

chapter

8

Nucleotides and Nucleic Acids

- 1. Nucleotide Structure** Which positions in the purine ring of a purine nucleotide in DNA have the potential to form hydrogen bonds but are not involved in Watson-Crick base pairing?

Answer All purine ring nitrogens (N-1, N-3, N-7, and N-9) have the potential to form hydrogen bonds (see Figs. 8-1, 8-11, and 2-3). However, N-1 is involved in Watson-Crick hydrogen bonding with a pyrimidine, and N-9 is involved in the *N*-glycosyl linkage with deoxyribose and has very limited hydrogen-bonding capacity. Thus, N-3 and N-7 are available to form further hydrogen bonds.

- 2. Base Sequence of Complementary DNA Strands** One strand of a double-helical DNA has the sequence (5')GCGCAATATTTCTCAAAATATTGCGC(3'). Write the base sequence of the complementary strand. What special type of sequence is contained in this DNA segment? Does the double-stranded DNA have the potential to form any alternative structures?

Answer The complementary strand is



(Note that the sequence of a single strand is always written in the 5' → 3' direction.) This sequence has a palindrome, an inverted repeat with twofold symmetry:



Because this sequence is self-complementary, the individual strands have the potential to form hairpin structures. The two strands together may also form a cruciform.

- 3. DNA of the Human Body** Calculate the weight in grams of a double-helical DNA molecule stretching from the Earth to the moon (~320,000 km). The DNA double helix weighs about 1×10^{-18} g per 1,000 nucleotide pairs; each base pair extends 3.4 Å. For an interesting comparison, your body contains about 0.5 g of DNA!

Answer

The length of the DNA is

$$(3.2 \times 10^5 \text{ km})(10^{12} \text{ nm/km})(10 \text{ Å/nm}) = 3.2 \times 10^{18} \text{ Å}$$

The number of base pairs (bp) is

$$\frac{3.2 \times 10^{18} \text{ Å}}{3.4 \text{ Å/bp}} = 9.4 \times 10^{17} \text{ bp}$$

Thus, the weight of the DNA molecule is

$$(9.4 \times 10^{17} \text{ bp})(1 \times 10^{-18} \text{ g}/10^3 \text{ bp}) = 9.4 \times 10^{-4} \text{ g} = 0.00094 \text{ g}$$

S-90 Chapter 8 Nucleotides and Nucleic Acids

- 4. DNA Bending** Assume that a poly(A) tract five base pairs long produces a 20° bend in a DNA strand. Calculate the total (net) bend produced in a DNA if the center base pairs (the third of five) of two successive $(dA)_5$ tracts are located **(a)** 10 base pairs apart; **(b)** 15 base pairs apart. Assume 10 base pairs per turn in the DNA double helix.

Answer When bending elements are repeated in phase with the helix turn (i.e., every 10 base pairs) as in **(a)**, the total bend is additive; when bending elements are repeated out of phase by one half-turn as in **(b)**, they cancel each other out. Thus, the net bend is **(a)** 40° ; **(b)** 0° .

- 5. Distinction between DNA Structure and RNA Structure** Hairpins may form at palindromic sequences in single strands of either RNA or DNA. How is the helical structure of a long and fully base-paired (except at the end) hairpin in RNA different from that of a similar hairpin in DNA?

Answer The RNA helix assumes the A conformation; the DNA helix generally assumes the B conformation. (The presence of the 2'-OH group on ribose makes it sterically impossible for double-helical RNA to assume the B-form helix.)

- 6. Nucleotide Chemistry** The cells of many eukaryotic organisms have highly specialized systems that specifically repair G-T mismatches in DNA. The mismatch is repaired to form a G=C (not A=T) base pair. This G-T mismatch repair mechanism occurs in addition to a more general system that repairs virtually all mismatches. Can you suggest why cells might require a specialized system to repair G-T mismatches?

Answer Many C residues of CpG sequences in eukaryotic DNA are methylated at the 5' position to 5-methylcytosine. (About 5% of all C residues are methylated.) Spontaneous deamination of 5-methylcytosine yields thymine, T, and a G-T mismatch resulting from spontaneous deamination of 5-methylcytosine in a G=C base pair is one of the most common mismatches in eukaryotic cells. The specialized repair mechanism to convert G-T back to G=C is directed at this common class of mismatch.

