Title

Online Content : Physical and Structural Techniques Applied to Nucleic Acids

ISBN

9781788019040

Authors

Tor, Yitzhak

Gehring, Kalle

Fabris, Daniele

et al.

Publication Date

2022-06-24

DOI

10.1039/9781837671328-000e1

15 Physical and Structural Techniques Applied to Nucleic Acids

Yitzhak Tor[a], Kalle Gehring[b], Daniele Fabris[c], Martin Egli*[d], Andrei A. Korostelev[e], Timothy D. Craggs[f], Alice Pyne[f], Keith T. Gagnon[g], Jonathan K. Watts[e], Thomas E. Cheatham III[h] and Nigel G. J. Richards[i]

[a]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA, 92093, USA
[b]Department of Biochemistry McGill University, Montreal, QC H3G 0B, Canada
[c]Department of Chemistry, University of Connecticut, Storrs, CT, 06269, USA
[d]Department of Biochemistry, Vanderbilt University, School of Medicine, Nashville, TN, 37232, USA
[e]RNA Therapeutics Institute, University of Massachusetts Medical School, Worcester, MA, 01605, USA
[f]Department of Chemistry, University of Sheffield, Sheffield S3 7HF, UK
[g]Department of Biochemistry and Molecular Biology, School of Medicine, Southern Illinois University, Carbondale, IL 62901, USA
[h]Department of Medicinal Chemistry, University of Utah, College of Pharmacy, Salt Lake City, UT 84112, USA
[i] School of Chemistry, Cardiff University, Park Place, Cardiff CF10 3AT, UK

*E-mail: martin.egli@vanderbilt.edu

Contents

Nucleic Acids in Chemistry and Biology, 4th Edition
Edited by G. Michael Blackburn, Martin Egli, Michael J. Gait and Jonathan K. Watts

Published by the Royal Society of Chemistry, www.rsc.org

15.1 Spectroscopic Techniques

15.1.1 UV Absorption

The light absorption characteristics of nucleic acids result from the combination of the strong ultraviolet absorption of the purine and pyrimidine nucleobases in the 240–280 nm range modulated by the structural, context and conformational influence of a ribose-phosphate backbone that is essentially transparent to light of that wavelength (Section 2.1.3).[1]

Oligonucleotides exhibit strong UV absorption maxima, λ_{max}, at approximately 260 nm and a molar extinction coefficient, ε, of the order of 10^4 ($dm^3\ mol^{-1}\ cm^{-1}$) (Table 2.2). This absorption arises almost entirely from complex electronic transitions in the purine and pyrimidine components. The intensity and exact position of λ_{max} is a function not only of the base composition of the nucleic acid but also of the state of the base-pairing interactions present, the salt concentration of the solution, and its pH. Most importantly, base-on-base stacking results in a decrease in ε known as **hypochromicity**. This has been proposed to arise from dipole–dipole interactions[2] and distinct delocalization of excitonic states in duplexes *vs.* single stranded oligomers.[3] The magnitude of the hypochromic effect depends on the three-dimensional structure of an oligonucleotide and ranges from 1–11% for deoxyribonucleoside phosphates to 30% for most helical polynucleotides. In practice, the effects of structure on the UV absorption of oligonucleotides are so complex that only the most basic interpretations can be made (Figure 15.1a).

In practical terms it is possible to estimate the molar quantity of an oligonucleotide, for example chemically synthesized (Section 7.1), by measuring the number of absorbance units (the optical density) at 260 nm (A_{260}) of its solution in a UV spectrometer and by relating the value obtained to the sum of the extinction coefficients of the individual nucleosides (*e.g.*, 8.8 for dT, 7.3 for dC, 11.7 for dG and 15.4 for dA for A in $cm^2\ \mu mol^{-1}$).[4] Such estimation makes no allowance for hypochromicity, which will depend on the particular oligomeric sequence and whether or not it forms a secondary structure. More accurate determinations can be made by enzymatically degrading the oligomers into their constituent nucleosides and using the individual extinction coefficients listed above.

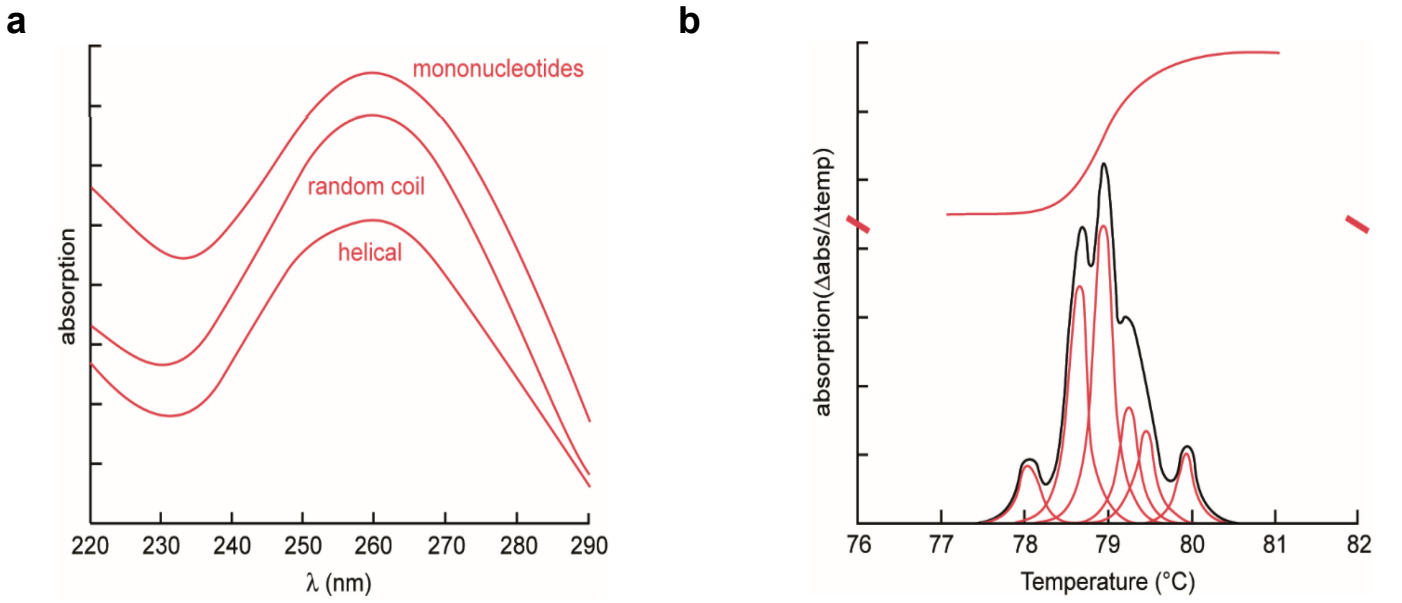

Figure 15.1 (a) Typical UV absorption curves for equimolar base concentrations of mononucleotides, of single-stranded (random coil) oligonucleotide, and of double-helical DNA. (b) An oligonucleotide melting curve (red), the derivative of the curve (black) and the deconvolution of the derivative into its composite components (red).

UV absorption is a sensitive and convenient way to monitor the 'melting behaviour' of DNA and RNA. When the UV absorption of a nucleic acid sample is measured as a function of temperature, the resulting thermal denaturation plot is known as a **melting curve**. The midpoint (or inflection point) in the increase in absorbance (**hyperchromic effect**) with increasing temperature is known as the **melting temperature**, T_m. This is dependent on the base-composition of the sample, the salt concentration of its solution, and even the type of counter-ion present (Sections 2.5.1 and 15.12). Such thermal melting is a co-operative phenomenon and the observed curves become progressively steeper with increasing length of the oligonucleotide. Simple sigmoidal melting curves are observed for many DNA samples, but it is also possible to observe more complex, multi-phasic melting in some cases (Figure 15.1b). The deconvolution of such multi-phase melting curves makes it possible to examine the effects of base-modification on the stability and nature of nucleic acid secondary structure in more detail.

15.1.2 Fluorescence

Fluorescence is defined as the emission of radiation as a molecule returns to its ground state from an excited electronic state.[5,6] To characterise the photoexcited emission from a molecule it is necessary to determine the spectral distribution, quantum yield, excited state lifetime and polarisation of the emission, all as a function of excitation wavelength. The precision with which fluorescence intensity can be measured is very high, since photon-counting techniques eliminate several sources of uncertainty and error. Excitation occurs in timescales of ~10^{-15} seconds, but the lifetime of the excited state of the fluorophore is around 10^{-9} seconds. Thus, any physical process that takes place on a similar timescale to the fluorescence lifetime can be analysed by examination of changes in the emission spectra, excited state lifetimes and polarisation.[5]

The fluorescence emission from nucleotides and dinucleoside phosphates is very weak at room temperature and can only be examined in frozen samples at ~80 K.[7] Fluorescence spectroscopy is nevertheless invaluable for examining nucleic acid–ligand interactions. Many DNA binding ligands (*e.g.*, ethidium salts, SYBR Green) have weak or no fluorescence in aqueous solution. However, after binding to a nucleic acid the ligand is in a hydrophobic environment and the solvent can no longer quench the intrinsic ligand fluorescence. Therefore, fluorescence emission is a direct probe of the concentration of bound ligand and can be used to directly or indirectly (*via* displacement measurements) assess binding events. More recent developments include the synthesis and implementation of emissive nucleoside surrogates that can frequently replace their native counterparts with minimal perturbations [8,9] and diverse approaches to fluorophore conjugation.[10]

Förster Resonance Energy Transfer (FRET) is a phenomenon in which the energy of an excited-state fluorescent donor molecule is transferred to an unexcited acceptor molecule *via* dipole–dipole coupling.[5] Importantly, the rate of energy transfer is dependent on the distance between donor and acceptor molecules, the spectral characteristics of the pair and the relative orientations of the donor and acceptor transition dipoles. FRET experiments have been used widely to determine proximity relationships in protein and nucleic acid systems since the 1940s when Förster derived the quantitative analysis for FRET and Stryer's demonstration of a FRET spectroscopic ruler in 1967.[11]

A key parameter for FRET is the efficiency of depopulation (E_T) and this is related to fluorescent lifetimes in the presence (τ_T) and absence (τ) of resonance energy transfer (eqn 15.1):

$$E_T = 1 - \frac{\tau_T}{\tau} \qquad (15.1)$$

The donor–acceptor distance (R) is related to E_T by the system constant R_o (the Förster radius), which is the distance at which there would be 50% of maximal energy transfer (equations 15.2 and 15.3):

$$E_T = \frac{R_0^6}{R^6 + R_0^6} \qquad (15.2) \qquad R = R^0 \left(\frac{1-E_T}{E_T}\right)^{1/6} \qquad (15.3)$$

From these relationships, FRET theory predicts that fluorescence transfer depends on the sixth power of the distance between partner molecules. In practice, the distances over which FRET can be measured vary between 40 and 100 Å, which depends, among other factors, on the nature of the donor and acceptor and their spectral overlap (Figure 15.2a). There are several factors that can influence the extent of fluorescence transfer and hence precautions need to be taken in order to account for all possible energy transfer pathways and to properly quantify FRET measurements. Despite these challenges, FRET techniques have been significantly advanced and widely used recently, quite commonly in single molecule measurements.[12]

A few naturally occurring fluorophores are found in nucleic acids, for example, **wyosine** is found in tRNA. More commonly, synthetic fluorophores can be introduced into oligonucleotides (*e.g.*, Section 8.3). Such modifications are extremely useful for studying nucleic acid structure and function. For FRET experiments in general, one or both of the donor and acceptor species is either covalently or non-covalently attached to a nucleic acid, often *via* chemical synthesis.[8–10]

Site-specific labelling of longer (thousands of base-pairs) DNA samples remains problematic, although recent advances in the use of sequence-specific methyl transferases to add specific labels to defined DNA sites are promising. Normally DNA methyl transferases (MTases) transfer an activated methyl group to the N-6 position of an adenine or C-5/N-4 position of a cytosine within a specific cognate sequence. By use of a fluorophore and a flexible linker, sequence-specifically labelled DNA can be obtained in a methyl transferase-catalysed reaction. This technique, developed by Elmar Weinhold, is known as **SMILing DNA**.[13] It has been expanded by Neely and Hofkens into **DNA Fluorocode**, a single molecule optical map of DNA.[14,15]

David Lilley and co-workers have used FRET to delineate the overall geometry of four-way DNA junctions[16] and the fold of the hammerhead ribozyme.[17] In addition FRET is a useful method for distinguishing between intercalation and groove-binding in small molecule–DNA interactions (Section 12.2), since in general FRET is observed during intercalation, but not during in groove-binding. Recent developments have expanded the utility of FRET measurements into high-resolution structural modelling and single-molecule measurements.[18,19,20]

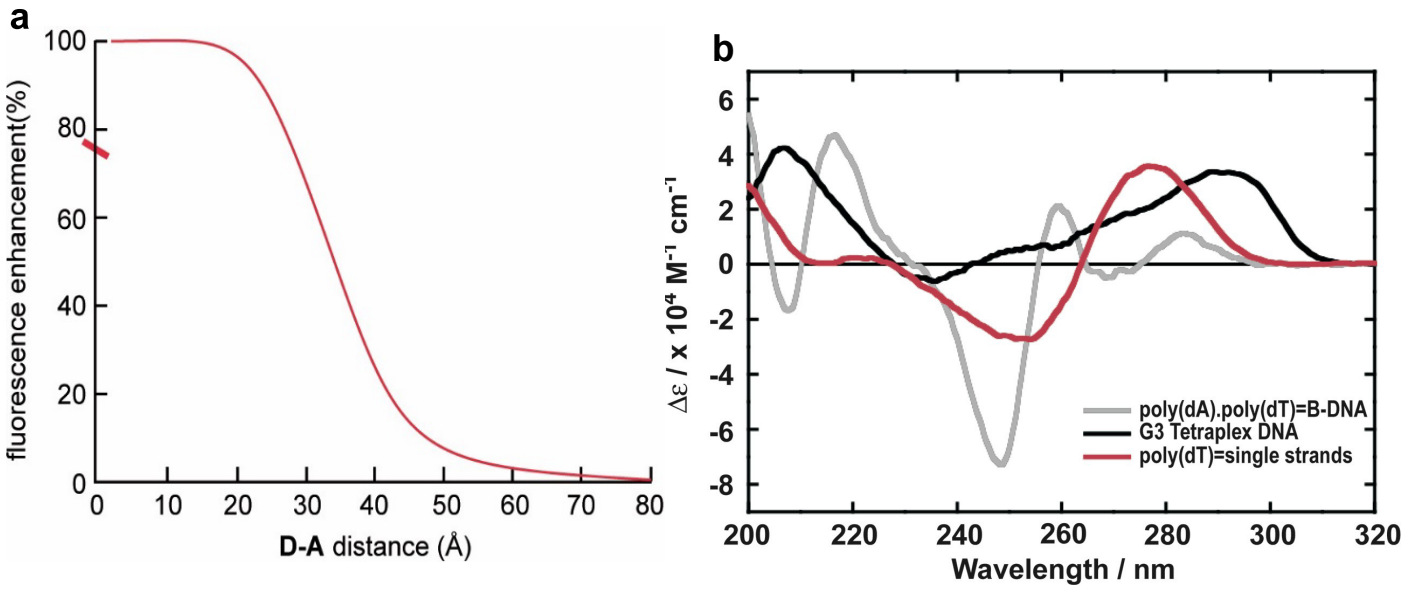

Figure 15.2 (a) A typical relationship between fluorescence enhancement, resulting from FRET, and the distance between donor and acceptor fluorophores. (b) Circular dichroism spectra for three different conformations of DNA, a poly(dA)•poly(dT) B-type duplex, a guanine quadruplex (tetraplex), and a single-stranded random coil.

15.1.3 Circular and Linear Dichroism

Circular dichroism (CD) is a widely used form of polarised light spectroscopy that has been applied to the study of nucleic acids.[21,22,23] A CD signal results from the differential absorption of left and right circularly polarized light. A chiral molecule will absorb left and right circularly polarized light to differing extents and this difference gives rise to the phenomenon of CD. Typically, data are presented as plots of $\Delta\varepsilon$, the difference in molar extinction coefficient for left and right circularly polarised light ($\varepsilon_L - \varepsilon_R$), *versus* wavelength.

Isolated purine and pyrimidine nucleobases are achiral planar heterocycles, intrinsically optically inactive and hence do not exhibit a CD signal. When incorporated into nucleosides and nucleotides, the glycosylic bond from the C1' atom of the chiral D-ribose to either the N-9 of a purine or N-1 of a pyrimidine gives rise to a chiral perturbation of the UV absorption of the base. The CD signal of a nucleic acid increases with its length because of the co-operativity of chiral interactions between contiguous bases. This occurs both as a result of sequence effects arising from nearest neighbour interactions as well as from overall gross secondary structure.

The information derived from CD spectra is complementary to other types of optical spectroscopy, such as UV, IR and LD, and it provides a quick, convenient and accurate picture of the overall conformation and secondary structure of a particular nucleic acid in solution. Reference CD spectra for single strands, B-DNA duplexes and quadruplexes show that each type of secondary structure (as well as additional polymorphs) exhibits its own signature spectrum (Figure 15.2b). A- and Z-type DNA duplexes also have unique CD spectra, as do many nucleic acid–ligand complexes. CD spectroscopy also allows inter-conversions between different secondary structures to be monitored. For example, titration of NaCl into a solution of poly(dG-dC) to salt concentrations above 4 M induces a structural transition from a standard right-handed B-helix to the left-handed Z-form. Increasing the temperature of a nucleic acid solution whilst measuring the CD signal allows DNA denaturation to be monitored directly and it is possible to observe the structural transition from folded duplex to random coil single strands.[24]

CD spectroscopy is a useful tool for studying nucleic acid–ligand interactions.[25] The CD spectrum of each component in solution is directly proportional to its concentration (**Beer's law**), and the total spectrum arises from the sum of all component spectra. If ligand binding induces extrinsic optical activity in the chromophores of the bound ligand, an *induced* CD signal is observed, which is directly proportional to the extent of nucleic acid–ligand complex formed. It can therefore be used to construct a binding isotherm. Alternatively, ligand binding may result in a conformational change in the nucleic acid, and the resultant change in the intrinsic CD signal of the macromolecule facilitates quantification of the binding event.

Modern CD spectropolarimeters can be equipped with Peltier temperature control, automated titration and stopped-flow accessories which allow CD to be used to provide information on structure, thermodynamics (binding affinity, van't Hoff enthalpy, free energy) and kinetics (association and dissociation rate constants).

A closely related technique to CD is **linear dichroism** (LD).[26,27] LD is a type of spectroscopy that yields useful information on DNA conformation in terms of base-inclination and flexibility as well as the binding geometries of drug–DNA complexes. In this technique, the differential absorption of linearly or plane-polarized light is measured. A key parameter in LD is the **transition moment**, which is a vectorial property of light absorption related to a particular direction in the molecule. Light that is polarized parallel to the transition moment has a high probability of absorption in the spectral region of interest, whereas if light is polarized perpendicular to the transition moment, no absorption takes place. In practice, this means that intercalators that stack closely to base-pairs have linear dichroism similar to the base-pairs themselves. However, the dichroism of groove-binders is frequently opposite to that of base-pairs, since they bind along the edges of the latter. Hence LD is a useful type of spectroscopy for assessing the binding mode of low MW drugs to DNA.

15.1.4 Infrared and Raman Spectroscopy

Infrared (IR) and **Raman spectroscopy** are often regarded as closely related techniques in which the vibrational frequencies of specific functional groups of molecules are observed.[28,29] Both techniques are largely non-destructive and can be used on small samples. A major advantage is that DNA can be analysed in crystals, gels, or fibres, as well as in solution, and this has supported direct correlations between the observed

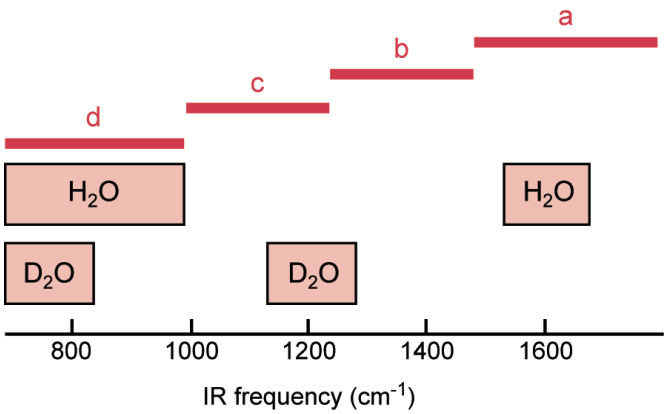

Figure 15.3 Schematic distribution of IR absorption bands of DNA and solvent. (a) 1800–1500 cm^{-1} corresponding to stretching of C=X double bonds. (b) 1500–1250 cm^{-1} corresponding to base-sugar entities (including glycosyl torsion angle effects). (c) 1250–1000 cm^{-1} corresponding to phosphate and sugar absorptions. (d) Below 1000 cm^{-1} associated with phosphodiester chain coupled with the sugar vibrations.

vibrational frequencies and three-dimensional structures derived by X-ray diffraction. Useful IR absorptions from nucleic acids are observed in the frequency range 1800 to 700 cm^{-1}. The problem of strong IR absorption by water near 1600 cm^{-1} and also below 1000 cm^{-1} can be circumvented using D_2O as solvent when the water signal at 1600 cm^{-1} is shifted to about 1200 cm^{-1} and the absorption at 1000 cm^{-1} is shifted by an almost equivalent amount (Figure 15.3). The use of D_2O also causes small but significant shifts in nucleic acid absorptions resulting from deuterium exchange, and these can be used to monitor **H–D exchange** processes.[30]

Fourier transform infrared absorption, **FTIR**, is so sensitive that it is possible to make measurements on very small crystals of nucleic acids, and needs about 100 μg of material. IR spectra are largely unaffected by the external environment and so FTIR has supported the identification of structurally significant bands by calibration with X-ray crystallography of nucleic acid samples. Such 'marker bands' are sensitive to helical type and can be used to show the existence of a conformational transition when different factors, such as temperature, hydration, concentration, or the nature and quantity of cations are varied (B → A, B → C, B → Z form, helix → coil, *etc*). FTIR is also used to support studies on the recognition of DNA sequences by a wide variety of molecules, such as oligonucleotides (triple-stranded structures), drugs, and proteins. Raman spectroscopy also depends on the vibrational frequencies of groups within the molecule and, as with IR, provides information concerning vibrational modes of nucleic acid components that are conformationally sensitive. The Raman technique has some useful advantages over IR. First, the incident radiation in Raman is not strongly absorbed and so does little damage to the sample. Secondly, water has weak scattering properties and lacks absorption at the irradiation frequencies used for sample irradiation, so its presence in the sample is not a problem. Thirdly, unlike IR, the intensity of the Raman bands is proportional to the concentration of the target species. As with IR, the accurate assignment of resonance lines can be simplified by calibration using samples of known X-ray structure. Raman spectroscopy has been used to examine nucleic acids in a wide variety of situations including microcrystals and even within living cells.[31]

15.2 Nuclear Magnetic Resonance

Nuclear magnetic resonance (**NMR**) spectroscopy is the method of choice for investigation of the conformation of short (up to about 25 base-pairs) nucleic acid fragments in solution. Compared to other spectroscopic techniques routinely used, NMR is rather insensitive and requires near millimolar sample concentrations. However, the structural information that may be gained from an NMR spectrum is much more detailed than that available from any other solution-phase technique and is complementary to that available from X-ray diffraction studies on solid samples (Section 15.4).[32,33] NMR spectra provide a unique window into the flexibility and dynamic behaviour of nucleic acids.[34]

The basis of NMR is that atomic nuclei are endowed with a property called **nuclear spin** and will align with an externally applied **magnetic field** (B_o).[35] The degree of this alignment is dependent on the type of nucleus and the strength of the magnetic field. Different elements and isotopes have different nuclear spins and give

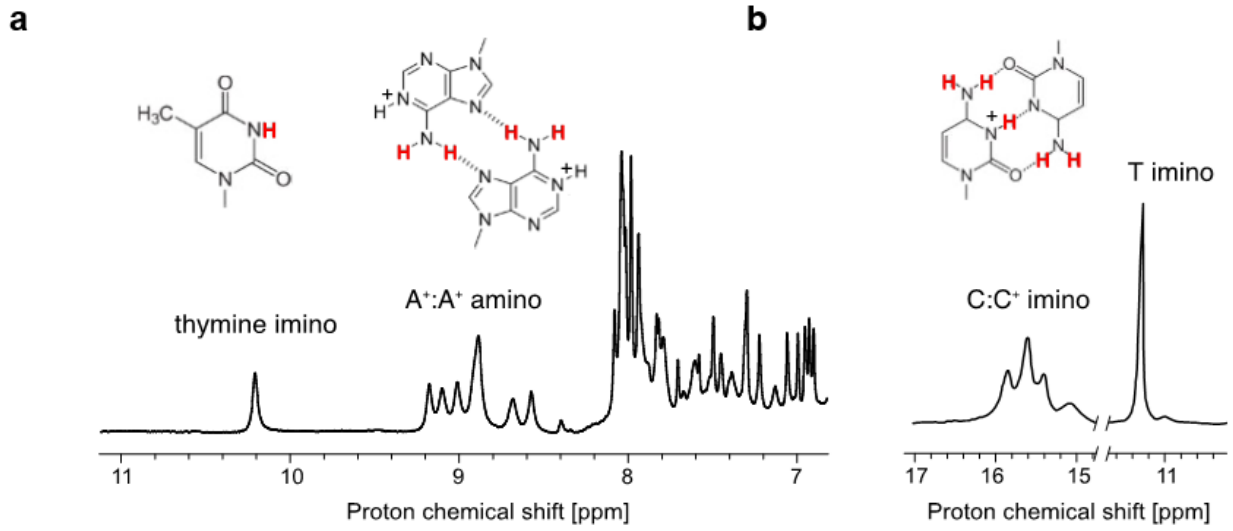

Figure 15.4 One-dimensional (1D) NMR spectra of exchangeable hydrogens detect formation of nucleic acid base-pairs and duplexes. (a) Amino resonances of adenine:adenine Hoogsteen base-pairs confirm the formation of a parallel duplex by dT(rA)8.[41] (b) Imino resonances of hemi-protonated cytosine: cytosine base-pairs indicate formation of the four-stranded i-motif by dTdC5.[42] Adapted from Ref. 42 with permission from Oxford University Press, Copyright © 2017.

rise to very different forms of NMR signal. The proton (^{1}H) is the most sensitive, non-radioactive NMR-active hydrogen isotope and the most widely studied with applications in chemistry, physics and medicine, including MRI medical imaging. Among other common elements, the most abundant isotopes of carbon and oxygen (^{12}C and ^{16}O) have zero spin and are therefore NMR silent. ^{14}N produces a weak NMR signal and with broad lines and is rarely used. Nucleic acids contain phosphorus as ^{31}P, which has excellent properties for NMR and is naturally 100% abundant. For NMR studies, ^{13}C and ^{15}N isotopes can be incorporated biosynthetically into nucleic acids. Both isotopes have good properties for NMR and can supplement ^{1}H NMR studies.[36] Fluorine (^{19}F) is a nucleus that is a very good probe of conformation and can be introduced into sugars and bases *via* chemical synthesis. ^{19}F has been deployed in metal fluorides as analogues of the phosphoryl group in enzyme **transition states** (Section 3.2.2).[37]

Commercial NMR spectrometers with fields up to 23.5 Tesla (roughly half a million times the earth's magnetic field) provide optimal sensitivity and resolution. NMR signals are detected by excitation of the nuclear spins through the application of radiation from the radio frequency region of the electromagnetic spectrum (typically 300 to 1000 MHz). This leads to the emission of radio signals from the nuclear spins that reveal information about the atom type, neighbouring atoms, and molecular motion.

For a proton (^{1}H) in a 14 Tesla magnet, excitation of the nuclear spin will lead to signals of 600 MHz within a relatively small window of 10 to 15 parts per million (10 to 15 kHz). NMR signals in liquids are generally long-lived, which gives rise to information-rich spectra with narrow lines. Signals from solid-state samples have broader lines but include orientation information that is missing in spectra of liquid samples. The frequency of the emitted signal depends on the chemical and magnetic environment of the nucleus. Electrons surrounding the nucleus shield the nucleus from the applied field and give rise to the small differences in frequency for the same type of nucleus when in different molecules or chemical environments.[35]

Processing of the radio-frequency signals by application of the **Fourier transform** is used to generate an **NMR spectrum**. All NMR spectra share common features regardless of the isotope or sample: an *x*-axis that contains the frequency of the detected radiation and a *y*-axis that reports intensity of the signal. The position or frequency of the spectral line is called the **chemical shift**. In order to remove the magnetic field dependency, the chemical shift is reported in terms of parts per million (ppm) as a difference from a reference compound that defines zero. For historical reasons, the left-side of the spectrum presents higher frequency signals (and is called **downfield**) and the right-side is **upfield**. An important property of an NMR signal is the **linewidth** which often provides information about molecular motions. The area under a signal relates to the concentration of the particular spin. The fine structure in lines is referred to as their multiplicity.

