–3375, 2006.

[195] Mykhailo Vybornyi, Alina L. Nussbaumer, Simon M. Langenegger, and Robert Häner. Assembling multiporphyrin stacks inside the dna double helix. *Bioconjugate Chemistry*, 25(10):1785–1793, 2014. PMID: 25186936.

[196] C. B. Winiger, S. M. Langenegger, O. Khorev, and R. Häner. Influence of perylenediimide-pyrene supramolecular interactions on the stability of dna-based hybrids: Importance of electrostatic complementarity. *Beilstein Journal of Organic Chemistry*, 10:1589–1595, July 2014.

[197] R. W. Sinkeldam, A. J. Wheat, H. Boyaci, and Y. Tor. Emissive nucleosides as molecular rotors. *Chemphyschem*, 12(3):567–570, February 2011.

[198] R. W. Sinkeldam, P. A. Hopkins, and Y. Tor. Modified 6-aza uridines: Highly emissive ph-sensitive fluorescent nucleosides. *Chemphyschem*, 13(14):3350–3356, October 2012.

[199] R. S. K. Lane, R. Jones, R. W. Sinkeldam, Y. Tor, and S. W. Magennis. Two-photon-induced fluorescence of isomorphic nucleobase analogs. *Chemphyschem*, 15(5):867–871, April 2014.

[200] M. S. Noe, R. W. Sinkeldam, and Y. Tor. Oligodeoxynucleotides containing multiple thiophene-modified isomorphic fluorescent nucleosides. *Journal of Organic Chemistry*, 78(16):8123–8128, August 2013.

[201] Patrycja A. Hopkins, Renatus W. Sinkeldam, and Yitzhak Tor. Visibly emissive and responsive extended 6-aza-uridines. *Organic letters*, 16(20):5290–3, October 2014.

[202] D. Shin, R. W. Sinkeldam, and Y. Tor. Emissive rna alphabet. *Journal of the American Chemical Society*, 133(38):14912–14915, September 2011.

[203] André Dallmann, Matthias Pfaffe, Clemens Mügge, Rainer Mahrwald, Sergey A. Kovalenko, and Nikolaus P. Ernsting. Local thz time domain spectroscopy of duplex dna via fluorescence of an embedded probe. *The Journal of Physical Chemistry B*, 113(47):15619–15628, 2009.

[204] F. Berndt, M. Sajadi, N. P. Ernsting, and R. Mahrwald. Covalent linkage of n-methyl-6-oxyquinolinium betaine to trehalose. *Carbohydrate Research*, 346(18):2960–2964, December 2011.

[205] M. Weinberger, F. Berndt, R. Mahrwald, N. P. Ernsting, and H. A. Wagenknecht. Synthesis of 4-aminophthalimide and 2,4-diaminopyrimidine c-nucleosides as isosteric fluorescent dna base substitutes. *Journal of Organic Chemistry*, 78(6):2589–2599, March 2013.

[206] G. Saroja, T. Soujanya, B. Ramachandram, and A. Samanta. 4-aminophthalimide derivatives as environment-sensitive probes. *Journal of Fluorescence*, 8(4):405–410, December 1998.

[207] John L Markley, Ad Bax, Yoji Arata, C.W Hilbers, Robert Kaptein, Brian D Sykes, Peter E Wright, and Kurt Wüthrich. Recommendations for the presentation of nmr structures of proteins and nucleic acids. *Journal of Molecular Biology*, 280(5):933 – 952, 1998.

[208] C. Altona and Sundaralingam M. Conformational-analysis of sugar ring in nucleosides and nucleotides - new description using concept of pseudorotation. *Journal of the American Chemical Society*, 94(23):8205–&, 1972.

[209] J. E. Kilpatrick, K. S. Pitzer, and R. Spitzer. The thermodynamics and molecular structure of cyclopentane. *Journal of the American Chemical Society*, 69(10):2483–2488, 1947.

[210] Albert W. Overhauser. Polarization of nuclei in metals. *Phys. Rev.*, 92:411–415, Oct 1953.

[211] F. A. L. Anet and A. J. R. Bourn. Nuclear magnetic resonance spectral assignments from nuclear overhauser effects. *Journal of the American Chemical Society*, 87(22):5250–&, 1965.