- 7. Spontaneous DNA Damage** Hydrolysis of the *N*-glycosyl bond between deoxyribose and a purine in DNA creates an AP site. An AP site generates a thermodynamic destabilization greater than that created by any DNA mismatched base pair. This effect is not completely understood. Examine the structure of an AP site (see Fig. 8-33b) and describe some chemical consequences of base loss.

Answer Without the base, the ribose ring can be opened to generate the noncyclic aldehyde form. This, and the loss of base-stacking interactions, could contribute significant flexibility to the DNA backbone.

- 8. Nucleic Acid Structure** Explain why the absorption of UV light by double-stranded DNA increases (the hyperchromic effect) when the DNA is denatured.

Answer The double-helical structure is stabilized by hydrogen bonding between complementary bases on opposite strands and by base stacking between adjacent bases on the same strand. Base stacking in nucleic acids causes a decrease in the absorption of UV light (relative to the non-stacked structure). On denaturation of DNA, the base stacking is lost and UV absorption increases.

- 9. Determination of Protein Concentration in a Solution Containing Proteins and Nucleic Acids** The concentration of protein or nucleic acid in a solution containing both can be estimated by using their different light absorption properties: proteins absorb most strongly at 280 nm and nucleic acids at 260 nm. Estimates of their respective concentrations in a mixture can be made by measuring the absorbance (A) of the solution at 280 nm and 260 nm and using the table that follows, which gives $R_{280/260}$, the ratio of absorbances at 280 and 260 nm; the percentage of total mass that is nucleic acid; and a factor, F , that corrects the A_{280} reading and gives a more accurate protein estimate. The protein concentration (in mg/ml) = $F \times A_{280}$ (assuming the cuvette is 1 cm wide). Calculate the protein concentration in a solution of $A_{280} = 0.69$ and $A_{260} = 0.94$.

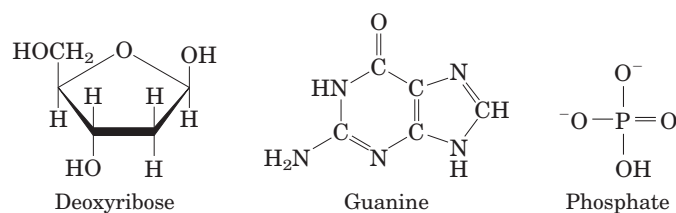
$R_{280/260}$	Proportion of nucleic acid (%)	F
1.75	0.00	1.116
1.63	0.25	1.081
1.52	0.50	1.054
1.40	0.75	1.023
1.36	1.00	0.994
1.30	1.25	0.970
1.25	1.50	0.944
1.16	2.00	0.899
1.09	2.50	0.852
1.03	3.00	0.814
0.979	3.50	0.776
0.939	4.00	0.743
0.874	5.00	0.682
0.846	5.50	0.656
0.822	6.00	0.632
0.804	6.50	0.607
0.784	7.00	0.585
0.767	7.50	0.565
0.753	8.00	0.545
0.730	9.00	0.508
0.705	10.00	0.478
0.671	12.00	0.422
0.644	14.00	0.377
0.615	17.00	0.322
0.595	20.00	0.278

Answer For this protein solution, $R_{280/260} = 0.69/0.94 = 0.73$, so (from the table) $F = 0.508$. The concentration of protein is $F \times A_{280} = (0.508 \times 0.69) \text{ mg/mL} = 0.35 \text{ mg/mL}$.

Note: the table applies to mixtures of proteins, such as might be found in a crude cellular extract, and reflects the absorption properties of average proteins. For a purified protein, the values of F would have to be altered to reflect the unique molar extinction coefficient of that protein.

- 10. Solubility of the Components of DNA** Draw the following structures and rate their relative solubilities in water (most soluble to least soluble): deoxyribose, guanine, phosphate. How are these solubilities consistent with the three-dimensional structure of double-stranded DNA?

Answer



S-92 Chapter 8 Nucleotides and Nucleic Acids

Solubilities: phosphate > deoxyribose > guanine. The negatively charged phosphate is the most water-soluble; the deoxyribose, with several hydroxyl groups, is quite water-soluble; and guanine, a hydrophobic base, is relatively insoluble in water. The polar phosphate groups and sugars are on the outside of the DNA double helix, exposed to water. The hydrophobic bases are located in the interior of the double helix, away from water.

