Structures of the Molecular Components in DNA and RNA with Bond Lengths Interpreted as Sums of Atomic Covalent Radii

Raji Heyrovska*

Institute of Biophysics of the Academy of Sciences of the Czech Republic, Královopolská 135, 61265 Brno, Czech Republic

Abstract: The author has found recently that the lengths of chemical bonds are sums of the covalent and or ionic radii of the relevant atoms constituting the bonds, whether they are completely or partially covalent or ionic. This finding has been tested here for the skeletal bond lengths in the molecular constituents of nucleic acids, adenine, thymine, guanine, cytosine, uracil, ribose, deoxyribose and phosphoric acid. On collecting the existing data and comparing them graphically with the sums of the appropriate covalent radii of C, N, O, H and P, it is found that there is a linear dependence with effectively unit slope and zero intercept. This shows that the bond lengths in the above molecules can be interpreted as sums of the relevant atomic covalent radii. Based on this result, the author has presented here the atomic structures of the above molecules in terms of the atomic radii (for the first time).

INTRODUCTION

Following the exciting discovery [1] of the molecular structure of nucleic acids, the question of the skeletal bond lengths in the molecules constituting DNA and RNA has continued to be of interest. The author was enthused to undertake this work by the recent findings [2a, b] that the length of the chemical bond between any two atoms or ions is, in general, the sum of the radii of the atoms and or ions constituting the bond. The additivity of radii was also found to hold for the hydration bonds [3a] with ions in aqueous solutions as well as for the lengths of the hydrogen bonds [3b] in the Watson-Crick [1b] base pairs of nucleic acids and in many other inorganic and biological compounds.

Pauling [4] dealt with the bond lengths in the purines: adenine (A) and guanine (G) and the pyrimidines: thymine (T), cytosine (C) and uracil (U), but he did not interpret the bond lengths in these molecules in terms of the covalent radii of the atoms, although he considered covalent bond lengths as sums of the covalent radii of the atoms. This article shows that the existing bond lengths [4-7] (see Tables **1** and **2**) in the above molecules and in ribose, deoxyribose and phosphoric acid, abbreviated here as Ri, De and Ph respectively, correspond to the sums of the appropriate covalent radii of the five atoms, carbon, nitrogen, hydrogen, oxygen and phosphorus.

ATOMIC COVALENT RADII AND THE CN BOND LENGTH IN ADENINE

It was noticed by Pauling [4] that in all the bases, A, T, C, G and U, (see Fig. (**1**)), the CC bond length is around 1.40 Å and the CN bond length has values around 1.32 Å and 1.36 Å. A survey of the existing bond length data [4-7] (see Tables **1** and **2**), shows that the CC, CN CH, NH, CO, OH and PO bonds in the above molecules do have some essentially fixed bond lengths, in addition to the three considered by Pauling.

For the interpretation here of these bond lengths, the atomic covalent radii, defined [4a] as $R_{cov} = d(A) = d(AA)/2$, where d(AA) is the covalent bond length between two atoms (A) of the same kind, have been used. In [4b], these radii are referred to as bonding atomic radii. Since R_{cov} is half the inter-atomic distance, it is actually a distance although written with an R. Considering carbon, Pauling [4] points out two types of single bond radii for C: the aliphatic single bond covalent radius of 0.77 Å (as in cubic diamond), and the aromatic single bond (involved in resonance) covalent radius of 0.72 Å (as in hexagonal graphite with delocalized charge). The covalent double bond radius of C, defined [4] as half the C=C double bond length, is 0.67 Å. Note that 0.72 = (0.77 + 0.67)/2, as if it is a midway 1.5 bond! In general, the covalent double bond radii of atoms are less than those for single bonds [4]. The above three C (with the subscripts, s.b: single bond, res: aromatic resonance bond and d.b: double bond) are represented by circles of radii R_{cov} in Fig. (**2**). Similarly, in Fig. (**2**), R_{cov} for $N_{s.b}$ & $N_{d.b}$ and $O_{s.b}$ & $O_{d.b}$ are halves of the single bond and double bond lengths of N and O respectively, R_{cov} for H is half the HH single bond distance [4] and R_{cov} for P was obtained by subtracting R_{cov} of O(I) from the PO(I) bond length [4-7].

