*INTERACTION ENERGIES IN HYDROGEN-BONDED PURINE-PYRIMIDINE TRIPLETS**

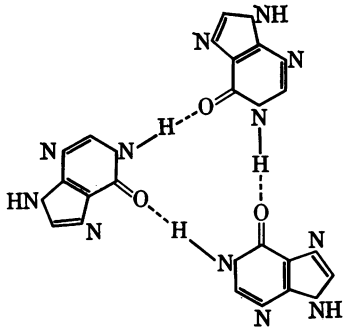
By Bernard Pullman, Pierre Claverie, and Jacqueline Caillet

INSTITUT DE BIOLOGIE PHYSICO-CHIMIQUE, UNIVERSITÉ DE PARIS, FRANCE

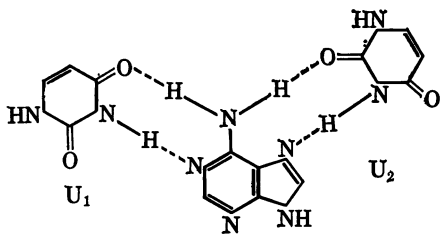
Communicated by Albert Szent-Györgyi, April 3, 1967

We have recently studied the Van der Waals-London interaction energies in hydrogen-bonded purine and pyrimidine base pairs.[1-5] In the present paper we are extending our previous studies to hydrogen-bonded purine and pyrimidine base-triplets.

The known triplets can be subdivided roughly into three groups: (*a*) Cyclic homo-triplets such as the triplet (I) of hypoxanthines that exists most probably in the three-stranded polyinosinic acid.[6] An analogous triplet of xanthines has been considered for polyxanthylic acid.[7] A somewhat similar but higher aggregate (a tetramer involving, moreover, two hydrogen bonds between each pair of bases) has been postulated to occur in gels of guanylic acid.[8] (*b*) Open hetero-triplets in which a central base is linked by two or (possibly) three hydrogen bonds to two complementary ones. This type of arrangement is exemplified by the triplets uracil-adenine-uracil (II or III) postulated to exist in the three-stranded helix poly (A + 2U),[9, 10] or by the triplet hypoxanthine-adenine-hypoxanthine (IV) postulated to exist in poly (A + 2I).[11] (*c*) Open triplets similar to the preceding ones, but in which the triplet can only be formed if one of the bases is *protonated*. Such a structure, exemplified by the combinations cytosine-guanine-cytosine (V) or (VI), has been postulated to exist in the three-stranded helix formed between poly C and guanosine mononucleotides and oligonucleotides.[12, 13]



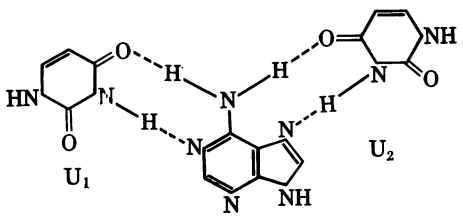
I



U_1 U_2

A

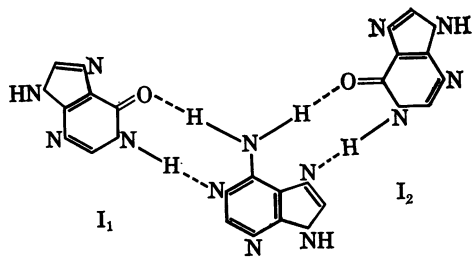
II



U_1 U_2

A

III



I_1 I_2

A

IV

It may be mentioned that besides their presence in the structure of observable three-stranded helices, similar triplets have been postulated to be involved as intermediates in certain biological processes, e.g., in gene repression[14] or in the interaction of RNA's with ribosomes.[15, 16]

Method.—The method of calculation is essentially similar to the one described in detail in our previous publications on base-pairs (see in particular ref. 1). The energies of interaction (E_M) are built up of three components: the electrostatic monopole-monopole energy ($E_{\rho\rho}$), the induction or polarization monopole-induced dipole energy ($E_{\rho\alpha}$), and the dispersion or London energy (E_L). It is necessary, however, to point out that for more than two interacting molecules a new feature appears, namely the nonpairwise additivity of the polarization energy. A quantum-mechanical perturbation treatment shows that while the electrostatic and dispersion energies are the sum of the pairwise contributions, this is no longer true for the polarization energy, and that this last nonadditivity is exactly the same as that which can be predicted from classical electrostatics.