- 11. Sanger Sequencing Logic** In the Sanger (dideoxy) method for DNA sequencing, a small amount of a dideoxynucleotide triphosphate—say, ddCTP—is added to the sequencing reaction along with a larger amount of the corresponding dCTP. What result would be observed if the dCTP were omitted?

Answer If dCTP is omitted from the reaction mixture, when the first G residue is encountered in the template, ddCTP is added and polymerization halts. Only one band will appear in the sequencing gel.

- 12. DNA Sequencing** The following DNA fragment was sequenced by the Sanger method. The asterisk indicates a fluorescent label.

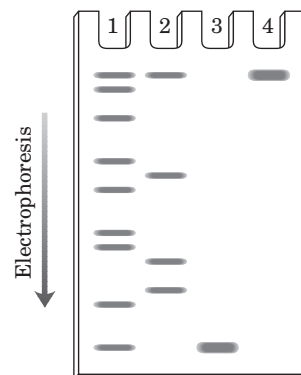


A sample of the DNA was reacted with DNA polymerase and each of the nucleotide mixtures (in an appropriate buffer) listed below. Dideoxynucleotides (ddNTPs) were added in relatively small amounts.

1. dATP, dTTP, dCTP, dGTP, ddTTP
2. dATP, dTTP, dCTP, dGTP, ddGTP
3. dATP, dCTP, dGTP, ddTTP
4. dATP, dTTP, dCTP, dGTP

The resulting DNA was separated by electrophoresis on an agarose gel, and the fluorescent bands on the gel were located. The band pattern resulting from nucleotide mixture 1 is shown below. Assuming that all mixtures were run on the same gel, what did the remaining lanes of the gel look like?



Answer

Lane 1: The reaction mixture that generated these bands included all the deoxynucleotides, plus dideoxythymidine. The fragments are of various lengths, all terminating where a ddTTP was substituted for a dTTP. For a small portion of the strands synthesized in the experiment, ddTTP would not be inserted and the strand would thus extend to the final G. Thus, the nine products are (from top to bottom of the gel):

5'-primer-TAATGCGTTCCTGTAATCTG
 5'-primer-TAATGCGTTCCTGTAATCT
 5'-primer-TAATGCGTTCCTGTAAT
 5'-primer-TAATGCGTTCCTGT
 5'-primer-TAATGCGTTCCT
 5'-primer-TAATGCGTT
 5'-primer-TAATGCGT
 5'-primer-TAAT
 5'-primer-T

Lane 2: Similarly, this lane will have four bands (top to bottom), for the following fragments, each terminating where ddGTP was inserted in place of dGTP:

5'-primer-TAATGCGTTCCTGTAATCTG
 5'-primer-TAATGCGTTCCTG
 5'-primer-TAATGCG
 5'-primer-TAATG

Lane 3: Because mixture 3 lacked dTTP, every fragment was terminated immediately after the primer as ddTTP was inserted, to form 5'-primer-T. The result will be a single thick band near the bottom of the gel.

Lane 4: When all the deoxynucleotides were provided, but no dideoxynucleotide, a single labeled product formed: 5'-primer-TAATGCGTTCCTGTAATCTG. This will appear as a single thick band at the top of the gel.

- 13. Snake Venom Phosphodiesterase** An exonuclease is an enzyme that sequentially cleaves nucleotides from the end of a polynucleotide strand. Snake venom phosphodiesterase, which hydrolyzes nucleotides from the 3' end of any oligonucleotide with a free 3'-hydroxyl group, cleaves between the 3' hydroxyl of the ribose or deoxyribose and the phosphoryl group of the next nucleotide. It acts on single-stranded DNA or RNA and has no base specificity. This enzyme was used in sequence

S-94 Chapter 8 Nucleotides and Nucleic Acids

determination experiments before the development of modern nucleic acid sequencing techniques. What are the products of partial digestion by snake venom phosphodiesterase of an oligonucleotide with the following sequence?



Answer When snake venom phosphodiesterase cleaves a nucleotide from a nucleic acid strand, it leaves the phosphoryl group attached to the 5' position of the released nucleotide and a free 3'-OH group on the remaining strand. Partial digestion of the oligonucleotide gives a mixture of fragments of all lengths, as well as some of the original, undigested strand, so the products are (P represents the phosphate group):



and the released nucleoside 5'-phosphates, GMP, UMP, AMP, and CMP.

