

Studies on Intermolecular Complex Formation.

I. Between Nucleotide Bases and Organic Compounds

Chihiro TAMURA, Noriko SAKURAI, and Sadao SATO

The Central Research Laboratories of Sankyo Co., Ltd., Hiromachi, Shinagawa-ku, Tokyo

(Received October 30, 1970)

The present investigation was focussed on a survey of the chemical compounds which are capable of binding with nucleotide bases. X-Ray powder patterns show that adenine and cytosine have a strong tendency to interact with many organic compounds having a carboxyl group and cyclic alternant CONH groups. A few compounds bind with thymine and uracil.

In view of the structure and function of DNA, many single crystal analyses of the complexes between the nucleotide bases have been analyzed.¹⁻⁷⁾ Several types of base-pairing were found, and in these complexes the principal binding force is due to complementary hydrogen bonds. We had expected that complexes exist not only between the nucleotide bases themselves but also between nucleotide bases and other organic compounds having some characteristic structure. By infrared spectrum, Kyogoku *et al.* studied some biologically active compounds which interact with nucleotide bases in solution.⁸⁾

The X-ray powder method was applied in surveying these compounds with respect to their ability to form complexes with nucleotide bases. A large number of combinations was found to give rise to a change in diffraction patterns which might indicate the formation of complex. Some of them were obtained in the form of single crystal and we succeeded in determining their crystal structures.^{9,10)} New types of intermolecular hydrogen bonding were found, the details of which will be published elsewhere. In the present paper we will describe the results of the survey and discuss the structural characteristics of the molecules which can form the complexes.

Experimental

The experimental procedure was as follows: equivalent moles (*ca.* 100 mg) of organic substance and the base were added to 70% aqueous ethanol (*ca.* 15 cc). The mixture was warmed on a water-bath until dissolution was complete and then allowed to stand for several days in a refrigerator. A glass holder with a depth of 0.2 mm was filled with crystalline material thus obtained, and X-ray diffractions were recorded up to $2\sin\theta=50^\circ$ with a Rigakudenki model D6C diffractometer. Spacings of diffraction peaks are listed in Table 4, the height of the maximum peaks being normalized

to one hundred.

The formation of new crystalline species was verified by comparing their diffraction patterns with those of components. An example of such comparison is shown in Fig. 1. All component samples were recrystallized by use of the same solvent in order to avoid an erroneous conclusion due to polymorphism of the components. The mass and infrared spectra of such crystals were sometimes useful in confirming the formation of new complexes. The stability of each crystal was examined by taking diffraction patterns of the same sample at intervals of several hours. Any changes in these patterns indicate that the substance is unstable at room temperature and humidity. Subsequently such samples were kept during measurements at a low temperature and at the same humidity as that in the refrigerator.

Some of the complexes showed only a small number of new peak superimposed over those of the diffraction patterns of the components. In such cases, the diffraction patterns were recorded repeatedly more than three times to confirm their reproducibility.