[212] S. Macura and R. R. Ernst. Elucidation of cross relaxation in liquids by two-dimensional nmr-spectroscopy. *Molecular Physics*, 41(1):95–117, 1980.

[213] G. C. K. Roberts, editor. *NMR of Macromolecules: A Practical Approach*. Oxford University Press, New York, USA, 1993.

[214] Richard R. Ernst, Geoffrey Bodenhausen, and Alexander Wokaun. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*. Oxford University Press, 1987.

[215] Angela M. Gronenborn and G. Marius Clore. Investigation of the solution structures of short nucleic acid fragments by means of nuclear overhauser enhancement measurements. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 17:1–32, 1985.

[216] Brian R. Reid, Kevin Banks, Peter Flynn, and Willy Nerdal. Nmr distance measurements in dna duplexes: sugars and bases have the same correlation times. *Biochemistry*, 28(26):10001–10007, December 1989.

[217] A. J. Birchall and A. N. Lane. Anisotropic rotation in nucleic acid fragments: significance for determination of structures from nmr data. *European Biophysics Journal*, 19(2):73–78, 1990.

[218] Giovanni Lipari and Attila Szabo. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. theory and range of validity. *Journal of the American Chemical Society*, 104(17):4546–4559, 1982.

[219] Giovanni Lipari and Attila Szabo. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. analysis of experimental results. *Journal of the American Chemical Society*, 104(17):4559–4570, 1982.

[220] Peter F. Flynn, Ramona J. Bieber Urbauer, Hui Zhang, Andrew L. Lee, and A. Joshua Wand. Main chain and side chain dynamics of a heme protein: 15n and 2h nmr relaxation studies of r. capsulatus ferrocytochrome c2. *Biochemistry*, 40(22):6559–6569, 2001. PMID: 11380250.

[221] Andrew N. Lane. Nmr studies of dynamics in nucleic acids. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 25(5):481 – 505, 1993.

[222] A. N. Lane. Influence of conformational averaging on ^{1}h-h noes and structure determination in dna. *Magnetic Resonance In Chemistry*, 34:S3–S10, December 1996.

[223] Eike Brunner. Residual dipolar couplings in protein nmr. *Concepts in Magnetic Resonance*, 13(4):238–259, 2001.

[224] J. H. Prestegard, C. M. Bougault, and A. I. Kishore. Residual dipolar couplings in structure determination of biomolecules. *Chemical Reviews*, 104(8):3519–3540, 2004. PMID: 15303825.

[225] Martin Blackledge. Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 46(1):23 – 61, 2005.

[226] D. MacDonald, K. Herbert, X. L. Zhang, T. Polgruto, and P. Lu. Solution structure of an a-tract dna bend. *Journal of Molecular Biology*, 306(5):1081–1098, March 2001.

[227] A. Saupe and G. Englert. High-resolution nuclear magnetic resonance spectra of orientated molecules. *Phys. Rev. Lett.*, 11(10):462, November 1963.

[228] A. Saupe. Kernresonanzen in kristallinen flussigkeiten + in kristallinflussigen losungen .i. *Zeitschrift Fur Naturforschung Part A-astrophysik Physik Und Physikalische Chemie*, A 19(2):161–+, 1964.

[229] J. R. Tolman, J. M. Flanagan, M. A. Kennedy, and J. H. Prestegard. Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proceedings of the National Academy of Sciences*, 92(20):9279–9283, 1995.

[230] Ad Bax and Nico Tjandra. High-resolution heteronuclear nmr of human ubiquitin in an aqueous liquid crystalline medium. *Journal of Biomolecular NMR*, 10(3):289–292, 1997.

[231] Nico Tjandra and Ad Bax. Direct measurement of distances and angles in biomolecules by nmr in a dilute liquid crystalline medium. *Science*, 278(5340):1111–1114, 1997.

[232] Marcel Ottiger and Ad Bax. Bicelle-based liquid crystals for nmr-measurement of dipolar couplings at acidic and basic ph values. *Journal of biomolecular NMR*, 13(2):187–91, February 1999.

[233] H. M. Al-Hashimi, H. Valafar, M. Terrell, E. R. Zartler, M. K. Eidsness, and J. H. Prestegard. Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings. *Journal of Magnetic Resonance*, 143(2):402–406, April 2000.

