BULLETIN OF THE CHEMICAL SOCIETY OF JAPAN, VOL. 45, 3254—3261 (1972)

Studies on Intermolecular Complex Formation. II. Crystal Structure of Cytosine-*N*-Benzoylglycine Complx Monohydrate

Chihiro TAMURA, Tadashi HATA, Sadao SATO, and Noriko SAKURAI

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

(Received December 28, 1971)

Many intermolecular complexes between *N*-acylamino acids and nucleotide bases have been found by means of X-ray powder diffractometry. Among these, the crystal structure of *N*-benzoylglycine cytosine complex monohydrate has been determined by the direct method. The crystal data are: $a=18.71$, $b=7.11$, $c=18.37$ Å, and $\beta=144.0°$. The space group is $P2_1/c$; $D_{obsd}=1.37$; $D_{calcd}=1.35$ g/cm^3. The final R factor was 0.131. In the crystal, a cytosinyl cation and a *N*-benzoylglycyl anion are joined by two NH···O hydrogen bonds, of which the shorter one is 2.652 Å long. Such a complex formation may exist *in vivo*.

Complementary hydrogen bonding between purine and pyrimidine bases in the Watson-Crick model of DNA is widely known, and several crystals of model complexes of nucleotide bases have been studied by single-crystal analysis.[1-7] Such base pairs are also assumed in explaining the structure of the stem in the clover model of *t*-RNA and the codon-anticodon interaction between *t*- and *m*-RNA. In biological systems, similar hydrogen bonding may exist not only between these bases, but also between nucleotides and peptides. However, no structural studies of the latter hydrogen bonding have yet been published. As we have previously reported,[8] a survey of the complex formation ability by means of X-ray diffractometry showed that none of the amino acids binds with any bases in spite of the fact that the carboxylic acid could give rise to an affinity for adenine and/or cytosine. It was assumed that the inertness of amino acids was due to their zwitterion structure. Subsequently, we have continued the survey using some modified amino acids which were presumed to suppress the zwitterion

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 Crystallogr.*, **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.*, **40**, 885 (1963).
6) A. E. Hashemeyer and H. M. Sobell, *Nature*, **202**, 969 (1964).
7) K. Hoogsteen, *Acta Crystallogr.*, **12**, 822 (1959); *ibid.*, **16**, 907 (1963).
8) C. Tamura, N. Sakurai, and S. Sato, This Bulletin, **44**, 1473 (1971).

formation; we found that many *N*-acylamino acids combine with adenine and/or cytosine. Herein we will describe the formation of the complexes of N-acylamino acids with nucleotide bases[9] and the crystal structure of one such complex: *N*-benzoylglycine-cytosine complex monohydrate.

Experimental

The ability of complex formation for nucleotide bases was examined in a similar manner, as has already been reported[8]

The procedure may be described briefly as follows: equivalent moles of the *N*-acylamino acid derivative and the base were dissolved in hot 70% aqueous ethanol, and the mixture was allowed to stand for several days in a refrigerator. X-ray diffractions of the materials obtained were measured up to $2\theta=50°$ (for CuK_α) with a Rigaku Denki model D6C diffractometer. These materials were generally fine crystalline powders which were not suitable for X-ray single-crystal analysis. The formation of new crystalline species was verified by comparing their diffraction patterns with those of the components. The only complex obtained as a single crystal at this stage was *N*-benzoylglycine-cytosine.

Because of the coexistence of the component crystals, single crystals of the complex were separated under a stereomicroscope; two components, cytosine and *N*-benzoylglycine, were detected by means of a mass spectrogram. The unit-cell dimensions and the space group were determined from Weissenberg photographs, and the density of the crystals was measured by the floatation technique. The crystal and physical data obtained are as follows: Monoclinic, $a=18.71$, $b=7.11$, $c=18.37$ Å, $\beta=144°$, Volume=1436.4 Å^3, $Z=4$, $D_{obsd}=1.37$, $D_{calcd}=1.35$ g/cm^3. Space group $P2_1/c$. Three-dimensional intensity data were collected from the equi-inclination Weissenberg photographs along the *b* axis through the layers of *h*0*l*—*h*6*l*. The intensities of 1993 independent reflections were measured by means of visual comparison with a standard scale. Lorentz-polarization corrections were made in the usual way, and the structure factors on an absolute scale were derived by Wilson's plot.