In order to make explicit the magnitude of this effect, we shall consider three molecules placed at the vertices of an equilateral triangle, each of them being related to the others by a rotation of $2\pi/3$ around the center O (Fig. 1). In order to easily obtain a qualitative picture of the phe-

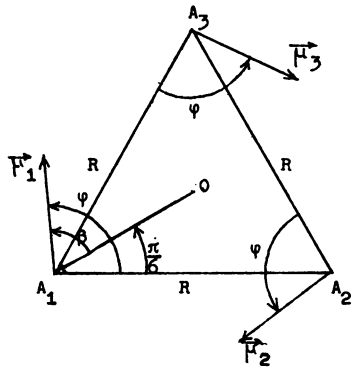


FIG. 1.—The geometry of a cyclic triplet.

nomenon, we shall discuss it in the dipole approximation. The geometry of the system is defined by $R = A_1A_2 = A_2A_3 = A_3A_4$ and the angle $\varphi = (\mathbf{A_1A_2}, \boldsymbol{\mu}_1)$. $\beta = \varphi - \pi/6$ will also be used. Moreover, $|\boldsymbol{\mu}_1| = |\boldsymbol{\mu}_2| = |\boldsymbol{\mu}_3| = \mu$. The electric fields created by $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ at the point A_3 are:

$$\mathbf{E}_{13} = \frac{1}{R^3}[\boldsymbol{\mu}_1 - 3(\boldsymbol{\mu}_1 \cdot \mathbf{u}_{13})\mathbf{u}_{13}], \qquad \mathbf{u}_{13} = \frac{\mathbf{A_1A_3}}{R}. \tag{1}$$

$$\mathbf{E}_{23} = \frac{1}{R^3}[\boldsymbol{\mu}_2 - 3(\boldsymbol{\mu}_2 \cdot \mathbf{u}_{23})\mathbf{u}_{23}], \qquad \mathbf{u}_{23} = \frac{\mathbf{A_2A_3}}{R}. \tag{2}$$

The addition of pairwise contributions would give, for the polarization energy of the molecule A_3, the sum: $E_{13} + E_{23}$, where $E_{13} = {}^1/_2\,\alpha\,\mathbf{E}_{13}^2$ and $E_{23} = {}^1/_2\,\alpha\,\mathbf{E}_{23}^2$ (α being the isotropic polarizability).

The true polarization energy E_3 of A_3 corresponds to the true inducing field, namely $\mathbf{E}_{13} + \mathbf{E}_{23}$, therefore

$$E_3 = \frac{1}{2}\,\alpha\,(\mathbf{E}_{13} + \mathbf{E}_{23})^2.$$

Let us introduce the ratio $\rho = \dfrac{E_3}{E_{13} + E_{23}}$ of the true energy to the pairwise energy (since this ratio is the same for the three molecules, it is sufficient to define it for the energies related to one molecule).

$$\rho = \frac{(\mathbf{E}_{13} + \mathbf{E}_{23})^2}{\mathbf{E}_{13}^2 + \mathbf{E}_{23}^2} = \frac{\mathbf{E}_{13}^2 + \mathbf{E}_{23}^2 + 2\,\mathbf{E}_{13}\cdot\mathbf{E}_{23}}{\mathbf{E}_{13}^2 + \mathbf{E}_{23}^2} = 1 + \frac{2\,\mathbf{E}_{13}\cdot\mathbf{E}_{23}}{\mathbf{E}_{13}^2 + \mathbf{E}_{23}^2}. \quad (3)$$