An interesting observation from the bond length data in [4-7] assembled in Table **1** is e.g., that the CN bond lengths in the aromatic ring in adenine (see col. 2, Table **1**) are the same [1.34 (+/-) 0.02Å] irrespective of whether they pertain to single (N1-C2 and N3-C4) or double (C2-N3 and C6-N1) bonds. This is similar to the finding in [4] that in the case of benzene, all the six CC bonds are of length 1.39 (+/-) 0.01 Å (with equal assigned [4] bond order of 1.5, due to resonance), although represented by three alternating single and double bonds. This CC bond length in the aromatic ring was interpreted [2a] as the sum (0.72 + 0.67 Å) of the radii for C_{res} and $C_{d.b}$ (see Fig. (**2**)). Similarly, in adenine, the CN bond distance of 1.34 Å is the sum of R_{cov} of C_{res} (= 0.72) and $N_{d.b}$ (= 0.62).

*Address correspondence to this author at the Institute of Biophysics of the Academy of Sciences of the Czech Republic, Královopolská 135, 61265 Brno, Czech Republic; E-mail: rheyrovs@hotmail.com

1874-1991/08

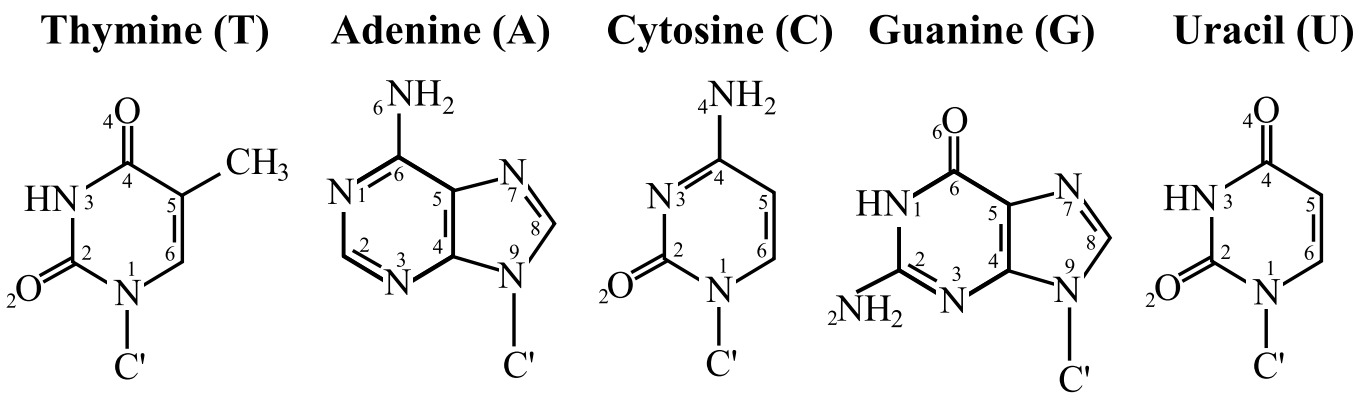

Fig. (1). Molecular structures [7] of thymine (T), adenine (A), cytosine (C), guanine (G) and uracil (U). Bond length data [4-7] in Tables **1** and **2** correspond to distances between two adjacent (numbered) atoms.

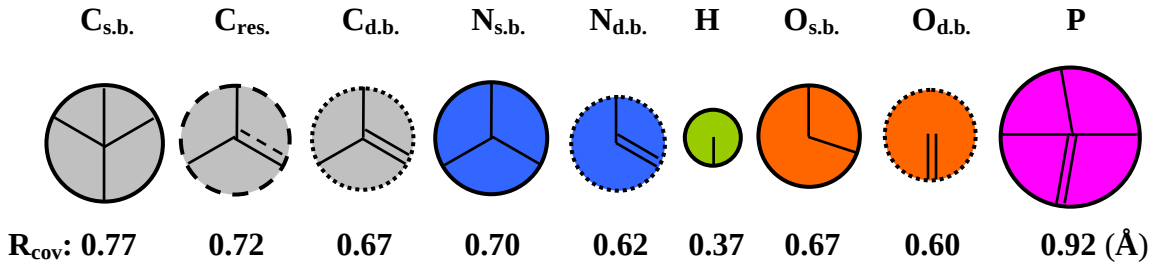

Fig. (2). The (nine) covalent radii R_{cov} (+/- 0.02Å) [4] of the five atoms constituting the molecular components of nucleic acids. C: gray, N: blue, O: orange, H: green and P: pink. Nature makes unique combinations of these five atoms (from mono to pentavalent) into the molecules of life (see Figs. (**4-6**)).