[234] L. G. Barrientos, C. Dolan, and A. M. Gronenborn. Characterization of surfactant liquid crystal phases suitable for molecular alignment and measurement of dipolar couplings. *Journal of biomolecular NMR*, 16(4):329–37, April 2000.

[235] Markus Ruckert and Gottfried Otting. Alignment of biological macromolecules in novel nonionic liquid crystalline media for nmr experiments. *Journal of the American Chemical Society*, 122(32):7793–7797, August 2000.

[236] M R Hansen, L Mueller, and a Pardi. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature structural biology*, 5(12):1065–74, December 1998.

[237] James Chou, Sander Gaemers, Bernard Howder, John Louis, and Ad Bax. A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles. *Journal of Biomolecular NMR*, 21(4):377–382, December 2001.

[238] Junhe Ma, Gregory I. Goldberg, and Nico Tjandra. Weak alignment of biomacromolecules in collagen gels: an alternative way to yield residual dipolar couplings for nmr measurements. *Journal of the American Chemical Society*, 130(48):16148–9, December 2008.

[239] Jens Wohnert, Katherine J Franz, Mark Nitz, Barbara Imperiali, and Harald Schwalbe. Protein alignment by a coexpressed lanthanide-binding tag for the mea-

surement of residual dipolar couplings. *Journal of the American Chemical Society*, 125(44):13338–13339, November 2003.

[240] Shawn M. Douglas, James J. Chou, and William M. Shih. Dna-nanotube-induced alignment of membrane proteins for nmr structure determination. *Proceedings of the National Academy of Sciences*, 104(16):6644–6648, 2007.

[241] J. H. Prestegard, H. M. Al-Hashimi, and J. R. Tolman. Nmr structures of biomolecules using field oriented media and residual dipolar couplings. *Quarterly reviews of biophysics*, 33(4):371–424, November 2000.

[242] Eva de Alba and Nico Tjandra. Nmr dipolar couplings for the structure determination of biopolymers in solution. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 40(2):175–197, 2002.

[243] J. A. Losonczi, M. Andrec, M. W. Fischer, and J. H. Prestegard. Order matrix analysis of residual dipolar couplings using singular value decomposition. *Journal of magnetic resonance (San Diego, Calif. : 1997)*, 138(2):334–42, June 1999.

[244] Axel T. Brünger. *X-PLOR (Version 4.0) A System for X-Ray Crystallography and NMR*. Yale University, 1996.

[245] Charles D. Schwieters, John J. Kuszewski, Nico Tjandra, and G. Marius Clore. The xplor-nih nmr molecular structure determination package. *J. Magn. Reson.*, 160:65–73, 2003.

[246] Scott J. Weiner, Peter A. Kollman, David A. Case, U. Chandra Singh, Caterina Ghio, Guliano Alagona, Salvatore Profeta, and Paul Weiner. A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society*, 106(3):765–784, 1984.

[247] Alexander D. MacKerell, Nilesh Banavali, and Nicolas Foloppe. Development and current status of the charmm force field for nucleic acids. *Biopolymers*, 56(4):257–265, 2000.

[248] J. L. Mergny and L. Lacroix. Analysis of thermal melting curves. *Oligonucleotides*, 13(6):515–537, 2003.

[249] Vyacheslav V. Filichev and Erik B. Pedersen. Stable and selective formation of hoogsteen-type triplexes and duplexes using twisted intercalating nucleic acids (tina) prepared via postsynthetic sonogashira solid-phase coupling reactions. *Journal of the American Chemical Society*, 127(42):14849–14858, 2005.

[250] Francis H. Martin, Olke C. Uhlenbeck, and Paul Doty. Self-complementary oligoribonucleotides: Adenylic acid-uridylic acid block copolymers. *Journal of Molecular Biology*, 57(2):201 – 215, 1971.

[251] Philip N. Borer, Barbara Dengler, Ignacio Tinoco Jr., and Olke C. Uhlenbeck. Stability of ribonucleic acid double-stranded helices. *Journal of Molecular Biology*, 86(4):843 – 853, 1974.

[252] Matthew Petersheim and Douglas H. Turner. Base-stacking and base-pairing contributions to helix stability: Thermodynamics of double-helix formation with ccgg, ccggp, ccggap, accggp, ccggup and accggupt. *Biochemistry*, 22:256–263, 1983.