Replacing $\mathbf{E}_{13}$ and $\mathbf{E}_{23}$ according to (1) and (2), ρ may be expressed as a function of φ only (after an evident simplification by $(1/R^3)^2$, the scalar products are expressed in terms of μ and φ, and a further simplification by μ^2 is possible). The final result is:

$$\rho(\varphi) = 1 + \frac{2 + 6\cos\varphi\cos(\pi/3 - \varphi)}{17/4 + 3\cos\varphi\cos(\pi/3 - \varphi)} = 3 - \frac{26}{17 + 12\cos\varphi\cos(\pi/3 - \varphi)}$$

$$= 3 - \frac{26}{20 + 6\cos(2\varphi - \pi/3)} = 3 - \frac{26}{20 + 6\cos 2\beta}.$$

The function $\rho(\beta) = 3 - \frac{26}{20 + 6\cos 2\beta}$ is represented in Figure 2.

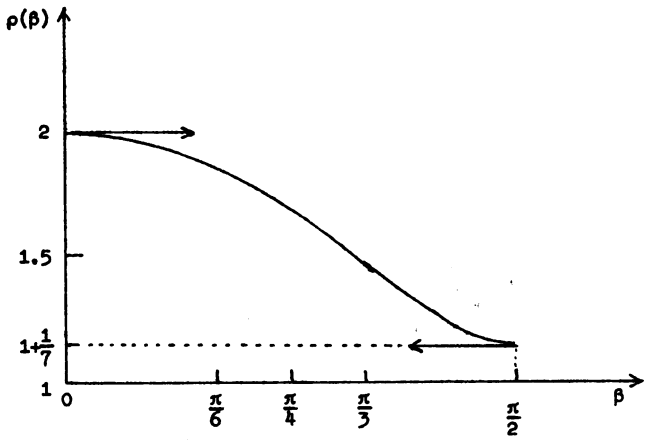


FIG. 2.—ρ as a function of β.

More generally, it appears from equation (3) that $0 \leqslant \rho \leqslant 2$ for every possible geometry, since

$$2 - \rho = \frac{\mathbf{E}_{13}^2 + \mathbf{E}_{23}^2 - 2\,\mathbf{E}_{13}\,\mathbf{E}_{23}}{\mathbf{E}_{13}^2 + \mathbf{E}_{23}^2} = \frac{(\mathbf{E}_{13} - \mathbf{E}_{23})^2}{\mathbf{E}_{13}^2 + \mathbf{E}_{23}^2} \geqslant 0.$$

Therefore, as long as the polarization energies are small with respect to the other energy differences which play a role in the problem, their nonadditivity is not expected to modify the *qualitative* picture, or even to modify the quantitative picture to any important degree.

As a numerical example, let us consider the triplet of hypoxanthines (I). The three contributions to the interaction in one pair are (in kcal/mole) (Table 1): $E_{\rho\rho} = -5.80$, $E_{\rho\alpha} = -0.27$, $E_L = -0.16$, hence $E_{\text{total}} = -6.23$. Here, $\beta \simeq 113°$ gives $\rho(\beta) \simeq 1.36$, hence, a corrected polarization energy $E_{\rho\alpha}^{\text{cor}} \simeq 0.38$ (on admitting, as it seems reasonable, that the factor ρ holds also for the "monopole" approximation) and $E_{\text{total}}^{\text{cor}} \simeq -6.34$, the relative error being about 1.7% for a difference of 0.11 kcal/mole. This cannot affect the comparison with, for example, the -5.42 kcal/mole corresponding to the interaction energy in a doubly bonded hypoxanthine pair (VII) (see next section).

For this reason, although the existence of such a correction must in principle be borne in mind, we have neglected it in the calculations described below. It must also be remembered that the calculations correspond to base-associations *in vacuo*.

Results and Discussion.—The results of the calculations are summed up in Table 1, and their examination leads to the following conclusions.