Structure Analysis

The phases of the structure factors were determined directly by the symbolic addition procedure for a centrosymmetric case. All the structure factors were converted to $|E|$'s. For the reflections with $E\geqq1.50$, triple-product sign relationships and the associated probabilities were listed by the use of the $\sum 2$ list program. As a starting set, three linearly independent reflections which had large $|E|$ magnitudes and a large number of interactions (9,1,-6; 5,1,9; 7,6,-3) were assigned to +, and four other reflections were given symbols (2.2,8;A: 6,0,-4;B: 12,3,1;C: 0,1,13;D). After the symbolic addition procedure using $|E|$'s greater than 1.50, the B,C, and D symbols were uniquely assigned to −, +, and − respectively. For two possible signs of A, the E-maps were synthesized, and a satisfactory molecular structure was obtained when A is +; the structure factor calculation on the basis of this approximate structure gave an *R* factor of 0.55. After three cycles of a full-matrix least-squares refinement with isotropic thermal parameters, the *R* factor decreased to 0.45. Since the temperature factors of some of the atoms became large, the Fourier synthesis was carried out. In the resulting map, the water molecule was found, and it was revealed that two of the atomic positions had been incorrectly assigned. Starting from the new set of atomic coordinates, six cycles of the full-matrix least-squares refinement with isotropic thermal parameters were carried out; the *R*-factor was thus reduced to 0.17. The hydrogen coordinates were given by the optimum positional calculation, assuming the distances of 1.05 Å for C–H and N–H bonds, since only a few hydrogen atoms were revealed on the difference map. Final cycles of the full-matrix least-squares, with anisotropic thermal parameters except for hydrogen atoms, dropped the *R* to 0.131. The structure factor amplitudes are listed in Table 1, while the final positional and thermal parameters are listed in Table 2(a) and (b).

Results and Discussion

General Ability of Complex Formation of N-*Acylamino acids for Nucleotide Bases.* Table 3 shows the ability of complex formation of *N*-acylamino acid with the nucleotide bases. A + sign denotes the formation of a new crystalline species, and a − shows that no change in diffraction patterns was observed under these conditions. It is remarkable that many *N*-acylamino acids produce complex crystals, contrary to the case of free amino acids. In a previous paper, the inertness of free amino acids for nucleotide bases was assumed to be due to their zwitterion form, which has been established in the crystal structures of many amino acids.[10–15] On the other hand, the crystal analysis of 2-amino-3-methyl-benzoic acid,[16] which can bind with adenine, shows that proton transfer from the carboxylic acid to the amino group does not occur, probably because the amino group is stabilized by π electrons in the benzene ring. Similarly, the *N*-acylation of the amino acids gives rise to a decrease in the proton-accepting ability of the amino group because of the formation of the peptide bond. Consequently, the carboxylic group can donate a proton to any proton-accepting group other than the amino group.

About 88 and 78% of the acidic and neutral *N*-acylamino acids form complex with adenine and/or cytosine under the present experimental conditions, while neither of the basic *N*-acyl derivatives, histidine or arginine, produces any complexes. Therefore, it can be assumed that the acidity of *N*-acylamino acid strongly affects the complex formation with nucleotide bases. How-

9) C. Tamura, N. Sakurai, and S. Sato, The proceeding of the Conference of Pharmaceutical Society of Japan, p. IV-131 (1970).

10) G. A. Jeffrey and Y. Kinoshita, *Acta Crystallogr.*, **16**, 20 (1963).
11) G. S. Parry, *ibid.*, **7**, 313 (1954).
12) R. Gerdil, *ibid.*, **14**, 333 (1961).
13) B. D. Sharman and J. F. McConnell, *ibid.*, **19**, 797 (1965).
14) P. J. Wheatley, *ibid.*, **13**, 80 (1960).
15) Y. Iitaka, *ibid.*, **13**, 35 (1960).
16) G. M. Brown and R. E. March, *ibid.*, **16**, 191 (1963).

TABLE 1. OBSERVED AND CALCULATED STRUCTURE FACTORS

Table 1. (Continue)