Hydrogens in nucleic acids can be exchangeable or non-exchangeable. In aqueous solvents, hydrogens on nitrogen and oxygen exchange positions with hydrogens from water molecules. This broadens their NMR

signals and in the case of hydrogens on oxygen (*e.g.* hydroxyl groups) makes them difficult or impossible to detect. The exchange of hydrogens on nitrogen is slower and a valuable source of information about the single-stranded or double-stranded conformation of nucleic acids (Figure 15.4). When base-paired, the nucleotide base imino and amino hydrogens are protected from solvent exchange. Measurements of their **exchange rates** provides information about the existence and the rate of opening of base-pairs.[38-40] Non-exchangeable hydrogens are found on both the sugar and base moieties. Spectra of non-exchangeable hydrogens are generally recorded in deuterated water (^2H_2O) to eliminate the large signal from solvent ^1H_2O that overwhelms the much weaker sample signals.

The NMR resonances of nucleic acids are well spread out and, to large degree, are identifiable simply based on their chemical shifts.[33] Exchangeable hydrogens are the most downfield in the range of 8 to 15 ppm (Figure 15.4). The aromatic hydrogens on the nucleotide bases are the next most downfield at 6 to 8 ppm. Methyl resonances of thymine are easily identified by their narrow linewidth, intensity (three hydrogens), and upfield chemical shift of ~1 ppm. In DNA, hydrogens H2' and H2'' on the deoxyribose also have distinct chemical shifts. That is lost in NMR spectra of RNA, which additionally have less chemical shift dispersion and are more difficult to analyse than DNA spectra. The ^1H spectrum of a small DNA/RNA hybrid duplex with ~300 hydrogens is shown (Figure 15.5).

A variety of experiments have been devised that enable the assignment of NMR signals to individual atoms. In general, these rely on interactions (termed couplings) between nuclei. There are two types of interactions: **scalar couplings** and **dipolar couplings**. Scalar couplings, also called ***J*-couplings**, are transmitted by electrons in bonds and give rise the multiplet structure used in chemistry to assign signals. *J*-couplings give information about the covalent structure of molecules. Dipolar couplings give rise to the **nuclear Overhauser effect** (**NOE**) and arise from through-space interaction between the magnetic fields of neighbouring nuclei. NOEs are independent of the covalent structure and depend on the distance separating the nuclei.

NMR couplings allow two-dimensional (2D) NMR spectra to be recorded with cross-peaks between two signals indicating the existence of an NMR coupling between the corresponding nuclei. Scalar couplings are typically detected by **correlation spectroscopy** (**COSY**) and **total correlation spectroscopy** (**TOCSY**) spectra. In liquids, dipolar couplings are measured using **nuclear Overhauser effect spectroscopy** (**NOESY**) spectra. 2D NMR spectra are typically shown as contour plots with signals appearing as peaks surrounded by lines of decreasing intensity. 2D spectra are classified as either homonuclear, with the same type of nucleus on both axes, or heteronuclear, with two different types of nuclei. The most common heteronuclear spectra are hydrogen–nitrogen (^1H–^{15}N), hydrogen–carbon (^1H–^{13}C) and hydrogen–phosphorus (^1H–^{31}P). Homonuclear spectra are almost always hydrogen–hydrogen and typically have a strong diagonal line connecting a signal on one axis with itself on the second axis. The useful information in 2D spectra is found in the cross-peaks that connect two different signals. Understanding the couplings between the signals allows their assignment to individual atoms and the determination of three-dimensional structures.

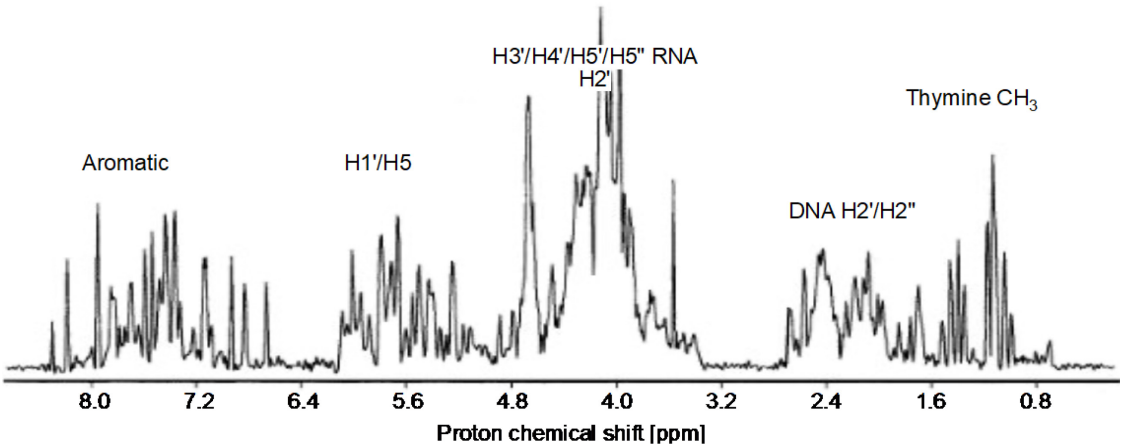

Figure 15.5 Non-exchangeable hydrogen signals of a mixed RNA/DNA duplex in ^2H_2O of 5′d(ACTCGATTTCATAGCC)3′//5′r(GGCUAUGAAAUCGAGA)3′. With the exception of thymine methyl groups, signals of nucleic acid bases lie between 5.5 and 8.5 ppm. Ribose residues of RNA (H1′ to H5′/5″) fall within a relatively narrow window of 3.2 to 6 ppm. DNA offers better dispersion as the deoxyribose H2′/H2″ resonances are shifted upfield, away from the other sugar signals. Spectra in ^2H_2O will often show a peak at ~4.8 ppm from residual ^1H_2O in the sample.

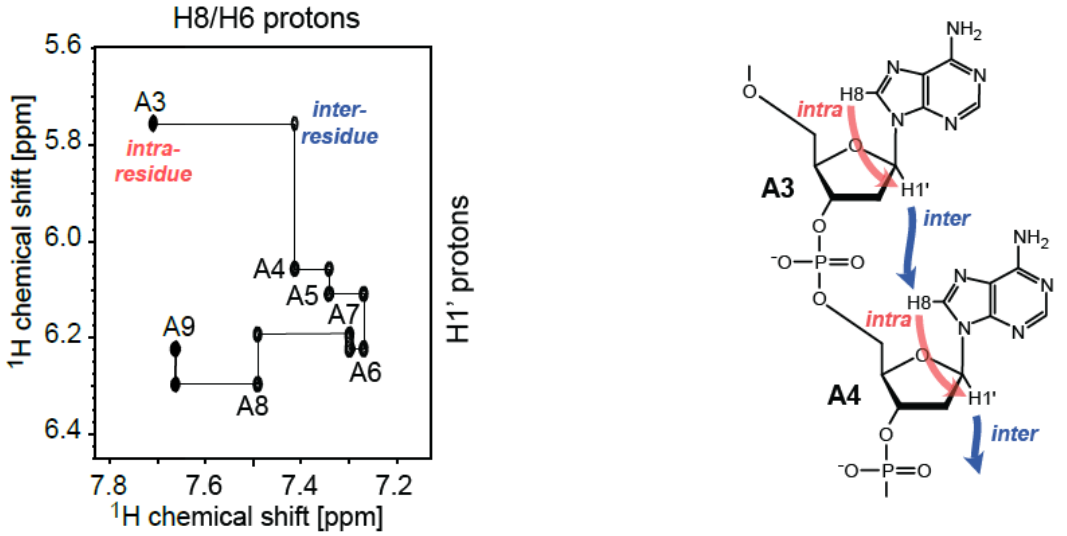

Figure 15.6 Part of a 2D NOESY spectrum of the non-exchangeable hydrogen signals of an RNA duplex.[41] The ribose H1′ hydrogens show NOEs with the base H8 and H6 hydrogens within the same nucleotide and with the following nucleotide. Tracing alternately between the intra- and inter-residue cross-peaks allows determination of the residue specific assignments of the NMR signals.

The first step in assigning NMR spectra is the identification of **nuclear spin systems**. These are sets of nuclei that are connected through *J*-couplings to each other. To be in the same spin system, the nuclei do not need to be directly *J*-coupled but do need to be coupled to other nuclei that are themselves *J*-coupled. The presence of more than three bonds between hydrogens typically prevents *J*-coupling and acts to separate hydrogens into separate spin systems. In nucleic acids, the hydrogens on each nucleotide sugar form an independent spin system as do the downfield H5–H6 hydrogens on cytosine. In proteins, each amino acid constitutes a spin system while some, such as phenylalanine, have two spin systems. COSY and TOCSY spectra detect *J*-couplings and hence are used to identify spin systems. The identification that signals from an H1' and H5'' are in the same spin system indicates that they are in the same ribose but does not identify that ribose within the nucleic acid. To obtain residue-specific signal assignments, NOESY spectra are used to connect the spin systems to one another using through-space dipolar couplings. The method makes assumptions about the three-dimensional structure and is most reliable for assigning regular structures such as duplexes. To resolve ambiguities, heteronuclear experiments can be used to connect hydrogen spin systems *via* hydrogen-phosphorus or hydrogen-carbon *J*-couplings.

To align spin systems, NOESY experiments are used to 'walk' along the nucleic acid backbone. Nearby nuclei give rise to NOE cross-peaks that can be used to connect adjacent nucleotides thus allowing the assignment of a sequence position to each signal.[43] In the example shown (Figure 15.6), the H1' signal of residue A3 (5.75 ppm) shows NOE cross-peaks with two H8 signals: the H8 of the directly attached adenine ring (7.7 ppm) and the H8 of the following nucleotide (7.4 ppm). Alternating between intra-residue and inter-residue NOE cross-peaks allows all signals to be specifically attributed. Other sets of NOE cross-peaks can be used to confirm and extend the assignments where overlaps or ambiguities occur. NOESY experiments with exchangeable hydrogens can be used to make assignments across base-pairs.

Beyond signal assignments, cross-peaks in NMR spectra contain valuable information that can be used to determine 3D structures. Information used to determine NMR structures comes in a wide variety of forms, typically formulated as specific constraints on angles and interatomic distances. COSY experiments permit *J*-coupling constants to be quantified, which give information about torsion angles. The variation of scalar couplings between atoms separated by three bonds as a function of the torsion angle between their bonds is empirically described by the Karplus equation.[35] The coupling is largest when hydrogens are opposite or on the same side and smallest when they form close to a right angle. In nucleic acids, measurement of the scalar couplings between ribose hydrogens, such as H1' and H2', can be used to constrain the nucleoside sugar pucker.[43] With isotopically-labelled samples and heteronuclear COSY experiments, it is possible to measure other angles to describe the geometry of the phosphate backbone and the glycosylic angle.

NOESY spectra are the most valuable source of information in NMR structure determination. NOEs provide a map of all the close contacts in a molecule by gauging the distance between atoms. Multiple factors,

such as dynamics, affect the size of the NOE cross-peak so the distances are often grouped into approximate ranges. The intensity of a NOESY cross-peak decreases strongly with distance (as $1/r^6$), which limits NOE cross-peaks to atoms separated by less than approximately 5 Å. For more accurate measurements, fixed distances such as H6–H5 in cytosine bases or H2'–H2'' in deoxyribose can be used for calibration. The large exponent means that a two-fold uncertainty in the size of the NOE peak is reduced to 10% uncertainty in internuclear distance.

As NMR structures rely on measurements of short distances, determining atom positions and long-range order of extended structures is a challenge. In some cases, the disorder in NMR structures reflects the inherent flexibility of the molecule studied, but it can also arise from a lack of structural information. One solution is the measurement of **residual dipolar couplings** (**RDCs**), which provide distance-independent information about the orientation of atom pairs relative to the external magnetic field. Dipolar couplings are present in solid-state NMR spectra – and responsible for the large line widths – but they are absent in solution NMR spectra due to averaging by molecular rotations. To measure the couplings in solution, the molecule needs to be partially aligned; incorporation in a **liquid crystal** is the most widely used method. The most common liquid crystals are **phospholipid bicelles** or filamentous phage particles. The **magnetic anisotropy** of the bicelles or phage particles align them in the magnetic field of the NMR spectrometer. Through collisions, some of this alignment is transferred to induce partial alignment of the macromolecule. The orientation information from RDC experiments is particularly valuable for studying nucleic acid structures, such as duplexes, that are spatially extended with relatively few long-range NOEs.[44]

Structural studies on larger nucleic acids are generally done using ^{13}C and ^{15}N labelled samples. These additional NMR-active nuclei allow the resolution of signal overlap in ^{1}H spectra through the use of heteronuclear NMR experiments. Most common is **2D HSQC** (**heteronuclear single quantum correlation**) that shows cross-peaks between ^{13}C or ^{15}N atoms and a directly bonded ^{1}H. Such a 2D experiment may be added to a COSY or NOESY type experiment to produce 3D and 4D spectra. These experiments allow many more constraints to be measured for determination of atomic structure.

Calculations of NMR structures are performed using restrained molecular dynamics (Section 15.13), which involves simulating the atomic motions of the molecule to search for the best fit with the NMR data.[37] Newton's equations of motion are used to simulate the physical movements of the atoms. In addition to the natural forces (van der Waals contacts, bond lengths, angles, electric charges) artificial forces that reflect **NMR constraints** are added and the simulation run at high temperature to allow good coverage of the conformational space. The temperature of the simulation is progressively decreased to simulate annealing of the molecule into the lowest energy state. Some simulations become trapped in local minima, so the calculations are repeated multiple times and the lowest energy structures superposed as a representation of the NMR structure (Figure 15.7).

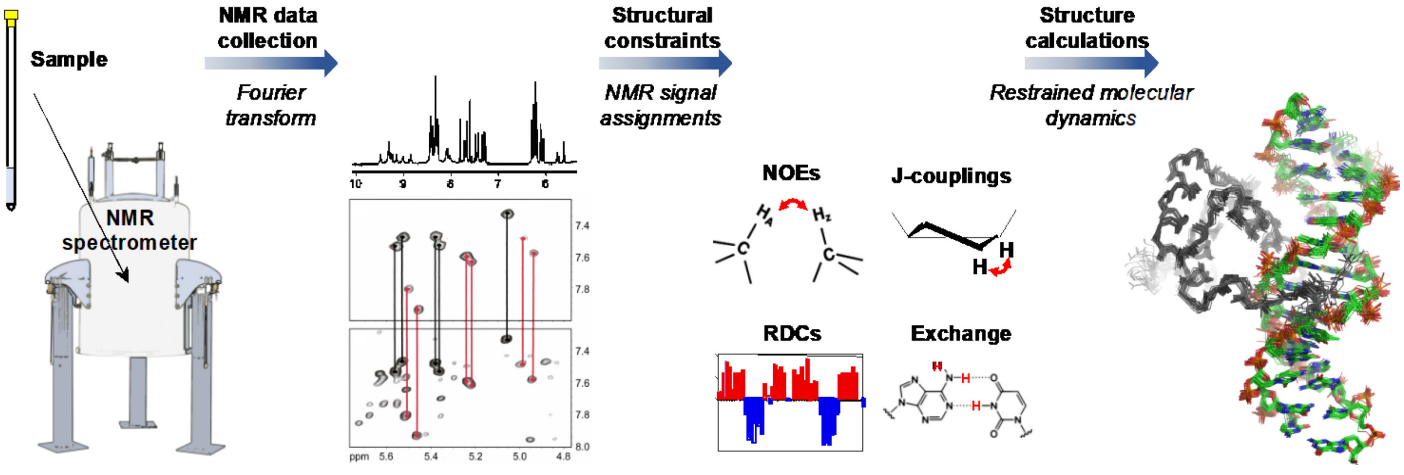

Figure 15.7 Workflow for structure determination by NMR spectroscopy. Typically, the sample consists of 400 QL at a high micromolar concentration (*e.g.* 300 QM) in $^{1}H_2O$ or $^{2}H_2O$. Over the course of several days, spectra are acquired to assign the NMR signals and measure various constraints on hydrogen-hydrogen distances, torsion angles, *etc.* Molecular dynamics calculations with artificial restraints are used to determine an ensemble of three-dimensional structures compatible with the experimental data.

Such approaches are now aiding the conformational analysis of nucleic acids. Although the molecular mass limit for routine NMR structure determination for nucleic acids is less than 25 kDa, technical advances in selective isotopic labelling, experimental design, and NMR spectrometers allow molecules as large as 100 kDa to be studied.[45] NMR structures have been generated for structural motifs, such as stem-loops, parallel duplexes,[41] G-tetraplexes, i-motifs[42] and triplexes (Chapter 2), as well as complexes of nucleic acids with drugs (Chapters 11 and 12) and proteins[46] (Chapters 3, 13 and 14). Of particular interest is the recent progress in characterizing molecular motions of nucleic acids by NMR, including rarely-populated transition states.[47] The application of methods first developed for proteins[48] has allowed invisible, transiently-populated states to be detected and characterized. Other areas of promising development are the introduction of labelled molecules into cells for *in vivo* NMR studies[49] and **solid-state NMR** of nucleic acids.

15.3 Mass Spectrometry

Specificity, sensitivity and speed make **mass spectrometry** (**MS**) one of the most versatile platforms for the characterization of both natural and synthetic nucleic acids.[52] The platform capitalizes on soft ionization techniques, such as **matrix-assisted laser desorption/ionization** (**MALDI**) and **electrospray ionization** (**ESI**), to preserve the integrity of labile analytes during transfer to the gas phase. Coupled with a wide range of mass analysers, these ionization techniques can afford the accurate mass determination of individual species, the composition of sample mixtures, and unambiguous sequence information.[53] In recent years, the advances observed in the field of nucleic acids analysis has followed in lockstep those made in protein analysis and proteomics, leading to the realization of similar capabilities and performance. Propelled by these advances, MS analysis has experienced ever expanding applications in nucleic acid sequencing,[54] genotyping,[55] detection of genetic variations,[56] microsatellites,[57] short tandem repeats,[58] transcriptomics,[59] and epitranscriptomics.[60]

15.3.1 Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry

Introduced by K. Tanaka, **MALDI-MS** involves mixing the analyte of interest with a small organic matrix to form crystals that are capable of absorbing the radiation of UV or IR lasers.[61] In nucleic acid analysis, 3-hydroxypicolinic acid (3HPA)[62] and a mixture of 2,3,4- and 2,4,6-trihydroxy-acetophenone (THAP)[63] have been shown to afford the best results among the dozen or so matrices employed with UV lasers. The matrix disperses the analyte throughout the entire volume of the crystal and absorbs the majority of the photons' energy. Under high vacuum conditions, the process produces a fast-expanding plume, which leads to analyte desorption into the gas phase and formation of intact molecular ions. The process can be carried out also at atmospheric pressure to further reduce the possibility of activating unwanted fragmentation of the species of interest.[64] Molecular ions can be in turn analysed on virtually any type of mass analyser, which can be selected according to the type of information and performance sought by the user. For example, **time-of-flight** (**TOF**) mass spectrometry is typically utilized to take advantage of a virtually unlimited mass range and excellent detection sensitivity.[65] Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometry is preferred instead for its high resolving power and accuracy.[66] Additionally, a variety of analysers are also capable of performing tandem mass spectrometry to generate characteristic fragment ions that reveal the sequence of the analyte of interest.[53]

Although nucleic acids can be readily detected in cationic form, the analysis is most often carried out in negative ion mode to take advantage of the highly acidic character of the phosphate groups. As an example, a representative MALDI-TOF spectrum is provided in Figure 15.8a, which was obtained from a crude oligodeoxynucleotide sample prepared by solid-phase synthesis (Section 7.1). The spectrum is dominated by the $[M - H]^-$ molecular ion detected at 5879.8 *m/z*, whereas the doubly charged counterpart can be also observed at half the *m/z* value. Indeed, this ionization technique tends to produce predominantly singly charged ions, but higher charge states become more abundant with increasing analyte size. In this respect, 100 nt (~30 kDa) can be considered as a practical upper limit for UV-MALDI analysis of oligodeoxynucleotides, whereas IR-MALDI with glycerol matrix is capable of reaching up to ~2000 nt (~600 kDa), as demonstrated in the analysis of products obtained by digesting a plasmid with restriction enzymes.[67] Mass resolution is influenced by both the resolving power of the available analyser and the incidence of matrix and cation adducts to the molecular ion, which becomes more prominent with analyte size. Similar considerations are valid also for mass accuracy, which is also affected by the ability to properly distinguish the molecular ion from possible

adducts. For example, the signal of the 19-mer (Figure 15.8a) provided a ~165 full-width half maximum (FWHM) resolution, which did not allow one to properly discriminate the cation adducts responsible for signal broadening. As a result, this molecular mass determination exhibited an accuracy of only ~663 ppb. This figure of merit could be greatly improved by eliminating the adducts or resolving their signal contributions. The excellent efficiency of the MALDI process and the ability to focus the laser light into very narrow spots account for the very thrifty sample consumption characteristic of this ionization technique. Ancillary strategies for depositing samples onto miniature stages or chip arrays, which may utilize piezoelectric pipets, microfluidics, or similar expedients, can further reduce sample consumption to sub-femtomole levels.[68] Excluding the time spent for sample preparation, a routine MALDI-TOF analysis can be accomplished in less than 10 s with only minimum training necessary.

A variety of approaches have been developed over the to obtain sequence information from MALDI-MS analysis. Classic tandem mass spectrometry (MS/MS) involves the activation of a selected precursor ion by using different forms of energy to induce gas-phase fragmentation.[69] In the case of nucleic acid samples, the process leads to the selective cleavage of the backbone's phosphodiester bond, which produces series of fragment ions differing from one another by one nucleotide.[70] Therefore, the mass difference between consecutive fragments identify unambiguously the corresponding nucleotide, and can reveal the sequence position of possible nucleotide modifications from corresponding mass shifts. A practical upper limit for MS/MS sequencing is approximately 20–25 nt, but successful analysis of samples up to ~100 nt has been also reported.[71] In the absence of an analyser with MS/MS capabilities, similar outcomes can be achieved by finely adjusting the laser power, and thus the internal energy imparted to the ions, in such a way as to induce in-source fragmentation. An approach called delayed extraction (DE) has been devised in which ions are allowed sufficient time to undergo all possible dissociative processes in the source region, before they are accelerated into the TOF region for mass analysis.[72] Alternatively, molecular ions activated by the MALDI process can experience post-source decay (PSD) after leaving the source and, thus, their fragments can be distinguished only by utilizing appropriately designed ion mirrors.[73] However, the most popular strategies associated with MALDI-TOF do not involve any gas-phase fragmentation step, but rely instead on the detection of oligonucleotide "ladders" obtained in solution.[74] Specific exonucleases and Maxam-Gilbert chemistry[i] have

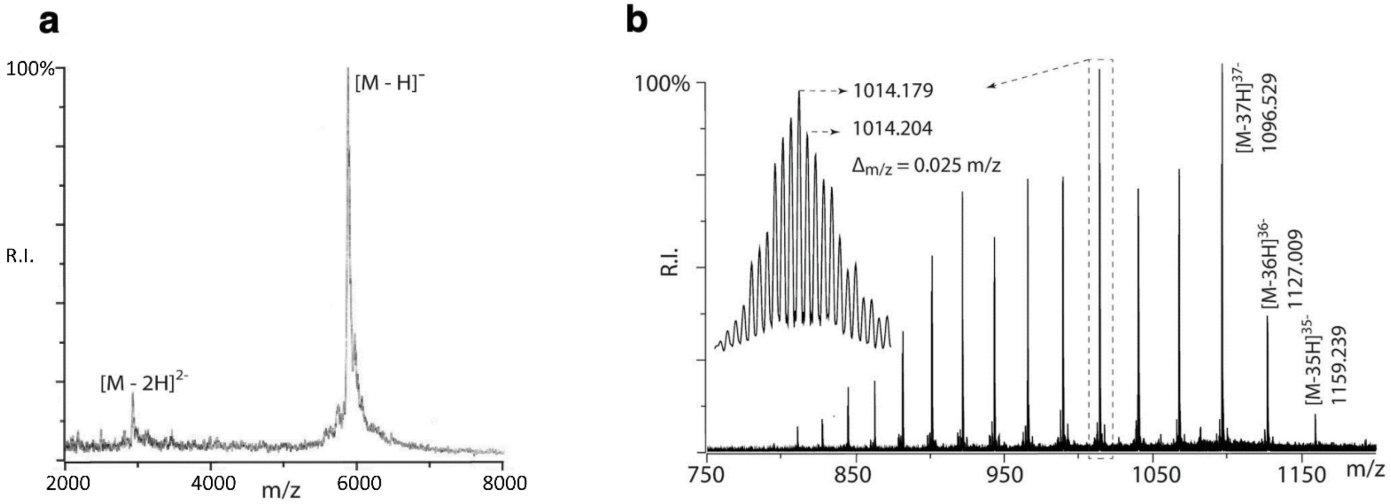

Figure 15.8 (a) Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrum of a crude sample generated by chemical synthesis. The target product, a 19-mer oligodeoxynucleotide with d(GGATTACAGGTATGAGCCA) sequence, was detected as a deprotonated ion at 5879.8 *m/z* (5875.9 u average mass calculated from sequence). The spectrum was obtained in linear, negative ion mode by using 3-hydroxypicolinic acid as matrix. (b) Electrospray ionization (ESI) mass spectrum of a recombinant 124-nt oligoribonucleotide that replicates the packaging signal of the HIV-1 genome (Ψ-RNA). The analysis was carried out in negative ion mode on a 12T Fourier transform-ion cyclotron resonance mass spectrometer (FT-ICR). The sample provided a charge state distribution spanning from 35- to 51-. The inset shows the fully resolved isotopic signals obtained at ~476K resolution for the 40- charge state. The independent determinations provided by all detected charge states afforded a 40,608.6 u average experimental mass, which compared favourably with the expected average mass of 40,608.5 u calculated from sequence, corresponding to a 287 ppb accuracy. The overall precision of the experiment was 3 0.3 u (or 7 ppm RSD).

[i] **This is now obsolete (see Section 11.5.3)**

been in turn utilized to induce consecutive cleavages of the initial strand, thus generating a series of products that differed from one another by one nucleotide. Alternatively, MALDI-TOF has been utilized also to analyse chain-termination products generated by the Sanger method (Section 8.3.1), with the corresponding masses providing a direct readout of the initial sequence. All strategies based on in source fragmentation and solution "ladders" are operationally straightforward but produce best results when individual targets are examined separately. Indeed, the simultaneous presence of multiple analytes in the sample may result in overlapping series of fragments/products, which can then be very difficult to tease apart. In this case, a separation step is typically recommended to isolate each analyte, which can be then analysed separately to eliminate any assignment ambiguity.

15.3.2 Electrospray Ionization Mass Spectrometry

Introduced by J. Fenn, electrospray ionization (ESI) involves driving a sample solution through a narrow metal tube (*i.e.*, the emitter), which is held at a high voltage versus a counter-electrode in the atmospheric pressure region of the ion source.[75] The strong electric field pulls the meniscus at the end of the emitter into a conical shape that breaks down into microscopic droplets. Similar outcomes can be achieved by thrusting nebulizer gas parallel to the emitter tip, or by applying mechanical or acoustic vibrations. In all cases, the droplets are entrained by voltage and pressure gradients into the vacuum region of the source, in which they undergo successive cycles of rapid evaporation and fission driven by the coulombic repulsion between like charges on the droplet surface, as well as collisions with other droplets and neutral solvent molecules. As a result, ions can be ejected directly into the gas phase by the strong columbic repulsion on the surface of rapidly shrinking droplets (ion-evaporation model).[76] Alternatively, the evaporation/fission process can cause limit elimination of all solvent from the droplet, which leaves ions solvent-free and ready for mass analysis (charged-residue model).[77] As observed in MALDI, protonated or deprotonated ions can be readily detected in positive or negative ion mode, respectively. The pH of sample solutions can be modified by addition of volatile acids or bases (*i.e.*, acetic and formic acid for positive ion mode, or ammonium hydroxide for negative) to adjust the protonation state and polarity of molecular ions. Unlike in MALDI, the analyte can produce a broad distribution of signals corresponding to different charge states. This outcome is exemplified by the ESI-MS spectrum of a 124-mer oligoribonucleotide shown (Figure 15.8b), which provided a distribution ranging from 35- to 51-.