BOND LENGTHS IN ALL THE ABOVE MOLECULES AS SUMS OF ATOMIC COVALENT RADII

The above additivity of atomic radii was then tested for the distances between any two atoms in the skeletal structures of the molecular components of nucleic acids by comparing the data from [4-7] with the sums R(sum) of the appropriate atomic covalent radii in Fig. (**2**). The data for A, G, T and C are given in Table **1** and for U, Ri, De and Ph in Table **2**. The average values of the bond lengths [4-7] for the above molecules (from Tables **1** and **2**) and the corresponding R(sum) for the various bonds are tabulated in Table **3**. On plotting these bond lengths *vs* R(sum) as shown in Fig. (**3**), and drawing a least square straight line through all the data points, it was found to have a slope of 1.01 and an intercept of -0.02. On using this slope and intercept, the calculated bond lengths (given in the last column in Table **3**) are found to be identical with R(sum) (col. 2, Table **3**). *This shows that the bond lengths in all the above molecules can be considered as the sums of the relevant atomic covalent radii.*

Note also that since this graph includes the theoretical values of bond lengths from [6], the above result shows that the latter values are good representations of the radii sum, R(sum).

The above finding has thus enabled the conventional molecular structures (see Fig. (**1**)) to be resolved (for the first time) into the atomic structures based on the individual covalent radii of the constituent atoms. Fig. (**4**) shows the atomic structures of the bases: T (U), A, C and G and Fig. (**5**), those for phosphoric acid, deoxyribose (and ribose). For these molecules, while confirming Pauling's [4] two CN bonds of lengths 1.32 and 1.36 (+/- 0.02 Å) and one CC bond length of 1.40 Å, it is found here that actually there are five CN bonds, three CC bonds, two CO bonds and two PO bonds (see Table **3**).

The bond lengths of C, O and N with H (R_{cov} = 0.37 Å, see Fig. (**2**)) are also given in Table **3**. Note that whereas the increase in the CH bond length in going from aromatic to aliphatic C is attributed by Pauling [4] to the change in the radius of H, here it is accounted for by those of C.

Fig. (**6**) shows how a combination of all the molecules in Figs. (**4**,**5**) fits into the known 34:20 Å dimension of the DNA nucleotides and a detailed description is given in the caption for this figure.

Support for the additivity of covalent atomic radii in bond lengths has been found also for the bonds in xanthine, caffeine and related compounds [3c].

Thus, it is concluded here that the skeletal bond lengths in the conventional molecular structures of the components of DNA and RNA are composed of the covalent radii (see Fig. (**2**)) of the two adjacent atoms constituting the bonds (see Figs. (**4**,**5**)).

ACKNOWLEDGEMENTS

The author is grateful to Prof. E. Palecek of the Institute of Biophysics for the moral support, and for the grants AVOZ50040507 of the Academy of Sciences of the Czech Republic and LC06035 of the Ministry of Education, Youth and Sports of the Czech Republic.

Table 1. Bond Lengths (Å) (+/- 0.02Å) [4-7] of Adenine (A), Guanine (G), Thymine (T) and Cytosine (C), Compared with Sums of Atomic Covalent Radii, R(sum) (Upper Values, in Bold, See Also Table 3 and Fig. (4))