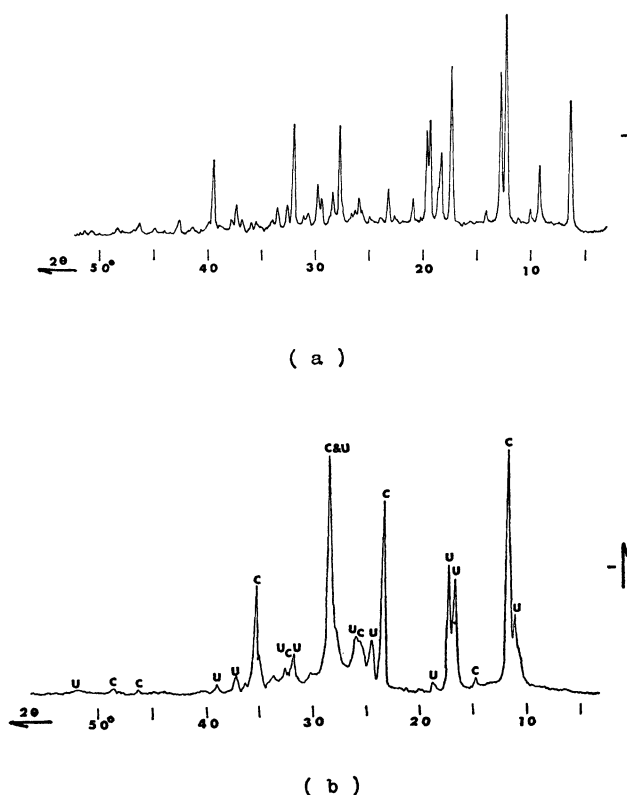


Fig. 1. X-ray powder pattern of the complex of uracil-cytosine (a), the pattern of a mechanical mixture of uracil (U) and cytosine (C), the original components, (b). $\lambda=1.5418$,

- 1) F. S. Mathews and A. Rich, *J. Mol. Biol.*, **8**, 89 (1964).
- 2) L. Katz, K. Tomita, and A. Rich, *ibid.*, **13**, 340 (1969).
- 3) A. E. V. Hashemeyer and H. M. Sobell, *Acta Cryst.*, **18**, 525 (1965).
- 4) E. J. O'Brien, *ibid.*, **23**, 92 (1967).
- 5) H. M. Sobell, K. Tomita, and A. Rich, *Proc. Natl. Acad. Sci. U.S.A.*, **40**, 885 (1963).
- 6) A. E. Hashemeyer and H. M. Sobell, *Nature*, **202**, 969 (1964).
- 7) K. Hoogsteen, *Acta Cryst.*, **12**, 822 (1959); *ibid.*, **16**, 907 (1963).
- 8) Y. Kyogoku, R. C. Lord, and A. Rich, *Nature*, **218**, 69 (1968).
- 9) C. Tamura, N. Sakurai, and S. Sato, *The Proceedings of the Conference of Chemical Society of Japan*, p. 423 (1970).
- 10) *Ibid.*, p. 424 (1970).

Results and Discussion

The results of the experiments, in part, are shown in Tables 1 and 2. A + sign denotes the formation of a new crystalline species and a 0 designates that no change for a pair was observed under these conditions. The

sign $\pm$ indicates that only a few new diffraction peaks were observed in addition to the original peaks. Some of the new crystalline species were established as being due to complex formation by means of single crystal analysis, and some were suggested to be complexes by infrared spectra or mass spectra. Therefore, it

TABLE 1. INTERMOLECULAR COMPLEX FORMATION BETWEEN SOME ACIDS AND NUCLEOTIDE BASES
With thymine and uracil all are not interactive

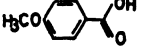
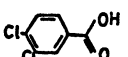
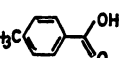
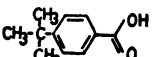
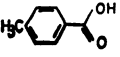
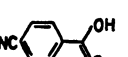
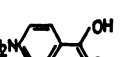
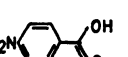
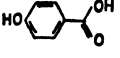
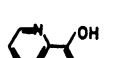
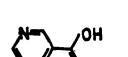
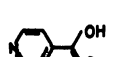
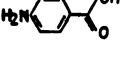
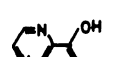
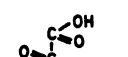
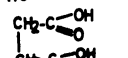
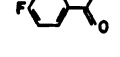
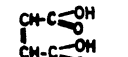
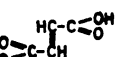
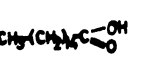
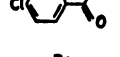
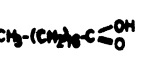
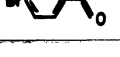
		Adenine	Cytosine			Adenine	Cytosine
	benzoic acid	+	+		anisic acid	+	0
	cyclohexanecarboxylic acid	+	+		3,4-dichlorobenzoic acid	+	+
	<i>o</i> -toluic acid	+	+		<i>p</i> -toluic acid	+	0
	<i>m</i> -toluic acid	+	+		<i>p</i> - <i>tert</i> -butylbenzoic acid	+	$\pm$
	<i>p</i> -toluic acid	+	0		<i>p</i> -cyanobenzoic acid	+	+
	salicylic acid	+	+		<i>p</i> -aminosalicylic acid	+	+
	<i>m</i> -hydroxybenzoic acid	+	+		<i>p</i> -nitrobenzoic acid	+	+
	<i>p</i> -hydroxybenzoic acid	+	+		picolinic acid	+	+
	anthranilic acid	+	0		nicotinic acid	+	+
	<i>m</i> -aminobenzoic acid	+	0		isonicotinic acid	0	+
	<i>p</i> -aminobenzoic acid	+	0		pyradinic acid	+	+
	<i>o</i> -fluorobenzoic acid	+	+		oxalic acid	+	+
	<i>m</i> -fluorobenzoic acid	+	+		succinic acid	+	+
	<i>p</i> -fluorobenzoic acid	+	+		maleic acid	+	+
	<i>o</i> -chlorobenzoic acid	+	+		fumaric acid	+	+
	<i>m</i> -chlorobenzoic acid	+	+		palmitic acid	$\pm$	0
	<i>p</i> -chlorobenzoic acid	0	$\pm$		stearic acid	$\pm$	0
	<i>o</i> -bromobenzoic acid	+	+				
	<i>m</i> -bromobenzoic acid	+	+				
	<i>p</i> -bromobenzoic acid	0	0				