Obtaining the experimental mass from an ESI charge distribution is a two-step process. The first step consists of assigning the charge state of each signal in the distribution. If (m/z) and $(m/z)_2$ represent the m/z values of two contiguous signals in the distribution, with $(m/z)_1 < (m/z)_2$, then

$$z_1 = \frac{(m/z)_2 + 1}{(m/z)_2 - (m/z)_1} \qquad (15.4)$$

For positive ions, the numerator must read $(m/z)_2 - 1$. In Figure 15.8b, if the signals at 1127.009 and 1159.239 m/z were assigned to $(m/z)_1$ and $(m/z)_2$, respectively, then the former would exhibit $z_1 = (1159.239 + 1)/(1159.239 - 1127.009) = 36$ negative charges. Consequently, the signal at 1159.239 m/z would exhibit a 35- charge state, 1096.529 m/z 37-, and so on. An alternative approach can be implemented when the available resolving power is sufficient to distinguish the individual isotopic contributions of a given charge state. If the spacing between contiguous isotopic signals on the m/z scale is indicated as $\Delta_{m/z}$, then the corresponding charge is:

$$z_1 = \frac{1 m/z}{\Delta_{m/z}} \qquad (15.5)$$

For example, the fully resolved isotopic signals in the inset of Figure 15.8b displayed a 0.025 m/z spacing between contiguous contributions, thus providing $z = (1\ m/z)/(0.025\ m/z) = 40$ negative charges for the corresponding isotopic envelope.

Once charge states are assigned, each signal in the distribution can provide a separate determination of the analyte's molecular mass M_i:

$$M_i = [(m/z)_i + 1.0079] \times z_i \qquad (15.6)$$

For positive ions, the average mass of the proton 1.0079 u must be subtracted rather than added. The

monoisotopic mass 1.007825 u must be utilized instead, if the ^{12}C signal of the species of interest can be resolved to determine the analyte's monoisotopic mass. In the example, the 35- signal at 1159.239 *m/z* provided $M_i = [1159.239 + 1.0079] \times 35 = 40{,}608.6$ u, whereas the 36- and 37- signals provided 40,608.6 u and 40,608.9 u, respectively. The individual masses obtained from all the detected charge states can be finally averaged together to obtain the experimental mass M of the analyte, which can be used to calculate the overall accuracy of the determination when a true value calculated from sequence is known. The standard deviation of the individual values provides instead a measure of the precision of the determination. Unique to ESI-MS, this figure of merit can be legitimately derived from a single experiment by virtue of the independent nature of the determinations of the various charge states.

Molecular ions produced by ESI can be analysed on virtually any type of mass spectrometer, in analogy with those produced by MALDI, thus bringing to full fruition the specific capabilities of the various analysers. For instance, the high-resolution afforded by FT-ICR and orbitrap instruments is particularly beneficial in the analysis of complex mixtures, in which discrete charge state distributions might not be immediately recognizable and the ability to assign charge states from isotope signals becomes very desirable. Tandem mass spectrometry capabilities are essential to accomplish the gas-phase sequencing approaches described above, which afford similar figures of merit regardless of ionization technique. However, ESI distinguishes itself for the ability to generate multiply charged ions, which translates in the possibility of detecting very large analytes on relatively inexpensive analysers with limited mass range. This favourable feature has been fundamental to the widespread adoption of ESI-MS for analysing progressively larger biopolymers. An indication of putative upper limits may be offered by the analysis of coliphage T4 DNA (~100 MDa), which was accomplished on an FT-ICR instrument.[78] In the case of protein-nucleic acid complexes, *E. coli* 70S ribosomes (~2.3 MDa) have been detected on a quadrupole (Q)-TOF mass spectrometer,[79] whereas tobacco mosaic virus particles (~40 MDa) were observed intact in charge detection TOF experiments.[80] As observed in MALDI-MS, however, the increasing incidence of adduct formation with analyte size may have significant repercussions on the resolution and accuracy attainable by ESI-MS. This fact may constitute a major obstacle to the investigation of samples obtained from natural sources, which come replete with the Na^+ and K^+ salts present in typical environments, or any functional nucleic acids requiring Mg^{2+} to retain optimal structure and activity. Early on, the MS community realized the benefits of replacing any alkaline or alkaline-earth salts in the sample with ammonium equivalents, which takes advantage of the ability of ensuing ammonium adducts to dissociate into H^+ and volatile NH_3.[81] To this effect, a variety of approaches have been devised over the years to perform desalting and ammonium replacement, including ethanol precipitation,[82] metal chelation,[83] ultrafiltration/dialysis[84] ion exchange,[85] and reversed-phase liquid chromatography.[86]

Another favourable characteristic of this ionization technique is the ability to handle samples in solution. This feature makes ESI immediately compatible with liquid chromatography (LC), capillary electrophoresis (CE), and other high-resolution separation techniques. Over the years, *ad-hoc* strategies have been developed to enable direct LC-MS and LC-MS/MS characterization of nucleic acids, which use ESI to interface a chromatographic system with the selected mass analyser. Ion-pairing chromatography is particularly well-suited for this purpose, which can be successfully accomplished on typical C18 columns with aqueous mobile phases containing triethylamine or diethyl-methylamine.[86] Organic additives, such 1,1,1,3,3,3-hexafluoro-isopropanol (HFIP), are included to modulate the hydrophobic character of mobile phases and to achieve the proper surface tension necessary to establish stable spray.[87] While the search for new improved methods for LC-MS and LC-MS/MS has been very successful, the same cannot be said for CE-MS analysis of nucleic acids, which is still hampered by the need of rather high electrolyte concentrations to achieve desirable separation. The key to overcome this challenge might be offered by the exploration of the properties of nano-flow ESI,[88] which operates at nL/min flow rates rather than the μL/min to mL/min range typical of conventional ESI. Indeed, it has been shown that quartz emitters with sub-micrometre internal diameter produce much lower incidence of cation adducts than conventional ESI or nanospray emitters with micrometre-size tips.[89] Great promise in this direction is also held by the development of microfluidic devices integrating separation and interface capabilities.

ESI has also emerged as an excellent tool for the investigation of non-covalent complexes of nucleic acids, which takes advantage of the favourable energetics of the desolvation process to preserve the specific interactions between components. The 70S ribosomes and intact viruses mentioned above constitute examples of protein-nucleic acid complexes that are stabilized by rather weak H-bonding and electrostatic interactions.[89,90] However, the MS community has successfully employed ESI to investigate also the binding of nucleic acids with other nucleic acids, drug-like compounds, and metals.[90] The recent commercial

introduction of ion mobility spectrometry (IMS) mass spectrometers has paved the way for investigating the structure and dynamics of nucleic acids in the gas phase,[91] which are defined by weak intermolecular interactions preserved during the ESI process. In this direction, a greater understanding of nucleic acid structure in a solvent-free environment and new computational approaches for aiding data interpretation will be expected to play determinant roles in expanding the application of ESI-MS to the investigation of the structure-function relationships in nucleic acids.

15.4 Diffraction Techniques

15.4.1 X-Ray Fibre Diffraction

The earliest work on the structure of the nucleic acids was based on X-ray **fibre diffraction** experiments.[92-97] Unlike oligonucleotides that were then not available, polynucleotides could be pulled into thin fibres such that DNA strands were predominantly aligned along the fibre direction and then exposed to an X-ray beam that was directed perpendicular to the fibre axis. Typical diffraction images resulting from A- and B-form DNA fibres associated with different levels of **humidity** are depicted in Figure 15.9. Before trying to glean the structure of DNA from such images, readers should take comfort in the fact that their interpretation is still absolutely not straightforward!

A few basics: The black dot in the centre is the result of the beam stop that prevents the incident X-ray beam from directly striking the detector or in this case the photographic paper. X-rays (high-energy photons) are diffracted by electrons and the function describing amplitude and phase of the diffracted X-rays (dark areas in the photographs indicate higher intensity diffraction) and electron density are related by a Fourier Transform. One characteristic of this transformation is that the dimensions in the fibre or the crystal lattice and in diffraction or **reciprocal space** are inversely related. For example, the B-DNA fibre diffraction image shows a very strong meridional peak at a resolution of 3.4 Å (as we move from the centre of the diffraction image to the periphery the resolution increases). This indicates the presence of repetitive groups of atoms in the fibre that are separated by a distance of 3.4 Å. These are of course the base-pairs in B-DNA. The same diffraction image shows recurring equidistant patches as we move away from the centre to the outside. These coincide with layer lines coloured in red: 0 (= equator), 1, 2, 3, ...10 (meridional peak) that are perpendicular to the (vertical) direction of the fibre. Layer lines in the diffraction image are separated by a distance (Figure 15.9, black double arrow) that is much smaller than the distance between the origin of the image and the meridional peak that we attributed to stacked base-pairs. In fact, the spacing is one tenth that between the centre and the meridional peak. Because of the inverse relationship between distances in the fibre and in reciprocal space, the distance is 10 times the stacking distance, *i.e.*, 34 Å, which corresponds to the pitch of a full helical turn (10 bp) in B-DNA.

An intriguing observation is that the dark patches on the layer lines in the diffraction image take the shape of a cross. From interpretations of X-ray photographs of poly-γ-methyl-L-glutamate[98] and the theoretical description of helices and helical macromolecules in terms of their structure factors in reciprocal space,[99] Francis Crick and others were aware that cross-shaped patterns in diffraction images were indicative of a helix (Figure 15.9, magenta). (Again, this is not obvious and the interested reader may turn to references 100 and 101 for a much more detailed description of the theory behind fibre diffraction). A missing layer line at position 4 (red) as a result of an interference in the diffracted X-rays is consistent with two strands that are offset one against another (giving rise to the major and minor grooves). Most importantly, with regard to the mechanism of copying DNA, the diffraction pattern reveals a **dyad** (Figure 15.9, blue) that is normal to the fibre axis, meaning that the two strands are aligned in an **antiparallel** fashion (in actuality, there are local dyads between adjacent base-pairs that relate the sugar-phosphate backbones but not the bases). The positions of the inner maxima (magenta dot) in the image are related to the diameter of the double helix, whereby B-DNA is slimmer than A-DNA. The appearance of diamond shaped patterns (orange) that vary with the level of humidity indicate extensive hydration of the outside of the DNA duplex, in line with a polyanionic form and phosphates mapping to the surface.

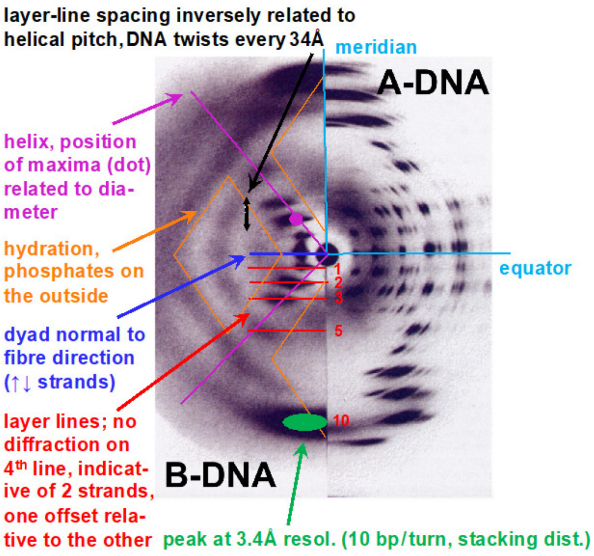

Figure 15.9 Fibre diffraction images provided key insights into the structure of DNA that allowed Watson and Crick to build the model of the double helix.

Fibre diffraction has now all but disappeared as an approach to characterize the structure of biopolymers. Single crystal X-ray crystallography of synthetic oligonucleotides alone and in complex with proteins and cryo-EM of higher order chromatin folds (Section 15.5) are now the main providers of structural information on DNA. But, in the early 1950s, physics and a simple theory of helical diffraction based on fibres afforded revolutionary insights into biology.

15.4.2 Single Crystal X-Ray Crystallography

In four decades of crystallographic structure determination, thousands of structures of DNA, RNA, and protein-nucleic acid complexes have been accumulated, as witnessed by the 13,980 X-ray structures deposited in the **Nucleic Acid Database** (**NDB**) as of February 2022 (http://ndbserver.rutgers.edu),[102] and the (total number of) 186,670 X-ray structures in the Research Collaboratory for Structural Bioinformatics/ **Protein Data Bank** (RCSB/**PDB**) as of February 2022 (https://www.rcsb.org).[103]

Dramatic advances have been made in virtually all areas of X-ray crystallography,[104] including (i) crystallization: sparse-matrix crystallization screens, liquid handing and drop-setting robots, crystallization plate storage hotels, and automatic imaging; (ii) crystal handling: mounting, cryo-protection, storage and shipping; (iii) speed of diffraction data acquisition and processing as well as data resolution: synchrotron radiation, microfocus beamlines, **X-ray free-electron laser** (**XFEL**), higher-flux home sources (Figure 15.10), charge-integrating pixel array detectors, shutterless data collection, sample-changing robots and introduction of new figure-of-merit statistical parameters; (iv) phasing: **multi-wavelength** and **single-wavelength anomalous dispersion** – **MAD** and **SAD**, respectively, soft X-rays, fragment-based searches; (v) electron density map interpretation and model building: computer graphics, automatic chain-tracing and real-space refinement; and (vi) structure refinement: computer power, simulated annealing, full-matrix least squares refinement, maximum likelihood estimation and valence geometry standards.[105–110]

Amplitude and phase of a diffracted X-ray are described by the **structure factor** (F_{hkl}) equation:

$$F_{hkl} = \sum_{j=1}^{n} f_j e^{-2\pi i(hx_i+ky_i+lz_i)} = \sum_{j=1}^{n} f_j e^{i\phi_j} \qquad (15.7)$$

where the **atomic form factor** f is the contribution of each atom j (for example, at scattering angle $\sin\theta/\lambda = 0$: 1 for H, 6 for C, 8 for O, 18 for Cl^-, *i.e.* the number of electrons), ϕ_j is the phase, (x_j,y_j,z_j) is the position of atom j in the unit cell, and *(hkl)* defines a **reciprocal lattice point** at (ha^*, kb^*, lc^*) (an individual dark spot on the diffraction frame depicted (Figure 15.11c) that corresponds to the **real space** plane defined by the **Miller indices** *(hkl)*. William Lawrence Bragg first introduced the concept of diffraction as resulting from waves that are scattered from crystal lattice planes separated by the interplanar distance d.

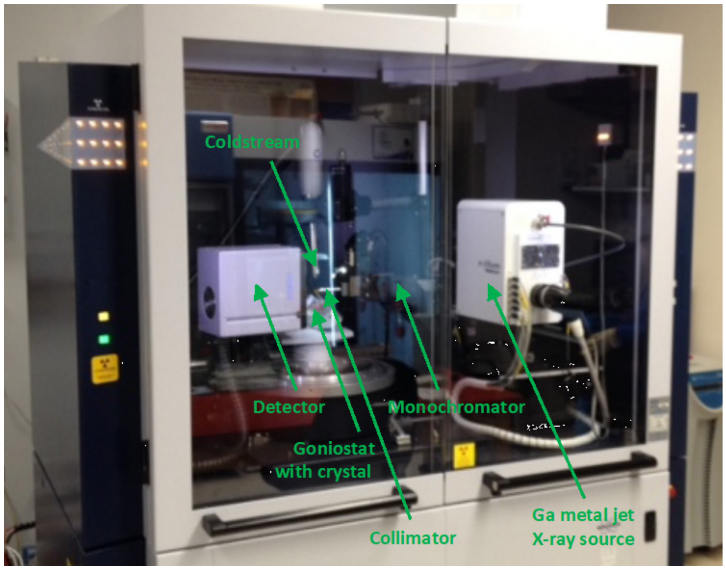

Figure 15.10 The Bruker AXS D8 Venture "MetalJet" X-ray diffraction system at Vanderbilt University. The MetalJet drives a thin stream of liquid gallium alloy at about 50 meters per second. The electron beam, produced using an LaB6 source, is placed on the gallium stream to generate the highest brilliance X-rays available for any non-synchrotron source. Ga Kα radiation has a wavelength of 1.34 Å, located between the lower energy of copper Kα 1.54 Å and the 1 Å X-rays typically used for native data collection at synchrotron sources.

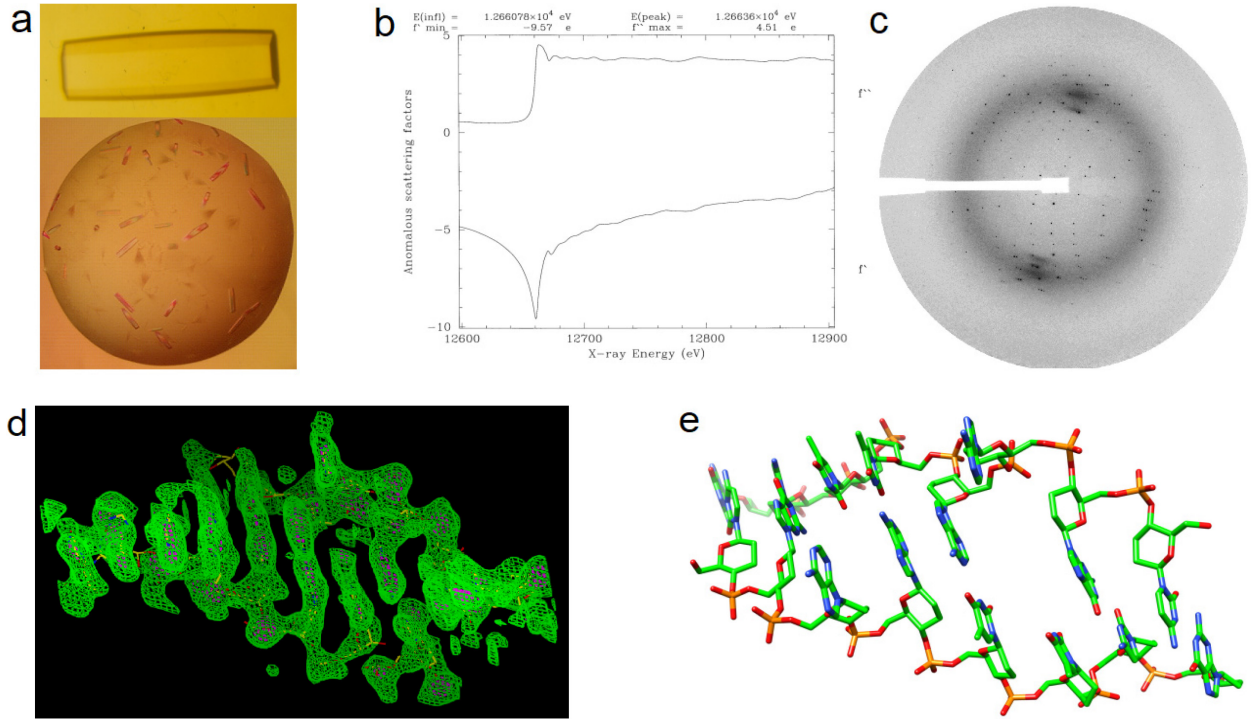

Figure 15.11 Steps in X-ray crystallographic structure solution. (a) Native crystal of the 2′,3′-ideoxyglucopyranosyl nucleic acid duplex (homo-DNA) [dd(CGAATTCG)]$_2$ (top) and crystals of the homo-DNA duplex with a single stereopure phosphoroselenoate (PSe) modification per strand (bottom). (b) X-ray fluorescence spectrum of the PSe-homo-DNA crystal, indicative of the Se K absorption edge (12.6578 keV, 0.9795 Å[ii]) and depicting the wavelength-dependent amplitudes of the real (*f*′) and imaginary (*f*″) anomalous components of the atomic scattering coefficient. (c) X-ray diffraction frame from a MAD data collection. (d) Se-MAD experimental electron density. (e) Final structural model of the homo-DNA duplex.[113]

[ii] **For a full list of anomalous absorption edges, see http://www.bmsc.washington.edu/scatter/**

Accordingly, constructive interference between individual waves only occurs if the path difference is an integral multiple of the wavelength λ. This is expressed by **Bragg's law** that relates the distance d_{hkl} between planes *(hkl)* that dissect the crystal lattice, the scattering angle θ and the X-ray wavelength λ:

$$2 d_{hkl} \sin\theta = \mathbf{n}\lambda \qquad (15.8)$$

where **n** is a positive integer. It establishes the condition on the scattering angle for the constructive interference between individual waves of scattered radiation to be at its strongest.

Single crystal diffraction data include the set of structure factors F_{hkl} extracted from the experimentally measured intensities I_{hkl} of individual diffraction spots (Figure 15.11c), but not the phase information, which is necessary to reconstruct the three-dimensional structure of the unit cell and molecules therein. As is evident from the above structure factor equation, calculating the phase ϕ requires knowledge of the whereabouts of the scattering atoms (x_j,y_j,z_j) – the very goal of a structure determination! – and that is known as the **phase problem**. Solving the phase problem can still present a formidable hurdle to determine certain crystal structures.[114] The basic approaches to determining phase information are **multiple isomorphous replacement** (**MIR**), MAD or SAD, and **molecular replacement** (**MR**). The first two techniques require the incorporation of "heavy" atoms or heavy atom-containing derivatives into specific positions within the molecular lattice. Such incorporation can be achieved by either soaking crystals in a solution containing the heavy atom, by crystallizing the compound in the presence of the **heavy atom**, or by incorporation of a heavy atom into the compound by chemical synthesis or modification, for example 5-bromodeoxyuridine, phosphoroselenoate (Figure 15.11a and b) or selenomethionine for protein bound to NAs.[115,116] For MIR to be successful, it is an absolute requirement that crystals containing heavy atoms are of the same form as the original (isomorphous), so that their X-ray diffraction patterns (Figure 15.11c) can be used together with the diffraction pattern from the non-derivatised crystal to identify the positions of heavy atoms.[117] The consequent **phase** and **structure factor** amplitude information can be used to calculate the phases of all other reflections (similar for MAD/SAD) and compute an experimental **electron density map** (Figure 5.11.d):

$$P(x,y,z) = \left(\frac{1}{V}\right) \sum_{hkl} F_{hkl}\, \mathrm{e}^{-2\pi i(hx_i + ky_i + lz_i)} \qquad (15.9)$$

This map serves as the basis for building the structure and refining atom coordinates and temperature factors to arrive at a final model (Figure 15.11e).

In many determinations of oligonucleotide structures, the phase problem can be circumvented by use of MR, where a related structure or a model structure is used to search for possible solutions to the phase problem. This approach involves the computation of the **Patterson function** *P(u,v,w)*, *i.e.* the **Fourier transform** of the intensities I_{hkl} (= F_{hkl}^2), rather than the structure factors F_{hkl}:

$$P(u,v,w) = \sum_{hkl} |F_{hkl}|^2 \mathrm{e}^{-2\pi i(hu + kv + lw)} \qquad (15.10)$$

As can be seen, no phase information is required to compute the Patterson function. This function is equivalent to the electron density convolved with its inverse and represents a vector map used to compute rotation and translation functions for correctly orienting and positioning, respectively, the search model in the **asymmetric unit** of the unknown crystal structure. This is the smallest volume element that, by applying the space group **symmetry**, generates the unit cell. If a DNA or RNA resists crystallization or if phasing of the crystal structure fails, it is sometimes possible to co-crystallize them with a scaffolding protein that traps the nucleic acid in a non-specific complex.[118,119] Phasing by MR whereby the protein component serves as the search model affords a crystal structure of the nucleic acid molecule that then serves as the search model for phasing the crystal structure of the DNA or RNA alone.

If two or more crystal forms are available and/or the crystallographic asymmetric unit of a specific form contains several independent subunits, it may be possible to sample multiple conformations of the same molecule. Thus, alternative packing forces may trap particular conformational states of a molecule to yield dynamic information. Such an analysis has provided insight into the conformational basis for a bulge-mediated RNA self-cleavage reaction.[120]

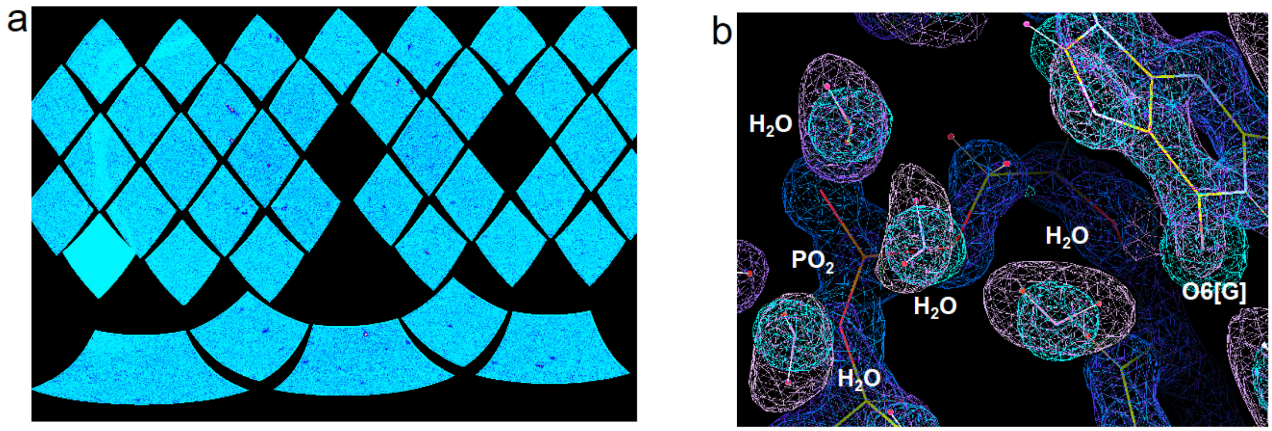

Figure 15.12 Cryo-neutron crystallography of a left-handed Z-DNA fragment. (a) Diffraction spots on position- and time-sensitive MaNDi/SNS Anger cameras of the spherical detector array. (b) Quality of the electron and neutron density (meshwork coloured in blue and lilac, respectively) around water molecules that link guanine and phosphate on the convex surface.

While we emphasize the importance of X-ray crystallography in research directed at the structure and function of macromolecules – the year 2020 marked the centennial of a publication that provided the initial spark of polymer/macromolecular science[121] – we should not forget the spectacular impact of crystallography on **small molecule** structure determination over many decades. The **Cambridge Structural Database** (**CSD**) was established in 1965 to curate crystal structures of small molecules systematically and is run by the **Cambridge Crystallographic Data Centre** (**CCDC**; http://www.ccdc.ac.uk/). Crystal structures of small molecules can now be determined often in a matter of minutes thanks to phasing by **Direct Methods**[122] (in rare cases oligonucleotide structures can also be phased by this approach[123]). The CSD currently contains *ca.* 1.1 million entries, including the structures of nucleobases, sugars, nucleosides and nucleotides (Chapter 3), and has become an indispensable tool in the chemical sciences, crystallography, biophysics, **drug discovery** and related fields.[124]

15.4.3 Neutron Diffraction

X-ray and **neutron diffraction** differ fundamentally as the former interact with electrons and the latter interact with the nucleus of the atom. Moreover, macromolecular crystals that typically have a high water content decay quite rapidly in an X-ray beam (a problem that is alleviated by cryoprotection), whereas neutrons are non-destructive (hence most macromolecular neutron structures are based on data collected at room temperature). The measure of the scattering intensity of a photon (wave) by an isolated atom, the above atomic form factor f is replaced by the **scattering length** b in neutron diffraction. Interestingly, the scattering lengths of hydrogen (^{1}H, $b = -3.7406$) and deuterium (^{2}H, $b = 6.671$) are drastically different ($f = 1$ for both). Thus, deuterium exhibits a scattering length that is not too different from those of carbon (^{12}C, $b = 6.6511$) or oxygen (^{16}O, $b = 5.803$); (For a full list of neutron scattering lengths, see https://www.ncnr.nist.gov/resources/n-lengths/). Therefore, neutron diffraction in combination with crystals of perdeuterated proteins or DNA/RNA grown from D_2O conditions (H/D exchange) offers a number of advantages over X-ray crystallography which is basically blind to hydrogen even at very high resolutions.[125] These include insights into H-bonding patterns as the orientations of hydroxyl and amide groups can be established, as well as the **protonation states** of side chains and even backbone phosphate groups.[126] Neutron crystallography is also able to capture the orientations of water molecules and ammonium ions at active sites, in channels and around DNA duplexes (Figure 15.12).[127] Because neutrons are scattered more weakly than X-rays, data collection can take a week or more even with big crystals rather than just minutes as with X-ray diffraction experiments. Neutron experiments are only available at national facilities. The Macromolecular Neutron Diffractometer on the Spallation Neutron Source (MaNDi/SNS) at Oak Ridge National Laboratory (ORNL, Oak Ridge, TN) currently offers constitutes one of the most powerful neutron setups for experiments with macromolecular crystals word wide. The flux at sample is 1.3 x 10**7 neutrons cm-2 s-1 and the instrument can resolve unit cell dimensions up to 300 Å.[128]

15.4.4 Electron Diffraction

The theoretical frameworks of X-ray and **electron diffraction** (ED) are very similar, but electron beams exhibit shorter wavelength (typically around 0.025 Å). Because the wavelength of an electron decreases as its velocity increases, electron diffraction typically causes significant damage to biological samples and applications have often been limited to 2D crystals, *e.g.* of membrane proteins. An important feature in the use of ED is the fact that electrons interact more strongly with matter than do X-rays; they deposit greatly reduced energy into crystals during a scattering event, two to three orders in magnitude below the energy deposited by X-rays.[129] This renders ED particularly useful for tiny crystals, as demonstrated by recent advances in application of **micro-ED** with extremely small crystals of small molecule compounds[130,131] and biological samples.[132]

15.5 Cryogenic Electron Microscopy (Cryo-EM)

15.5.1 The Basics of Cryo-EM

Cryo-EM commonly refers to **single-particle** *Cryo*genic *E*lectron *M*icroscopy performed using *t*ransmission *e*lectron *m*icroscopes (**TEM**; Fig. 15.13ab). As a broader term, cryo-EM may also refer to other electron microscopy methods, including electron tomography (cryo-ET; see below), microcrystal electron diffraction (**micro-ED**),[133,134] and scanning electron microscopy (**SEM** or cryo-SEM).[135] This chapter focuses on single-particle cryo-EM, which allows visualization of macromolecules at near-atomic resolution (3.5 Å or better). Thanks to recent technical advances, cryo-EM has become a central method of structural biology, making macromolecular structural analyses accessible to the broader scientific community.