Adenine	Bond Lengths	Guanine	Bond Lengths	Thymine	Bond Lengths	Cytosine	Bond Lengths
N1-C2	**1.34 = 0.62 + 0.72**	N1-C2	**1.37 = 0.67 + 0.70**	N1-C2	**1.37 = 0.67 + 0.70**	N1-C2	**1.37 = 0.67 + 0.70**
*	1.34,1.34,1.32,1.34		1.32,1.37,1.34		1.38,1.38,1.36		1.41,1.40,1.36
C2-N3	**1.34 = 0.62 + 0.72**	C2-N3	**1.29 = 0.67 + 0.62**	C2-N3	**1.37 = 0.67 + 0.70**	C2-N3	**1.37 = 0.67 + 0.70**
*	1.34,1.33,1.32,1.34		1.36,1.32,1.32		1.38,1.37,1.36		1.37,1.35,1.34
N3-C4	**1.34 = 0.62 + 0.72**	N3-C4	**1.29 = 0.67 + 0.62**	N3-C4	**1.37 = 0.67 + 0.70**	N3-C4	**1.29 = 0.67 + 0.62**
*	1.34,1.34,1.36,1.34		1.42,1.35,1.32		1.40,1.38,1.36		1.31,1.34,1.32
C4-C5	**1.39 = 0.67 + 0.72**	C4-C5	**1.39 = 0.67 + 0.72**	C4-C5	**1.39 = 0.67 + 0.72**	C4-C5	**1.39 = 0.67 + 0.72**
*	1.40,1.38,1.40,1.44		1.44,1.38,1.38		1.46,1.45,1.40		1.44,1.43,1.40
C5-C6	**1.39 = 0.67 + 0.72**	C5-C6	**1.39 = 0.67 + 0.72**	C5-C6	**1.39 = 0.67 + 0.72**	C5-C6	**1.39 = 0.67 + 0.72**
*	1.39,1.41,1.40,1.38		1.39,1.42,1.40		1.35,1.34,1.40		1.36,1.34,1.40
C6-N1	**1.34 = 0.62 + 0.72**	C6-N1	**1.37 = 0.67 + 0.70**	C6-N1	**1.37 = 0.67 + 0.70**	C6-N1	**1.37 = 0.67 + 0.70**
*	1.34,1.35,1.32,1.34		1.36,1.39,1.36		1.37,1.38,1.36		1.35,1.37,1.34
C5-N7	**1.29 = 0.67 + 0.62**	C5-N7	**1.34 = 0.62 + 0.72**	C2-O2	**1.27 = 0.67 + 0.60**	C2-O2	**1.27 = 0.67 + 0.60**
*	1.37,1.39,1.32,1.32		1.37,1.39,1.32		1.22,1.22,1.23		1.22,1.24,1.23
N7-C8	**1.29 = 0.67 + 0.62**	N7-C8	**1.29 = 0.67 + 0.62**	C4-O4	**1.27 = 0.67 + 0.60**	C4-N4	**1.37 = 0.67 + 0.70**
*	1.32,1.31,1.32,1.28		1.32,1.31,1.32		1.22,1.23,1.23		1.36,1.34,1.35
C8-N9	**1.37 = 0.67 + 0.70**	C8-N9	**1.37 = 0.67 + 0.70**	C5-M5	**1.49 = 0.72 + 0.77**	N1-C1'	**1.47 = 0.70 + 0.77**
*	1.37,1.37,1.33,1.37		1.37,1.37,1.32		1.49,1.50,1.53		-,1.47,1.53
N9-C4	**1.42 = 0.70 + 0.72**	N9-C4	**1.37 = 0.67 + 0.70**	N1-C1'	**1.47 = 0.70 + 0.77**		
*	1.38,1.37,1.36, -		1.37,1.38,1.34		-,1.47,1.53		
C6-N6	**1.42 = 0.70 + 0.72**	C2-N2	**1.37 = 0.67 + 0.70**				
*	-,1.34,1.35,1.34		1.38,1.34,1.35				
N9-C1'	**1.47 = 0.70 + 0.77**	C6-O6	**1.27 = 0.67 + 0.60**				
*	-,1.46,1.53,-		1.22,1.24,1.23				
		N9-C1'	**1.47 = 0.70 + 0.77**				
*			-,1.46,1.53				
*Values:		*Values:		*Values:		*Values:	
Refs. 6,7,4,5		Refs. 6,7,4		Refs. 6,7,4		Refs. 6,7,4	

The numberings of the atoms are as in [7].

Table 2. Bond Lengths (+/- 0.02Å) [4-7] of Uracil (U), Ribose (Ri), Deoxyribose (De), Phosphoric Acid/Phosphate (Ph), Compared with Sums of Atomic Covalent Radii, R(sum) (Upper Values, in Bold, See Also Table 3 and Figs. (4,5)