TABLE 2. INTERMOLECULAR COMPLEX FORMATION BETWEEN NUCLEOTIDE BASES AND RELATED COMPOUNDS
OF NUCLEOTIDE BASES
With thymine and uracil all are not interactive

		adenine	cytosine			adenine	cytosine
	uracil	±	+		benzoylene urea	+	0
	thymine	+	0		phthalazine-1,4-dione	0	0
	5-bromouracil	+	+		2-3-dihydroxyquinoxaline	0	0
	5-nitouracil	+	+		1-(2H)-phthalazinone	+	+
	4-amino-1,3-dimethyluracil	0	0		4-ketobenzotriazine	0	0
	alloxane	+	+		isatonicanhydride	0	0
	barbituric acid	+	+		phthalimide	0	0
	barbital	+	0		3-indazolinone	0	0
	cyanuric acid	+	0		theophylline	+	+
	orotic acid	+	+		theobromine	+	0
	4,6-dihydroxy-2-mercaptopyrimidine	+	+		caffeine	0	0
	2-amino-4-hydroxy-6-methylpyrimidine	0	+		8-bromotheophylline	0	0
	2-amino-4-hydroxy-5,6-dimethylpyrimidine	+	0		4-hydroxypteridine	0	+
	3,6-dioxohexahydro-1,2,4,5-tetrazine	0	0				

seems most probable to assume that the new crystalline-species correspond to those of the complexes, but not another modification of the components. Table 3 shows the number of substances forming complexes

TABLE 3. COMPLEX-FORMATION ABILITY OF ORGANIC COMPOUNDS WITH ADENINE AND CYTOSINE
number of substances/total number of
forming complexes / substances examined

Category	For adenine		For cytosine	
1 Containing -COOH	33/36	91.8%	28/36	77.8%
2 Containing -CH-COO [⊖] NH ₃ [⊕]	0/18	0.0	0/18	0.0
3 Containing -CONH ₂	1/17	5.9	0/17	0.0
4 Linear molecules containing CO & NH	0/10	0.0	0/10	0.0
5 Cyclic molecules containing CO & NH	6/39	15.4	5/39	12.8
6 Cyclic molecules containing -CONHCONH-	11/12	91.6	5/12	41.6

and the total number examined; they can be classified into six categories. It has been observed that adenine and cytosine form complexes with many organic compounds, while thymine and uracil form hardly any such complexes with those examined. Guanine was not included because of its poor solubility. The spacings listed in Table 4 are for the complexes and their components.

Carboxylic Acids. Table 3 shows that 92% of the carboxylic acids examined can bind with adenine and 78% with cytosine. Even long chain fatty acids, which are structurally different from the planar molecule of adenine, form the complexes as shown in Table 1. The structure of single crystal analysis,⁹ in which the combination of the acid and base is of the type (acid₄-base₂) through the NH-O and OH-N hydrogen bonds is shown in Fig. 2. It seems that most of the other acid-adenine complex crystals are of a similar type.