In cryo-EM, macromolecules distributed in a thin frozen sample are bombarded with high-energy electrons (100 keV to 300 keV) (Figure 15.13ab). Freezing helps preserve the sample from evaporation in a vacuum, from damage by electrons, and from moving during imaging. Each particle in the sample scatters the electrons, projecting an electrostatic or Coulomb potential (or "projection") that is magnified through electromagnetic lenses and recorded by the electron detector (Figure 15.13b). Hundreds or thousands of **micrographs** recorded from different areas of the sample yield projections of tens to hundreds of thousands of individual particles ("single particles").

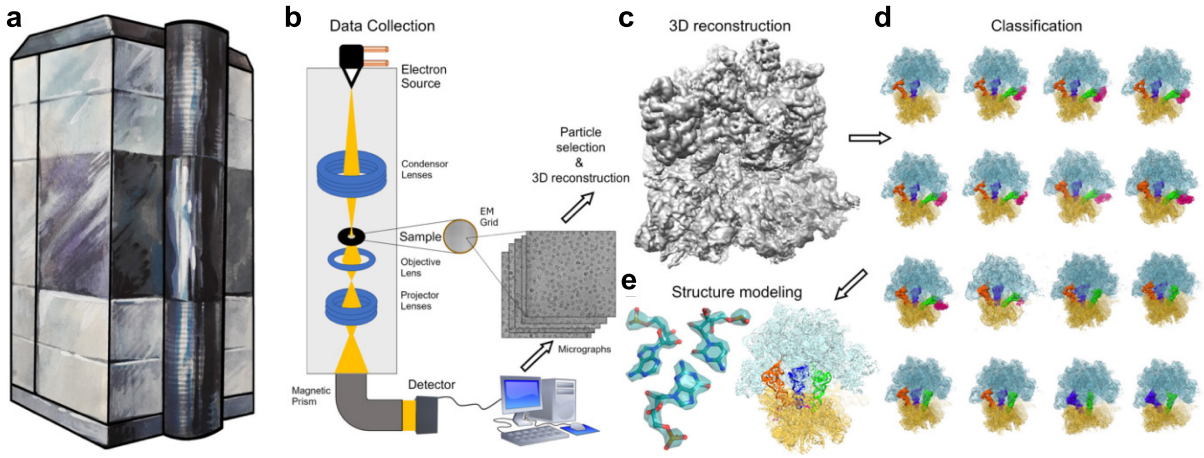

Figure 15.13 Single-particle cryo-EM. (a) Modern transmission electron microscopes are encased to improve the optical performance of the microscope and enhance sample stability. (b) Schematic of a transmission electron microscope. The electron column pumped to vacuum is shown as a grey rectangle. It contains a series of electromagnetic lenses to form particle images (micrographs) from the electron beam (yellow) interacting with a sample on an EM grid. (c) Three-dimensional reconstruction (map of electrostatic potential) is calculated from projections of particles representing different orientations of a macromolecule in the sample. (d) Classification (sorting) of particles improves the resolution and reveals distinct structural states of a macromolecule in the sample. Cryo-EM maps of *E. coli* 70S ribosome functional states during translation elongation[140] are shown (50S and 30S subunits are in cyan and yellow; elongation factor EF-Tu in magenta, transfer RNAs in green (A site), blue (P-site) and orange (E-site). (e) Cryo-EM maps (cyan surface) are used to build atomic structural models (shown in ball and stick representation on the left, and as secondary structure on the right).

Macromolecules in the frozen hydrated sample adopt different orientations (Figure 15.13b: micrographs). Single-particle 2D projections are combined to reconstruct a high-resolution **3D map of electrostatic potential**[136] (Figure 15.13c). 3D classification then enables separation of different compositional and conformational states within the sample (Figure 15.13d). Cryo-EM maps differ from **X-ray crystallographic** maps, which represent **electron density** arising from interaction of **X-ray photons** with electron clouds (Section 15.4). Nevertheless, atomic structural models can be built into the maps using conceptually similar computational methods[137-139] (Figure 15.13e).

15.5.2 Brief History and Recent Advances that Propelled Cryo-EM as a Method.

Electron microscopy has a long history, dating back to 1930s when first electron microscopes were built.[141] Electron microscopy of macromolecules was advanced by the introduction of sample freezing (hence, cryo-EM) in the 1980s.[142] Initial studies allowed viruses to be visualized at 35 Å resolution.[143] The quality of cryo-EM reconstructions continued to improve with advances in computational analyses of imaging data, including the averaging of 2D projections[144] and sorting of distinct macromolecular species in a single sample.[145] Nevertheless, the resolution of cryo-EM reconstructions (10–20 Å) continued to lag behind that of X-ray crystallographic structures. The low resolution of cryo-EM maps was largely due to *indirect* scintillation-based CCD/CMOS detectors that convert electron signals into photons.[146] The slow readout and beam-induced macromolecular drift produced noisy low-contrast projections of "smeared" particles.[147]

The introduction of **direct electron detectors**[146] and improved computational methods between 2010 and 2015 has transformed cryo-EM into a high-resolution method.[148] The rapid direct detection of electrons increases the signal-to-noise ratio[149] and allows for computational correction of beam-induced particle movement.[150] As a result, cryo-EM studies can be used quickly to solve structures at resolutions that rival or exceed those attained by X-ray crystallography — *e.g.*, 1.2 to 2 Å for proteins[151] and 2 Å for nucleoprotein complexes.[152] The number of cryo-EM structures in the **RCSB** macromolecular structure database has grown dramatically in the last 5 years: from ~600 structures (determined between 1975 and 2015) to approaching 10,000 in February 2022. While some macromolecules require extensive sample optimization delaying cryo-EM structure determination, the number of previously inaccessible structures is growing exponentially, illustrating that cryo-EM has revolutionized the field of structural biology.

15.5.3 The Power Of Single-Particle Cryo-EM: High Resolution And Structural Ensembles Resolve Individual Nucleotides and Reconstruct Molecular Mechanisms.

Cryo-EM has several key advantages over X-ray crystallography:

1. Molecules do not need to be crystallised (often the rate-limiting step in X-ray crystallography);
2. Samples can be formed in buffers that mimic cellular conditions or *in vitro* biochemical experiments;
3. A wide range of macromolecular sizes (from ~50 kDa to more than 10 MDa) can be analysed; and
4. Numerous conformations of macromolecules (structural ensembles) can be captured in a single sample (Figure 15.13d).

Perhaps most important, macromolecular **dynamics** and compositional **heterogeneity** have long been considered an obstacle for structural biology. Cryo-EM, by contrast, can resolve multiple distinct structures or even a continuum of different conformations in a single sample. These structural ensembles enable the reconstruction of macromolecular dynamics.[154] Capturing the evolution of structural dynamics over time (also known as **time-resolved cryo-EM**) allows movie-like visualization of biochemical reactions.

Cryo-EM studies are shedding light on nearly every step of gene expression, revealing detailed mechanisms of DNA replication[154-156] and transcription,[157-161] and RNA splicing and translation (Chapter 5). Translation complexes were among the first nucleoprotein structures to be determined at near-atomic resolution by cryo-EM.[162] More recent studies have visualized transient **ribosome** states represented by as few as 1% of the particles in a sample.[140,163] The cryo-EM structures of low-population translation intermediates resolve individual nucleotides in the ribosome core, revealing how the ribosome accurately translates mRNA.[140,164-168] The **spliceosome** — refractory to crystallography — had eluded structural analyses for decades (Section 5.2), but in the last five years, cryo-EM studies have captured spliceosome complexes at different stages of intron splicing.[169-172] Unprecedented details of virus biology (Section 6.4) have also emerged from cryo-EM studies, providing nearly complete views of nucleoprotein capsid structures[173-177] and visualizing how viral mRNAs hijack the translational machinery of the host.[163,172] During the **SARS-CoV-2 pandemic**, cryo-EM

has been the method of choice for fast visualization of **coronavirus** mechanisms and therapeutics.[178,179] Cryo-EM studies are also offering insight into the mechanisms of **CRISPR**-mediated genome editing[180] (Sections 9.4 and 9.5) and small RNA-mediated gene silencing.[181] Furthermore, cryo-EM enables resolution of the structures of post-transcriptional **nucleotide modifications**[182] (Section 5.3) and ligand-binding nucleic acids (*e.g.*, riboswitches (Section 6.5.2)).[183]

15.5.4 Negative-Stain EM and Electron Tomography

In addition to single-particle cryo-EM, two notable applications of transmission electron microscopy are widely used.

Negative-stain EM refers to the use of high-molecular-weight salt (*e.g.*, uranyl acetate) to accentuate low-resolution features of macromolecules. Negative-stain EM can help to assess macromolecule size and shape, and to distinguish monomers, oligomers, and irregular aggregates. As an auxiliary approach to cryo-EM, negative-stain EM is therefore used to optimize sample conditions that reduce macromolecular aggregation or improve structural homogeneity. Importantly, for samples refractory to cryo-EM, negative-stain EM can be used to help characterize macromolecular interactions or structural dynamics at low resolution.[184]

Cryogenic electron tomography (cryo-ET) is used to visualize features of multi-component samples, including cells and tissues. Projection images of the sample are recorded at different tilt angles (usually covering a range from −70° to +70°), producing a "tilt series" that allows 3D reconstruction of a sample segment. The resolution is lower than single-particle cryo-EM because the thick samples analysed by cryo-ET result in a low signal-to-noise ratio that cannot be improved by the averaging of unique micrograph segments, in contrast to single-particle cryo-EM. Nevertheless, in the studies of repeating molecular structures, *sub-tomogram averaging* can be used, which resembles single-particle cryo-EM and yields near-atomic resolution for viruses.[185] Recent advances in electron detection and computational methods enable sub-nanometre resolution for entire cellular compartments[186] and allow individual macromolecules to be identified within cells.[187]

15.6 Optical Microscopy of Nucleic Acids

15.6.1 Fluorescence Microscopy of Nucleic Acids

'Seeing is believing': This old adage reflects the drive of scientists over the centuries physically to look into the natural world, with ever greater detail across the extremes of length scales by developing new instruments and methods, from telescopes to microscopes. Imaging of nucleic acids in their cellular contexts, especially during cell division, has produced iconic images of mitosis, in which **chromatin** (so called because of its ability to become 'coloured' by various stains) condenses to form chromosomes.

Light microscopy of DNA, and its development to **fluorescence microscopy**, relied on the discovery of molecules that selectively stain or label nucleic acids, in order to provide contrast or fluorescent makers. Fluorescent moieties absorb photons of light (promoting an electron to an excited state) (Section 15.1.2), and then emit photons at a longer wavelength as the molecule relaxes back to the ground electronic state (Figure 15.14a). Fluorescence microscopy takes advantage of the different excitation and emission wavelengths of fluorescent dyes to produce high quality images (high contrast, low signal to noise), by filtering out the excitation light before detection (Figure 15.14b). A key advantage of fluorescence microscopy is the ability to selectively label specific molecules of interest in the cell. Several probes have been developed that significantly increase their fluorescence when bound to nucleic acids, and so provide excellent markers for visualization of DNA by fluorescence microscopy. One such commonly used probe is **DAPI**,[188] 4',6-diamidino-2-phenylindole (Section 12.7.2), which binds to the minor groove of DNA (Figure 15.14c). It is frequently used as a DNA marker, but also as a more general label for the nucleus of cells.

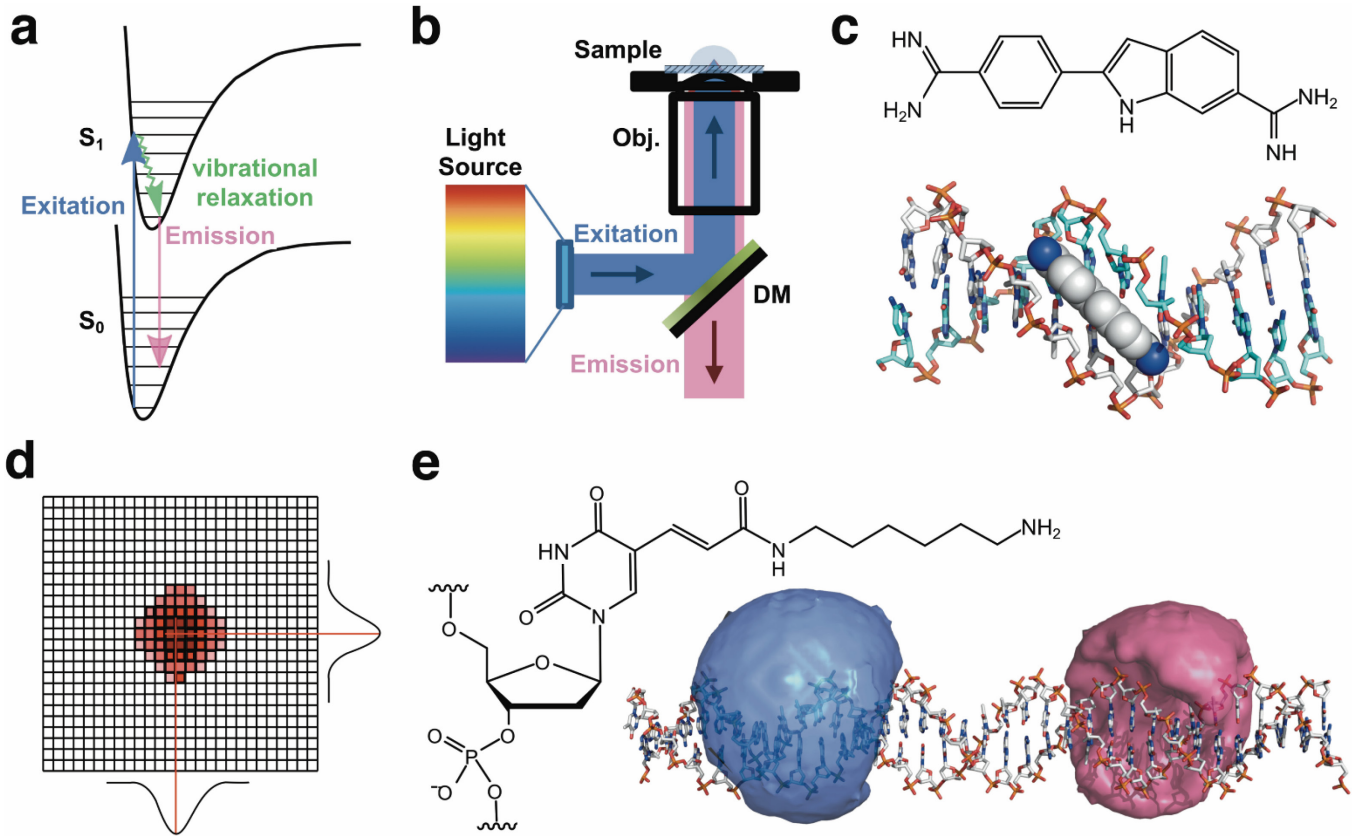

Figure 15.14 Optical microscopy approaches. (a) Fluorescence occurs when a photon of light is absorbed, exciting an electron from the ground state (S_0) to the first excited state (S_1). After vibrational relaxation, a photon can be emitted as the molecule returns to the ground state. Emission occurs at a longer wavelength than excitation, due to the lower energy of the emitted photon. (b) A simple schematic of an epi-fluorescence microscope setup. A specific wavelength of excitation light is selected by a filter and reflected by a dichroic mirror (DM) into the objective lens (Obj.). The emitted light is collected by the same objective, passes through the dichroic mirror (as it is a longer wavelength) and can then be detected on a camera. (c) Top: The chemical structure of DAPI (4′,6-diamidino-2-phenylindole). Bottom: DAPI (space-fill spheres) is shown binding to the minor groove of DNA (PDB structure 1D30). (d) A schematic showing the pixels of a camera with an example PSF. Sub-pixel localisation is achieved by fitting the PSF (*e.g.* to a 2D Gaussian function – projected black lines) allowing the centre of the PSF to be precisely determined (red lines). (e) Left: The chemical structure of the modified 2'-deoxythymidine nucleotide, with a common primary amine linker (dT-C6-amino) used for labelling DNA at internal positions with NHS-derivatised dyes. Right: duplex DNA (sticks) with the accessible volumes (AVs) of attached dyes (blue and pink). The AVs show the sterically allowed positions of the dyes with respect to the DNA as they diffuse around, tethered to the DNA by a flexible linker.

15.6.2. Fluorescence *in situ* Hybridization (FISH)

While DAPI is selective for DNA, there are many applications which benefit from a probe with sequence specificity. This is achieved in the method known as **Fluorescence *in situ* hybridisation** (FISH). Here, a fluorescently-labelled single-stranded DNA oligonucleotide hybridises to the target nucleic acid *via* complementary base-pairing interactions, and is subsequently detected, visualized and quantitated by fluorescence microscopy. Both DNA and RNA can be targeted, enabling the study of genomic sequences and transcribed mRNA in individual cells. This technique has proved particularly powerful in the field of cytogenetics (which studies the structure of DNA within the nucleus) as probes can be introduced into cells, and then passed to the nucleus where they hybridise to their target sequence, revealing both the positions and the number of occurrences of specific sequences on the **chromosomes**. As such FISH can be used to detect a range of chromosomal abnormalities, including deletions, duplications and translocations, that aid in the diagnosis of diseases such as cystic fibrosis, muscular dystrophy and leukaemia.[189]

15.6.3. Super-Resolution Microscopy – DNA-PAINT

The resolution in traditional fluorescence microscopy is limited to half the wavelength of the emitted light (the diffraction limit, ~200 nm). Importantly, resolution beyond the diffraction limit can be achieved by fitting the point-spread function (PSF) generated by photons emitted from a single fluorophore (Figure 15.14d) to determine its position with sub-pixel precision (typically down to ~20 nm). This is termed **single-molecule localisation microscopy** (SMLM). To achieve it, only a small number of fluorophores in the field of view can be emitting photons at a given time, otherwise the PSFs from adjacent fluorophores overlap and cannot be fitted precisely if at all. Many super-resolution methods work by switching the fluorescence of individual molecules on and off (exploiting a photophysical property of some dyes termed 'blinking') and taking sequential images, from all of which the single-molecule localisations are then combined to produce a final super-resolved image.

DNA-Points Accumulation for Imaging in Nanoscale Topography (DNA-PAINT), provides an alternative method for localisation-based super-resolution microscopy, which does not rely on blinking from dye photophysics.[190] Instead, DNA-PAINT makes use of the transient hybridization of short fluorescently-labelled oligonucleotides ('imager' strands) to complementary sequences (docking strands), which are often chemically attached to antibodies, for specific labelling of target proteins in cells. The imager strands are washed onto the cells, and prior to docking diffuse rapidly, spreading out their fluorescence emission over many camera pixels in a single frame, rather than forming a detectable PSF. In contrast, when they bind to a docking strand, they are fixed in place for a time, emitting many photons that produce a well-defined PSF, yielding a precise localisation. The number of PSFs present in a given frame can be finely tuned by modulating the on- and off-rates for imager binding. Increasing the concentration of the imager strand will increase the on-rate, while the off-rate can be modulated by changing the strand length, CG content (Figure 2.1), temperature, or ionic strength of the imaging buffer. DNA-PAINT has been used to image **DNA origamis** (Section 2.6.3), by direct hybridisation to target sequences, and cellular proteins, *via* antibody-attached docking strands, achieving up to ~1 nm localization precision.

15.6.4. Förster Resonance Energy Transfer (FRET) in Nucleic Acids

FRET is a photophysical effect whereby a 'donor' dye is excited and transfers its energy in a non-radiative process to an 'acceptor' dye, which subsequently emits a photon (Section 15.1.2). The efficiency of this process scales with the inverse sixth power of the inter-dye distance, resulting in a useful dynamic range of ~3-10 nm for many popular donor-acceptor pairs, well-matched to the study of nucleic acid conformations and dynamics. As such, FRET has been used to study the dynamic conformations of many DNA structures, both alone, and in complexes with binding proteins. Key to FRET experiments, as for most of the fluorescence methods discussed above, is the production of dye-labelled oligonucleotides. These can either be bought direct from suppliers, or amino-modified bases can be incorporated into synthetic oligonucleotides and later coupled to NHS-ester derivatives of the required dyes (Figure 15.14e). While relative measures of FRET efficiency have been used to infer the conformations and dynamics of many key DNA structures, *e.g.*, Holliday junctions,[191] more recently it has become possible to measure absolute FRET efficiencies.[192] Coupled with careful modelling of the dye position relative to the biomolecule (Figure 15.14e) and the development of single-molecule detection, a number of labs have now used FRET to determine structural models of DNA-protein complexes, for example, HIV-1 reverse transcriptase bound to a DNA primer template[193] and DNA Polymerase I bound to gapped-DNA.[194]

15.7 Atomic Force Microscopy

The field of **scanning probe microscopy** (**SPM**) refers to a class of techniques, devised originally during the 1980s, which provide images of samples through the use of a minute probe (tip) to 'feel' the topography of a surface. Members of this class include the scanning tunnelling microscope (STM), the **atomic force microscope** (**AFM**), the scanning near-field optical microscope and the scanning capacitance microscope. Of these, it is the atomic force microscope which has contributed most to the structural analysis of nucleic acids.[195]

In contrast to more conventional microscopies (including cryo-EM), the high signal-to-noise of an AFM measurement allows molecular imaging resolution of the dimension of the DNA double helix *without*

averaging or labelling. A further advantage of an AFM for the structural analysis of biological samples is that it can image non-conducting surfaces in liquid environments. Unsurprisingly, soon after its invention the AFM was applied to analysis of many biological systems, including membrane proteins, DNA and RNA, protein-nucleic acid complexes, and more complex structures such as chromosomes.[196]

The instrumental setup of an AFM consists of a sharp probe (typically silicon or silicon nitride, Si_3N_4), which is deposited or etched at the apex of a flexible cantilever of defined stiffness (spring constant). As the probe is scanned over the surface of the sample, the cantilever bends or twists in response to forces acting between the probe and the sample. The position of the probe is calculated using a laser beam which is reflected off the back surface of the cantilever onto a position-sensitive photodiode. The cantilever is raster-scanned across the surface, and the variation in bending of the cantilever in response to surface changes is used to produce an image of the surface topography (Figure 15.15a).

The AFM can be used in multiple imaging modes that include contact, tapping, and non-contact, with each describing the way in which the probe interacts with the surface during imaging.[197] For imaging of biological samples, tapping mode (including the more recently developed PeakForce™ tapping mode) is more commonly used. In this mode the probe intermittently comes into and out of contact with (*i.e.* lightly taps) the sample as is it scanned over the surface. This reduces the strong lateral forces associated with contact mode, which can denature or sweep away soft or poorly immobilized samples.

Since 2010, substantial advances have been made in controlling the forces applied during imaging in tapping mode. In particular, rapid force-distance imaging modes such as PeakForce™ tapping, have allowed for imaging of biomolecules at higher resolution.[198] PeakForce™ tapping operates by performing continuous sinusoidal force curves across the sample, by rapidly oscillating a piezoelectric scanner in the *z* direction at 1–32 kHz, and using feedback to maintain constant the maximum applied force. This allows for imaging at forces as low as 50 pN, with imaging scans on the order of seconds.[199]

Up to 2020, AFM investigations of nucleic acids have revealed structural variation and conformational heterogeneity in non-crystalline, topologically complex molecules (Figure 15.15b).[200,201] AFM also allows for conformational changes within DNA molecules and between DNA and proteins to be investigated through time-lapse microscopy.[202] Such studies have included tracking DNA condensation and the processing and/or degradation of DNA by enzymes.[203] One of the disadvantages of conventional AFM imaging is the speed of imaging for capture of dynamic events (normally 30–60 s per image). This problem has been overcome using small cantilevers which are able to image faster and capture more dynamic events (< 2 s per image).[204,205] While these cantilevers were initially used only on bespoke systems, advances in hardware have allowed them to be more widely adopted, with the result that they are now used routinely in commercial AFMs.

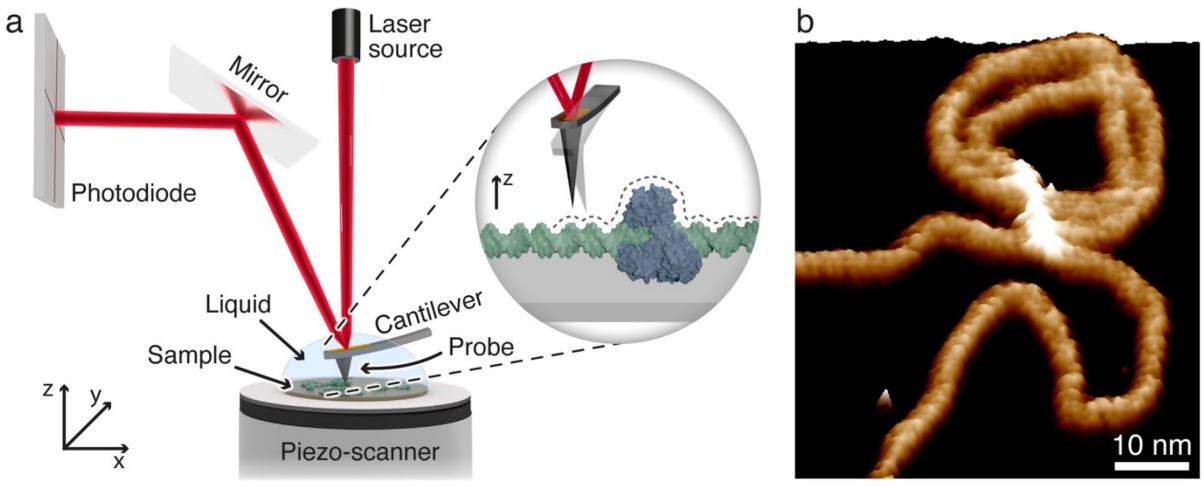

Figure 15.15 (a) Schematic depicting AFM imaging of DNA in solution. A sharp tip is scanned line-by-line across the sample surface to build up an image of the surface topography (features not to scale). The topography is read out as a function of the bending of the cantilever *vs.* the position of a laser beam deflected onto a position-sensitive detector (4-quadrant photodiode). Inset: dotted line depicts the AFM probe tracking the topography of a DNA-protein complex. (b) High resolution scan of supercoiled plasmid DNA imaged in liquid in PeakForce™ tapping mode.[201] Corrugation along the molecule corresponds to the major and minor grooves of the DNA double helix.