Uracil	Bond Length	Ribose, Deoxyribose		**Phosphoric Acid, Phosphate	
		Bonds	Bond Length	Bonds	Bond Length
N1-C2	**1.37 = 0.67 + 0.70**	C1'-C2'	**1.54 = 0.77 + 0.77**	**P=O	**1.52 = 0.92 + 0.60**
*	1.38,1.34		1.53,1.52		1.49,1.52
C2-N3	**1.37 = 0.67 + 0.70**	C2'-C3'	**1.54 = 0.77 + 0.77**	**P-O(H)	**1.59 = 0.92 + 0.67**
*	1.37,1.38		1.53,1.52		1.61,1.57
N3-C4	**1.37 = 0.67 + 0.70**	C3'-C4'	**1.54 = 0.77 + 0.77**	P-O3'	**1.59 = 0.92 + 0.67**
*	1.38,1.37		1.52,1.53		1.61,1.56
C4-C5	**1.39 = 0.67 + 0.72**	C4'-O4'	**1.44 = 0.77 + 0.67**	P-O5'	**1.59 = 0.92 + 0.67**
*	1.43,1.41		1.45,1.45		1.59,1.56
C5-C6	**1.39 = 0.67 + 0.72**	O4'-C1'	**1.44 = 0.77 + 0.67**	(P)O5'-C5'	**1.44 = 0.77 + 0.67**
*	1.34,1.41		1.41,1.42		1.44,-
C6-N1	**1.37 = 0.67 + 0.70**	C3'-O3'	**1.44 = 0.77 + 0.67**	(P)O3'-C3'	**1.44 = 0.77 + 0.67**
*	1.38,1.34		1.42,1.43		1.43,-
C2-O2	**1.27 = 0.67 + 0.60**	C5'-C4'	**1.54 = 0.77 + 0.77**		
*	1.22,1.23		1.51,1.51		
C4-O4	**1.27 = 0.67 + 0.60**	C2'-O2'	**1.44 = 0.77 + 0.67**		
*	1.23,1.24		1.41,-		
N1-C1'	**1.47 = 0.70 + 0.77**	C1'-N1/N9	**1.47 = 0.70 + 0.77**		
*	1.47,-		1.47,1.47		
		O5'-C5'	**1.44 = 0.77 + 0.67**		
*			1.42,1.42		
*Values:		*Values: ribose, deoxyribose		*Values:	
Refs. 7,4		Ref. 7		Refs. 7,4	

The numberings of the atoms are as in [7].

Table 3. Average Bond Lengths (+/- 0.03 Å) [4-7] Corresponding to Sums of Atomic Covalent Radii, R(sum)

Bonds	R(sum)	Average Bond Lengths from Tables 1 and 2							Calculated
		A	G	T	C	U	Ri, De	Ph	Bond Length
$C_{d.b}$ - $O_{d.b}$	**1.27**		1.24	1.23	1.23	1.23			**1.27**
$C_{d.b}$ - $N_{d.b}$	**1.29**	1.33	1.32		1.32				**1.29**
C_{res} - $N_{d.b}$	**1.34**	1.34	1.37						**1.34**
$C_{d.b}$ - $N_{s.b}$	**1.37**	1.36	1.36	1.37	1.36	1.37			**1.37**
C_{res} - $C_{d.b}$	**1.39**	1.4	1.39	1.4	1.44	1.4			**1.39**
C_{res} - $N_{s.b}$	**1.42**	1.35							**1.42**
$C_{s.b}$ - $O_{s.b}$	**1.44**						1.42	1.44	**1.44**
$C_{s.b}$ - $N_{s.b}$	**1.47**	1.5	1.5	1.45	1.5	1.47	1.47		**1.47**
$C_{s.b}$ - C_{res}	**1.49**			1.48					**1.49**
P - $O_{d.b}$	**1.52**							1.51	**1.52**
$C_{s.b}$ - $C_{s.b}$	**1.54**						1.52		**1.54**
P - $O_{s.b}$	**1.59**							1.58	**1.59**
$C_{d.b}$ - H	**1.04** (A,T,U,C,G)								
C_{res} - H	**1.09** (A,U,C)								
$C_{s.b}$ - H	**1.14** (T,Ri,De)								
$N_{s.b}$ - H	**1.07** (A,T,U,C,G)								
$O_{s.b}$ - H	**1.04** (Ri,DE,Ph)								

Last Column: Values Using the Least Square Line in Fig. (**3**).