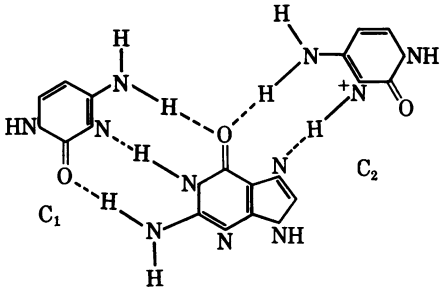
(1) *The triplet of hypoxanthines (poly I):* The total energy of interaction in the triplet is −18.69 kcal/mole, evenly distributed (−6.23 kcal/mole) per hydrogen bond or per hypoxanthine present. It may be particularly significant to compare these values with those predictable for a hypothetical hypoxanthine-hypoxanthine base pair (VII) linked together by two hydrogen bonds, which represents probably the most stable such a pair. This energy would be equal to −10.83 kcal/mole, which represent −5.42 kcal/mole per hypoxanthine ring. *It therefore appears that in this case the formation of the triplet represents a more advantageous organization by about* 1.2 *kcal per base.*

TABLE 1

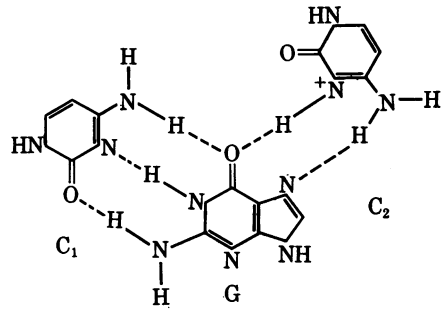
INTERACTION ENERGIES (KCAL/MOLE)

Base-triplet (or pair)	Interaction	$E_{\rho\rho}$	$E_{\rho\alpha}$	E_L	E_M
I	Per pair of bases	−5.79	−0.27	−0.16	−6.23
	Total	−17.37	−0.81	−0.48	−18.69
II	A — U_1	−4.64	−0.25	−0.69	−5.58
	A — U_2	−5.63	−0.17	−0.94	−6.74
	U_1 — U_2	+0.57	−0.01	−0.03	+0.53
	Total	−9.70	−0.43	−1.66	−11.79
III	A — U_1	−4.64	−0.25	−0.69	−5.58
	A — U_2	−5.86	−0.22	−0.88	−6.96
	U_1 — U_2	+0.96	−0.02	−0.03	+0.91
	Total	−9.54	−0.49	−1.60	−11.63
IV	A — I_1	−6.45	−0.29	−0.37	−7.11
	A — I_2	−4.90	−0.34	−0.41	−5.65
	I_1 — I_2	+2.05	−0.05	−0.03	+1.97
	Total	−9.30	−0.68	−0.81	−10.79
V	G — C_1	−15.91	−2.02	−1.25	−19.18
	G — C_2	−26.20	−2.85	−0.30	−29.35
	C_1 — C_2	+1.26	−0.53	−0.06	+0.67
	Total	−40.85	−5.40	−1.61	−47.86
VI	G — C_1	−15.91	−2.02	−1.25	−19.18
	G — C_2	−25.10	−2.77	−0.30	−28.17
	C_1 — C_2	+0.49	−0.47	−0.07	−0.05
	Total	−40.52	−5.26	−1.62	−47.40
VII		−10.20	−0.31	−0.32	−10.83
VIII		−12.28	−1.24	−0.72	−14.24
IX		−26.52	−7.20	−1.17	−34.89

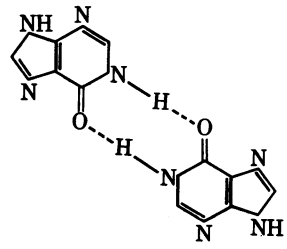
(2) *The triplet uracil-adenine-uracil (poly (A + 2U)):* The calculations concerning these triplets (and also the remaining open triplets) have been decomposed into their constituent elements, representing the partial interactions between each pair of bases. It is interesting to observe that while the interactions between the linked bases correspond to attractions (equal in our approximation to the attractions between the corresponding isolated pairs), *the interaction between the nonlinked terminal bases introduces a repulsion.* Although the value of this repulsive term is relatively small, it seems to have in this particular case an important structural consequence. Thus, it may be noted that if this repulsive term was neglected, configuration III would be favored over configuration II, this result being due to the greater energy of interaction between A and U_2 in the arrangement adopted for this interaction in III, and which is the arrangement observed in cocrystals of adenine and uracil[17, 18] (it may be added that the arrangement adopted in II is observed in cocrystals of adenine and bromouracil).[19, 20] Now it is the slightly greater value of the