Two main factors influence the level of resolution attainable in AFM images of nucleic acids. The first is the sample preparation method employed for the immobilization of DNA to the underlying surface and the second is the character of the AFM probe (spring constant and sharpness).[201] For immobilisation of DNA onto a mica surface, a surface modification must be employed to bridge the electrostatic repulsion between the negative charges of the DNA and the mica. A range of immobilization methods are used, however the most widely adopted is the divalent cation method.[206] This treats an unmodified mica surface with a buffer containing divalent metal ions (usually Ni^{2+} or Mg^{2+}) to bridge the negative charges on the DNA and the mica surface.[207]

The AFM probe is the point of contact between DNA and the surface and so is central to achieving high resolution. To ensure adequate force control for high resolution imaging, the probe should be highly flexible, with a spring constant < 0.25 N m^{-1}. Furthermore, one of the notable drawbacks for AFM is the probe convolution/broadening effect. This results in a difference between the observed dimensions of the image feature and the actual molecular dimensions. It arises when the radius of curvature of the imaging probe (typically ~10 nm) is greater than the size of the feature to be imaged. Recent advances in AFM manufacturing have produced probes of ~1 nm tip radius, reducing the uncertainty in the imaging, and allowing for imaging at sub-molecular resolution.

In addition to imaging, AFM can also be used to investigate the mechanical properties of single biopolymeric molecules including DNA with **single molecule force spectroscopy** (**SMFS**).[208] In SMFS, the opposite ends of a linear double-stranded DNA molecule are tethered between the AFM probe and substrate, and then the molecule is extended/stretched as the probe-substrate separation is increased. Pulling on single linear DNA molecules (with random sequence) produces force-extension curves (Figure 15.16). The plateau region at around 65 pN corresponds to the overstretching transition of DNA in which DNA is stretched from its B-DNA state to an overstretched state (around 1.7 times its B-DNA contour length). Upon further extension, a second transition occurs during which the DNA can be melted into two single strands. Researchers have investigated the influence of different base sequences on this 'mechanical fingerprint', which has been applied to the investigation of more complex molecular architectures including DNA/RNA hairpins. The effects of DNA-binding drugs on the mechanical properties of DNA have also been investigated (Figure 15.16), showing promise for applications of SMFS in fields such as drug development.[208]

Advances in the availability and usability of AFM over the recent years have led to AFM becoming a routine tool for investigation of nucleic acid structure and interactions. This has been driven by the advantages it holds over mainstream microscopies (including cryo-EM and fluorescence microscopy) for molecular resolution imaging of single biomolecules in solution. It is limited by inherently being a surface-sensitive technique, with temporal resolution limited by the speed of the scanner. However, the combination of AFM with other rapidly developing experimental approaches, such as single molecule fluorescence, also holds considerable promise for deepening our understanding of the structure and function of nucleic acids.

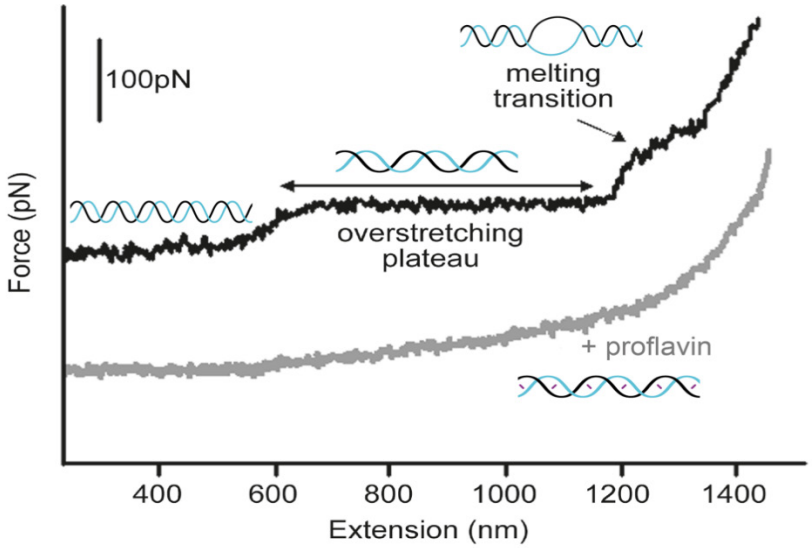

Figure 15.16 Representative force-extension curves of linear double-stranded DNA fragments recorded in 10 mM Tris buffer containing 1 mM EDTA and 150 mM NaCl (black) and the same buffer containing intercalating drug proflavine at a saturating concentration (*ca.* 3 μg ml^{-1}) (grey). Schematics at each point of the force curve illustrate the structural changes observed.

15.8 Electrophoresis

In free solution the movement of DNA in an electric field is independent of shape and molecular mass and dependent only on charge. However, when DNA is exposed to an electric field in a gel matrix, the movement is dependent on size and shape as well as charge. Gels commonly used for nucleic acid electrophoresis are made of **agarose** or **polyacrylamide**. Both types of material consist of three-dimensional networks of cross-linked polymer strands, which contain pores whose size varies according to the concentration of polymer used. The mobility of DNA in such gels is dependent mainly on size and shape since the charge per unit length of DNA is effectively constant.

15.8.1 Principles of Electrophoresis

For linear DNA, there is an inverse relationship between size and rate of migration, such that mobility is approximately inversely proportional to $\log_{10}$ of the molecular mass. Horizontal **agarose gel electrophoresis**[209] is useful for the separation of linear DNA molecules up to 2000 kilobase-pairs (kb). Staining with nucleic acid-binding dyes, like **ethidium bromide** or **SYBR gold**,[210] can reveal the position of DNA in the gel after UV illumination (Section 15.1). Alternatively, the DNA is labelled: traditionally this label was radioactive ^{32}P,[211] in which case the DNA is detected by use of a phosphor-imager. But now for many applications, a fluorescent dye may be a suitable label, which is much easier and cleaner to use than radioactivity and in which case the gel can be imaged using a gel documentation system.

Polyacrylamide gel electrophoresis (**PAGE**) is typically used for smaller DNA or RNA fragments. It can be used preparatively for species up to 1 kb and forms the basis of gel-based nucleic acid sequencing and analysis methods when resolved under denaturing conditions, such as 6–8 M urea.[212]

One proposed explanation for the theoretical basis of size selection in gel electrophoresis is that the mobility is proportional to the volume fraction of the pores that can be entered. This theory predicts that mobility would decrease with increasing gel concentration and also with increasing molecular mass. This is consistent with the observation that very small nucleic acid fragments move independently of molecular mass. An alternative explanation posits **reptation**,[213] which is an end-to-end movement of the DNA through the pores of the gel. This theory explains the observed movement of DNA through gels more accurately, especially at high voltages where mobility becomes independent of molecular mass. In practice, the analysis of observed mobility in gels is made by comparison with and extrapolation from known standard samples (often one lane of the gel is loaded with a **ladder** consisting of bands of known size for comparison).

15.8.2 Electrophoresis and Topology

The topological conformation of a DNA molecule can have a marked effect on the rate of migration and the response of the sample to gel concentration and applied voltage. Relaxed **circular DNA** and **supercoiled DNA** migrate anomalously in agarose gel electrophoresis in comparison to linear DNA of similar molecular mass.[214] The mobility of supercoiled DNA (Section 2.3.5) is also dependent on the **linking number**; as the linking number increases, so does the mobility of the sample. In addition, its mobility changes with the concentration of ethidium bromide because increased intercalation changes the number of superhelical turns. Resolving RNA on native polyacrylamide gels is also dependent upon folding and shape but in less predictable ways due to the three-dimensional folding of single-stranded RNA.

The shape of linear DNA also has a strong effect on electrophoretic mobility. Bends or curvature in the DNA helix have the effect of slowing the migration relative to non-bent DNA. The extent of retardation depends on the end-to-end distance of the DNA and so is dependent on the location of the bend within the DNA helix (Figure 15.17). A qualitative theory, derived by Lumpkin and Zimm, supports the general rule that mobility is related to the mean square end-to-end distance.[215] In addition to the above examples, many elegant experiments in electrophoresis have been performed to examine the properties of poly(A) tracts, extra-base bulges, protein binding (gel-shift) and Holliday junctions.

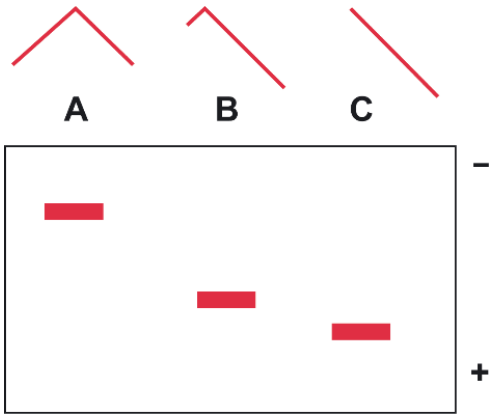

Figure 15.17 The electrophoretic mobility of curved DNA is dependent on the exact location of the bend.

15.8.3 Electrophoretic Mobility Shift Assay (EMSA)

Electrophoretic mobility shift assays (**EMSA**), also known as gel shift or band shift assays, can be used to observe binding of proteins to nucleic acids under native conditions. These assays often employ radiolabelled or fluorescently labelled nucleic acids, although label-free techniques have been described.[216] The labelled and purified nucleic acid, at concentrations below the expected or known dissociation constant (K_d), is typically titrated with increasing concentrations of the protein of interest. Protein binding causes an increase in the effective size of the complex, which leads to reduced mobility through the gel matrix. The number of shifted bands that appear and their relative size can help determine binding stoichiometry and cooperativity, while their quantification and concentration dependence can provide K_d value approximations.[217] The appropriate use of competitor nucleic acids or proteins can confirm specificity. For complexes composed of multiple proteins, *e.g.*, ribonucleoprotein particles (RNPs), their order of addition can further reveal the order of RNP assembly.

15.8.4 Pulsed Field Electrophoresis

The fact that DNA of very high molecular mass does not separate well under normal electrophoretic conditions has led to the development of **pulsed field electrophoresis**.[218] There are several variations of the technique, but in general the use of a voltage pulse with a resting period in between pulses changes the mobility of large DNA fragments and this can be used to separate DNA species 1000 kb or larger. The use of an alternating polarity field, with the positive pulse being longer than the negative pulse, enables the efficient resolution of very large DNA fragments. The application of inhomogeneous, perpendicular electric fields has been used to separate whole yeast chromosomes. In all these applications, the size and duration of pulses can be tuned to obtain the specific resolution required.

15.8.5 Capillary Electrophoresis

Electrophoresis is not limited to gels. Indeed, nucleic acids researchers often carry out electrophoresis in buffer-filled capillaries. Capillary electrophoresis presents two significant advantages relative to gels: the narrow internal diameter of the capillaries allows the heat generated during electrophoresis to dissipate quickly, even at high electric fields. Secondly, and perhaps even more importantly, capillary electrophoresis is easier to automate than gel electrophoresis.

For example, in automated Sanger sequencing based on fluorescently labelled dideoxynucleoside triphosphates (Section 8.3), the flexible capillaries can readily be built into an instrument and coupled with a fluorescence detector.[219] A different implementation of capillary electrophoresis, in single-use microfluidic chips, is the Bioanalyzer instrument commonly used to check the size of DNA or RNA fragments in a next-generation sequencing library, or to check the quality of an RNA sample.[220]

15.9 Chromatographic Methods

Chromatography is a family of separation techniques based on passing a solution containing multiple components over a solid phase **adsorbent**. The dissolved components are separated based on differences in their relative affinities for the adsorbent and the **mobile phase** (**eluent**). Compounds with higher affinity for the adsorbent will elute at a longer **retention time**; those that spend a higher fraction of their time in the mobile phase will elute more quickly (shorter retention time). The three types of chromatography most commonly used in nucleic acids research are reverse-phase HPLC, ion-exchange HPLC, and size exclusion chromatography. Chromatography is used both for preparation and for analysis of therapeutic and research-grade oligonucleotides.[221]

In all of these types of chromatography, the nucleic acids of interest can be detected using their UV absorbance at 260 nm (Section 15.1). When using a multi-wavelength detector, it is common to set a second detection wavelength to a region where nucleic acids show lower but non-zero UV absorbance (such as 280 or 290 nm) – this in essence allows a detector to work with a higher dynamic range since a peak that is too intense for the detector to measure accurately at 260 nm might still be a well-defined peak when detected at 290 nm. Detecting at multiple wavelengths can also help in quickly identifying peaks corresponding to impurities that are not nucleic acids by their ratio of absorbance at the different wavelengths. In addition to absorbance, researchers may also employ detectors such as fluorescence (in the case of a fluorescently-labeled oligonucleotide; Section 15.1) or conductance (in the case of desalting, to identify salts which may be invisible to other detection methods).

15.9.1 Reverse-Phase HPLC

Historically in chemistry, liquid chromatography was originally carried out using a polar stationary phase such as alumina, cellulose or acetylated cellulose, or silica gel with a low to moderate polarity organic solvent as mobile phase. For this reason, when chromatography is carried out with a mobile phase that is more polar than the stationary phase, it is called reverse-phase. The most common columns used for reverse-phase chromatography contain silica functionalized with a hydrophobic group such as a C18 hydrocarbon chain. Nucleic acids are added to the column in buffered aqueous solution and then eluted using increasing percentage gradients of organic solvent (typically acetonitrile or methanol).[222]

Nucleic acids are highly polar, and ion-pairing reagents such as triethylammonium acetate are used to improve their separation on **reverse-phase HPLC**. The triethylammonium cation associates with the anionic phosphate backbone, increasing its affinity for the hydrophobic stationary phase.

The secondary structure of a nucleic acid analyte can modify its interactions with the stationary phase. This can confound analysis and make separation less efficient. Thus, reverse-phase HPLC is often carried out at elevated temperature (*e.g.,* 60 °C) by warming the HPLC column in a small oven integrated with the HPLC instrument.

A major challenge is the separation of a full-length oligonucleotide from its "n–1" neighbour – *i.e.*, species missing a single nucleotide. To aid such separations, it can be helpful to have the 5′-*O*-terminal dimethoxytrityl (5′-*O*-DMT) group in place, to be removed after the HPLC separation. The presence of a hydrophobic DMT group on the oligonucleotide leads to increased retention time relative to its impurities (including the "n–1" species), and this enables cleaner separation. The DMT group is subsequently removed by treatment with 80% acetic acid followed by desalting.

15.9.2 Ion-Exchange HPLC

Nucleic acids are polyanions and therefore interact strongly with cationic solid supports. After the oligonucleotide mixture has been loaded onto the column in a low-salt buffer, they associate strongly with the cationic column, but can be eluted using solutions with increasing concentration gradients of salt (typically either sodium perchlorate or sodium chloride). This approach is called **ion-exchange** or **anion-exchange** (**AEX**) chromatography.[223] Longer sequences (containing more phosphate linkages) bind more tightly to the cationic support than shorter sequences. Thus, AEX chromatography is useful for separating oligonucleotides

from shorter 'failure' sequences. Compared to reverse-phase HPLC, the AEX approach is typically not as sensitive to modifications that do not affect length. It may be necessary to use both approaches to resolve or remove all impurities.

When this approach is used for sample preparation, the purified nucleic acid necessarily contains salt. This can be removed by ultrafiltration, by precipitation (if using perchlorate, which is relatively soluble in acetone or ethanol), by size-exclusion chromatography, or by using a volatile buffer (triethylammonium bicarbonate).

15.9.3 Size-Exclusion Chromatography

The two techniques discussed above achieve separation of compounds based on their hydrophobicity and total charge. A third technique often used in nucleic acids research separates compounds based on size. Such **size-exclusion chromatography**, also known as **gel-filtration chromatography**, makes use of a porous solid support such as a dextran polymer (Sephadex). Small molecules freely diffuse into and out of the pores, while larger molecules are excluded from the pores. Thus, the larger molecules experience a smaller effective column volume, and therefore elute more quickly from the column. Meanwhile the smaller molecules, experiencing a larger accessible volume, are retained longer on the column.[224]

A common application of size-exclusion chromatography in nucleic acids research is desalting or buffer exchange of synthetic or gel-purified nucleic acids, *i.e.* the removal of excess salt or other small molecules in a nucleic acid sample. A conductance detector can help to identify the salt peak to ensure its full separation from the nucleic acid peak. In such a case, the porous support would be packed into a column and fitted to a medium-pressure chromatography instrument. Alternatively, for routine removal of excess salts from oligonucleotides, it is common to use pre-packed columns of a Sephadex gel with very small pores and use gravity or centrifugal force to drive the elution. The behaviour of these columns is sufficiently predictable that the elution volumes of nucleic acid and salt fractions can normally be assumed rather than measured directly.

In addition to its routine use in desalting, size-exclusion chromatography has been used to resolve nucleic acid secondary structures[225] and for denaturing and native RNA purification for non-coding RNA (Chapter 6).[226]

15.10 Centrifugation

Analytical Ultracentrifugation (**AUC**) is a powerful analytical tool that allows the hydrodynamic or thermodynamic, solution state behaviour of proteins and nucleic acids to be characterised.[227,228] The AUC technique should not be confused with preparative procedures carried out by use of standard ultracentrifuges.

Classical experiments of Meselson and Stahl in 1957 utilised **equilibrium density gradient centrifugation** in caesium chloride (CsCl) to establish unambiguously the semi-conservative nature of DNA replication.[227] A CsCl density gradient can also be established by diffusion under the influence of a centrifugal force to give a range of densities that encompass the buoyant densities of the molecular species to be separated. When centrifuged in such a gradient, a sample of DNA migrates until it reaches the point at which its own buoyant density equals that of the CsCl gradient. A number of different factors affect the buoyant density of DNA, for example the G + C content, the degree of hydration, the presence of other ions, pH, or the presence of DNA binding or intercalating agents. The fact that intercalating dyes (Section 12.5) alter the buoyant density of DNA has been exploited to provide an experimental method for the preparative purification of DNA. The concentration of ethidium bromide bound to DNA is dependent on the superhelical state of the DNA. For example, closed, circular plasmid DNA binds at a higher density than either nicked circular or linear DNA. Other cellular components are removed as a pellet at the bottom of the tube (*e.g.* RNA) or as a surface layer (proteins).[229]

Modern, computer controlled, AUC instruments (such as the Beckman XL-I, which is equipped with both absorption and interference optics) allow characteristics such as molecular mass, stoichiometry of complexes, conformation and shape, diffusion and sedimentation, self-association and equilibrium constants to be studied quantitatively (Figure 15.18). The analytical ultracentrifuge spins the sample in a vacuum at a set spin speed and fixed temperature. At predefined time intervals, the instrument records the concentration distribution in the sample by measuring the absorbance. It is possible to attain speeds as high as 60,000 rpm (~250,000 × *g*). The sample itself is a solution in its native state and under biologically relevant conditions. It is held in specialised cells designed to withstand the high gravitational forces, which at the same time allow transmittance of light through the sample in the wavelength range 200 to 800 nm. The centrifuge cells

accommodate both the sample and reference buffer, and the cells are assembled prior to each experiment from a cell housing, quartz or sapphire windows and one of several different types of centrepieces. Different types of rotors are available that can be used at different spin speeds or that contain either four or eight cell housings, one of which is used for the reference buffer in each case.

In general, there are two different approaches to AUC: (1) **sedimentation velocity** and (2) **sedimentation equilibrium**. Sedimentation velocity is a hydrodynamic technique that is a probe for both mass and shape of a macromolecule.[230] A uniform solution of the sample is placed into the centrifuge cell, which is spun at high speed. This drives sedimentation of the macromolecule towards the bottom of the cell. This means that the macromolecule is depleted at the top of the cell (*i.e.* near the meniscus) and there is a sharp boundary, in terms of solute concentration, between the depleted region and the uniform concentration of the sedimenting molecule. A series of scans, where the sample concentration is measured as a function of radial distance, is taken at discreet time intervals and these measurements allow the rate of movement and broadening of the boundary to be monitored over time. From these data, the sedimentation coefficient, ($\boldsymbol{s}$) can be determined and is directly proportional to the mass ($\boldsymbol{m}$) of the solute and inversely proportional to the frictional coefficient ($\boldsymbol{f}$), which is in effect a measure of size.

These relationships are summarised in the Svedberg equation (15.11):

$$S = \frac{v}{\omega^2 r} = \frac{m(1-\bar{v}\rho_{sol})}{f} \qquad f = \frac{RT}{ND} \qquad (15.11)$$

where v (= dr/dt) is the velocity of the molecule, $\omega^2 r$ is the strength of the centrifugal field ($\omega = 2\pi \cdot \text{rpm}/60$ and r is the radial distance from the centre of rotation), ρ_{solv} is the density of the solvent, N_A is Avogadro's number, D is the diffusional coefficient, R is the gas constant and T is temperature.

Sedimentation equilibrium is a thermodynamic technique that is sensitive to mass but not to shape of the molecular species.[230] The principle difference between sedimentation velocity and sedimentation equilibrium is that in the latter the initially uniform sample is spun at lower angular velocities than those employed in sedimentation velocity experiments. At the slower speeds used here, the macromolecule again starts to sediment towards the bottom of the cell and thus the concentration towards the bottom begins to increase. However, diffusion processes begin to oppose sedimentation and after a suitable period of time the two opposing forces reach equilibrium. If several different species with different molecular weights exist, for example DNA that is free and bound to a ligand, or different association states of multi-subunit proteins, then each of the species will be distributed over the solution until it is at equilibrium. Higher molecular weight species will be nearer the bottom of the cell, whilst lower molecular weight species will be present at the top of the cell.

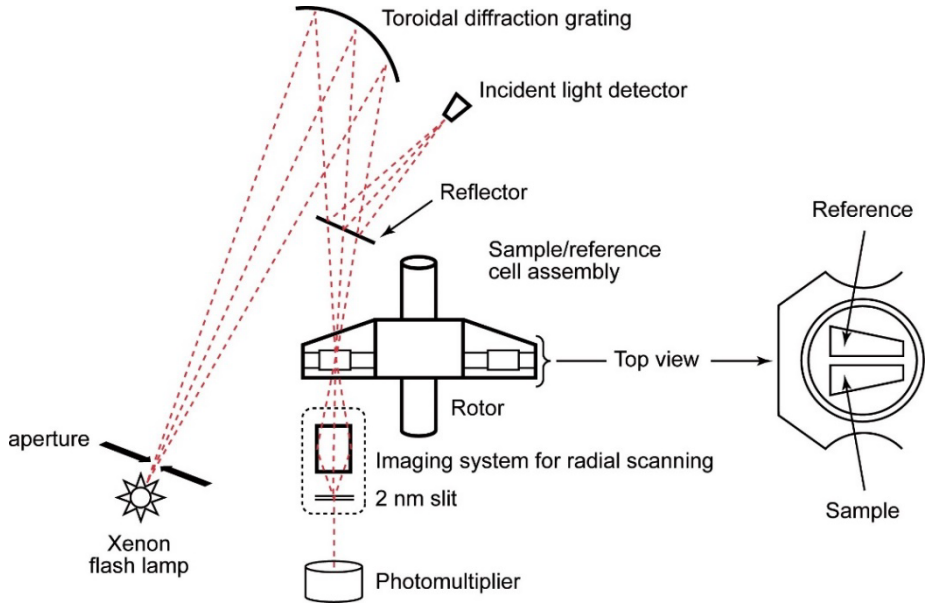

Figure 15.18 Schematic diagram of the optical arrangement of an analytical ultracentrifuge (this diagram is based on the Beckman Optima XL-A).

Sedimentation equilibrium is particularly useful for evaluation of the equilibrium association constant for reversible interactions such as ligand binding, or protein self-association. The technique is sensitive to K_{obs} values in the range 10–100 M^{-1} but can also be used to measure affinities up to 10^7 M^{-1}.[231,232] The detailed mathematical theory that underlies both sedimentation velocity and equilibrium can be very complex, but there are several useful data analysis software packages that allow a non-expert to analyse and extract useful information.[233]

Another type of ultracentrifugation useful for examination of viscosity changes employs **sucrose density gradients**. This method can be used for studying various RNA species where separation depends largely on the size of the RNA molecule, or for separating different classes of ribozymes.[234] All these nucleic acid species have a higher buoyant density than the sucrose gradient and hence equilibrium is never reached; therefore separation depends on the different rates of migration of molecules through the sucrose gradient.

15.11 Light Scattering Techniques

Light scattering is another useful technique for determination of the molecular weight and size of biological macromolecules and evaluation of their homogeneity. Two approaches can be differentiated, **static light scattering** (**SLS**) and **dynamic light scattering** (**DLS**).[235-238] Both rely on laser light in the visible range with a typical wavelength of 589 nm (sodium D-line). Light scattering occurs on the outer electrons of atoms in a molecule, whereby the scattering intensity is proportional to the product of molecular weight and concentration and thus stronger for larger molecules. SLS measures the scattering intensity per unit time, the absolute mean intensity, that reflects the molecular weight of molecules. Thus, SLS allows measurements of the **molecular weight** (M; 1000 to 10^9 g mol^{-1} range), and size (**radius of gyration** R_g; 10 to 1000 nm range), as well as molecular interactions, *i.e.*, the **second virial coefficient**. DLS uses an increased sampling time that permits the observation of fluctuations in the intensity over time that reflect the diffusion coefficient of molecules. DLS allows measurements of the **hydrodynamic radius** R_h (1 to 1000 nm range) and relaxation times in gel systems. The sample volume in SLS and DLS is typically 1 mL and concentrations need to be optimized (higher concentrations for lower molecular weights). Solutions have to be filtered (no dust) and completely transparent (no turbidity) and should not absorb light of the wavelength used. Further, the refractive index of molecules in solution should differ from that of the solvent, the solution should be thermodynamically stable during the measurement (no reaction should occur on the time scale of the experiment), and the temperature needs to be controlled.

A complementary method that involves scattering of high-energy photons (X-rays with a typical wavelength of 0.1 to 0.2 nm) on all of the electrons of atoms in a molecule is **small angle X-ray scattering** (**SAXS**).[239,240] X-ray scattering is useful in that it can help to determine the molecular size and low-resolution structure (shape) of a macromolecule. SAXS experiments are typically conducted on synchrotron beamlines. A related approach is **small angle neutron scattering** (**SANS**)[241-243] that involves scattering of neutrons on the nuclei of atoms in a molecule. Since different isotopes of the same element display distinct scattering efficiencies, SANS experiments can be used in combination with contrast variation, *e.g.* to study a complex between a hydrogenated protein and a perdeuterated protein and using different H_2O : D_2O ratios.[244] As with neutron diffraction experiments (Section 15.4.3), SANS experiments take much longer than SAXS or SLS/DLS measurements as neutron beams are weaker than X-ray beams or visible lasers. Other approaches that are complementary to SLS are **osmometry** (determination of molecular weight and second virial coefficient), **mass spectrometry** (high precision molecular weight determination; Section 15.3), and **analytical ultracentrifugation** (determination of molecular weight, diffusion coefficient and second virial coefficient; Section 15.10).

15.11.1 Static Light Scattering

Light scattering occurs when polarizable particles are exposed to an oscillating electric field present in a light beam. The varying field induces oscillating dipoles in the particles, which radiate light in all directions. For globular proteins with molecular weights of ~500 kDa or less, light scattering is uniform in all directions. Therefore, the amount of light scattered by a solution is measured at some angle relative to the incident laser beam. However, for nucleic acid duplexes or other rod-like macromolecules, scattering varies significantly

with angle. Hence multi-angle laser light scattering must be employed where the intensity of the scattered light is measured at a number of different angles. This allows the absolute molecular weight as well as the geometric size to be determined.

The relationship between molecular weight M and I_{scat}, the intensity of the light scattered by a solution of molecules of concentration C and measured at the angle θ, is described by the following equation:

$$\frac{KC}{R(\theta)} = \frac{1}{MP(\theta)} + 2A_2C \qquad (15.12)$$

where $R(\theta)$ is the Rayleigh ratio (V_{scat} is the scattering volume from angle θ and r is the sample-detector distance)

$$R(\theta) = \frac{I_{scat}}{I_{laser}} \bullet \frac{r^2}{V_{scat}(\theta)} \qquad (15.13)$$

and the constant K can be calculated as

$$K = \frac{(2\pi n_0)^2 \left(\frac{dn}{dC}\right)^2}{N_A \lambda^4} \qquad (15.14)$$

where n_0 is the refractive index of the pure solvent, dn/dC is the refractive index increment of the solute/solvent system (= 0.185 mL g^{-1} for proteins in water and 0.145 mL g^{-1} for carbohydrates in water), λ is the wavelength of the laser, and N_A is Avogadro's number 6.022×10^{23} mol^{-1}. $P(\theta)$ is a correction factor, also called the **form factor** of the particles (P is 1 at $\theta = 0°$ and becomes smaller at higher scattering angles) and A_2 is the so-called **second virial coefficient**. This coefficient is a measure for the nonideality of the solution, *i.e.* a measure of the interaction forces that exist between molecules in solution. A_2 is positive when these forces are repulsive, A_2 is negative when the forces are attractive, and it is zero when forces are absent (ideal solution).