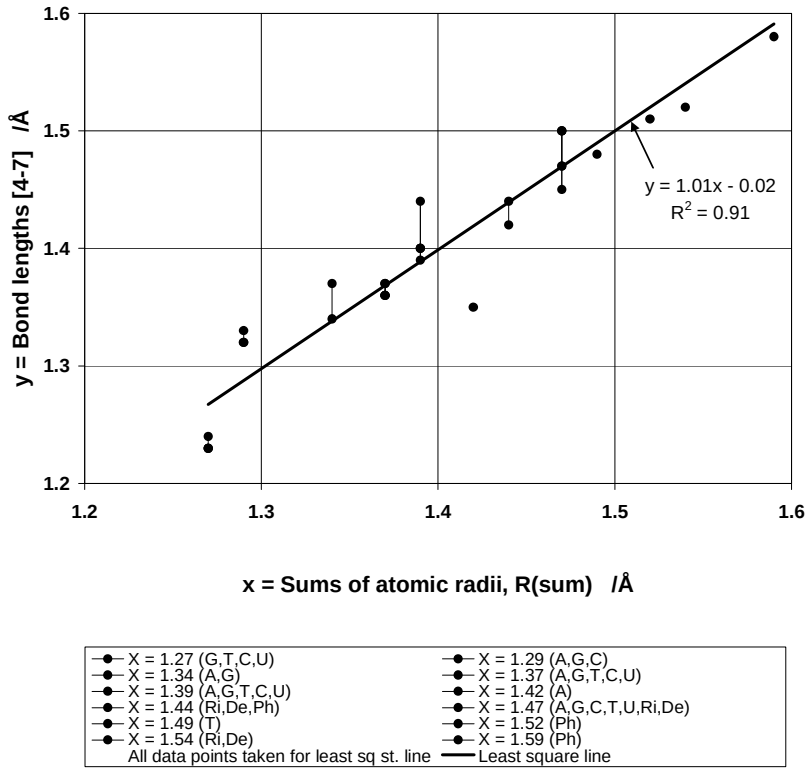

Fig. (3). Graph of the bond lengths from [4-7] *vs* the sums of atomic covalent radii, R(sum), data in Table **3**. The unit slope and zero intercept of the least squares line show that the bond lengths are equal to R(sum).