H	FO	FC
-12	25	38
-14	54	61
-15	32	43
-16	22	31
-17	33	33
L= 7, K= 3		
0	113	131
1	119	123
2	139	131
3	129	119
4	35	39
5	23	23
7	14	19
-1	39	47
-3	116	126
-4	222	221
-5	145	154
-6	226	193
-7	348	315
-8	76	60
-9	152	151
-10	9	20
-11	36	33
-12	82	84
-16	21	29
L= 8, K= 3		
0	159	180
1	46	50
3	54	50
-1	83	91
-3	89	94
-4	162	185
-6	179	169
-7	89	118
-8	186	193
-9	45	52
-10	117	117
-11	126	127
-13	41	55
-14	90	106
-17	54	52
L= 9, K= 3		
0	91	92
2	51	42
-1	130	119
-2	121	119
-3	132	142
-5	86	88
-6	60	74
-7	50	65
-9	100	100
-10	200	181
-11	37	55
-12	53	63
-14	42	57
-15	23	29
-16	12	26
-17	63	67
-18	16	28
L= 10, K= 3		
0	33	32
2	21	19
-1	41	40
-4	24	33
-5	33	37
-7	10	17
-8	119	133
-9	115	128
-10	51	60
-11	105	104
-12	37	53
-13	31	45
-15	11	17
L= 11, K= 3		
0	46	43
1	14	13
-1	39	34
-2	24	23
-3	53	61
-5	34	39
-6	159	156
-7	93	103
-8	79	87
-9	117	124
-10	53	65
-11	29	30
-12	56	98
-13	89	104
-15	43	61
-16	42	55
-18	38	41
L= 12, K= 3		
-2	17	16
-5	117	113
-6	56	54
-7	112	112
-8	61	69
-10	165	198
-12	59	131
-13	24	28
-14	33	44
-15	63	75
-17	32	42
-19	36	46
L= 13, K= 3		
0	27	17
-1	40	25
-2	42	31
-3	21	18
-5	59	53
-6	52	49
-7	80	74
-8	16	23
-9	37	45
-10	55	58
-11	32	44
-12	25	40
-13	46	43
-14	34	39
-18	46	48
-19	16	21
L= 14, K= 3		
-4	34	23
-6	36	36
-7	42	41
-8	52	57
-9	43	57
-10	58	61
-12	52	51
-13	42	54
-14	55	66
-15	90	92
-17	27	33
-18	27	39
-19	24	32
L= 15, K= 3		
-6	36	32
-7	25	22
-8	17	19
-9	32	41
-12	182	163
-13	51	52
-15	27	33
-19	50	48
L= 16, K= 3		
-9	30	27
-10	27	29
-12	117	99
-13	50	45
-15	57	79
-16	43	51
-17	25	19
-19	12	12
-20	22	24
L= 17, K= 3		
-6	40	23
-9	26	19
-11	30	30
-12	21	22
-13	73	63
-14	33	35
-15	17	21
-16	53	60
-17	53	53
-18	17	16
L= 18, K= 3		
-9	14	14
-13	30	27
-14	12	9
-16	12	16
-20	25	31
-21	16	20
L= 19, K= 3		
-11	14	13
-15	16	16
-16	42	35
-18	66	56
-21	9	15
L= 20, K= 3		
-15	14	15
-17	15	24
-18	39	36
-19	19	20
-20	25	27
-22	17	23
L= 21, K= 3		
-17	15	13
-19	31	26
L= 22, K= 3		
-19	15	13
-20	6	14
L= 0, K= 4		
1	135	122
2	91	82
3	104	83
4	51	41
5	113	111
6	66	76
7	33	43
8	26	30
9	25	25
10	16	21
11	12	14
12	6	3
L= 1, K= 4		
0	41	40
1	11	15
2	54	52
3	122	125
4	131	126
5	156	152
6	40	53
7	19	30
8	29	34
9	93	92
10	49	49
11	45	44
-1	260	293
-2	254	165
-3	277	238
-4	30	30
-5	7	19
-6	107	115
-7	61	70
-8	46	53
-9	41	42
-10	43	50
-11	16	20
-13	50	51
L= 2, K= 4		
0	53	60
1	75	66
2	17	22
4	42	53
5	175	158
6	12	17
7	36	34
8	114	105
9	21	20
10	12	14
11	19	24
-1	21	23
-2	28	34
-3	140	113
-4	172	147
-5	6	12
-7	40	44
-8	43	45
-10	13	18
-11	26	25
-12	9	11
-13	40	35
-14	35	39
L= 3, K= 4		
0	131	118
1	266	209
2	111	114
3	32	50
5	115	121
6	13	18
7	116	106
8	24	26
10	30	29
-1	56	63
-2	226	209
-3	124	116
-4	30	34
-5	37	40
-6	72	76
-7	23	21
-8	62	59
-9	36	39
-10	17	26
-11	16	22
-12	26	30
-14	36	35
-15	36	44
L= 4, K= 4		
0	54	54
1	85	93
2	120	140
4	19	22
5	166	101
6	21	16
-1	115	114
-2	69	71
-3	166	97
-4	77	85
-5	29	23
-7	25	14
-8	42	56
-9	66	64
-10	43	39
-11	37	35
-12	26	34
-13	9	11
-15	11	14
L= 5, K= 4		
0	7	8
1	23	26
3	181	178
4	50	44
5	24	25
-1	226	230
-2	59	57
-3	151	159
-5	32	39
-6	142	175
-7	59	57
-8	200	187
-9	43	55
-10	34	42
-12	32	39
-13	62	61
-15	39	44
-16	11	15
L= 6, K= 4		
1	96	101
2	69	90
3	42	39
4	40	36
5	25	25
6	7	5
-1	52	59
-2	62	73
-3	26	34
-5	62	88
-6	16	14
-7	35	22
-8	81	71
-10	56	60
-11	16	22
-12	40	53
-13	23	28
-15	19	20
-17	16	22
L= 7, K= 4		
0	129	141
1	102	103
2	26	24
3	39	36
5	49	40
6	16	15
-3	36	36
-4	74	73
-5	159	161
-6	36	45