Finally, the particle size, *i.e.* the **radius of gyration R_g** is related to the form factor P, whereby P in the Guinier approximation is a function of the length of the scattering vector q ($P(q)$) rather than a function of the scattering angle θ ($P(\theta)$) above:

$$P(q) \approx e^{-\frac{1}{3} \bullet R_g{}^2 \bullet q^2} \qquad (15.15)$$

where

$$q = \frac{4\pi n}{\lambda} \bullet \sin\left(\frac{\theta}{2}\right) \qquad (15.16)$$

15.11.2. Dynamic Light Scattering

Dynamic light scattering is a technique in which the time dependence of the light scattering from a small, focused region of the solution is measured over a time scale ranging from tenths of a milli-second to milli-seconds. The time-dependent fluctuations in the intensity of scattered light are related to diffusion of molecules in and out of the region under study, which occur by Brownian motion. Therefore, the fluctuations are related to the solution viscosity η and the **diffusion coefficient** D of the macromolecule. The diffusion coefficient is directly related to the hydrodynamic radius R_h (**Stokes radius**) as shown in the equation:

$$R_h = \frac{k_B T}{6\pi\eta D} \qquad (15.17)$$

where k_B is the Boltzmann constant and T is the temperature.

The data can be analysed to determine either diffusion coefficients, or a distribution of diffusion coefficients if multiple species are present. In most cases, data is presented not in terms of the diffusion coefficient but rather in terms of particle size (Figure 15.19). The Stokes radius derived from this analysis gives the size of a spherical particle that would have a diffusion coefficient equal to that observed for protein or nucleic acid. However, most biological molecules, especially nucleic acids, are not spherical and their apparent Stokes radii are dependent on their shape, *i.e.* conformation, and mass. Their diffusion is also affected by the hydration state. Hence the hydrodynamic size determined from DLS can differ significantly from the true physical size observed in X-ray crystal or NMR structures. In solutions of large DNA molecules, one can observe the translational diffusion coefficients and also the rotational coefficients of the species. Also, because large DNA molecules exhibit a certain degree of flexibility, it is possible to determine the motions of small internal segments of these molecules.

DLS has found useful applications in macromolecular crystallization experiments to evaluate the crystallizability of proteins, nucleic acids and their complexes by examining sample **dispersity** (**monodisperse** *vs.* **polydisperse**) and the potential presence of aggregation.[246–250] Accordingly, polydisperse behaviour and formation of aggregates negatively affect the chance to obtain viable crystals, and monodisperse droplets are critical to successful crystallization in the majority of cases. Monodispersity is also of importance for other biophysical approaches such as NMR and SAXS/SANS.

Table 15.1 SLS *versus* DLS Experiments.

Input / Output	Needs to be known beforehand				Can be determined experimentally				
Parameter	C [g L^{-1}]	dn/dC	η	T	M	R_g	A_2	Shape	R_h
SLS	yes	yes	no	no	yes	yes	yes	yes	no
DLS	no	no	yes	yes	no	no	no	no	yes

Table 15.1 provides a comparison of the requirements for SLS and DLS experiments and the particular parameters that can be extracted. The use of 96-well plates for DLS, crystallization experiments and more recently even for X-ray diffraction data collection on synchrotron beamlines in combination with suitable robotics allow applications of these approaches in a high-throughput mode.

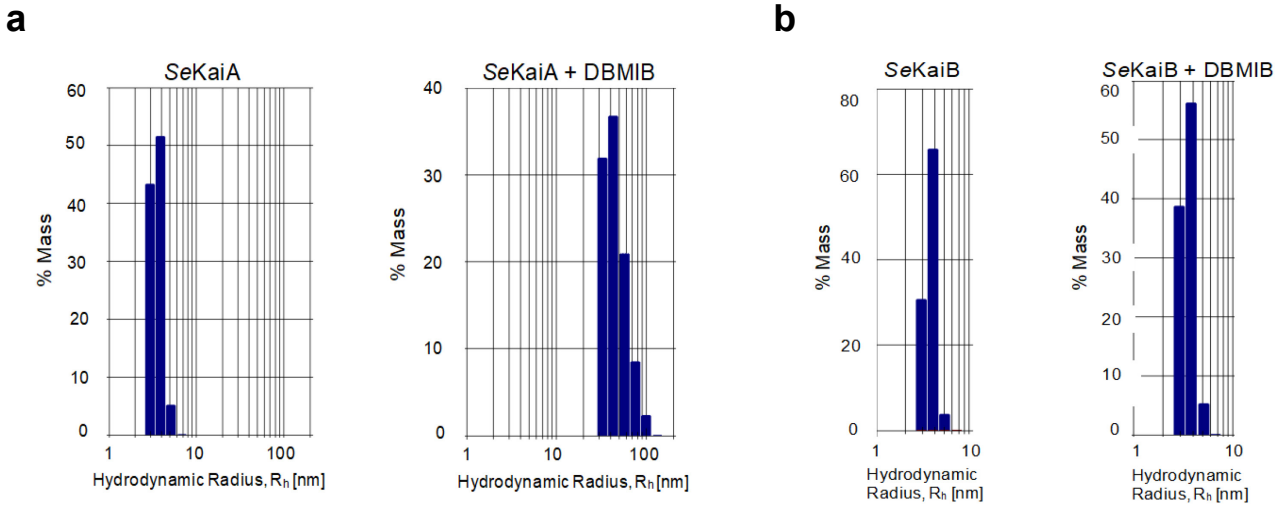

Figure 15.19 DLS-based hydrodynamic radii Rh for *Synechococcus elongatus* (Se) circadian clock proteins (a) KaiA alone (left) and in the presence of dibromothymoquinone (DBMIB, right), and (b) KaiB alone (left) and in the presence of DBMIB (right), indicative of DBMIB-induced aggregation in the case of KaiA and a lack of interaction between DBMIB and KaiB.[245] Selected DLS statistics: KaiA, D = 708 x 10^{-9} cm^2 s^{-1}, R_h = 3.50 nm, M = 63.3 kDa (dimer); KaiA + DBMIB, D = 38 x 10^{-9} cm^2 s^{-1} (avg.), R_h = 54.9 nm (avg.), M = 4.0 x 10^4 kDa (avg.); KaiB, D = 685 x 10^{-9} cm^2 s^{-1}, R_h = 3.54 nm, M = 64.9 kDa (tetramer); KaiB + DBMIB, D = 699 x 10^{-9} cm^2 s^{-1}, R_h = 3.51 nm, M = 63.6 kDa.

15.11.3 Small Angle X-ray Scattering

In a SAXS experiment, the scattering intensity $I(q)$ of a solution is measured as a function of resolution, *i.e.* the q range, $q = (4\pi\sin\theta)/\lambda$, where 2θ is the scattering angle and λ is the X-ray wavelength. The shape of the scattering curve should change only minimally as a function of the concentration (Figure 15.20a). Moreover, it is important to ascertain that the so-called Guinier analysis, a plot of $\log(I)$ *vs.* q^2 in the region closest to the zero scattering angle, displays linearity (Figure 15.20a, inset). A non-linear plot indicates aggregation or other concentration-dependent phenomena that preclude a meaningful scattering analysis.[251]

The so-called Kratky plot $q^2I(q)$ *vs.* q (not shown here) provides a useful concept to distinguish between macromolecules and their assemblies with a globular fold compared to unfolded or partially folded species. Thus, a globally folded protein will exhibit a bell-shaped curve, whereas lack of folding or partial unfolding will be manifested by a plateau or increasing values for $q^2I(q)$ in the upper q range. The pairwise distribution function $P(r)$ is the SAXS equivalent of the Patterson function in X-ray crystallography (Section 15.4). It can be directly computed as the Fourier transform of the scattering curve $I(q)$. $P(r)$ provides useful information about the distances between electrons (atoms) in the scattering sample. In theory $P(r)$ is zero at $r = 0$ and $r \geq D_{max}$, where D_{max} corresponds to the maximum linear dimension of the scattering particle (Figure 15.20a). Because scattering curve and 3D shape of the sample are related, one can compute a molecular envelope,[252] back-calculate the scattering curve based on it, and then compare calculated and experimental scattering curves, to arrive eventually at an improved shape model after several iterations. Such SAXS envelopes can be combined with X-ray or EM-based models of components to generate complete 3D structures of molecular assemblies (Figure 15.20b).[253-255]

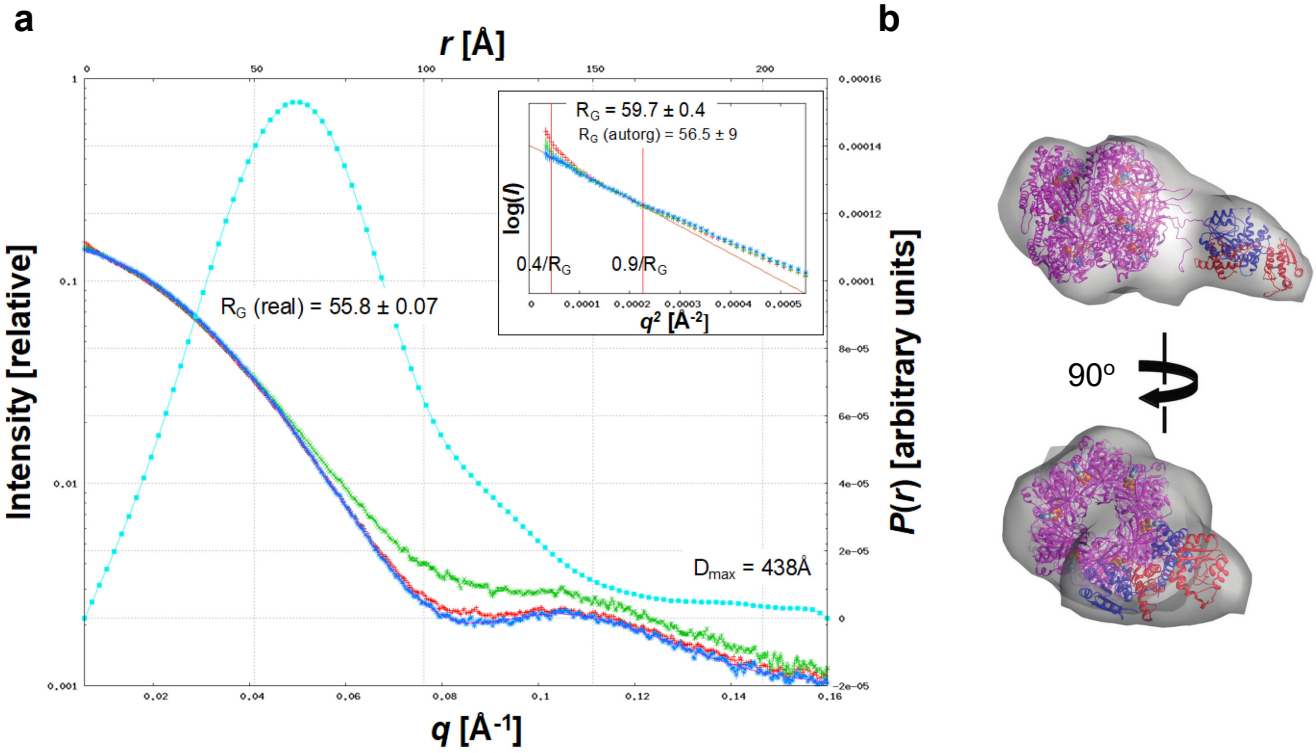

Figure 15.20 SAXS analysis of the dimension and shape of the KaiA-KaiC circadian protein complex. (a) Scattering curves (green, blue and red indicate low, medium and high concentration, respectively), Guinier plot (inset), and pairwise function (cyan). (b) The SAXS-based envelope (grey surface) shows a good fit for the model of the KaiA-KaiC complex (proteins coloured in red/blue and magenta, respectively) obtained using cryo-EM. Adapted from Ref. 254, https://doi.org/10.1371/journal.pone.0023697, under the terms of the CC BY 4.0 license, https://creativecommons.org/licenses/by/4.0/.

15.12 Thermodynamic Analysis of Nucleic Acids

Measurements of the thermal stability and thermodynamic parameters of nucleic acids alone and in complex with small-molecule ligands or proteins are commonplace and fundamentally important. Knowledge of stability provides insights into sequence-specific effects: the consequences of mutation and chemical modification, or damage and a better understanding of the interactions with drugs, and recognition and/or processing by a plethora of proteins, enzymes and receptors. This section provides an overview of fundamental approaches to extract stability and thermodynamic data from oligo- and polynucleotides.

15.12.1 UV Melting Assay

The simplest way to determine the stability of a nucleic acid molecule is by **UV melting**. This assay (introduced in Section 2.5.1 and 15.1.1) is based on the **hypochromicity**, the reduced absorption of UV light – typically measured at 260 nm – of double helices compared to random coils. Melting profiles for DNAs of different base composition are depicted (Figure 2.41) and demonstrate that the **melting temperature** (T_m) is increased for double helices with higher (per cent) G•C content.[256,257] The thermal melting assay is, of course, equally useful for RNA[258] and can be applied to unimolecular (hairpin-to-coil) and bimolecular (duplex-to-coil) transitions as well as to those involving higher molecularity, *e.g.* **triplexes**[259,260] and **quadruplexes**.[261,262] In order to conduct the assay, it is necessary first to determine the molar **extinction coefficient** (ε_{260}) of an oligonucleotide (Sections 2.3.1 and 15.1.1, Table 2.2). This can be done by using a colorimetric phosphate assay, whereby phosphate is enzymatically released from the oligonucleotide.[263] Alternatively, extinction coefficient and oligonucleotide concentration can be calculated by a nearest-neighbour approximation.[258] For a typical melting experiment with a single strand (*e.g.* a DNA or RNA hairpin molecule), a duplex either constituted by a self-complementary sequence or by a pair of oligonucleotides (for the latter a 1 : 1 mixture needs to be carefully established[263]), a 2 μM solution of the oligo in a buffer of defined concentration and pH (*e.g.* 10 mM sodium cacodylate, pH 7.4) and ionic strength (*e.g.* 100 mM NaCl) is then prepared. Using a UV-Visible spectrometer, absorbance *versus* temperature profiles are recorded at 260 nm, whereby the temperature is typically increased at 0.5 °C per minute.[264] The result is an absorbance *versus* temperature curve as depicted in Figures 2.41 and 15.1b. Next, this melting profile is converted into an α *versus* temperature curve (α represents the fraction of strands that remain hybridized) by fitting it to a two-state transition model with lower and upper base lines that are sloping linearly (Figure 15.21).[265,266] The T_m is the temperature at α = 0.5. The experiment needs to be repeated at least three times.

Thermodynamic parameters, ΔH, ΔS and ΔG (= $\Delta H - T\Delta S$), can be obtained from (i) the width at the half-height of a differentiated melting curve (ΔH), (ii) concentration-dependent melting experiments (ΔH, ΔS), or (iii) differential scanning calorimetry (DSC; ΔH, ΔS) experiments (Section 15.12.2.2).

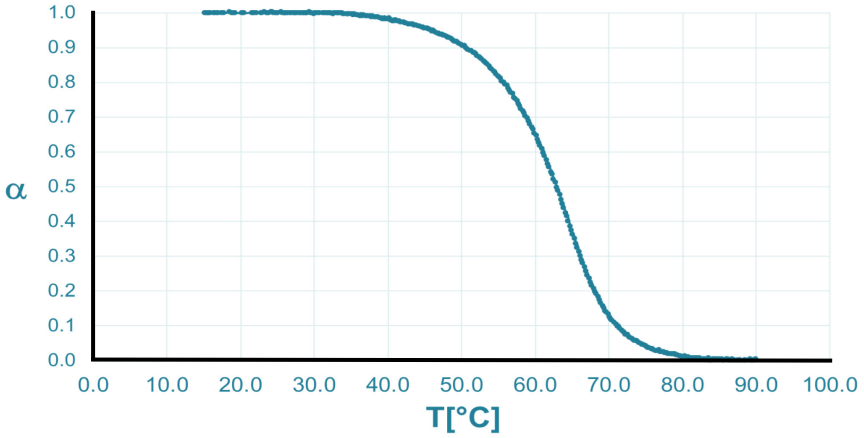

Figure 15.21 α *versus* temperature curve for a sarcin/ricin loop rRNA 27mer.

The first approach entails converting the above α *versus* T melting curve into a differentiated $\delta\alpha/\delta(T^{-1})$ *versus* T curve. The width of this curve at half-height is inversely proportional to the van't Hoff transition enthalpy, *i.e.* for a bimolecular transition $\Delta H = 10.14/ (T_1^{-1} - T_2^{-1})$, where T_1 and T_2 are the lower and upper temperatures, respectively, at half-height.[267]

Spectrometers equipped with a Peltier-thermostatted multi-cell holder readily enable the recording of concentration-dependent melting curves. A typical experiment may use five to six oligo concentrations in the range between 1 and 16 μM. For a bimolecular system (self-complementary sequence), the following equation applies: $1/T_m = (R/\Delta H)\ln c + \Delta S/\Delta H$, where R is the universal gas constant 1.986 cal mol^{-1} K^{-1} and c is the total strand concentration.[267] The plot of $1/T_m$ *versus* $\ln c$ is linear with $R/\Delta H$ as the slope.

For a monomolecular process such as formation or melting of a hairpin, the concentration-dependent approach cannot be used. Instead, we rely on the following general expression for calculating the transition enthalpy: $\Delta H_{vH} = (2 + 2n)RT_m^2(\delta\alpha/\delta T)_{T=T_m}$.[265] For the monomolecular process $(2 + 2n) = 4$, T_m is T at $\alpha = 0.5$, and $\delta\alpha/\delta T$ is the slope of the α *versus* T curve at T_m. The Gibbs free energy under standard conditions is obtained as follows: $\Delta G^\circ = \Delta H_{vH}(1 - T/T_m)$ at $T = 298.15$ °C.[265]

The spectroscopic approach is also suitable for the determination of binding thermodynamics of ligands such as groove-binders or intercalators to oligonucleotides or nucleic acids of higher molecular weight.[268] A further application of the UV melting assay concerns so-called osmotic stressing or steric crowding experiments.[269,270] Their use, in conjunction with a thermal melt of co-solutes of an oligonucleotide (small molecules such as acetamide, ethylene glycol or glycerol) that congregate around a duplex without binding firmly, permits the determination of the changes in the number of water molecules Δn_w associated with the transition to a random coil.[264,271] Accordingly, the number of waters released can be calculated as follows: $\Delta n_w = (-\Delta H/R)[d(T_m^{-1})/d(\ln a_w)]$, where ΔH is the enthalpy change obtained from either a UV melting or a DSC experiment and a_w is the experimentally determined value of water activity at given concentration of cosolutes.[271]

In summary, the optical melting assay remains a valuable and highly versatile technique to measure Tm values and thermodynamic parameters of oligonucleotide-size DNA and RNA fragments as well as longer nucleic acids. The approach has several advantages over the more accurate calorimetric assays.[272] They include the need for only small amounts of material, they are fast (the experiments can be completed in a few hours) and the equipment is relatively inexpensive and easy to maintain.

15.12.2 Calorimetry

The major calorimetric techniques for characterizing the thermodynamics of nucleic acids are **isothermal titration calorimetry** (**ITC**) and **differential scanning calorimetry** (**DSC**).[273] ITC is used to measure the heat, ΔQ, released from hybridizing two nucleic acid molecules at constant temperature or from forming a complex between a nucleic acid and either a small molecule ligand or a protein.[274,275] ITC cannot be used for monomolecular reactions (*e.g.* hairpin-to-coil transitions) or self-complementary oligonucleotides. Moreover, it should not come as a surprise that thermodynamic data derived from a thermal melting assay can deviate from the data generated by ITC, as a nucleic acid molecule may adopt residual structures at the lower (constant) temperature used for ITC that may not exist at temperatures near its T_m.[276] DSC is used to measure the excess heat capacity, C_p^{ex}, of a nucleic acid or protein as a function of temperature in a buffered solution relative to a reference consisting of the same volume of buffer solution alone.[277,278] It is noteworthy that the thermodynamic data gained from a DSC experiment are independent of a particular model of unfolding, *i.e.* two-state *versus* multi-state. Thus, a DSC experiment can reveal intermediate states in the melting process that may pave the way to a more detailed model of individual species participating in the transition along with their thermodynamic properties.

15.12.2.1 Isothermal Titration Calorimetry

An ITC experiment to determine the thermodynamics of the pairing of two oligonucleotides or the binding of a ligand to an oligonucleotide has one strand, referred to as analyte nucleic acid (A), present in the sample cell as a buffered solution of concentration 0.1 to 1.0 mM. Small aliquots (μL) of a solution of the other strand, referred to as the titrant nucleic acid (T), at a concentration 10 times that of A are then injected repeatedly into the sample cell (the concentration of titrant can be even higher for a ligand according to the expected stability

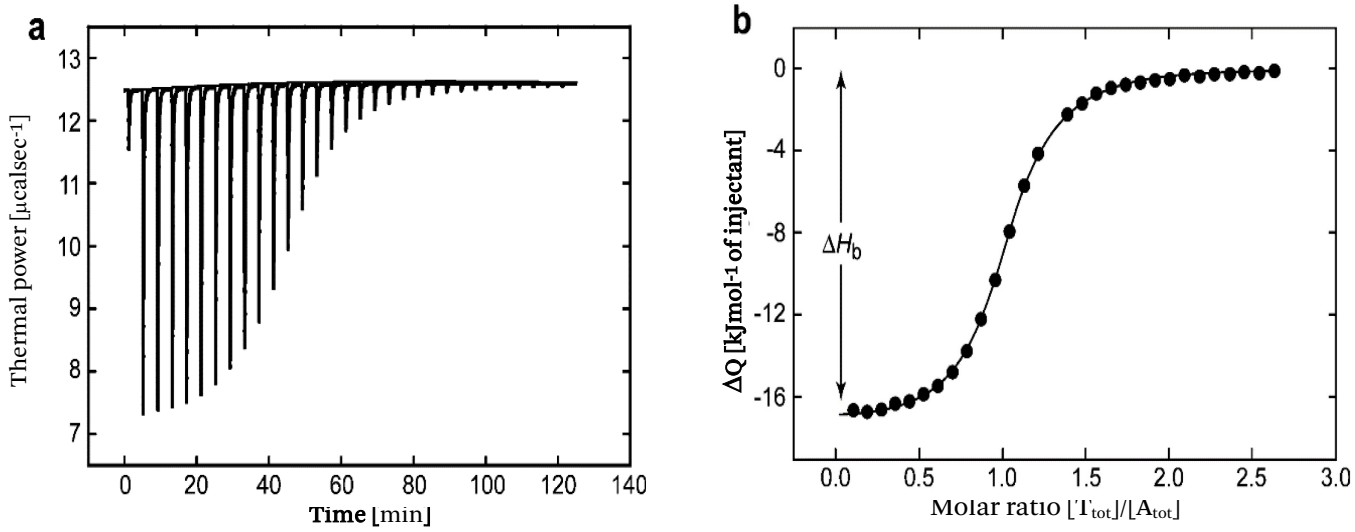

Figure 15.22 ITC data for a ligand binding to DNA. (a) Individual heat bursts resulting from an initial 3 μL pre-injection that is followed after 5 min by the first experimental injection of 15 μL and then by an additional 29 injections with an equilibration time of 4 min in between each. (b) Integration of the peaks in panel a with respect to time and correction to a per mole basis produces the binding isotherm.

of the complex). Each injection generates a heat burst with a thermal power expressed in μcal sec^{-1} (Figure 15.22a). The heat ΔQ, expressed in kcal mol^{-1} or kJ mol^{-1} of injected T, that results from each injection can be obtained by integration over the individual heat burst curves and eventually plotted as a function of the molar $[T_{tot}]/[A_{tot}]$ ratio (Figure 15.22b). Non-linear least-squares fitting of this titration curve yields the stoichiometry n of the reaction (*i.e.* 1 for a duplex or a 1:1 ligand–oligo complex), the enthalpy of the formation of the complex ΔH, and the equilibrium association constant K.[273-275] Once K has been determined, the free energy of binding ΔG can be derived.

From the equation $\Delta G = -RT\ln K$, with ΔG and ΔH known, the entropy of binding can be calculated using $\Delta S = (\Delta H - \Delta G)/T$. A further parameter that can be derived from ITC experiments conducted at different temperatures is the heat capacity change ΔC_p of the reaction: $\Delta C_p = \delta\Delta H/\delta T = (\Delta H_{T2} - \Delta H_{T1})/(T_2 - T_1)$, where T_1 and T_2 are the temperatures for the two ITC experiments.

One limitation of the ITC experiment concerns tight binding constants K that exceed 10^9 M^{-1}. In such cases the sharpness of the titration curve precludes accurately fitting for K and it is better to use DSC for determining the thermodynamics of the binding reaction. Alternative approaches that can be applied with Ks that lie

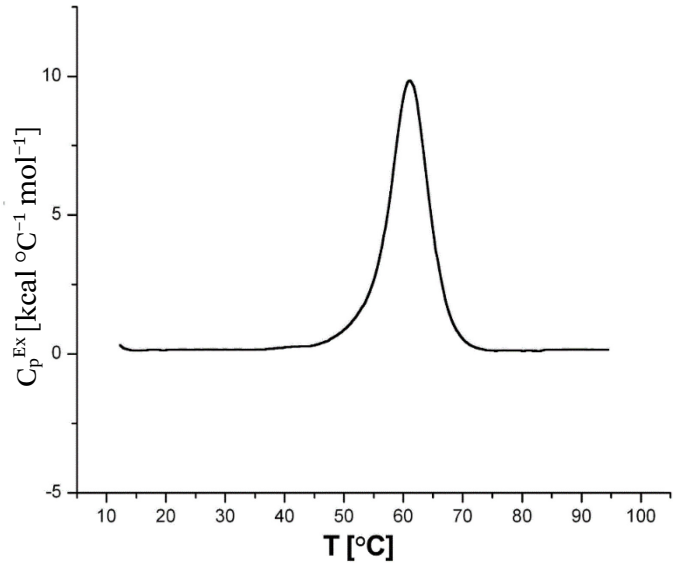

Figure 15.23 DSC profile for the thermal denaturation of bovine RNase A.

beyond 10^9 M^{-1} are nitrocellulose filter binding, gel electrophoresis, fluorescence spectroscopy, and surface plasmon resonance.[279] However, neither these methods nor biolayer interferometry, another approach suitable for characterizing very tight binding interactions,[230] provide thermodynamic data. Because of the relatively high concentrations needed to derive reliable data from ITC, solubility issues may arise. Aggregation, a problem not uncommon with binding experiments involving protein and nucleic acid, can ruin an ITC experiment. Moreover, compared to the low amounts of material needed for thermal melts, running an ITC experiment repeatedly will consume significant quantities of nucleic acid and protein.

15.12.2.2. Differential Scanning Calorimetry

DSC instruments contain two cells with a volume of 0.5 to 1.5 mL, whereby the reference cell is filled with buffer, and the sample cell is filled with buffer containing the macromolecule. The choice of buffer for a DSC experiment with a nucleic acid is very important. Thus, buffers whose pK_a values exhibit large temperature dependencies, *e.g.,* Tris/HCl, should be avoided. Suitable DSC buffers include phosphate, citrate and acetetate.[273] DSC involves the continuous measurement of the apparent specific heat of a system as a function of temperature and can be used to examine physicochemical processes triggered by increases or decreases in temperature, such as phase transitions or conformational changes. Integration of the experimental curve of the excess heat capacity C_p^{Ex} *versus* T (Figure 15.23) yields the transition enthalpy $\Delta H = \int C_p dT$. The temperature at which excess heat capacity is at a maximum defines the transition temperature (T_m). Another parameter that can be directly derived from this curve is the heat capacity change ΔCp which is the difference between the pre- and post-translation baselines.[281] In the example shown (Figure 15.23), $\Delta C_p \approx 0$. The curve C_p^{Ex} *versus* T can be converted into C_p^{Ex}/T *versus* T and the newly obtained curve integrated over T to yield the transition entropy: $\Delta S = \int (C_p/T) dT$.[265] Therefore, from a single DSC experiment, one obtains ΔH, ΔS and ΔC_p leading to ΔG.