- 14. Preserving DNA in Bacterial Endospores** Bacterial endospores form when the environment is no longer conducive to active cell metabolism. The soil bacterium *Bacillus subtilis*, for example, begins the process of sporulation when one or more nutrients are depleted. The end product is a small, metabolically dormant structure that can survive almost indefinitely with no detectable metabolism. Spores have mechanisms to prevent accumulation of potentially lethal mutations in their DNA over periods of dormancy that can exceed 1,000 years. *B. subtilis* spores are much more resistant than are the organism's growing cells to heat, UV radiation, and oxidizing agents, all of which promote mutations.
- (a) One factor that prevents potential DNA damage in spores is their greatly decreased water content. How would this affect some types of mutations?
- (b) Endospores have a category of proteins called small acid-soluble proteins (SASPs) that bind to their DNA, preventing formation of cyclobutane-type dimers. What causes cyclobutane dimers, and why do bacterial endospores need mechanisms to prevent their formation?

Answer

- (a) Water is a participant in most biological reactions, including those that cause mutations. The low water content in endospores reduces the activity of mutation-causing enzymes and slows the rate of nonenzymatic depurination reactions, which are hydrolysis reactions.
- (b) UV light induces the condensation of adjacent pyrimidine bases to form cyclobutane pyrimidine dimers. The spores of *B. subtilis*, a soil organism, are at constant risk of being lofted to the top of the soil or into the air, where they are subject to UV exposure, possibly for prolonged periods. Protection from UV-induced mutation is critical to spore DNA integrity.

- 15. Oligonucleotide Synthesis** In the scheme of Figure 8-35, each new base to be added to the growing oligonucleotide is modified so that its 3' hydroxyl is activated and the 5' hydroxyl has a dimethoxytrityl (DMT) group attached. What is the function of the DMT group on the incoming base?

Answer DMT is a blocking group that prevents reaction of the incoming base with itself.

Biochemistry on the Internet

- 16. The Structure of DNA** Elucidation of the three-dimensional structure of DNA helped researchers understand how this molecule conveys information that can be faithfully replicated from one generation to the next. To see the secondary structure of double-stranded DNA, go to the Protein Data Bank website (www.rcsb.org). Use the PDB identifiers listed below to retrieve the structure summaries for the two forms of DNA. Open the structures using Jmol (linked under the Display Options), and use the controls in the Jmol menu (accessed with a control-click or by clicking on the Jmol logo in the lower right corner of the image screen) to complete the following exercises. Refer to the Jmol help links as needed.
- (a) Obtain the file for 141D, a highly conserved, repeated DNA sequence from the end of the HIV-1 (the virus that causes AIDS) genome. Display the molecule as a ball-and-stick structure (in the control menu, choose Select > All, then Render > Scheme > Ball and Stick). Identify the sugar-phosphate backbone for each strand of the DNA duplex. Locate and identify individual bases. Identify the 5' end of each strand. Locate the major and minor grooves. Is this a right- or left-handed helix?
 - (b) Obtain the file for 145D, a DNA with the Z conformation. Display the molecule as a ball-and-stick structure. Identify the sugar-phosphate backbone for each strand of the DNA duplex. Is this a right- or left-handed helix?
 - (c) To fully appreciate the secondary structure of DNA, view the molecules in stereo. On the control menu, Select > All, then Render > Stereographic > Cross-eyed or Wall-eyed. You will see two images of the DNA molecule. Sit with your nose approximately 10 inches from the monitor and focus on the tip of your nose (cross-eyed) or the opposite edges of the screen (wall-eyed). In the background you should see three images of the DNA helix. Shift your focus to the middle image, which should appear three-dimensional. (Note that only one of the two authors can make this work.)