It should be noted that many *p*-substituted benzoic acid derivatives do not form the complexes.

Aminoacids. In biological processes some interactions are expected to exist between nucleic acids

ADENINE-A *		2,023	2	D	INT	4,191	2	M-CHLORO- BENZOIC ACID	2,698	10	D	INT	3,576	11	2,564	2	3,986	3
		2,071	2	5,216	100	2,858	1		4,004	9	3,232	100	3,969	10			1,884	1
D INT		2,321	2	6,067	93				3,507		2,495	50	3,978	9			1,884	1
3,198		100	2	3,038	71						2,950	17	5,068	8	PARMITEIC			
6,758		80		6,058	68	UREOLEVENE- BENZENE	6,417	100			3,520	12	2,691	7				
5,574		63		4,058	50				4,6-DIHYDROXY 2-MERCAPTO- PYRIMIDINE	2,309	8	5,340	7					
3,770		54		11,946	4					2,495	50	2,495	50					
2,998		21	ALLOXANE	3,427	35					6,109	5	2,374	5					
3,401		16		2,032	18	10,526	100	2,449	15			2,481	5					
3,969		15		2,259	16	5,340	93	4,506	14			2,529	5					
5,155		15		7,500	10	3,562	21	3,181	13	9,211	100	2,652	5					
2,780		12		2,007	77	2,488	10	2,459	11	3,089	24	2,292	4					
3,562		10		3,708	25	3,604	9	2,894	11	3,069	11	2,921	3					
2,912		7		4,022	17	2,436	8	3,351	11	3,363	10	4,040	3					
4,928		5		2,969	14	3,480	9	3,480	9	3,480	9	3,480	9					
4,108		12		2,979	14	5,039	8	3,576	9	3,619	6							
2,938		4		3,132	14	2,692	7	2,161	7	4,770	6							
1,985		3		2,468	11	2,969	7	3,058	3	3,376	5							
				2,921	11	2,007	6	3,754	3	2,270	6							
				3,723	10	2,755	3	8,425	4	2,455	3							
				2,840	10	2,380	4	11,787	3	3,164	3							
				3,326	9	2,950	4	2,667	2	3,314	3							
				3,900	9	2,614	4	3,132	2	2,763	3							
				2,650	9	2,422	1	2,222	1	2,493	2							
				5,539	4	7,437	4	2,321	1	2,040	2							
				1,855	5	4,484	1	2,607	2	2,988	2							
				2,171	5			1,929	1	7,375	2							
				3,267	48	4,291	5			2,132	1							
				5,340	5					4,353	22							
				3,132	32	5,680	5			5,717	20							
				4,004	32	2,281	4			2,481	18							
				5,471	25	4,822	4			3,818	19							
				6,196	73	2,944	70			1,922	16							
				3,590	18	2,304	3			3,028	14							
				2,722	16	2,607	3			2,771	7							
				3,900	11	2,085	2			2,127	1							
				1,957	9					2,109	7							
				2,156	9	O-AMINO- BENZOIC ACID	3,079	12	11,632	12	5,405	39	2,067	7				
				2,894	9					3,648	15	4,095	35	2,637	7			
				1,988	9	2,142	1	3,351	8	7,138	5	5,505	14	2,298	3			
				3,440	9	D INT	2,755	6	3,278	8	3,255	10	2,176	12	1,873	6		
				4,332	9					6,559	9	2,462	12	2,691	6			
				2,585	7	5,829	98	1,945	5	2,287	7	3,153	12	9,408	6			
				3,151	7	7,004	9	3,000	5	3,058	5	2,058	11	2,979	2			
				2,674	6	4,351	62	3,723	4	2,675	4	3,290	11	2,739	5			
				6,657	6	3,786	19	2,304	3	2,714	4	2,488	8	2,151	4			
				2,686	5	2,840	9	2,350	3	2,867	4	3,663	8					
				2,522	5	3,453	9	2,522	2	2,912	4	3,018	8					
				2,635	5	2,912	2	2,713	2	2,217	3	3,232	8					
				4,062	4	2,607	5	6,109	2	D INT	3,363	6						
				2,998	4	5,842	25	12,450	100	3,427	3	1,866	5					
				3,389	4	2,196	4	6,326	63	2,161	2	9,983	3					
				4,575	4	2,675	5	4,231	3			3,619	1068					
						3,048	4	3,834	32			2,698	2					
						6,417	4	3,079	28	CYANURIC- ACID	3,802	1						
						2,002	3	2,176	23			3,314	54					
						2,141	3	2,848	23			2,464	6					
						2,238	3	5,868	25			4,004	16					
						4,720	3	6,607	44			3,427	14					
						3,198	48	3,153	39			2,080	13					
						3,378	43	3,934	16			3,033	12					
						4,439	28	2,315	11			2,404	11					
						3,619	22	3,619	22			2,495	3					
						2,449	2	2,585	12			3,534	3					
						3,019	24	2,585	12			2,270	9					
						2,626	22	10,782	19			2,675	9					
						3,618	24	2,117	11			2,488	7					
						2,619	12	3,302	2			5,405	17					
						4,439	1	2,107	7			3,834	58					
								4,152	7			2,036	46					
								3,376	9			2,529	5					
								3,440	9			3,739	5					
								4,874	7			2,463	23					
								2,495	8			2,156	22					
								2,522	6			7,309	5					
								3,048	6			9,309	4					
								5,754	6			2,356	3					
								6,194	6			2,823	3					
								2,071	5			3,058	10					
								2,233	4			2,432	29					
								2,392	4			2,506	29					
								2,763	4			2,895	2					
								2,805	4			2,496	25					
								2,876	4			2,036	2					
								3,590	4			2,780	1					
								4,623	100			2,782	3					
								4,955	50			3,952	22					
								4,396	40			3,480	21					
								5,405	17			2,840	20					
								4,462	1			2,171	4					
								8,043	32			5,539	16					
								3,314	16			2,614	15					
								3,900	12			2,894	13					
								3,568	58			3,194	73					
								3,400	6			6,511	51					
								2,018	5			2,254	49					
								6,916	5			3,326	17					
								2,592	2			2,049	8					
								3,100	3			2,958	16					
								2,206	2			1,881	7					
								2,309	2			12,277	11					
								2,617	45			3,278	45					
								1,945	1			3,243	9					
								2,488	1			2,171	4					
								2,637	1			3,279	9					
								2,683	7			4,114	9					
								2,823	1			2,543	8					
								2,921	1			2,667	9					
								2,550	4			3,414	8					
								3,453	4			4,396	8					
								5,644	4			2,338	7					
								2,122	3			4,212	8					
								1,814	2			5,906	7					
								2,332	2			11,946	6					
								2,196	2			6,215	5					
								2,298	10			3,008	17					
								2,780	2			2,959	8					
								4,375	2			2,103	5					
								4,720	2			2,127	5					