-7	136	171
-8	37	35
-10	43	42
-13	29	32
-14	62	66
-15	16	21
-17	40	43
L= 8, K= 4		
0	27	23
1	26	22
2	19	11
3	35	31
4	28	16
5	13	7
-1	76	71
-3	24	30
-4	34	38
-5	44	43
-7	176	163
-8	12	6
-9	26	18
-10	95	92
-11	14	22
-12	14	18
-13	17	18
-14	41	47
-16	36	36
-17	15	19
-18	5	9
L= 9, K= 4		
1	16	14
2	17	18
3	30	17
4	11	5
-1	27	19
-2	24	26
-3	151	139
-6	76	89
-7	74	83
-8	20	11
-9	96	81
-10	49	47
-11	26	20
-12	47	52
-13	17	19
-14	42	55
-16	9	15
-17	23	25
-18	9	9
-19	3	13
L= 10, K= 4		
0	31	27
1	37	28
2	26	14
3	14	11
-1	37	41
-3	49	50
-4	131	126
-5	62	62
-6	6	11
-7	103	120
-8	13	18
-9	29	37
-10	87	92
-11	110	122
-13	44	37
-14	31	40
-16	16	23
-17	12	14
-19	10	19
L= 11, K= 4		
0	31	20
1	9	5
-1	16	14
-2	37	28
-4	49	50
-5	146	148
-6	63	90
-7	54	56
-8	66	100
-9	16	26
-10	87	84
-11	62	62
-12	103	136
-13	47	46
-14	20	19
-15	34	30
-16	19	19
-17	16	19
-18	14	17
-19	33	35
L= 12, K= 4		
0	24	9
-1	33	19
-2	9	16
-4	29	25
-6	47	40
-7	61	82
-8	12	17
-10	99	98
-11	27	24
-12	70	66
-14	46	55
-16	13	23
-18	17	8
L= 13, K= 4		
-1	16	14
-2	33	23
-3	14	5
-4	32	22
-6	35	29
-9	94	96
-11	15	15
-12	105	103
-13	47	60
-14	37	46
-17	33	35
L= 14, K= 4		
-2	9	9
-3	24	9
-4	52	36
-6	16	17
-8	39	36
-9	29	24
-10	83	66
-11	63	55
-12	15	23
-14	13	20
-17	16	18
L= 15, K= 4		
-5	25	17
-6	31	22
-8	23	14
-10	82	70
-11	73	62
-12	51	50
-14	23	25
-15	45	46
-17	33	31
-18	35	35
-19	33	33
-20	26	23
-21	15	6
L= 16, K= 4		
-6	11	6
-8	15	10
-11	30	20
-13	47	45
-14	19	23
-15	20	24
-18	19	17
-19	11	19
L= 17, K= 4		
-9	56	31
-10	6	3
-11	16	16
-13	27	26
-14	34	34
-15	9	12
-16	22	23
L= 18, K= 4		
-9	19	11
-10	28	16
-11	13	8
-13	32	25
-15	36	28
-16	46	39
-18	16	13
-19	16	14
L= 19, K= 4		
-14	15	12
-15	11	12
-16	29	27
-17	39	39
-18	12	13
-19	13	15
-21	6	13
-21	19	22
L= 20, K= 4		
-12	6	6
-16	24	16
-17	14	10
-18	15	13
-19	30	29
L= 21, K= 4		
-15	15	13
-17	10	10
-19	6	7
L= 0, K= 5		
1	75	62
2	112	191
3	149	134
4	74	79
5	80	86
6	25	32
7	20	23
9	31	32
10	6	11
L= 1, K= 5		
0	66	91
2	20	11
4	60	54
5	33	42
7	18	18
9	25	23
11	16	19
-1	53	59
-2	43	35
-3	51	36
-4	86	79
-5	167	143
-6	106	97
-7	77	82
-8	67	65
L= 2, K= 5		
0	119	122
1	49	48
2	41	47
3	54	56
5	71	74
6	30	27
7	16	11
8	24	20
9	23	19
-1	33	39
-2	41	36
-3	49	43
-4	156	139
-5	27	28
-7	43	48
-8	45	47
-10	10	16
-11	14	19
L= 3, K= 5		
1	86	87
2	44	54
3	56	62
6	7	6
9	13	11
-1	47	51
-2	66	74
-3	34	32
-4	26	30
-5	20	17
-6	71	78
-7	41	35
-8	56	53
-9	37	44
-10	27	22
-12	7	8
-13	6	8
-14	11	15
L= 4, K= 5		
0	8	7
1	6	4
2	64	68
3	35	33
4	7	15
5	26	23
6	37	32
7	10	12
8	9	8
-2	28	34
-3	42	47
-4	101	85
-5	21	19
-6	25	23
-7	31	33
-8	29	38
-9	63	67
-11	10	9
L= 5, K= 5		
1	40	47
2	17	12
4	50	50
7	6	5
-1	24	23
-2	47	39
-3	30	31
-4	40	43
-5	10	6
-6	8	8
-7	10	20
-8	10	16
-9	22	19
-12	25	33
-13	10	10
-14	25	25
L= 6, K= 5		
1	47	44
2	51	46
3	15	15
4	15	12
5	16	9
-1	34	42
-2	19	23
-3	33	39
-4	26	31
-5	39	47
-6	124	97
-7	31	28
-8	56	48
-9	74	73
-10	6	13
-11	65	53
-15	18	10
-16	9	10
L= 7, K= 5		
0	33	32
2	29	24
3	44	39
-4	55	55
-6	34	43
-7	48	36
-8	30	40
-9	125	100
-10	63	59
-12	27	26
-13	7	11
-14	24	20
L= 8, K= 5		
0	28	33
1	15	15
4	7	9
-1	33	29
-2	7	3
-3	71	76
-4	37	40
-5	9	13
-6	16	21
-7	43	45
-8	39	39
-9	11	15
-10	62	128
-11	65	68
-12	7	11
-13	7	8
-14	70	62
-15	12	12
-17	13	13
L= 9, K= 5		
0	7	13
2	15	13
3	9	7
-1	29	33
-2	12	11
-3	12	15
-4	47	59
-6	6	13
-7	11	13
-8	57	58
-9	27	35
-10	81	89
-11	82	81
-12	37	38
-13	36	35
-14	21	22
-15	6	12
L= 10, K= 5		
0	6	7
1	31	22
-2	13	11
-3	25	20
-4	37	43