[253] D. H. Turner, N. Sugimoto, and S. M. Freier. Rna structure prediction. *Annual Review of Biophysics and Biophysical Chemistry*, 17(1):167–192, 1988.

[254] Vassili Ivanov, Yan Zeng, and Giovanni Zocchi. Statistical mechanics of base stacking and pairing in dna melting. *Phys. Rev. E*, 70:051907, Nov 2004.

[255] John SantaLucia. A unified view of polymer, dumbbell, and oligonucleotide dna nearest-neighbor thermodynamics. *Proc Natl Acad Sci*, 95(4):1460–1465, 1998.

[256] John SantaLucia. The thermodynamics of dna structural motifs. *Annual Review of Biophysics and Biomolecular Structure*, 33:415–440, 2004.

[257] R. S. Lipsitz and N. Tjandra. Residual dipolar couplings in nmr structure analysis. *Annual Review of Biophysics and Biomolecular Structure*, 33:387–413, 2004.

[258] Curt M. Breneman and Kenneth B. Wiberg. Determining atom-centered monopoles from molecular electrostatic potentials. the need for high sampling density in formamide conformational analysis. *Journal of Computational Chemistry*, 11(3):361–373, 1990.

[259] R. L. J. Keller. *Optimizing the process of Nuclear Magnetic Resonance spectrum analysis and Computer Aided Resonance Assignment (Cara)*. PhD thesis, Swiss Federal Institute Of Technology Zurich, 2004.

[260] André Dallmann. *Structure and dynamics of fluorophore-labelled DNA helices probed by NMR-spectroscopy*. PhD thesis, Humboldt-Universität zu Berlin, 2010.

[261] Marcel Ottiger and Ad Bax. How tetrahedral are methyl groups in proteins? a liquid crystal nmr study. *Journal of the American Chemical Society*, 121(19):4690–4695, May 1999.

[262] Annaleen Vermeulen, Hongjun Zhou, and Arthur Pardi. Determining dna global structure and dna bending by application of nmr residual dipolar couplings. *Journal of the American Chemical Society*, 122(40):9638–9647, 2000.

[263] Olivier Mauffret, Georges Tevanian, and Serge Fermandjian. Residual dipolar coupling constants and structure determination of large dna duplexes. *Journal of biomolecular NMR*, 24(4):317–28, December 2002.

[264] R. Boelens, T. M. G. Koning, and R. Kaptein. Determination of biomolecular structures from proton-proton noe's using a relaxation matrix approach. *Journal of Molecular Structure*, 173:299–311, 1988.

[265] B. A. Borgias and T. L. James. Mardigras - a procedure for matrix analysis of relaxation for discerning geometry of an aqueous structure. *J. Magn. Res.*, 87:475–487, 1990.

[266] Lars Dehmel. Die entwicklung eines lua-skripts zur automatischen integration von noesy-spektren. Master's thesis, Humboldt-Universität zu Berlin, 2009.

[267] Jean-Luc Pons, T. E. Malliavin, and Marc A. Delsuc. Gifa v. 4: A complete package for nmr data set processing. *Journal of Biomolecular NMR*, 8(4):445–452, December 1996.

[268] Markus Zweckstetter. Nmr: prediction of molecular alignment from structure using the pales software. *Nat. Protocols*, 3(4):679–690, 2008.

[269] R. Manoharan and Sneh K. Dogra. The absorption and fluorescence spectra of 2- and 4-fluorenecarboxylic acids. *Journal of Photochemistry and Photobiology A: Chemistry*, 43(2):119 – 137, 1988.

[270] K. Huang, J. M. Louis, L. Donaldson, F. L. Lim, A. D. Sharrocks, and G. M. Clore. Solution structure of the mef2a-dna complex: structural basis for the modulation of dna bending and specificity by mads-box transcription factors. *Embo Journal*, 19(11):2615–2628, June 2000.

[271] M. Nakata, G. Zanchetta, B. D. Chapman, C. D. Jones, J. O. Cross, R. Pindak, T. Bellini, and N. A. Clark. End-to-end stacking and liquid crystal condensation of 6-to 20-base pair dna duplexes. *Science*, 318(5854):1276–1279, November 2007.