Answer

- (a) The DNA fragment modeled in file 141D, from the human immunodeficiency virus, is the B form, the standard Watson-Crick structure (although this particular structure is a bent B-form DNA). This fragment has an adenine at the 5' end and a guanine at the 3' end; click on the bases at each end of the helix to identify which is the 5' end. When the helix is oriented with the 5' adenine at the upper left-hand side of the model, the *minor groove* is in the center of the model. Rotating the model so that the 5' adenine is at the upper right-hand side positions the *major groove* in the center. The spiral of this helix runs upward in a counterclockwise direction, so this is a right-handed helix.
- (b) The model of DNA in the Z conformation includes a shell of water molecules around the helix. The water molecules are visible when the complex is viewed in ball-and-stick mode. Turn off the display of the water molecules using the console controls Select > Nucleic > DNA. Then Select > Display Selected Only. The backbone of DNA in the Z conformation is very different from that in the B conformation. The helix spiral runs upward in a clockwise direction, so this is a left-handed helix.
- (c) Viewing the structures in stereo takes a bit of practice, but perseverance will be rewarded! Here are some tips for successful three-dimensional viewing:
 - (1) Turn off or lower the room lighting.
 - (2) Sit directly in front of the screen.
 - (3) Use a ruler to make sure you are 10 to 11 inches from the screen.
 - (4) Position your head so that when you focus on the tip of your nose, the screen images are on either side of the tip (i.e., look down your nose at the structures).
 - (5) Move your head slightly closer to or farther away from the screen to bring the middle image into focus. Don't look directly at the middle image as you try to bring it into focus.

S-96 Chapter 8 Amino Acids, Peptides, and Proteins

- (6) If you find it uncomfortable to focus on the tip of your nose, try using the tip of a finger (positioned just beyond the tip of your nose) instead.
- (7) Relax as you attempt to view the three-dimensional image.

Note that many people, including one of the text authors, have some trouble making this work!

Data Analysis Problem

17. Chargaff's Studies of DNA Structure The chapter section "DNA Is a Double Helix that Stores Genetic Information" includes a summary of the main findings of Erwin Chargaff and his coworkers, listed as four conclusions ("Chargaff's rules"; p. 278). In this problem, you will examine the data Chargaff collected in support of these conclusions.

In one paper, Chargaff (1950) described his analytical methods and some early results. Briefly, he treated DNA samples with acid to remove the bases, separated the bases by paper chromatography, and measured the amount of each base with UV spectroscopy. His results are shown in the three tables below. The *molar ratio* is the ratio of the number of moles of each base in the sample to the number of moles of phosphate in the sample—this gives the fraction of the total number of bases represented by each particular base. The *recovery* is the sum of all four bases (the sum of the molar ratios); full recovery of all bases in the DNA would give a recovery of 1.0.

Base	Molar ratios in ox DNA					
	Thymus			Spleen		Liver
	Prep. 1	Prep. 2	Prep. 3	Prep. 1	Prep. 2	Prep. 1
Adenine	0.26	0.28	0.30	0.25	0.26	0.26
Guanine	0.21	0.24	0.22	0.20	0.21	0.20
Cytosine	0.16	0.18	0.17	0.15	0.17	
Thymine	0.25	0.24	0.25	0.24	0.24	
Recovery	0.88	0.94	0.94	0.84	0.88	

Base	Molar ratios in human DNA				
	Sperm		Thymus	Liver	
	Prep. 1	Prep. 2	Prep. 1	Normal	Carcinoma
Adenine	0.29	0.27	0.28	0.27	0.27
Guanine	0.18	0.17	0.19	0.19	0.18
Cytosine	0.18	0.18	0.16		0.15
Thymine	0.31	0.30	0.28		0.27
Recovery	0.96	0.92	0.91		0.87

Base	Molar ratios in DNA of microorganisms		
	Yeast		Avian tubercle bacilli
	Prep. 1	Prep. 2	Prep. 1
Adenine	0.24	0.30	0.12
Guanine	0.14	0.18	0.28
Cytosine	0.13	0.15	0.26
Thymine	0.25	0.29	0.11
Recovery	0.76	0.92	0.77

- (a) Based on these data, Chargaff concluded that “no differences in composition have so far been found in DNA from different tissues of the same species.” This corresponds to conclusion 2 in this chapter. However, a skeptic looking at the data above might say, “They certainly look different to me!” If you were Chargaff, how would you use the data to convince the skeptic to change her mind?
- (b) The base composition of DNA from normal and cancerous liver cells (hepatocarcinoma) was not distinguishably different. Would you expect Chargaff’s technique to be capable of detecting a difference between the DNA of normal and cancerous cells? Explain your reasoning.