DSC data can also provide important information concerning the cooperativity of a thermal transition. This can be achieved by comparison of the magnitudes of the model-dependent van't Hoff enthalpy (obtained by shape analysis of the calorimetric data)[267,282] and the calorimetric enthalpy. If $\Delta H_{vH} = \Delta H_{cal}$ and there are no chemical or instrumental kinetic limitations, such that T_m is independent of both scan rate and concentration, then the transition proceeds in a two-state manner and meaningful thermodynamic data can be obtained by examination of the temperature dependence of an equilibrium property. If $\Delta H_{vH} < \Delta H_{cal}$, then the transition must involve a significant number of intermediate states. Conversely, if $\Delta H_{vH} > \Delta H_{cal}$, then aggregation is implicated. The $\Delta H_{vH}/\Delta H_{cal}$ ratio provides quantitative insight into the nature of the transition. Specifically, it provides a measure of the fraction of the structure that melts as a single thermodynamic entity (*i.e.*, it defines the size of the cooperative unit). This is a unique advantage of DSC in the study of biological molecules and when combined with ITC, thermodynamic studies over a wide temperature range and under a variety of solution conditions can be carried out. These techniques offer a powerful tool to probe the energetics of macromolecule stability and ligand association reactions.

15.13 Molecular Mechanics and Dynamics

Several experimental methods are available for elucidating the structure and dynamics of nucleic acids, including their interactions with other molecules. Computational **molecular modelling** offers a complementary approach to simulate molecular behaviour. The methods it employs – accurate "**force fields**" and appropriate sampling of energetically accessible structures – can fully elucidate the **conformational ensemble** of DNA helices, reproducing the best NMR structures to an accuracy of better than 0.5 Å for internal base-pairs. As simulation methods and force fields have improved, researchers have demonstrated the ability to fold RNA tetraloops *de novo*, sample rare and transient Hoogsteen base-pairs in DNA helices, and spontaneously observe drug and ligand binding to DNA and RNA including observing intercalation and ligand-driven conformational changes (Chapter 12).

15.13.1 Molecular Mechanics and Nucleic Acid Force Fields

Modelling in nucleic acid studies has a proud history dating from Jim Watson's historic base-pair modelling in 1953 (Chapter 1). Most modelling is now performed using computational methods. The most accurate

predictions of molecular structure ought, in theory, to come from **quantum mechanical** (**QM**) calculations, which treat the electronic structure of molecules in detail from first principles (Section 15.14). However, QM methods cannot easily model systems larger than 500 atoms, do not yet represent the solvent environment satisfactorily, and are computationally very expensive for handling dynamics on a time scale necessary for modelling processes such as RNA folding.[283,284] A simpler yet surprisingly accurate approach is **molecular mechanics** (**MM**), while we note that QM methods are routinely applied to optimize the MM collection of parameters or force field.[285]

With MM, individual atoms are modeled as spheres with both intramolecular (bond, angle, dihedral) interactions for covalent connectivity and intermolecular interactions among all atoms (to model the electrostatics, atomic repulsion, and van der Waals attraction). Bonds and angles are modeled as springs, so there are no electrons in the MM model which means that it cannot model the chemistry of bond breaking and forming or of electron transfer. The dihedrals, also referred to as the torsions, model the rotation around the central bond of three connected bonds with a periodic function. The MM approach for intermolecular interactions uses a pairwise potential that includes a coulombic term (r^1) to model electrostatics and a Lennard Jones term ($-r^6$ and r^{-12} for attraction and repulsion, respectively; r is the interatomic distance). Most MM implementations are based on fixed point charges on each atom, which are not formal charges (such as the overall −1 charge on an oxyanion) but are often non-integral and related to the electronegativity and environment of the atom. Charges can be derived using QM methods on residues or molecular building blocks with van der Waals parameters derived either using QM methods or by empirical fitting liquid densities of smaller components (Section 15.14). The spring force constant, dihedral parameters, charges and van der Waals parameters define a force field which is used to compute the energy of the system as a function of its atomic coordinates: lower energy implying a "better" or more favourable structure. The MM energies between different molecules are only comparable and for use between equivalent molecular systems. A key assumption of MM is **transferability**, such that a C–C single bond in ethane is the same as those in cyclohexane and defines a transferable C–C single bond parameter. Atom charges are not transferable and are typically fixed, although polarizable force fields have emerged for nucleic acids that ease away from the fixed charge model to allow dynamic response to changes in the surroundings that result from conformational or environmental change.

15.13.2 Conformational Ensemble and Energy Minimization

The MM force field provides an energy that is dependent on the molecular conformation and its environment of surrounding solvent, ions, and other molecules. This implies a highly multidimensional **energy surface** or landscape that depends on the coordinates of all the atoms. To illustrate this, we use a simple 2D representation of an energy surface (Figure 15.24) where the *y*-axis is energy and the *x*-axis is an arbitrary **reaction coordinate** that characterizes the conformational space of the molecule (*e.g.* the number of Watson–Crick base-pairs, **radius of gyration**, native contacts, *etc.*). Lower energies are more favourable, so the wells in a free-energy surface represent more favourable structures. The simplest application of MM is the evaluation of the energy of a single conformation or structure. However, since it is not possible to compare energies between different molecules, and since only a single energy with an arbitrary zero is calculated, it cannot be used to evaluate the structure. To compare energies, one needs to find other conformations of the same molecule.

A simple approach to evaluate a structure is to minimize the MM force field. The lowest energy structure of a given conformation or structure corresponds to a situation in which there is no net force acting on any atom, *i.e.* to a point on this surface with zero slope (at the bottom of a well). Starting from an informed guess for this structure, the process of energy minimization involves calculation of the forces and moving 'downhill' on this surface until a minimum is reached. In practice, for biomolecular structures, the energy surfaces have multiple minima (Figure 15.24). The most favoured conformation is expected to adopt the **global minimum** energy conformation (at low temperature), but energy minimization usually leads to a **local minimum**, of which there are a vast set. Thus minimization rarely finds the global minimum structure for the molecule. The complexity of real energy surfaces makes the process of searching for the true global energy minimum conformation very difficult and time-consuming, it calls for examination of the entire **conformational ensemble** or energy surface. However, convergence of sampling that ensemble is elusive because of the difficulty in fully exploring rare states with high energy barriers.

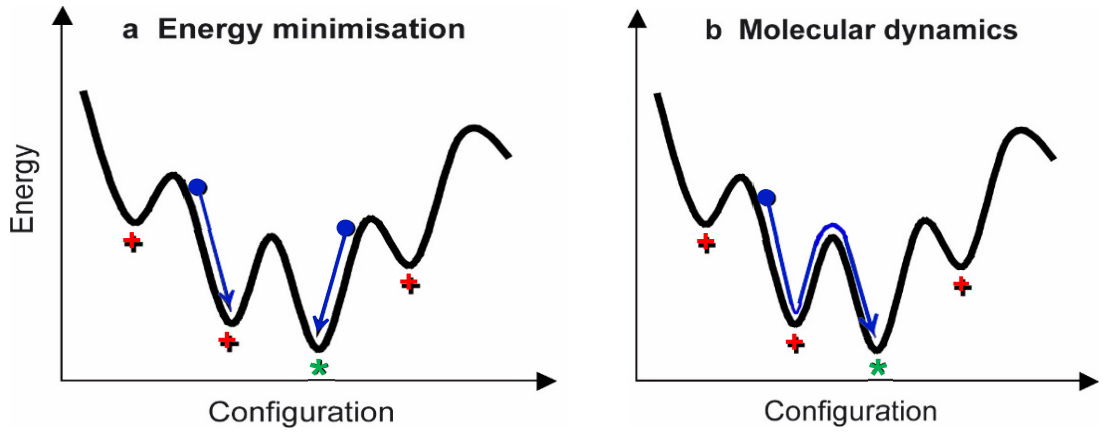

Figure 15.24 Each configuration of a molecule has an associated energy. (a) In energy minimization, different initial guesstimates (•) for the true structure may be optimized to local (+) rather than the true (*) energy minimum. (b) Using an MD simulation the lowest energy minimum may be more reliably identified.

15.13.3 Molecular Dynamics Can Elucidate the Full Conformational Ensemble

Molecular dynamics (**MD**) simulation is one of the most widely used methods to resolve the problem of elucidating the conformational ensemble of a structure. Masses and velocities, representative of a particular temperature, are assigned to each atom in some starting conformation. Using Newton's equations of motion, the conformation of the system at some later time-point is predicted. The forces acting on each atom, and thus the velocities, are then recalculated at this new conformation, the motion over the next time step is determined and so on. This process allows 'uphill' motion over the energy surface, and barriers between minima may be overcome depending on their height relative to the simulation temperature. To avoid larger initial forces at the start of MD, energy minimization is performed prior to MD.

The two main limitations of the MD simulation approach are the timescale that is accessible in the simulation and accuracy of the MM force field. For the calculations to be stable, the time-step between which the forces acting on the system are recalculated must be very small – just a few femtoseconds since bond stretching is very fast. Each calculation is computationally expensive and typically limits such simulations to hundreds of milliseconds at best (in 2020) for systems of up to 25,000 atoms. That is about the size of a 12-mer to 18-mer helix or an RNA stem-loop structure in solution. The set of available nucleic acid force fields is constantly evolving as improvements are made, with common ones associated with large MD simulation code packages including AMBER,[286] CHARMM,[287] GROMACS,[288] among others. Recently optimized nucleic acid force fields perform really well for nucleic acid helices.[289,290] More complex structures, such as non-helical regions of DNA and RNA, still need refinement to more accurately model these and avoid population of anomalous conformations.[291]

The basic output from an MD simulation is a trajectory file, *i.e.* the predicted behaviour of a single molecule as a function of time. However, individual MD 'snapshots' can alternatively be regarded as a Boltzmann-weighted ensemble of structures from which thermodynamic quantities may be calculated. Modelling methods based on this approach are particularly valuable in that they provide a link between the microscopic behavior of individual molecules – easy to simulate but difficult to observe experimentally – and the system's macroscopic properties. These are much easier to measure but can be difficult to interpret in terms of atomic-scale features. As noted, the methods have improved to a level where we can use MD simulation to observe many important nucleic acid related biological processes including **drug binding** and **intercalation** (Figure 15.25), folding and **base-pair opening**, and we can almost fully elucidate the conformational ensemble of the internal part of a nucleic acid helix.[289,290]

To probe larger size molecules and timescales **mesoscopic modelling** methods must be applied in which, for example, a length of DNA is treated as a uniform elastic rod, or as a set of disks (each representing a base or base-pair) connected to its neighbours by springs.[293] These approaches have been particularly used to study DNA supercoiling (Section 2.3.5) and packaging (Section 2.6), but are also being improved upon by new coarse-grained approaches for the simulation of duplex DNA.[294]

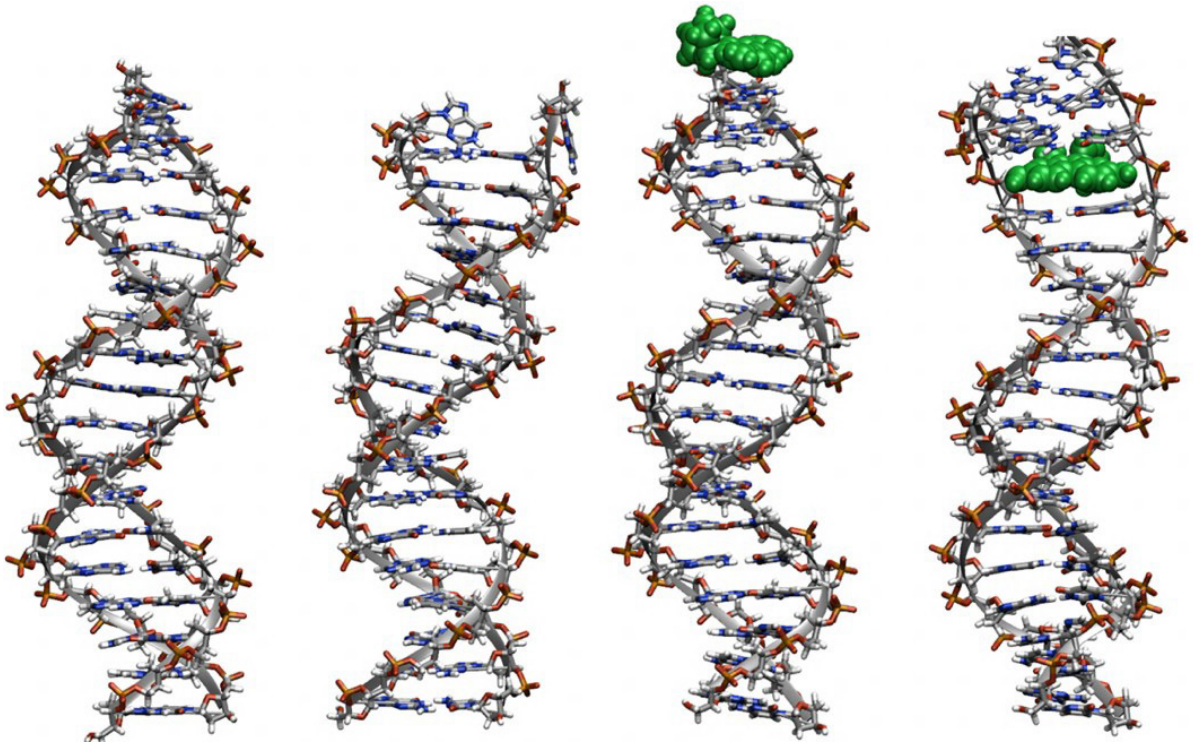

Figure 15.25 Molecular graphics representations for four snapshots from a 20 μsec MD trajectory of a DNA duplex with sequence d(GCATGAACGAACGAACGC) and ethidium bromide to highlight molecular configureations sampled. Waters and ions used in the simulation are not shown. The leftmost structure is the native duplex, next right is the duplex with terminal base-pair fraying, next ethidium bromide (green) end-stacks on the duplex, and the right-most structure shows a spontaneously intercalated ethidium bromide.[292]

15.14 QM/MM Methods for Modelling Nucleic Acids Reactions

Notwithstanding the undoubted success of force field-based MD simulations for a wide variety of applications,[295] the difficulty of developing empirical parameters for chemical bond breaking and formation precludes their use in modelling reactions.[296] By contrast, **quantum mechanical** (**QM**) calculations provide energetic and structural insights along the reaction coordinate but are limited to molecular systems comprised of, at best, a few hundred atoms. However, as first demonstrated by Warshel and Levitt,[297] combining QM calculations with **force field**-based (MM) energy descriptions permits the study of reactions that take place in condensed phases. The aim of the QM/MM methodology is to retain a quantum-mechanical description only for those atoms that require QM, while describing their wider environment with a molecular-mechanics force field. The development of high-quality force fields for modelling proteins and nucleic acids (Section 15.13) has led to **QM/MM** methods being routinely used to understand the chemical reactivity of nucleic acids in an aqueous environment, and the catalytic mechanisms employed by ribozymes and enzymes involved in DNA and RNA metabolism (Section 3.2). Here the use of QM/MM calculations is illustrated by recent work on the chemistry of base-excision repair (BER) (Section 11.1.1).[298]

15.14.1 Overview of the QM/MM Method

Students of chemistry are familiar with the Schrödinger equation and "first principles" methods for its solution. While a variety of QM methods have been developed, most calculations in organic and biological chemistry employ Kohn-Sham **density functional theory** (**DFT**).[299] In this model, which is complementary to approaches based on many-electron wavefunctions, the ground state electron density distribution, $\mathbf{n}(\mathbf{r})$, is calculated. Knowing $\mathbf{n}(\mathbf{r})$ then yields the energy of the system together with all other molecular properties of the system, including the energy and molecular geometry of the optimised structure, vibrational force constants and frequencies, electrical polarisability and dipole moment.[300] Moreover, DFT calculations yield

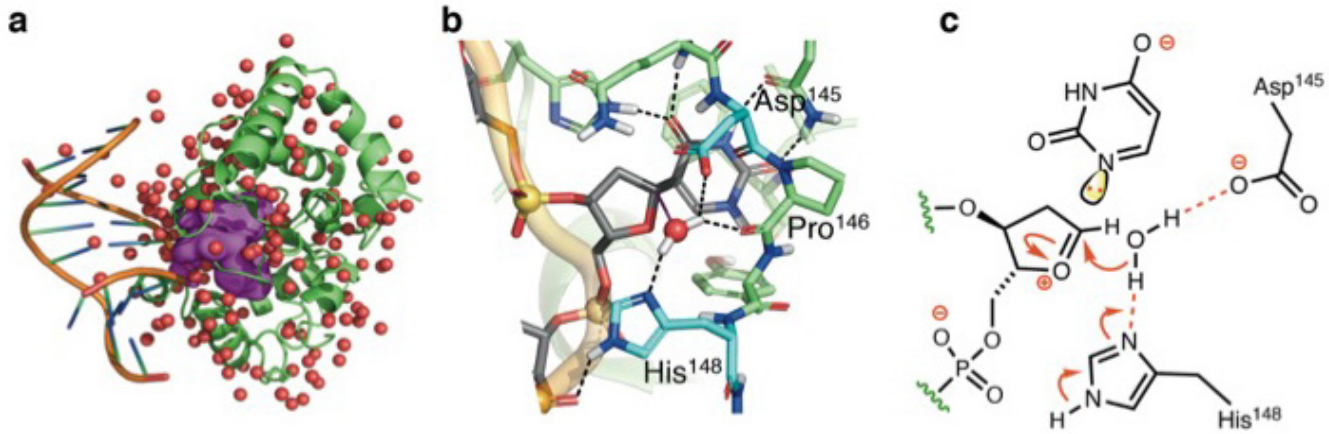

Figure 15.26 (a) Cartoon representation of UDG bound to uracil showing atoms in the QM region (magenta spheres). All other atoms of DNA and the protein (both shown in cartoon representations) and water (red spheres) are defined to be in the MM region of the system. Similar decompositions are used in all QM/MM calculations on enzyme catalysed reactions. (b) Close-up of residues in the active site of UDG showing the location of the water molecule (red sphere) that is held in position by Asp145 and His148 (highlighted in cyan). The water is located close to C-1' of the pseudo-2'-deoxyuridine nucleoside (grey carbons) in the crystal structure (PDB structure 1EMH).[303] (c) Activation of water by deprotonation by His148 in the second step of the dissociative SN1 reaction mechanism.

good estimates of structures and energies as reactant(s) transform into product(s) along a defined **reaction coordinate**. Although modern computational resources mean that DFT calculations can be performed for systems composed of several hundred atoms, introducing the effects of protein environment and water molecules generally requires the inclusion of thousands of atoms. Obtaining free energy values also requires extensive sampling of accessible conformational states in the protein and the arrangement of waters about the protein surface. Finally, applying DFT methods to systems containing DNA and RNA is technically challenging because of the highly charged nature of these biopolymers in water. These problems can be overcome by including the effects of the protein and solvent environment around the reactant(s) by constructing a QM/MM description of the system.[301] Thus, atoms in the system of interest assigned into two regions. The QM region is composed of atoms undergoing reaction together with catalytically important protein residues or water molecules and is treated using DFT methods. The remaining atoms in the system, including solvent waters and appropriate counter-ions, are assigned to the MM region and represented by force field-based potential energy functions (Figure 15.26a).[302]

In performing QM/MM calculations on enzymes, one must include catalytically important networks of hydrogen bonds and non-covalent electrostatic interactions within the QM region. Unfortunately, the number of atoms in this region is limited by the computing resources needed to converge the wavefunction. It is therefore important to examine how the results of QM/MM calculations depend on the number of atoms in the QM region to obtain reliable conclusions. On a practical note, we also recommend that X-ray crystal structures of better than 2.5 Å resolution are used when building the initial structure so as to obtain accurate positions for non-hydrogen atoms.

Atoms in the QM and MM regions of the system are "coupled" using a variety of approaches to obtain the energy of the complete molecular system.[304] For example, the electrostatic properties of the MM region are included within the QM **Hamiltonian**, and atoms in the QM region are assigned van der Waals parameters. As a result, atoms in the QM and MM regions are prevented from "collapsing" into each other during energy minimisation and **molecular dynamics** simulations. Entropic contributions to reaction energetics can also be included in QM/MM calculations by sampling different conformations of the MM region.[305] As a result, calculated activation free energies[306] or kinetic isotope effects[307] can be compared directly to experimental measurements.

15.14.2 QM/MM Studies of DNA Repair

Work on the enzyme uracil DNA glycosylase (UDG) provides a good example of how QM/MM calculations provide insight into reaction mechanism and the functional roles of active site residues. There is general

agreement substrate distortion facilitates cleavage of the C-N bond to give a uracilate anion (p*K*a 9.6) and an oxocarbenium ion intermediate (Section 11.11.3).[308,309] The cationic intermediate is then captured by water to yield an abasic site in the DNA strand, which is subsequently repaired by APE1 and Pol β.[310] QM/MM calculations were used to determine the function of His148, which could not be assigned unambiguously by experiment. Thus, QM/MM calculations were needed to determine whether His148 or Asp415 might promote nucleophilic attack by water (Figure 15.26b).[311]

After determining that the lactam tautomer of uracil is energetically preferred in the UDG active site, the QM region was chosen to comprise atoms in residues Gln144, Asp145, Pro146, Tyr147, His148, Phe158, Asn204 and His268, three active site waters (Wat600, Wat655, Wat727) and the dU nucleotide (Figure 15.26a).[311] The resulting QM region consists of 170 atoms, meaning that DFT calculations were the best choice for computing QM/MM free energies along the reaction coordinate. The rest of the system (DNA, protein, counter-ions and water) was described by the DL-POLY force field.[312] Proton transfer to His148 and Asp145 proceeds with similar QM/MM activation free energies (45.1 kJ mol^{-1}), which is consistent with the experimental k_{cat} value. We note that this is important constraint must be satisfied in any QM/MM calculation. The calculated reaction energy for protonating His148, however, is more favourable than that for protonating Asp145. These results show that His148 is the likely general base although the similarity of the QM/MM free energy barriers means that this role can be performed by Asp145 in the absence of His148 (Figure 15.26c). This finding explains why replacing His148 by other amino acids does not suppress the catalytic activity of these UDG variants.[313] These QM/MM calculations also show that both residues are important in **stereoelectronic** control to orientate the water molecule so that the lone pair can overlap with the vacant π^*-orbital of the oxocarbenium cation.

As is evident from a burgeoning literature, the conceptual basis of QM/MM calculations is well understood and many technical aspects of this computational method are unlikely to change in the future. Access to computers with enhanced processing capabilities and improved software implementations will, however, permit QM/MM calculations for very large molecular systems, such as the ribosome (Section 5.4.2) and nucleosome (Section 13.6.1). Additional targets will also become accessible given the power of cryo-EM to obtain high-resolution models of large protein/nucleic acid complexes.[314] Better methods for describing the QM region are also on the horizon, particularly for modeling metalloenzymes and molecular radicals, which should improve the accuracy of calculated energies and electronic properties.[315]

References

1. V.A. Bloomfield, D. Crothers and I. Tinoco, Physical Chemistry of the Nucleic Acids, Harper and Row, New York, 1974.
2. I. Tinoco, Jr., J. Am. Chem. Soc., 1960, 82, 4785.
3. M. D'Abramo, C.L. Castellazzi, et al. J. Phys. Chem., B 2013, 117, 8697.
4. A.O. Nwokeoji, P.M. Kilby, et al. Anal. Chem. 2017, 89, 13567.
5. J.R. Lakowicz, Principles of Fluorescence Spectroscopy, 3rd edition, Springer, New York, 2006.
6. L. Brand and M.L. Johnson, Meth. Enzymol., 1997, 278, 570.
7. D. Onidas, D. Markovitsi, et al., J. Phys. Chem., B 2002, 106, 11367.
8. R.W. Sinkeldam, N.J. Greco and Y. Tor, Chem. Rev., 2010, 110, 2579.
9. W. Xu, K.M. Chan and E.T. Kool, Nat. Chem., 2017, 9, 1043.
10. N. Klöcker, F.P. Weissenboeck, and A. Rentmeister, Chem. Soc. Rev., 2020, 49, 8749.
11. L. Stryer, Annu. Rev. Biochem., 1978, 47, 819.
12. B. Hellenkamp, S. Schmid, et al., Nat. Methods, 2018, 15, 669.
13. G. Pljevaljcic, F. Schmidt and E. Weinhold, ChemBioChem, 2004, 5, 265.
14. R.K. Neely, P. Dedecker, et al., Chem. Sci., 2010, 1, 453.
15. V. Goyvaerts, S. Van Snick, et al., Chem. Commun., 2020, 56, 3317.
16. D.M.J. Lilley, in Nucleic Acids and Molecular Biology, D.M.J. Lilley and F. Eckstein, eds., Springer-Verlag, Berlin, 1990, 4, 55.
17. J.C. Penedo, T.J. Wilson, et al., RNA, 2004, 10, 880.
18. T. Sabir, G.F. Schröder, et al., J. Am. Chem. Soc., 2011, 133, 1188.
19. A.K. Woźniak, G.F. Schröder, et al., Proc. Natl. Acad. Sci. USA, 2008, 105, 18337.
20. S. Kalinin, T. Peulen, et al., Nat. Methods, 2012, 9, 1218.
21. M. Eriksson and B. Nordén, Meth. Enzymol., 2001, 340, 68.
22. B. Nordén and T. Kurucsev, J. Mol. Recogn., 1994, 7, 141.
23. J. Kypr, I. Kejnovska, et al., in Circular Dichroism Spectroscopy of Nucleic Acids, 2012, Ch. 17, 10.1002/9781118120392.ch17.
24. J. Kypr, I. Kejnovska, et al., Nucleic Acids Res., 2009, 37, 1713.
25. T. Šmidlehner, I. Piantanida and G. Pescitelli, Beilstein J. Org. Chem., 2018, 14, 84.
26. B. Norden, M. Kubista and T. Kurucsev, Q. Rev. Biophysics, 1992, 25, 51B.
27. B. Nordén, A. Rodger and T. Dafforn, Linear Dichroism and Circular Dichroism: A Textbook on Polarized-Light Spectroscopy, RSC Publishing, 2010.
28. E. Taillandier and J. Liquier, Methods Enzymol., 1992, 211, 307.
29. G.J. Thomas and A.H.-J. Wang, in Nucleic Acids and Molecular Biology, D.M.J. Lilley and F.Eckstein, eds., Springer-Verlag, Berlin, 1988, 2, 1.
30. A.W. Parker and S.J. Quinn, in Encyclopedia of Biophysics, G.C.K Roberts, ed., Springer, Berlin, Heidelberg, 2013. https://doi.org/10.1007/978-3-642-16712-6_112
31. G.J. Thomas, Jr., Annu. Rev. Biophys. Biomol. Struct., 1999, 28, 1. 38
32. D. Marion, Mol. Cell Proteomics, 2013, 12, 3006.
33. K. Wüthrich, NMR of Proteins and Nucleic Acids, J. Wiley & Sons, New York, 1986.
34. L.R. Ganser, M.L. Kelly, et al., Nat. Rev. Mol. Cell. Biol., 2019, 20, 474.
35. H. Gunther, NMR Spectroscopy - Basic Principles, Concepts, and Applications in Chemistry, 3rd Ed., John
36. Wiley & Sons Ltd, Chichester, 2013.
37. R. Schnieders, S. Keyhani, et al., Chemistry, 2020, 26, 102.
38. Y. Jin, N.G. Richards, et al., Angew. Chem. Int. Ed., 2017, 56, 4110.
39. A. Dallmann and M. Sattler, Curr. Protoc. Nucleic Acid Chem., 2014, 59, 7.22.1.
40. M.W. Szulik, M. Voehler and M.P. Stone, Curr. Protoc. Nucleic Acid Chem., 2014, 59, 7.20.1.
41. M.W. Szulik, P.S. Pallan, et al., Biochemistry, 2015, 54, 1294.
42. W.Copp, A.Y. Denisov, et al., Nucleic Acids Res., 2017, 45, 10321.
43. K. Gehring, J.L. Leroy and M. Gueron, Nature, 1993, 363, 561.
44. A.M. Spring-Connell, M. Evich and M.W. Germann, Curr. Protoc. Nucleic Acid Chem., 2018, 72, 7.28.1.
45. J. Marchant, A. Bax and M.F. Summers, J. Am. Chem. Soc., 2018, 140, 6978.
46. R.P. Barnwal, F. Yang and G. Varani, Arch. Biochem. Biophys., 2017, 628, 42.
47. S. Campagne, M. Krepl, et al., Methods Enzymol., 2019, 614, 393.
48. A. Rangadurai, E.S. Szymaski et al., Prog. Nucl. Magn. Reson. Spectrosc., 2019, 112-113, 55.
49. A. Sekhar and L.E. Kay, Annu. Rev. Biophys., 2019, 48, 297.
50. Y. Yamaoki, T. Nagata, et al., Biophys Rev., 2020, 12, 411.