3.278	12	P-TOLUIC ACID	2.014	4	2.979	12	ADENINE	2.950	6	2.442	23	3.198	52	4.439	91	5.680	34			
3.279	12		1.870	4	8.425	9		3.401	6	1.976	22	3.649	46	7.568	33	7.257	21			
3.232	11	D INT	3.079	4			BARBITAL	4.848	8	2.429	19	3.834	31	3.834	31	2.557	24			
5.277	11	5.126	100	3.604	2	2.885	4	4.955	5	5.277	19	2.747	20	3.648	20	2.356	22			
6.151	10	3.339	51			3.121	4	D INT	1.937	4	2.222	17	2.652	21	4.152	17	3.493	22		
3.277	25	3.562	25			4.027	4	5.906	100	2.238	4	3.187	17	3.187	17	2.315	21			
2.265	5	6.026	21	ADENINE	4.040	7	4.562	5	1.961	6	3.100	16	4.022	13	4.004	12	5.126	21		
2.940	5	6.371	7			2.667	6	3.401	49	2.161	2	2.151	15	3.934	10	4.770	11	2.840	20	
3.376	5	2.564	5	O-AMINO- BENZOIC ACID	3.480	6	3.232	17			3.818	14	3.818	8	5.867	10	2.979	19		
3.501	8	3.754	4			4.217	7	2.950	16			3.322	13	3.322	13	4.842	18			
2.108	4	3.934	3			5.155	6	ADENINE	21.479	13	10.782	8	2.607	8	2.607	8	2.622	16		
2.998	4	5.680	3			2.166	4	5.539	15	*	2.356	12	2.141	7	4.623	8	3.089	15		
7.375	4	1.831	2	D INT	5.791	100	3.619	4	2.979	14		5.155	7	2.557	7	5.011	14			
2.460	2	2.460	2			5.24	81	3.110	14			3.527	12	3.527	12	2.315	21			
2.659	3	3.069	7			3.187	77	2.099	3	4.439	12	2.032	5	3.278	6	3.209	13			
6.511	3	3.198	2			11.481	52	2.137	3	4.671	9	5.372	100	2.508	5	4.901	5	3.414	13	
2.683	2					7.314	50	2.522	3	5.185	9	1.681	8	11.946	4	5.068	5	3.633	13	
2.722	2					9.118	25	3.255	7	3.900	8	2.780	8	1.222	3	4.901	5	3.767	12	
		2.446				2.747	45	2.747	2	2.196	9	8.346	8	1.863	2	2.344	4	2.442	12	
		5.574				2.780	2	2.380	5	4.464	26	5.187	5			6.151	4	3.143	11	
		6.559	34			1.041	1			5.738	4	3.278	7			2.488	3	1.957	9	
		6.707	81			1.918	8	2.698	5	3.187	13	ADENINE	3.927	3	3.927	3	2.767	13		
		3.562	31			2.023	1	3.754	5	3.576	30	4.874	6	3.028	3	2.501	8			
		3.754	32			2.062	1	7.762	5	3.100	29	7.138	9	M-CHLORO- BENZOIC ACID	8.346	3	2.683	8		
		4.506	30			2.287	1	2.309	4	6.109	28	9.213	6			2.675	2	3.008	8	
		3.466	69			9.118	25	2.788	4	3.163	7			5.372	2	4.901	5	2.767	13	
		5.011	63			3.633	22	3.754	1	3.708	26	1.941	4	D INT				2.730	7	
		3.100	42			10.282	21	10.04												