[272] Mohsen Sajadi, Yathrib Ajaj, Ilya Ioffe, Hermann Weingärtner, and Nikolaus P. Ernsting. Terahertz absorption spectroscopy of a liquid using a polarity probe: A case study of trehalose/water mixtures. *Angewandte Chemie International Edition*, 49(2):454–457, 2010.

[273] Mark K. Schlegel, Lars-Oliver Essen, and Eric Meggers. Atomic resolution duplex structure of the simplified nucleic acid gna. *Chem. Commun.*, 46:1094–1096, 2010.

[274] Andrew T. Johnson, Mark K. Schlegel, Eric Meggers, Lars-Oliver Essen, and Olaf Wiest. On the structure and dynamics of duplex gna. *The Journal of Organic Chemistry*, 76(19):7964–7974, 2011. PMID: 21838272.

[275] Eric Meggers and Lilu Zhang. Synthesis and properties of the simplified nucleic acid

glycol nucleic acid. *Accounts of Chemical Research*, 43(8):1092–1102, 2010. PMID: 20405911.

[276] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes 2nd Edition: The Art of Scientific Computing*. Cambridge University Press, 1992.

[277] H. E. Gottlieb, V. Kotlyar, and A. Nudelman. Nmr chemical shifts of common laboratory solvents as trace impurities. *Journal of Organic Chemistry*, 62(21):7512–7515, October 1997.

List of Abbreviations

2AP	**2**-**A**mino**p**urine
4AP	**4**-**A**mino**p**hthalimide
6HQ	**6**-**H**ydroxy-**q**uinolinium
13mer4AP-DAP	**13mer** DNA double strand incorporating **4AP** and **DAP**
13mer6HQ	**13mer** DNA double strand incorporating **6HQ**
13merHCF	**13mer** DNA double strand incorporating **HCF**
13merHNF	**13mer** DNA double strand incorporating **HNF**
13merRef	**13mer** DNA double strand as **ref**erence with central AT pair
13merRef(GC)	**13mer** DNA double strand as **ref**erence with central GC pair
ACMA	9-**A**mino-6-**c**hloro-2-**m**ethoxy**a**cridine
A	**A**denine
B3LYP	**B**ecke, **3**-parameter, **L**ee-**Y**ang-**P**arr, a DFT method
CARA	**C**omputer **A**ided **R**esonance **A**ssignment
C	**C**ytosine
CPG	**C**ontrolled **P**ore **G**lass
CSD	**C**hemical **S**hift **D**eviation
DQF-COSY	**D**ouble **Q**uantum **F**iltered **Co**rrelated **S**pectroscop**y**
DAP	2,4-**Dia**mino**p**yrimidine
DETEQ	**D**etection by **E**lectron **T**ransfer-controlled **E**mission **Q**uenching
DFT	**D**ensity **F**unctional **T**heory
DNA	**D**eoxyribose **N**ucleic **A**cid
ddNTPs	**d**i-**d**eoxy**N**ucleotide**T**ri**p**hosphates
FID	**F**ree **I**nduction **D**ecay

FIT	**F**orced **I**ntercalation **T**O-PNA probes
FRET	**F**luorescence **R**esonance **E**nergy **T**ransfer
GNA	**G**lycol **N**ucleic **A**cid
G	**G**uanine
HCF	2-**H**ydroxy-7-**C**arboxy**f**luorene
HMQC	**H**eteronuclear **M**ultiple **Q**uantum **C**oherence
HNF	2-**H**ydroxy-7-**N**itro**f**luorene
HOD	**H**ydrogen **O**xygen **Deuterium**, half deuterated water
ISPA	**I**solated **S**pin **P**air **A**pproximation
LNA	**L**ocked **N**ucleic **A**cid
MD	**M**olecular **D**ynamics
MQ	N-**m**ethyl-6-**q**uinolone
NMR	**N**uclear **M**agnetic **R**esonance
NOESY	**N**uclear **O**verhauser **E**nhancement (**E**ffect) **S**pectroscop**y**
qPCR	Real-time **q**uantitative **P**olymerase **C**hain **R**eaction
PAS	**P**rincipal **A**xis **S**ystem
PCA	**P**rincipal **C**omponent **A**nalysis
PCR	**P**olymerase **C**hain **R**eaction
PELDOR	**P**ulsed **El**ectron-electron **Do**uble **R**esonance
Pf1	**P**hage **f1**, a filamentous bacteriophage
PNA	**P**eptide **N**ucleic **A**cid
RDC	**R**esidual **D**ipolar **C**oupling
RNA	**R**ibose **N**ucleic **A**cid
RMSD	**R**oot-**M**ean-**S**quare **D**eviation
SBW	**S**pectral **B**and**w**idth
SA	**S**imulated **A**nnealing
SNP	**S**ingle **N**ucleotide **P**olymorphism
SVD	**S**ingular **V**alue **D**ecomposition