51. A. Marchanka and T. Carlomagno, Methods Enzymol., 2019, 615, 333.
52. P.F.Crain, and J.A. McCloskey, Curr. Opin. Biotechnol., 1998, 9, 25.
53. E. Nordhoff, F. Kirpekar and P. Roepstorff, Mass Spectrom. Rev., 1996, 15, 67.
54. D.-J. Fu, K. Tang, et al., Nat. Biotechnol., 1998, 16, 381.
55. P. Ross, L. Hall, et al., Nat. Biotechnol., 1998, 16, 1347.
56. A. Braun, D.P. Little and H. Köster, Clin. Chem., 1997, 43, 1151. 52
57. A. Braun, D.P., et al., Genomics, 1997, 46, 18.
58. P.L. Ross, and P. Belgrader, Anal. Chem., 1997, 69, 3966.
59. M. Beverly, C. Hagen and O. Slack, Poly A tail length analysis of in vitro transcribed mRNA by LC-MS. (2018).at <https://pubag.nal.usda.gov/catalog/5900193>
60. P.A. Limbach and M.J. Paulines, Wiley Interdiscip. Rev. RNA, 2017, 8.
61. F. Hillenkamp, M. Karas, et al., Anal. Chem., 1991, 63, 1193A.
62. K.J. Wu, A. Steding and C.H. Becker, Rapid Commun. Mass Spectrom., 1993, 7, 142.
63. Y.F. Zhu, C.N. Chung, et al., Rapid Commun. Mass Spectrom., 1996, 10, 383.
64. V.V. Laiko, M.A. Baldwin,and A.L. Burlingame, Anal. Chem., 2000, 72, 652.
65. U. Boesl, Mass Spectrom. Rev., 2017, 36, 86.
66. E.N. Nikolaev, Y.I. Kostyukevich and G.N. Vladimirov, Mass Spectrom. Rev., 2016, 35, 219.
67. S. Berkenkamp, F. Kirpekar and F. Hillenkamp, Science, 1998, 281, 260.
68. D.P. Little, T.J. Cornish, et al., Anal. Chem., 1997, 69, 4540.
69. P. Bayat, D. Lesage and R.B. Cole, Mass Spec.trom. Reviews 2020, 39(5-6), 680.
70. S.A. McLuckey, G.J. Van Berkel and G.L. Glish, J. Am. Soc. Mass Spectrom., 1992, 3, 60.
71. D.P. Little, T.W. Thannhauser and F.W. McLafferty, Proc. Natl. Acad. Sci. USA, 1995, 92, 2318.
72. R.S. Brown and J.J. Lennon, Anal. Chem., 1995, 67, 1998.
73. B. Spengler, D. Kirsch and R. Kaufmann, Rapid Commun. Mass Spectrom., 1991, 5, 198.
74. P.A. Limbach, Mass Spectrom. Rev., 1996, 15, 297.
75. M. Yamashita and J.B. Fenn, J. Phys. Chem., 1984, 88, 4671.
76. J.V. Iribarne and B.A. Thomson, J. Chem. Phys., 1976, 64, 2287.
77. M. Dole, L.L. Mack et al., J. Chem. Phys., 1968, 49, 2240.
78. R. Chen, X. Cheng et al., Anal. Chem., 1995, 67, 1159.
79. A.A. Rostom, P. Fucini et al., Proc. Natl. Acad. Sci. USA, 2000, 97, 5185.
80. S.D. Fuerstenau, W.H. Benner et al., Angew. Chem. Int. Ed. Engl., 2001, 40, 9822.
81. M.H. Amad, N.B. Cech, et al., J. Mass Spectrom., 2000, 35, 784.
82. J.T. Stults, J.C. Marsters and S.A. Carr, Rapid Commun. Mass Spectrom., 1991, 5, 359.
83. P.A. Limbach, P.F. Crain and J.A. McCloskey, J. Am. Soc. Mass Spectrom., 1995, 6, 27.
84. C. Liu, S.A. Hofstadler et al., Anal. Chem., 1998, 70, 1797.
85. E. Nordhoff, A. Ingendoh, et al., Rapid Commun. Mass Spectrom., 1992, 6, 771.
86. C.G. Huber and A. Krajete, Anal. Chem., 1999, 71, 3730.
87. R. Erb,and H. Oberacher, Electrophoresis, 2014, 35, 1226.
88. M. Wilm and M. Mann, Anal. Chem. 1996, 68, 1.
89. T. Kenderdine, Z. Xia, et al., Anal. Chem., 2018, 90, 13541.
90. D. Fabris, Anal. Chem., 2011, 83, 5810.
91. A. Garabedian, D. Butcher et al., Phys. Chem. Chem. Phys. PCCP, 2016, 18, 26691.
92. J.D. Watson and F.H.C. Crick, Nature, 1953, 171, 737.
93. M.H.F. Wilkins, A.R. Stokes and H.R. Wilson, Nature, 1953, 171, 738.
94. R.E. Franklin and R.G. Gosling, Nature, 1953, 171, 740.
95. A. Rich and D.R. Davies, J. Am. Chem. Soc., 1956, 78, 3548. 79
96. A. Varshavsky, Cell, 2006, 127,1295.
97. G. Felsenfeld, D.R. Davies and A. Rich, J. Am. Chem. Soc., 1957, 79, 2023.
98. W. Cochran and F.H.C. Crick, Nature, 1952, 169, 134.
99. W. Cochran, F.H.C. Crick and V. Vand, Acta Cryst., 1952, 5, 581.
100. A.A. Kornyshev, D.J. Lee and S. Leikin, Rev. Modern Physics, 2007, 79, 943.
101. T. Latychevskaia and H.-W. Fink, Optics Express, 2018, 26, 30991.
102. B. Coimbatore Narayanan, J. Westbrook, et al., Nucleic Acids Res., 2014, 42, D114.
103. D.S. Goodsell, C. Zardecki, et al., Protein Sci., 2020, 29, 52.
104. A. Förster and C. Schulze-Briese, Struct. Dyn., 2019, 6, 064302.
105. M. Egli, Curr. Protoc. Nucleic Acid Chem., 2016, 65, 17.13.1.
106. A. Ducruix and R. Giegé, eds, Crystallization of Nucleic Acids and Proteins: A Practical Approach, 2nd ed, Oxford University Press, Oxford, 1999.
107. B. Rupp, Biomolecular Crystallography, 1st edition, Garland Science, New York, 2010.
108. S. Doublié, ed, Macromolecular Crystallography Protocols: Preparation and Crystallization of Macromolecules, Meth. Mol. Biol., 363, vol. 1, Humana Press, Totowa, NJ, 2007.

109. S. Doublié, ed, Macromolecular Crystallography Protocols: Structure Determination, Meth. Mol. Biol., 364, vol. 2, Humana Press, Totowa, NJ, 2007.
110. E. Ennifar, ed, Methods and Protocols: Nucleic Acid Crystallography, Meth. Mol. Biol., 1320, Humana Press, New York, 2015.
111. J.M. Vargason, K. Henderson and P.S. Ho., Proc. Natl. Acad. Sci. USA, 2001, 98, 7265.
112. M.I. Aroyo, ed, International Tables for Crystallography, Vol. A, 6th ed, Space-group Symmetry, Wiley, 2017.
113. M. Egli, P.S. Pallan, et al., J. Am. Chem. Soc., 2006, 128, 10847.
114. M. Egli, P. Lubini and P.S. Pallan, Chem. Soc. Rev., 2007, 36, 31.
115. P.S. Pallan and M. Egli, Nat. Protoc., 2007, 2, 640.
116. P.S. Pallan and M. Egli, Nat. Protoc., 2007, 2, 647.
117. T.L. Blundell and L.N. Johnson, Protein Crystallography, Academic Press, London, 1976.
118. M. Egli and P.S. Pallan, in: Nucleic Acid Crystallography: Methods and Protocols, Meth. Mol. Biol., 1320, E. Ennifar, ed, Humana Press, New York, 2015, 111.
119. D. Jovanovic, P. Tremmel, et al., Angew. Chem. Int. Ed., 2020, 59, 15875.
120. V. Tereshko, S.T. Wallace, et al., RNA, 2001, 7, 405.
121. H. Staudinger, Ber. Dtsch. Chem. Ges., 1920, 53, 1073.
122. H. Hauptman, Curr. Opin. Struct. Biol., 1997, 7, 672.
123. M. Egli, V. Tereshko, et al., Biopolymers (Nucleic Acid Sciences), 2000, 48, 234.
124. P. Willett, J.C. Cole and I.J. Bruno, CrystEngComm, 2020, 22, 7233.
125. H. Kwon, P.S. Langan, et al., Acta Cryst. D, 2018, 74, 792.
126. V. Gopal Vandavasi, M.P. Blakeley, et al., Structure, 2018, 26, 1645.
127. J.M. Harp, L. Coates, B. Sullivan and M. Egli, Nucleic Acids Res., 2021, 49, 4782.
128. L. Coates, H.B. Cao, et al., Rev. Sci. Instrum., 2018, 89, 092802.
129. B.L. Nannenga and T. Gonen, Curr. Opin. Struct. Biol., 2014, 27, 24.
130. T. Gruene, J.T.C. Wennmacher, et al., Angew. Chem. Int. Ed., 2018, 57, 16313.
131. C.G. Jones, M.W. Martynowycz, et al., ACS Cent. Sci., 2018, 4, 1587.
132. J.A. Rodriguez, M. Ivanova, et al., Nature, 2015, 525, 486.
133. R.K. Hite, S. Raunser and T. Walz, Curr. Opin. Struct. Biol, 2007, 17, 389.
134. B.L. Nannenga and T. Gonen, Nat. Methods, 2019, 16, 369.
135. A. Bogner, P.H. Jouneau et al., Micron, 2007, 38, 390.
136. E. Nogales and S.H. Scheres, Mol. Cell., 2015, 58, 677.
137. M.S. Chapman, A. Trzynka and B.K. Chapman, J. Struct. Biol., 2013, 182, 10.
138. R. A. Nicholls, M. Tykac, et al., Acta. Crystallogr. D. Struct. Biol., 2018, 74, 492.
139. D. Liebschner, P.V. Afonine et al., Acta. Crystallogr. D. Struct. Biol., 2019, 75, 861.
140. B. Loveland, G. Demo and A.A. Korostelev, Nature, 2020, 584, 640.
141. E. Ruska, Biosci. Rep., 1987, 7, 607.
142. J. Dubochet, F.P. Booy et al., Annu Rev Biophys Bioeng, 1981, 10, 133.
143. R.H. Vogel, S. W. Provencher et al., Nature, 1986, 320, 533.
144. J. Frank, W. Goldfarb, et al., Ultramicroscopy, 1978, 3, 283.
145. S.H. Scheres, H. Gao et al., Nat. Methods., 2007, 4, 27.
146. S. Wu, J.P. Armache and Y. Cheng, Microscopy (Oxf), 2016, 65, 35.
147. F. Brilot, J.Z. Chen et al., J. Struct. Biol., 2012, 177, 630.
148. Y. Cheng, Cell, 2015, 161, 450.
149. R.S. Ruskin, Z. Yu and N. Grigorieff, J. Struct. Biol., 2013, 184, 385.
150. M.G. Campbell, A. Cheng et al., Structure, 2012, 20, 1823.
151. T. Nakane, A. Kotecha et al., Nature, 2020, 587, 152.
152. Z.L. Watson, F. R. Ward et al., Elife, 2020, 9, e60482.
153. L.A. Earl, V. Falconieri, et al., Curr. Opin. Struct. Biol., 2017, 46, 71.
154. R. Jain, W.J. Rice et al., Nat. Struct. Mol. Biol., 2019, 26, 955.
155. Y. Gao, Y. Cui et al., Science, 2019, 363, 814.
156. P. Goswami, F. Abid Ali et al., Nat Commun, 2018, 9, 5061.
157. G. Abascal-Palacios, E.P. Ramsay et al., Nature, 2018, 553, 301.
158. L. Farnung, S.M. Vos and P. Cramer, Nature. Commun., 2018, 9, 1.
159. C. Grimm, H. S. Hillen et al., Cell, 2019, 179, 1537.
160. J.Y. Kang, R.A. Mooney et al., Cell, 2018, 173, 1650.
161. L. Tafur, Y. Sadian et al., Elife, 2019, 8, e43204.
162. A. Amunts, A. Brown et al., Science, 2014, 343, 1485.
163. P.D. Abeyrathne, C.S. Koh et al., Elife, 2016, 5, e14874.
164. B. Loveland, G. Demo, et al., Nature, 2017, 546, 113.
165. S. Kaledhonkar, Z. Fu et al., Nature, 2019, 570, 400.
166. N. Fischer, P. Neumann et al., Nature, 2016, 540, 80.

167. T. Hussain, J.L. Llácer et al., Cell, 2016, 167, 133.
168. Z. Fu, G. Indrisiunaite et al., Nat. Commun., 2019, 10, 2579.
169. W.P. Galej, M. E. Wilkinson et al., Nature, 2016, 537, 197.
170. K. Bertram, D.E. Agafonov et al., Cell, 2017, 170, 701.
171. R. Bai, C. Yan et al., Cell, 2017, 171, 1589.
172. C. Plaschka, P.-C. Lin and K. Nagai, Nature, 2017, 546, 617.
173. Y. Sugita, H. Matsunami et al., Nature, 2018, 563, 137.
174. Y.-T. Liu, J. Jih et al., Nature, 2019, 570, 257.
175. J. Xu, N. Dayan, et al., Prof. Natl. Acad. Sci. USA, 2019, 116, 5493.
176. F. Wang, Y. Liu et al., Prof. Natl. Acad. Sci. USA, 2019, 116, 22591.
177. T. Yokoyama, K. Machida et al., Mol. Cell, 2019, 74, 1205.
178. C.O. Barnes, A.P. West, Jr. et al., Cell, 2020, 182, 828.
179. W. Yin, C. Mao et al., Science, 2020, 368, 1499.
180. T.W. Guo, A. Bartesaghi et al., Cell, 2017, 171, 414.
181. Z. Liu, J. Wang et al., Cell, 2018, 173, 1191.
182. Z. Liu, C. Gutierrez-Vargas et al., Protein. Sci., 2017, 26, 82.
183. K. Zhang, S. Li et al., Nat. Commun., 2019, 10, 5511. 90
184. M.M. Golas, C. Bohm et al., EMBO. J., 2009, 28, 766. 91
185. F.K. Schur, M. Obr et al., Science, 2016, 353, 506.
186. A. von Kügelgen, H. Tang et al., Cell, 2020, 180, 348.
187. J.P. Rickgauer, N. Grigorieff and W. Denk, Elife, 2017, 6, e25648.
188. J. Kapuscinski, Biotechnic & Histochemistry, 1995, 70, 220.
189. J.M. Levsky and R.H. Singer, J. Cell Science, 2003, 116, 2833.
190. J. Schnitzbauer, M. Strauss, et al., Nat. Protoc., 2017, 12, 1198.
191. R.M. Clegg, A.I.H. Murchie, et al., Biochemistry, 1992, 31, 4846.
192. B. Hellenkamp, S. Schmid, et al., Nat. Methods., 2018, 15, 669.
193. S. Kalinin, T. Peulen, et al., Nat. Methods., 2012, 9, 1218.
194. T.D. Craggs, M. Sustarsic, et al., Nucleic Acids Res., 2019, 47, 10788.
195. G. Binnig and C.F. Quate, Phys. Rev. Lett., 1986, 56, 930. 99.
196. D. Fotiadis, Micron, 2002, 33, 385.
197. Y.F. Dufrêne, T. Ando, et al., Nat. Nanotechnol., 2017, 12, 295.
198. Y.F. Dufrêne, D. Martinez-Martin, et al., Nat. Methods, 2013, 10, 847.
199. A.P. Nievergelt, C. Kammer, et al., Small Methods, 2019, 114, 1900031.
200. H.G. Hansma, M. Bezanilla, et al., Nucleic Acids Res., 1993, 21, 505.
201. A. Pyne, R. Thompson, et al., Small, 2014, 10, 3257.
202. A. Miyagi, T. Ando and Y.L. Lyubchenko, Biochemistry, 2011, 50, 7901.
203. F. Moreno-Herrero, M. de Jager, et al., Nature, 2005, 437, 440.
204. T. Uchihashi, H. Watanabe, et al., Ultramicroscopy, 2016, 160, 182.
205. A.P. Nievergelt, N. Banterle, et al., Nat. Nanotechnol., 2018, 13, 696.
206. A.J. Lee, M. Szymonik, et al., Nano Res., 2015, 8, 1811.
207. H.G. Hansma and D.E. Laney, Biophys. J., 1996, 70, 1933.
208. R. Krautbauer, L H. Pope et al., FEBS Lett., 2002, 510, 154.
209. P. Serwer, Electrophoresis, 1983, 4, 375.
210. L.R. Williams, Biotech. Histochem., 2001, 76, 127.
211. D.C. Rio, Cold Spring Harb. Protoc., 2014, 687.
212. L.M. Albright and B.E. Slatko, Curr. Protoc. Nucleic Acid Chem., 2001, Appendix 3: Appendix 3B.
213. J.L. Viovy, Mol. Biotechnol., 1996, 6, 31.
214. A.D. Bates and A. Maxwell, DNA Topology, 2nd ed., Oxford University Press, Oxford, 2005.
215. O.J. Lumpkin and B.H. Zimm, Biopolymers, 1982, 21, 2315.
216. M. Seo, L. Lei and M. Egli, Curr. Protoc. Nucleic Acid Chem., 2019, 76, e70.
217. K.T. Gagnon and E.S. Maxwell, Methods Mol. Biol., 2011, 703, 275.
218. D.C. Schwartz and C.R. Cantor, Cell, 1984, 37, 67.
219. B.L. Karger and A. Guttman, Electrophoresis, 2009, 30 (Suppl. 1), S196.
220. A. Schroeder, O. Mueller, et al., BMC Mol. Biol., 2006, 7, 3.
221. A. Goyon, P. Yehl and K. Zhang, J. Pharmaceut. Biomed. Anal., 2020, 182, 113105.
222. H. Cramer, K.J. Finn and E. Herzberg, Purity analysis and impurities detection by reversed-phase high-performance liquid chromatography. In Handbook of Analysis of Oligonucleotides and Related Products, J.V. Bonilla and G.S. Srivatsa, Eds, CRC Press, Boca Raton, 2011, 1.
223. J. Thayer, V. Murugaiah, and Y. Wu, Y., Purity analysis and impurities detection by AEX-HPLC. In Handbook of Analysis of Oligonucleotides and Related Products, J.V. Bonilla and G.S. Srivatsa, Eds, CRC Press, Boca Raton, 2011, 47.

224. M.F. Chan and I. Roymoulik, Purity analysis and molecular weight determination by size exclusion HPLC. In Handbook of Analysis of Oligonucleotides and Related Products, J.V. Bonilla and G.S. Srivatsa, Eds, CRC Press, Boca Raton, 2011, 105.
225. E. Largy and J.-L. Mergny, Nucleic Acids Res., 2014, 42, e149.
226. F. Kanwal and C. Lu, C., J. Chromatog. B, 2019, 120, 71.
227. M. Meselson, F.W. Stahl, and J. Vinograd, Proc. Natl. Acad. Sci. USA, 1957, 43, 581.
228. P. Hensley, Structure, 1996, 4, 367.
229. S.P. Modak, M.-T. Imaizumi, et al., Mol. Biol. Reports, 1978, 4, 55.
230. H. Zhao, C.A. Brautigam, et al., Curr. Protoc. Protein Sci., 2013, 71, 20.12.1.
231. J.L. Cole, J.J. Correia, and W.F. Stafford, Biophys. Chem., 2011, 159, 120.
232. S.E. Harding, and B.Z. Chowdhry, eds, Protein Ligand Interactions: Hydrodynamics and Calorimetry. Oxford University Press, Oxford, UK, 2001.
233. Savelyev, G.E. Gorbet, et al., PLoS Comput. Biol., 2020, 16, e1007942.
234. S. Mitra, Methods Enzymol., 2009, 469, 209.
235. J. Stetefeld, S.A. McKenna and T.R. Patel, Biophys. Rev., 2016, 8, 409.
236. L. Øgendal, Light scattering. A brief introduction. https://www.nbi.dk/~ogendal/personal/lho/LS_brief_intro.pdf.
237. R. Xu, Particuology, 2015, 18, 11.
238. C. Jeffries, Static and dynamic light scattering. https://www.presentica.com/doc/11148992/static-and-dynamic-light-pdf-document.
239. C.D. Putnam, M. Hammel, et al., Q. Rev. Biophys., 2007, 40, 191.
240. R.P. Rambo and J.A. Tainer, Nature, 2013, 496, 477.
241. G. Zaccaï and B. Jacrot, Ann. Rev. Biophys. Bioeng, 1983, 12, 139.
242. M.J. Hollamby, Phys. Chem. Chem. Phys., 2013, 15, 10566.
243. E. Mahieu and F. Gabel, Acta Crystallogr. D Struct. Biol., 2018, 74, 715.
244. D.A. Jacques and J. Trewhella, Protein Sci., 2010, 19, 642.
245. R. Pattanayek, S.K. Sidiqi and M. Egli, Biochemistry, 2012, 51, 8050.
246. M. Zulauf and A. D'Arcy, J. Crystal Growth, 1992, 122, 102.
247. A.R. Ferré-D'Amaré and S.K. Burley, Structure, 1994, 2, 357.
248. A. Matte and M. Cygler, Am. Laboratory, June 1 2007, p.1; https://americanlaboratory.com/914-Application-Notes/35112.
249. K. Dierks, A. Meyer, et al., Crystal Growth & Design, 2008, 8, 1628.
250. B. Lorber, F. Fischer, et al., Biochem. Mol. Biol. Educ., 2012, 40, 372.
251. M.A. Graewert, S. Da Vela, et al., Crystals, 2020, 10, 975.
252. V.V. Volkov and D.I. Svergun, J. Appl. Cryst., 2003, 36, 860.
253. M.V. Petoukhov and D.I Svergun, Biophys. J., 2005, 89, 1237.
254. R. Pattanayek, D.R. Williams, et al., PLoS ONE, 2011, 6, e23697.
255. J.E. Burke and S.E. Butcher, Curr. Protoc. Nucleic Acid Chem., 2012, 51, 7.18.1.
256. J. Marmur and P. Doty, J. Mol. Biol., 1962, 5, 109.
257. P. Yakovchuk, E. Protozanova and M.D. Frank-Kamenetskii, Nucleic Acids Res., 2006, 34, 564.
258. J.D. Puglisi and I. Tinoco Jr., Meth. Enzymol., 1989, 180, 304.
259. G.E. Plum, D.S. Pilch, et al., Annu. Rev. Biophys. Biomol. Struct. Dyn., 1995, 24, 319.
260. A.S. Boutorine and C. Escudé, Curr. Protoc. Nucleic Acid Chem., 2007, 29, 7.12.1.
261. J.-L. Mergny and L. Lacroix, Protoc. Nucleic Acid Chem., 2009, 37, 17.1.1.
262. R.D. Gray and J.B. Chaires, Curr. Protoc. Nucleic Acid Chem., 2011, 45, 17.4.1.
263. G.E. Plum, Curr. Protoc. Nucleic Acid Chem., 2000, 7.3.1.
264. P.S. Pallan, E. Greene, et al., Nucleic Acids Res., 2011, 39, 3482.
265. L.A. Marky and K.J. Breslauer, Biopolymers, 1987, 26, 1601.
266. P.S. Pallan, T.P. Lybrand, et al., Biochemistry, 2020, 59, 4627.
267. K.J. Breslauer, Meth. Enzymol., 1995, 259, 221.
268. P.C. Dedon, Curr. Protoc. Nucleic Acid Chem., 2000, 8.2.1.
269. E. Rozners, Curr. Protoc. Nucleic Acid Chem., 2010, 43, 7.14.1.
270. H. Tateishi-Karimata, S. Nakano and N. Sugimoto, Curr. Protoc. Nucleic Acid Chem., 2013, 53, 7.19.1.
271. C.H. Spink and J.B. Chaires, Biochemistry, 1999, 38, 496.
272. S.J. Schroeder and D.H. Turner, Meth. Enzymol., 2009, 468, 371.
273. E. Rozners, D.S. Pilch and M. Egli, Curr. Protoc. Nucleic Acid Chem., 2015, 63, 7.4.1.
274. A.L. Feig, Meth. Enzymol. 2009, 468, 409.
275. N.N. Salim and A.L. Feig, Methods, 2009, 47, 198.
276. H. Naghibi, A. Tamura and J.M. Sturtevant, Proc. Natl. Acad. Sci. USA, 1995, 92, 5597.
277. C.H. Spink, Differential scanning calorimetry. In Biophysical Tools for Biologists, Vol. 1, In Vitro Techniques, J. Correia and H. Detrich (eds.), Elsevier, New York, 2008, 115.
278. C.H. Spink, Methods, 2015, 76, 78.

279. M. Jing and M.T. Bowser, Anal. Chim. Acta, 2011, 686, 9.
280. X. Lou, M. Egli and X. Yang, Curr. Protoc. Nucleic Acid Chem., 2016, 67, 7.25.
281. J.T. Edsall and H. Gutfreund, Calorimetry, heat capacity, and phase transitions. In Biothermodynamics: The Study of Biochemical Processes at Equilibrium, John Wiley & Sons, New York, 1983, 210.
282. J. Gralla and D.M. Crothers, J. Mol. Biol., 1973, 78, 301.
283. T. Schlick and A.M. Pyle, Biophys J., 2017, 113, 225.
284. L. Huang, H. Zhang, et al., Bioinformatics, 2019, 35, i295.
285. J.A. Maier, C. Martinez, et al., J. Chem. Theory Comp., 2015, 11, 3696.
286. D.A. Case, T.E. Cheatham, 3rd, et al., J. Comput. Chem., 2005, 26, 1668.
287. B.R. Brooks, C.L. Brooks, 3rd, et al., J. Comput. Chem., 2009, 30, 1545.
288. D. Van Der Spoel, E. Lindahl, et al., J. Comput. Chem., 2005, 26, 1701.
289. R. Galindo-Murillo, D.R. Roe and T.E. Cheatham, 3rd, Biochim. Biophys. Acta, 2015, 1850, 1041.
290. R. Galindo-Murillo, J.C. Robertson, et al., J. Chem. Theory Comp., 2016, 12, 4114.
291. J. Šponer, G. Bussi, et al., Chem. Rev., 2018, 118, 4177.
292. R. Galindo-Murillo, J.C. Garcia-Ramos, et al., Nucleic Acids Res., 2015, 43, 5364.
293. V. Minhas, T. Sun, et al., J. Phys. Chem. B, 2020, 124, 38.
294. J. Walther, P.D. Dans, et al., Nucleic Acids Res., 2020, 48, e29.
295. M. Karplus and J.A. McCammon, Nat. Struct. Biol., 2002, 9, 646.
296. K. Ando, K.R. Condroski, et al., J. Org. Chem., 1998, 63, 3196.
297. A. Warshel and M. Levitt, J. Mol. Biol., 1976, 103, 227.
298. E.A. Mullins, A.A. Rodriguez, et al., Biochem. Sci., 2019, 44, 765.
299. W. Kohn and L-J. Sham, Phys. Rev. A, 1965, 140, 1133.
300. W. Kohn, A.D. Becke and R. G. Parr, J. Phys. Chem., 1996, 100, 12974.
301. S. Ahmadi, L. Barrios Herrera, et al., Int. J. Quantum Chem., 2018, 118, e25558.
302. R. Lonsdale, J.N. Harvey and A.J. Mulholland, Chem. Soc. Rev., 2012, 41, 3025.
303. S.S. Parikh, G. Walcher, et al., Proc. Natl. Acad. Sci., USA, 2000, 97, 5083.
304. H.M. Senn and W. Thiel, Angew. Chem. Int. Ed., 2009, 48, 1198.
305. S.C.L. Kamerlin, M. Haranczyk and A. Warshel, J. Phys. Chem. B, 2009, 113, 1253.
306. A. Warshel, P. K. Sharma, et al., Chem. Rev., 2006, 106, 3210.
307. M. Garcia-Viloca, J. Gao, et al., Science, 2004, 303, 186.
308. A.R. Dinner, G.M. Blackburn and M. Karplus, Nature, 2001, 413, 752.
309. R. Werner and J.T. Stivers, Biochemistry, 2000, 39, 14054.
310. T. Lindahl and R.D. Wood, Science, 1999, 286, 1897.
311. E. Naydenova, S. Roßbach and C. Ochsenfeld, J. Chem. Theory Comput., 2019, 15, 4344.
312. W. Smith and T.R. Forester, J. Mol. Graph., 1996, 14, 136.
313. C.D. Mol, A.S. Arvai, et al., Cell, 1995, 80, 869.
314. E. Nogales and B.J. Greber, Curr. Op. Struct. Biol., 2019, 59, 188.
315. H.-P. Cheng, and E. Deuments, et al., Front. Chem., 2020, 8, 587143.