-5	29	29
-6	6	6
-7	25	30
-8	9	15
-9	30	28
-10	9	12
-12	19	23
-13	43	48
-14	20	20
-15	53	47
-16	24	24
-18	9	3
L= 11, K= 5		
0	23	14
-1	9	6
-2	17	11
-3	7	5
-4	33	33
-5	26	30
-6	21	25
-7	10	10
-8	12	22
-9	11	11
-10	7	11
-14	25	25
-15	13	19
-16	20	20
-17	23	22
-18	7	12
L= 12, K= 5		
-1	6	5
-2	19	14
-3	19	16
-4	7	6
-5	25	24
-6	19	20
-7	36	24
-8	66	63
-10	26	21
-12	39	34
-15	25	27
-16	25	20
-18	6	13
-19	7	9
L= 13, K= 5		
-2	27	13
-3	25	15
-5	7	9
-6	40	36
-7	13	10
-8	27	28
-9	18	22
-10	7	13
-11	7	9
-13	25	27
-14	25	30
-15	13	13
-17	50	41
L= 14, K= 5		
-3	4	6
-4	24	12
-5	13	4
-8	20	15
-9	26	21
-10	37	38
-11	15	14
-12	16	19
-13	27	27
-16	12	10
L= 15, K= 5		
-6	10	5
-7	16	14
-9	14	11
-10	25	28
-11	15	12
-12	36	35
-14	16	12
-15	25	23
-16	12	16
-17	9	13
L= 16, K= 5		
-7	26	19
-12	21	22
-13	16	18
-15	32	29
-16	13	9
-20	6	6
L= 17, K= 5		
-9	42	5
-12	16	15
-14	13	14
-15	25	22
-20	8	12
L= 18, K= 5		
-11	32	20
-13	11	18
-14	10	9
-16	26	25
-17	5	6
-18	6	8
-19	18	13
L= 19, K= 5		
-12	4	3
-16	9	5
-17	4	4
L= 0, K= 6		
1	86	78
2	60	57
3	52	54
4	13	15
5	16	19
6	25	27
7	11	13
8	16	20
9	8	8
10	10	11
L= 1, K= 6		
0	122	127
1	63	70
4	14	19
5	27	28
6	25	23
7	62	63
8	16	17
9	14	15
-1	84	82
-2	20	23
-3	75	79
-4	129	132
-5	14	14
-6	7	10
-7	17	21
-8	30	23
-9	6	8
-11	6	12
L= 2, K= 6		
0	106	117
1	33	30
2	28	22
3	61	61
4	10	11
5	16	15
6	40	37
8	9	10
-1	113	114
-2	16	16
-3	16	18
-4	36	27
-5	121	114
-6	43	41
-7	59	56
-8	26	25
-9	32	30
-10	13	12
-12	9	11
L= 3, K= 6		
0	26	36
1	13	12
2	22	23
3	44	45
4	9	11
5	32	30
6	18	15
7	9	10
-1	63	71
-2	135	145
-3	26	30
-4	76	85
-5	46	43
-6	96	93
-7	133	127
-8	42	38
-9	17	16
-10	32	27
-11	25	22
-12	32	28
-13	6	15
L= 4, K= 6		
1	13	13
2	29	28
4	16	14
5	15	13
-1	9	9
-2	28	27
-3	45	52
-4	39	40
-5	44	41
-7	52	49
-8	52	57
-9	25	21
-10	5	3
-11	16	16
-12	41	39
L= 5, K= 6		
0	52	58
1	47	47
2	34	36
3	17	18
4	36	28
5	17	18
-1	63	69
-2	28	26
-3	13	12
-4	52	63
-5	102	96
-6	188	104
-7	13	13
-8	120	111
-9	24	26
-10	16	18
-11	66	64
-12	21	16
-14	29	29
L= 6, K= 6		
0	32	28
1	8	10
-2	37	39
-3	36	42
-4	9	9
-5	30	37
-6	123	130
-7	62	58
-8	4	2
-9	35	32
-10	23	24
-11	48	48
-12	41	38
-13	27	28
-14	24	23
-15	7	8
L= 7, K= 6		
1	11	11
2	7	8
3	7	5
-1	26	29
-2	16	14
-3	10	15
-4	13	8
-5	28	23
-6	29	30
-7	54	54
-8	116	112
-9	30	27
-10	16	19
-11	84	75
-12	64	61
-13	42	43
-14	7	4
L= 8, K= 6		
0	6	11
2	26	16
-2	16	15
-3	22	20
-4	41	41
-5	9	7
-6	9	12
-7	7	6
-8	4	6
-9	12	10
-10	25	20
-11	22	21
-12	32	32
-14	45	42
-15	6	6
-16	4	5
L= 9, K= 6		
0	11	9
1	20	21
-1	41	35
-2	10	9
-3	26	31
-4	16	13
-5	51	58
-6	61	78
-7	19	19
-10	13	13
-11	32	29
-12	47	43
-13	46	43
-14	22	23
-16	29	32
L= 10, K= 6		
-1	10	8
-6	16	16
-7	6	7
-8	36	40
-9	10	9
-10	34	29
-11	25	26
-12	79	78
-13	41	37
-14	14	14
-16	9	8
-17	14	18
L= 11, K= 6		
-2	14	14
-3	9	8
-6	17	15
-7	6	6
-8	20	21
-9	14	15
-10	16	12
-11	16	14
-12	14	16
-13	38	29
-14	34	33
-15	27	26
-16	23	27
-17	26	30
L= 12, K= 6		
-3	4	7
-4	10	6
-5	6	5
-6	9	12
-8	27	23
-10	34	27
-12	16	15
-14	6	6
-15	16	15
-16	15	18
-17	15	19
L= 13, K= 6		
-4	4	7
-6	21	14
-7	13	11
-8	23	20
-9	15	12
-10	17	18
-11	36	35
-12	47	41
-13	9	10
-14	4	7
-15	7	6
-16	13	11
-17	14	16
-18	28	29
L= 14, K= 6		
-6	3	4
-7	17	11
-9	6	9
-12	14	14
-14	15	13
-15	16	16
-16	19	19
-17	12	14
-18	23	26
L= 15, K= 6		
-7	4	9
-9	6	5
-10	4	6
-11	11	5
-14	6	7
-15	3	6
-17	12	12
L= 16, K= 6		
-10	9	6
-13	5	3
-16	5	9
-17	3	4
-18	3	7
L= 17, K= 6		
-13	27	28
-15	9	18
-16	16	17
-17	11	13
L= 18, K= 6		
-14	3	6
-15	3	3