TDSS	**Time-D**ependent **S**tokes **S**hift
T	**T**hymine
TNA	**T**hreose **N**ucleic **A**cid
TO	**T**hiazole **O**range
TR	**T**hiazole **R**ed
TOCSY	**To**tal **C**orrelation **S**pectroscop**y**
TZVP	**T**riple **Z**eta **V**alence plus **P**olarisation
UV/Vis	**U**ltra**v**iolet and **vis**ible range

List of Figures

List of Tables

Danksagung

Als Erstes möchte ich *Prof. Nikolaus P. Ernsting* für das Thema und die Betreuung meiner Doktorarbeit danken. Dazu zähle ich nicht nur die Unterstützung bei wissenschaftlichen Fragestellungen oder die Gelegenheit in Eigenregie experimentell zu arbeiten und neue Ideen zu verwirklichen, sondern auch das Vertrauen und die Möglichkeit mich neben den Herausforderungen eines Wissenschaftlers auch denen eines zweifachen Vaters zu stellen und diese unter einen Hut zu bekommen.

Besonders danken möchte ich *Dr. André Dallmann* für die Einführung in das Gebiet der Strukturaufklärung mittels NMR-Spektroskopie. Er stand mir immer mit Rat und Tat zur Seite, denn der Weg zur ersten Struktur ist mit vielen Stolpersteinen gepflastert.

Ebenso möchte ich *Prof. Dr. Clemens Mügge* für das Vertrauen und die Unterstützung am NMR-Spektrometer danken. Ohne die Freiheit lange und auch vielfältige Experimente am Spektrometer durchzuführen, wäre die Strukturbestimmung von 13mer4AP-DAP nicht möglich gewesen. An dieser Stelle möchte ich auch *Angela Thiesies* und *Katharina Pfaff* für die Kooperation und Unterstützung am Spektrometer danken.

Heiderose Steingräber und Sabrina Penn möchte ich für Hilfe bei den vielen kleinen Dingen im Alltag eines Wissenschaftlers danken sowie die angenehme Arbeitsatmosphäre.

Mein Dank gilt auch den anderen Mitgliedern des Arbeitskreises. *Dr. Horst Hennig* danke ich für die vielen anregenden Diskussionen, *Dr. Sergey Kovalenko* und *Dr. Alexander Dobryakov* für die Ratschläge zur wissenschaftlichen Arbeit. *Martin Quick* und *Mario Gerecke* für die gute Zusammenarbeit und stete Hilfsbereitschaft. Auch bei den ehemaligen Mitgliedern der Arbeitsgruppe möchte ich mich dafür bedanken. Im Einzelnen sind dies *Iris Suter*, *Dr. Luis Perez-Lustres*, *Dr. Mohsen Sajadi Hezaveh* und *Xinxing Zhang*.

Weiterhin möchte ich mich bei *Prof. Dr. Rainer Mahrwald* und *Dr. Matthias Pfaffe* für die Synthese des HCF-Bausteins bedanken. Ebenso bei *Felix Hövelmann* und *Prof. Dr. Oliver Seitz* für die Bereitstellung des 6HQ Stranges. Der 13mer4AP-DAP Doppelstrang ist in Kooperation mit der Arbeitsgruppe von *Prof. Dr. Hans-Achim Wagenknecht* entstanden. Ihm und *Dr. Michael Weinberger* danke ich für die Synthese des 4AP Stranges. Aus unserer Gruppe danke ich *Dr. Falko Berndt* für die Herstellung des DAP-Gegenstücks.

Schließlich möchte ich bei meiner Frau und unseren Eltern für die fortwährende Unterstützung bedanken, ohne die ich diese Arbeit nicht hätte fertigstellen können.

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 15.12.2014 Lars Dehmel