Table 4. continue

[illegible]

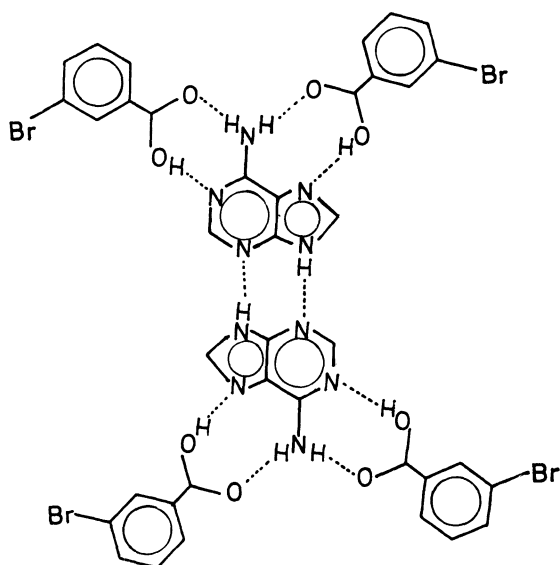
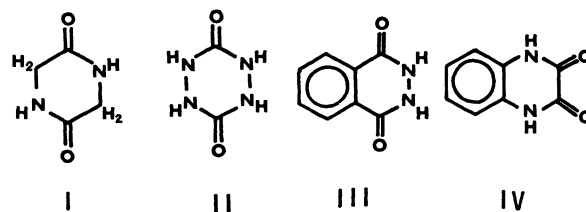


Fig. 2. The hydrogen bonding system of the complex crystal between *m*-bromobenzoic acid and adenine.

and amino acids. However, no amino acids were found to interact with these bases under the present conditions. The zwitter-ion form of amino acid might be related to this inertness. The substances tested in the present study were glycine, α -alanine, β -alanine, L-serine, L-threonine, L-valine, L-leucine, L-isoleucine, L-aspartic acid, L-histidine, L-proline, arginine, tryptophan, phenylalanine, methionine, cysteine, creatine, and aminocaproic acid.