TABLE 2(a). POSITIONAL AND THERMAL PARAMETERS OF NON-HYDROGEN ATOMS AT THE END OF THE LEAST-SQUARES REFINEMENT

Thermal parameters are in the form $\exp[-(\beta_{11}h^2+\beta_{22}k^2+\beta_{33}l^2+\beta_{12}hk+\beta_{13}hl+\beta_{23}kl)]$

	x	y	z	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}
C 1	0.0298	0.2223	0.4229	0.009440	0.041027	0.012760	−0.001028	0.009408	−0.001358
N 2	0.0599	0.4026	0.4665	0.008729	0.032611	0.012819	−0.000650	0.008456	−0.001447
C 3	0.0115	0.5566	0.3960	0.009027	0.034110	0.014488	0.000088	0.009703	−0.000135
N 4	−0.0663	0.5264	0.2769	0.008621	0.041154	0.011100	0.000678	0.007539	0.001027
C 5	−0.0993	0.3508	0.2300	0.011079	0.042139	0.013539	0.000406	0.009612	−0.000921
C 6	−0.0549	0.2008	0.2968	0.010665	0.036418	0.014486	−0.001966	0.009988	−0.004521
N 7	0.0840	0.0865	0.5045	0.013986	0.035738	0.016550	0.000783	0.012416	0.001230
O 8	0.0404	0.7116	0.4400	0.012334	0.029898	0.018218	−0.000672	0.011098	−0.001541
C 9	0.2659	0.3198	0.7750	0.008003	0.040756	0.012200	0.000808	0.007993	0.001240
O10	0.2388	0.1521	0.7397	0.012562	0.036416	0.013705	0.000753	0.009682	0.000685
O11	0.2255	0.4562	0.7092	0.011351	0.038137	0.012554	0.000657	0.009121	0.001800
C12	0.3544	0.3598	0.9087	0.009957	0.036054	0.013597	−0.001510	0.008842	−0.000132
N13	0.3900	0.2042	0.9827	0.008718	0.039390	0.012153	−0.000378	0.007793	0.000097
C14	0.3119	0.1251	0.9643	0.008587	0.037040	0.011234	−0.000049	0.007280	−0.000335
O15	0.2107	0.1879	0.8930	0.008629	0.040653	0.014995	0.001273	0.008615	0.002498
C16	0.3530	−0.0454	1.0360	0.008204	0.040764	0.008935	−0.000543	0.006457	−0.000031
C17	0.4419	−0.1654	1.0806	0.011505	0.044265	0.012424	0.002275	0.009952	0.002192
C18	0.4717	−0.3277	1.1402	0.014799	0.044156	0.015478	0.004897	0.012590	0.004688
C19	0.4179	−0.3732	1.1584	0.013295	0.049619	0.012459	0.000169	0.009706	0.004184
C20	0.3324	−0.2569	1.1188	0.010457	0.057842	0.013718	−0.001265	0.009233	0.003567
C21	0.3022	−0.0920	1.0596	0.009271	0.052496	0.011961	0.001616	0.008100	0.001508
C22	0.3568	−0.3050	1.3331	0.009346	0.043034	0.014898	−0.001109	0.009199	−0.000518