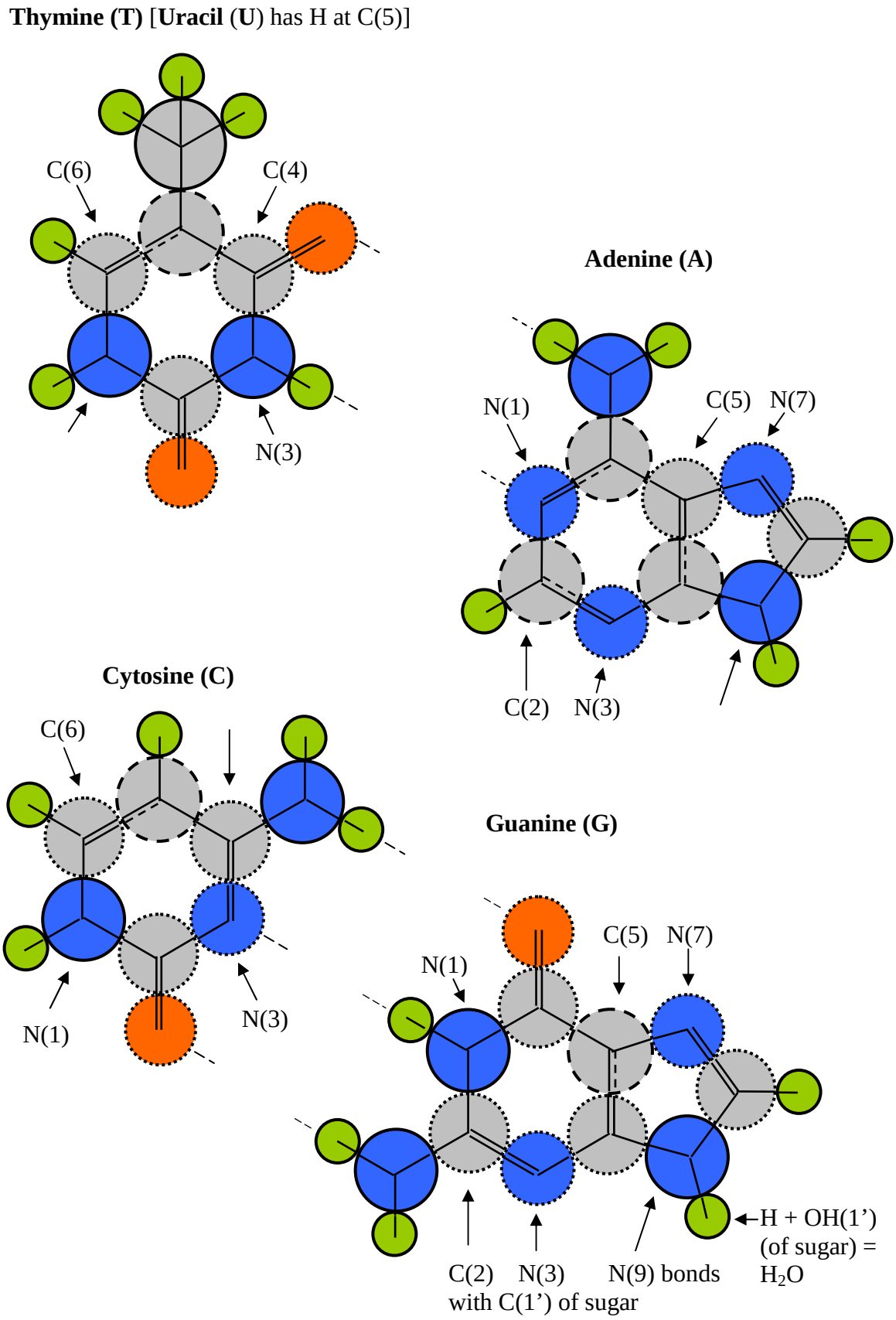

Fig. (4). Atomic structures of the purines: A and G and pyrimidines: T, (U) and C. All bond lengths are sums of the covalent radii of the adjacent atoms (see Table **3**). Base pairing [1a, b] occurs when hydrogen bonds connect T with A, and C with G along the broken lines, with bond lengths as in [3b].

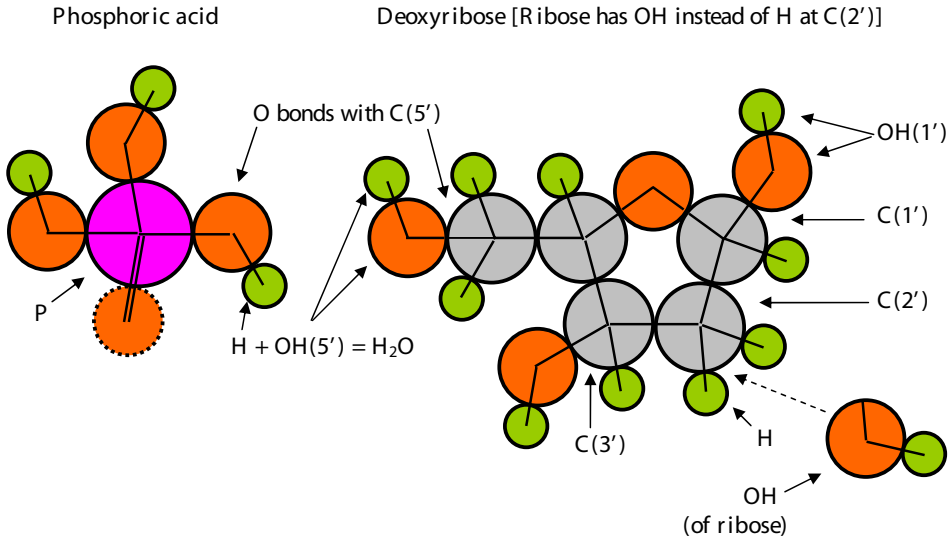

Fig. (5). Atomic structures of phosphoric acid and deoxyribose (and ribose). All bond lengths are sums of the covalent radii of the adjacent atoms (see Table **3**). By the elimination of water molecules as shown, the bases combine with sugars and the latter with phosphoric acid to form nucleotides [4-7].