Acid Amides and Lactams. In view of complex-forming ability, acid amides and lactams could be classified into four categories (3 to 6) as shown in Table 3. Acid amides are inert; compounds tested are benzamide, *o*-, *m*-, and *p*-toluamide, *o*-, *m*-, and *p*-hydroxy benzamide, *o*-, *m*-, and *p*-chlorobenzamide, picolinamide, nicotinamide, and isonicotinamide. The only exception in amide group is pyrazinamide which can bind with adenine. Lactam group are inactive; the compounds tested are butyrolactam, valerolactam, ϵ -caprolactam, 2-azacyclononane, tetrahydro-2-pyrimidone, succinamide, maleimide, phthalimide, glycinanhydride, 1,3-dimethylurea, acetylurea, 1-acetyl-3-methylurea, 1-phenyl-3-acetylurea, 1,3-diphenylurea, phenylurea, benzylurea, benzoylurea, and acetanilide.

Although the linear molecules containing CO and NH groups (category 4) do not interact with the bases, the corresponding atomic arrangement in cyclic molecules bind with adenine and/or cytosine; examples are benzoylurea/benzoyleneurea, and acetylurea/hydantoin. However even in cyclic molecules containing CO and NH groups, *e.g.*, anhydroglycine I, 3,6-dioxohydro-1,2,4,5-tetrazine II, 2,3-dihydroxyquinoxaline III and phthalazine-2,4-dione IV, binding



was not observed with these bases. This indicates that the atomic arrangement required for binding with adenine and cytosine would be cyclic alternant -CONH- groups. It should be noted that uracil¹¹⁾ and thymine are regarded as typical members of category 5. In such a system, the proton donor and acceptor groups will be in the proper orientation with those in the bases to reinforce individual hydrogen bonds. In fact, the hydrogen bond system in the adenine-phenobarbital complex solved by Kim and Rich¹³⁾ is just the case as shown in Fig. 3.

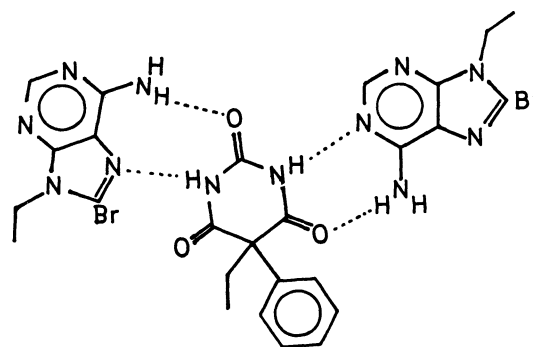


Fig. 3. The hydrogen bonding system of the complex crystal between phenobarbital and 8-bromo-9-ethyladenine, by Kim and Rich.

There are some other compounds which also bind with adenine and cytosine: 1-(2*H*)-phthalazinone, cycloserine, and 4,6-dihydroxy-1-mercaptopyrimidine. However, it is difficult to recognize any common features among them.

Amino and Nitrogen Heterocycles: A few compounds were found to bind with uracil and/or thymine; these are 1,3,5-triamino-2,4,6-triazine, azaadenine, and bromouracil. But many other related compounds, such as 2-methyl-4-amino-5-carbinopyrimidine, 2-aminopyridine, 2-aminopyrimidine, and 3-amino-1,2,4-triazole, do not form complexes.

We thank Professor Yoshio Sasada of Tokyo Institute of Technology for his valuable discussions and suggestions.

11) The complex of cytosine-uracil was found in this survey and its structure was determined by means of single crystal analysis.¹⁰⁾ The non-Watson-Crick pairing of uracil-cytosine found is almost the same as that of 5-fluorouracil-cytosine complex reported by Voet and Rich.¹²⁾

12) D. Voet and A. Rich, *J. Amer. Chem. Soc.*, **91**, 3069 (1969).

13) S. Kim and A. Rich, *Proc. Natl. Acad. Sci. U.S.A.*, **60**, 402 (1969).