TABLE 2 (b). POSITIONAL AND THERMAL PARAMETERS OF SOME HYDROGEN ATOMS WHICH ARE OBTAINED FROM D-SYNTHESIS

	x	y	z	B		x	y	z	B
H 1	0.136	0.420	0.579	3.12	H10	0.456	−0.148	1.042	5.57
H 3	−0.112	0.599	0.215	7.47	H11	0.525	−0.426	1.160	9.85
H 4	−0.161	0.344	0.132	6.72	H12	0.464	−0.479	1.123	9.37
H 5	−0.060	0.098	0.282	9.59	H13	0.277	−0.286	1.129	8.89
H 7	0.316	0.470	0.907	7.17	H14	0.221	0.017	1.015	9.98
H 8	0.427	0.423	0.945	5.25					

TABLE 3. INTERMOLECULAR COMPLEX FORMATION BETWEEN NUCLEOTIDE BASES AND *N*-ACYLAMINOACIDS

		Free amino acids		*N*-Formyl-		*N*-Acetyl-		*N*-Benzoyl-		*N*-Carbobenzoxy-	
		A	C	A	C	A	C	A	C	A	C
Neutral Amino acid	Glycine	−	−	+	+	+	−	+	+	+	+
	α-Alanine	−	−	+	+	+	−	+	/	+	+
	β-Alanine	−	−	/	/	−	−	+	+	−	−
	Valine	−	−	+	+	+	+	−	+	−	+
	Leucine	−	−	+	+	+	+	+	+	/	/
	Isoleucine	−	−	−	+	−	+	−	+	/	/
	Phenylalanine	−	−	+	+	+	−	+	+	−	/
	Tyrosine	−	−	+	+	+	+	+	?	+	/
	Serine	−	−	/	/	/	/	/	/	+	+
	Threonin	−	−	/	/	−	?	/	/	+	+
	Methionine	−	−	+	+	+	+	+	+	−	/
	Cysteine	−	−	/	/	+	+	/	/	/	/
	Proline	−	−	?	?	+	+	/	/	−	?
	Tryptophane	−	−	/	/	+	+	−	+	/	/
Acidic Amino acid	Glutamic acid	−	−	/	/	+	+	+	+	+	/
	Aspartic acid	−	−	/	/	/	/	+	?	+	−
Basic Amino acid	Arginine	−	−	/	/	−	−	−	−	/	/
	Histidine	−	−	−	−	−	−	−	−	/	/

A: adenine, C: cytosine, for thymine and uracil all are —.
? : For this combination, the only gel-like material is obtained.
\: No *N*-Acylation is available.

Table 4. Bond lengths and angles

Bond lengths (Å)				Bond angles (°)			
C 1–N 2	1.368	C12–N13	1.430	C 2–C1–C 6	116.8	C 9–C12–N13	117.0
C 1–C 6	1.399	N13–C14	1.331	N 2–C1–N 7	116.5	C14–N13–C12	120.9
C 1–N 7	1.322	C14–O15	1.220	C 6–C1–N 7	126.8	N13–C14–O15	123.3
N 2–C 3	1.341	C14–C16	1.469	C 1–N2–C 3	124.3	C16–C14–O15	119.8
C 3–N 4	1.341	C16–C17	1.401	N 2–C3–N 4	116.0	C16–C14–N13	116.9
N 4–C 5	1.350	C17–C18	1.361	N 2–C3–O 8	120.9	C17–C16–C14	122.6
C 5–C 6	1.298	C18–C19	1.332	N 4–C3–O 8	123.2	C14–C16–C21	119.4
O 8–C 3	1.206	C19–C20	1.374	C 3–N4–C 5	121.5	C17–C16–C21	118.0
C 9–O10	1.252	C20–C21	1.372	N 4–C5–C 6	123.0	C16–C17–C18	120.4
C 9–O11	1.223	C16–C21	1.366	C 1–C6–C 5	118.5	C17–C18–C19	120.4
C 9–C12	1.508	C12–H 8	1.01	O10–C9–O11	125.2	C18–C19–C20	120.9
N 2–H 1	1.24	C17–H10	0.95	O10–C9–C12	118.3	C19–C20–C21	119.4
N 4–H 3	0.85	C18–H11	1.01	O10–C9–C12	116.5	C16–C21–C20	120.8
C 5–H 4	1.10	C19–H12	1.08				
C 6–H 5	0.80	C20–H13	1.21				
C12–H 7	1.05	C21–H14	1.25				

ever, it should be noticed that, in some cases, a small structural differece in neutral *N*-acylamino acids may play a major role in forming a complex; for instance, *N*-acetylleucine and *N*-acetylisoleucine have a similar acidities, but the former binds only with adenine, and the latter, with cytosine.