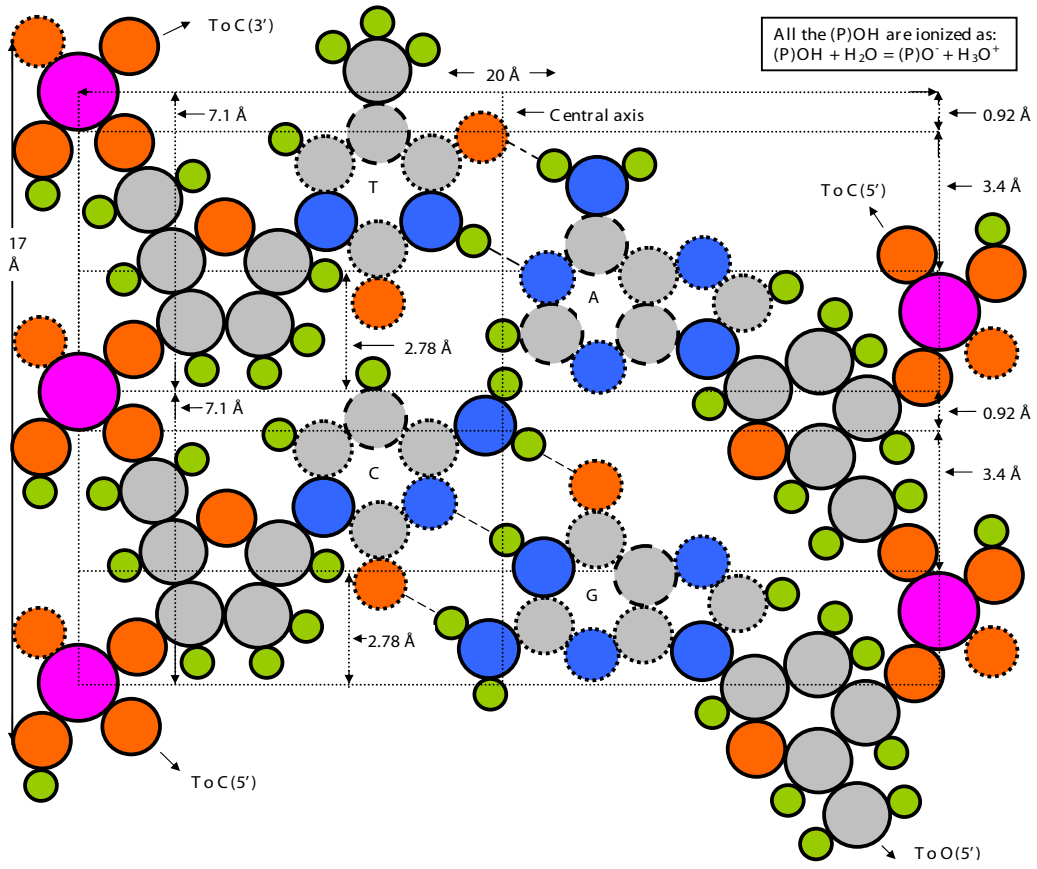

Fig. (6). Atomic structures of the DNA nucleotides with the Watson-Crick base pairs (T, A and C, G connected by H-bonds as shown) attached to deoxyribose and phosphate molecules. As per the known crystallographic distances [1c], the two P atoms (pink) attached *via* O(5') and O(3') (orange) to the sugar molecules are 7.1 Å apart and the distance of each P atom from the central axis is 10 Å. The base pairs are perpendicular to the plane of the sugar-phosphate back bone and form the steps of the ladder of the double helix. The steps are 3.4 Å apart [1c]. The distance of 17 Å containing 5 P atoms is half of the 34 Å turn (with 10 P atoms) of the helix and the distance 2.78 Å can be considered as the thickness of the steps. (For the colors, radii and bonds of the atoms see Figs. (**2**,**4**,**5**)).

Note: The above Figure holds also for RNA, but with Uracil (U) and Ribose in place of Thymine (T) and deoxyribose respectively.

REFERENCES

[1] Watson JD, Crick FHC. a) A structure for deoxyribonucleic acid. Nature 1953; 171: 737-738 and b) Genetic implications of the structure of deoxyribonucleic acid. Nature 1953; 171: 964-967, c) Franklin RE, Gosling RG. Molecular configuration in sodium thymonucleate. Nature 1953; 171: 740-741; see for full papers: http://www.nature.com/nature/dna50/archive.html

[2] Heyrovska R. a) The Golden ratio, ionic and atomic radii and bond lengths. Mol. Phys. 2005; 103: 877-882 and b) The Golden ratio, atomic, ionic and molecular capacities and bonding distances in metallic hydrides**.** International Joint meeting of ECS, USA and Japanese, Korean and Australian Societies, Honolulu, Hawaii, 2004; Extd. Abs. C2-0551. http://www.electrochem.org/dl/ma/206/pdfs/0551.pdf.

[3] Heyrovska R. a) Dependences of ion-water distances on covalent radii, ionic radii in water and distances of oxygen and hydrogen of water from ion/water boundaries. Chem Phys Lett 2006; 429: 600-605, b) Dependence of the length of the hydrogen bond on the covalent and cationic radii of hydrogen, and additivity of bonding distances. Chem Phys Lett 2006; 432: 348-351. c) Heyrovska R, Narayan S. 10th Eurasia conference on chemical sciences, Manila, Philippines, 7-11 January 2008, Book of Abstracts, p. 330.

[4] a) Pauling L. The Nature of the Chemical Bond. Cornell Univ. Press, Ithaca, New York, 1960; b) http://wps.prenhall.com/wps/media/objects/3311/3390919/blb0702.html

[5] Lehninger AL. Biochemistry. Worth Publishers, Inc., New York, 1975.

[6] Preuss M, Schmidt G, Seino K, *et al.* Ground- and excited-state properties of DNA base molecules from plane-wave calculations using ultrasoft psuedopotentials. J Comput Chem 2004; 25: 112-122.

[7] Nucleic acid database. http://ndbserver.rutgers.edu.

Received: October 19, 2007 Revised: December 7, 2007 Accepted: December 10, 2007