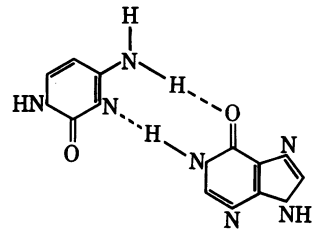
V



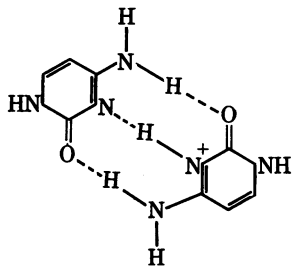
VI



VII



VIII



IX

repulsive term in configuration III over configuration II which determines the slightly greater over-all stability of configuration II. Although originally[9] configuration III was supposed to be the one present in poly (A + 2U), more recent and careful investigation of the shifts of infrared frequencies upon helix formation leads to the conclusion that it is in fact configuration II which is being observed.[10] Although the triplet is a part of a polynucleotide helix, in solution, the over-all stability of which involves also similar Van der Waals-London interactions between stacked bases[21, 22] and is influenced by the effect of the solvent, it is possible that the preference for configuration II over configuration III springs at least in part from the factor analyzed here.

(3) *The triplet hypoxanthine-adenine-hypoxanthine (poly (A + 2I):* The striking feature of this triplet is the relatively moderate attraction energy between the linked adenine-hypoxanthines which, together with the relatively high repulsion between the two hypoxanthines, leads to an over-all moderate interaction energy per purine base. The triplet may therefore be expected not to be a particularly stable one. Although, as indicated above, data from polynucleotide helices in solution cannot, of course, be directly compared with the results of calculations on the

isolated triplets, it is nevertheless interesting to observe that the T_m of poly (A + $2I$) is relatively low, and that the triple helix easily undergoes a displacement reaction with poly C leading to the formation of the more stable double-helix poly (I + C).[23] In connection with these findings, results are indicated in Table 1 on the interaction energies for the hypoxanthine-cytosine pair VIII such as may be expected to exist in poly (I + C). This energy is relatively high, although far from being as great as that calculated for the interaction of guanine with cytosine that amounted to −19.2 kcal/mole.[1] In connection with this last situation, it may again be useful to mention that poly (G + C) seems to be markedly more stable than poly (I + C).[24–26]

(4) *The triplets cytosine-guanine-cytosine cation:* This great stability of the interactions between guanine and cytosine is again illustrated in the results of the calculations for the triplets G + C_2, formed through the interaction between poly C and guanosine mononucleotides and oligonucleotides. A distinctive characteristic of these triplets is the presence of one of the cytosines (the one linked to N_7 of guanine) in the form of a cation (protonated at its N_3). From the theoretical point of view, the most striking feature of these triplets is the very high value of the interaction energies, which springs both from the high value of the separate G-C interactions and from the insignificant repulsion between the two cytosines. The pronounced effect of the protonation of the cytosine upon the strength of its interaction with guanine, although probably overestimated in the calculation, is worthwhile stressing. A similar effect may be considered in connection with the structure of poly C, at low pH, which probably consists of a double helix corresponding to the association of protonated and unprotonated cytosines following scheme IX.[27] The energy evaluated for such a configuration is appreciably greater than that calculated for a pair of doubly hydrogen-bonded unprotonated cytosines (≈13 kcal/mole).[1]