As you might expect, Chargaff’s data were not completely convincing. He went on to improve his techniques, as described in a later paper (Chargaff, 1951), in which he reported molar ratios of bases in DNA from a variety of organisms:

Source	A:G	T:C	A:T	G:C	Purine:pyrimidine
Ox	1.29	1.43	1.04	1.00	1.1
Human	1.56	1.75	1.00	1.00	1.0
Hen	1.45	1.29	1.06	0.91	0.99
Salmon	1.43	1.43	1.02	1.02	1.02
Wheat	1.22	1.18	1.00	0.97	0.99
Yeast	1.67	1.92	1.03	1.20	1.0
<i>Haemophilus influenzae</i> type c	1.74	1.54	1.07	0.91	1.0
<i>E. coli</i> K-12	1.05	0.95	1.09	0.99	1.0
Avian tubercle bacillus	0.4	0.4	1.09	1.08	1.1
<i>Serratia marcescens</i>	0.7	0.7	0.95	0.86	0.9
<i>Bacillus schatz</i>	0.7	0.6	1.12	0.89	1.0

- (c) According to Chargaff, as stated in conclusion 1 in this chapter, “The base composition of DNA generally varies from one species to another.” Provide an argument, based on the data presented so far, that supports this conclusion.
- (d) According to conclusion 4, “In *all* cellular DNAs, regardless of the species . . . $A + G = T + C$.” Provide an argument, based on the data presented so far, that supports this conclusion.

Part of Chargaff’s intent was to disprove the “tetranucleotide hypothesis”; this was the idea that DNA was a monotonous tetranucleotide polymer $(AGCT)_n$ and therefore not capable of containing sequence information. Although the data presented above show that DNA cannot be simply a tetranucleotide—if so, all samples would have molar ratios of 0.25 for each base—it was still possible that the DNA from different organisms was a slightly more complex, but still monotonous, repeating sequence.

To address this issue, Chargaff took DNA from wheat germ and treated it with the enzyme deoxyribonuclease for different time intervals. At each time interval, some of the DNA was converted to small fragments; the remaining, larger fragments he called the “core.” In the table below, the “19% core” corresponds to the larger fragments left behind when 81% of the DNA was degraded; the “8% core” corresponds to the larger fragments left after 92% degradation.

Base	Intact DNA	19% Core	8% Core
Adenine	0.27	0.33	0.35
Guanine	0.22	0.20	0.20
Cytosine	0.22	0.16	0.14
Thymine	0.27	0.26	0.23
Recovery	0.98	0.95	0.92

S-98 Chapter 8 Amino Acids, Peptides, and Proteins

- (e) How would you use these data to argue that wheat germ DNA is not a monotonous repeating sequence?

Answer

- (a) It would not be easy! The data for different samples from the same organism show significant variation, and the recovery is never 100%. The numbers for C and T show much more consistency than those for A and G, so for C and T it is much easier to make the case that samples from the same organism have the same composition. But even with the less consistent values for A and G, (1) the range of values for different tissues does overlap substantially; (2) the difference between different preparations of the same tissue is about the same as the difference between samples from different tissues; and (3) in samples for which recovery is high, the numbers are more consistent.
- (b) This technique would not be sensitive enough to detect a difference between normal and cancerous cells. Cancer is caused by mutations, but these changes in DNA—a few base pairs out of several billion—would be too small to detect with these techniques.
- (c) The ratios of A:G and T:C vary widely among different species. For example, in the bacterium *Serratia marcescens*, both ratios are 0.4, meaning that the DNA contains mostly G and C. In *Haemophilus influenzae*, by contrast, the ratios are 1.74 and 1.54, meaning that the DNA is mostly A and T.
- (d) Conclusion 4 has three requirements:
 A = T: The table shows an A:T ratio very close to 1 in all cases. Certainly, the variation in this ratio is substantially less than the variation in the A:G and T:C ratios.
 G = C: Again, the G:C ratio is very close to 1, and the other ratios vary widely.
 (A + G) = (T + C): This is the purine:pyrimidine ratio, which also is very close to 1.
- (e) The different “core” fractions represent different regions of the wheat germ DNA. If the DNA were a monotonous repeating sequence, the base composition of all regions would be the same. Because different core regions have different sequences, the DNA sequence must be more complex.

References

- Chargaff, E.** (1950) Chemical specificity of nucleic acids and mechanism of their enzymic degradation. *Experientia* **6**, 201–209.
- Chargaff, E.** (1951) Structure and function of nucleic acids as cell constituents. *Fed. Proc.* **10**, 654–659.