Summary.—Quantum-mechanical calculations have been performed on the interaction energies in hydrogen-bonded purine-pyrimidine triplets. In the cyclic homo-triplet of hypoxanthines as in poly I, the formation of the triplets represents a more advantageous organization than that of a dimer by about 1.2 kcal per base. In the open hetero-triplets A + 2U and A + 2I, the interaction between the nonlinked terminal bases introduces a repulsion, which may be significant for the geometry of the adopted configuration. The interaction energy is relatively small in A + 2I. It is, on the contrary, relatively very great in G + 2C. The involvement of a protonated cytosine in hydrogen bonding with guanine (or with a nonprotonated cytosine) greatly increases the energy of interaction.

* This work was supported by grant GM 12289-02 of USPHS (National Institute of General Medical Sciences), and grant 62-00-532 of the Délégation Générale à la Recherche Scientifique et Technique (Comité de Biologie Moléculaire).

[1] Pullman, B., P. Claverie, and J. Caillet, these PROCEEDINGS, **55**, 904 (1966).
[2] Pullman, B., P. Claverie, and J. Caillet, *J. Mol. Biol.*, **22**, 373 (1966).
[3] Pullman, B., P. Claverie, and J. Caillet, *Compt. Rend.*, **263**, 2006 (1966).
[4] Nash, H. A., and D. F. Bradley, *J. Chem. Phys.*, **45**, 1380 (1966).
[5] Pullak, M., and R. Rein, *J. Theoret. Biol.*, **11**, 490 (1966).
[6] Rich, A., *Biochim. Biophys. Acta*, **29**, 502 (1958).
[7] Michelson, A. M., and C. Monny, *Biochim. Biophys. Acta*, **129**, 460 (1966).
[8] Gellert, M., M. N. Lipsett, and D. R. Davies, these PROCEEDINGS, **48**, 2013 (1962).
[9] Felsenfeld, G., D. R. Davies, and A. Rich, *J. Am. Chem. Soc.*, **79**, 2023 (1957).
[10] Miles, H. T., these PROCEEDINGS, **51**, 1105 (1964).

[11] Rich, A., *Nature*, **181**, 521 (1958).
[12] Howard, F. B., J. Frazier, M. N. Lipsett, and T. H. Miles, *Biochem. Biophys. Res. Commun.*, **17**, 93 (1964).
[13] Lipsett, M. N., *J. Biol. Chem.*, **239**, 1256 (1964).
[14] Miller, J. H., and H. M. Sobell, these PROCEEDINGS, **55**, 1201 (1966).
[15] Spirin, A. S., *Dokl. Akad. Nauk SSSR*, **44**, 1963 (1963).
[16] Kruglyak, Yu. A., V. I. Danilov, and O. V. Shramko, *Biofizika*, **10**, 399 (1965).
[17] Hoogsteen, K., *Acta Cryst.*, **12**, 822 (1959).
[18] *Ibid.*, **16**, 907 (1963).
[19] Haschemeyer, A. E. V., and H. M. Sobell, these PROCEEDINGS, **50**, 872 (1963).
[20] Katz, L., K. Tomita, and A. Rich, *J. Mol. Biol.*, **13**, 340 (1965).
[21] De Voe, H., and I. Tinoco, Jr., *J. Mol. Biol.*, **4**, 500 (1962).
[22] Claverie, P., B. Pullman, and J. Caillet, *J. Theoret. Biol.*, **12**, 419 (1966).
[23] Sigler, P. B., D. R. Davies, and H. T. Miles, *J. Mol. Biol.*, **5**, 709 (1962).
[24] Inman, R. B., and R. L. Baldwin, *J. Mol. Biol.*, **8**, 452 (1964).
[25] Pochon, F., and A. M. Michelson, these PROCEEDINGS, **53**, 1425 (1965).
[26] Massoulié, J., and A. M. Michelson, *Biochim. Biophys. Acta*, **134**, 22 (1967).
[27] Langridge, R., and A. Rich, *Nature*, **198**, 725 (1963).