The Geometry of the Complex. From the structure determination of the present crystal, it is established that the complex is formed by two hydrogen bonds between cytosine and *N*-benzoylglycine. The interatomic distances and angles in the complex between cytosine and *N*-benzoyl glycine were calculated from the parameters in Table 1, as is shown in Fig. 1(a) and (b), and in Table 4. The maximum standard deviation was 0.017 Å for the bond length in cytosine and the carboxyl group and 0.033 Å for that in the benzoyl group.

The bond lengths and angles of the cytosine moiety of the complex do not agree well with those found in the crystal of cytosine hydrate. Table 5 shows a comparison with related structures. The significantly large bond angle of 124.3° for C1–N2–C3 implies that the hydrogen atom might be attached to N2, since the mean bond angles of C–NH–C and C=N–C are 125° and 116° respectively.[17] In the pyrimidine ring, the rather short distance, 1.298 Å might indicate that the C5–C6 bond is a double bond. The short bond distance, 1.322 Å, for C1–N7 might be attributed to the double-bond character. The C–O bond lengths of 1.223 and 1.252 Å, and the O10–C9–O11 angle of 125° are close to those observed in many amino acids and suggest the formation of a carboxyl anion. Although the difference electron density map is not good enough

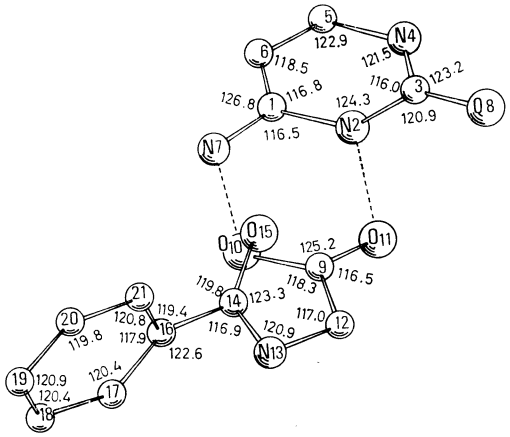


Fig. 1. (b) Bond angles of the complex.

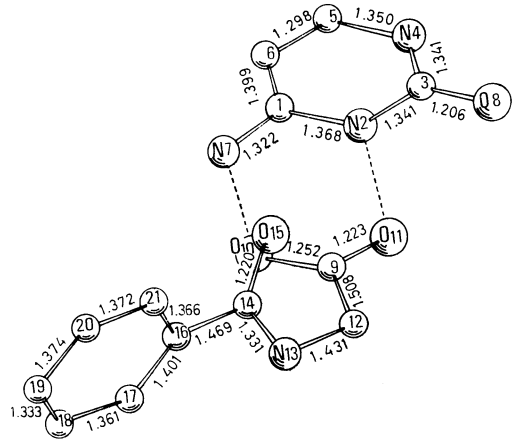


Fig. 1. (a) Bond distances of the complex.

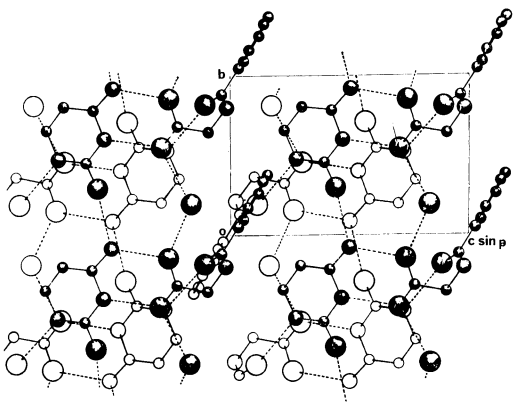
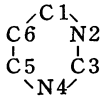


Fig. 1. (c) A view through two adjacent molecular layers, showing the stacking arrangement of the cytosine and carboxyl group viewed down the *c*-axis.

17) C. Singh, *Acta Crystallogr.*, **19**, 861 (1965).

TABLE 5. BOND LENGTHS AND ANGLES OF PYRIMIDINE DERIVATIVES

	This work	Cytosine[10]	Uracil[25]	Thymine[26]	Isocytosine[27]		Pyrimidine[28]
C1–N2	1.368Å	1.351	1.371	1.391	1.363	1.375	1.332
N2–C3	1.341	1.354	1.384	1.361	1.333	1.369	1.317
C3–N4	1.341	1.376	1.344	1.355	1.357	1.330	1.313
N4–C5	1.350	1.361	1.341	1.382	1.358	1.350	1.343
C5–C6	1.298	1.348	1.408	1.349	1.331	1.356	1.358
C6–C1	1.399	1.432	1.417	1.447	1.438	1.442	1.387
C6–C1–N2	116.8°	122.0	118.1	115.6	118.7	114.8	122.5
C1–N2–C3	*124.3	118.9	*124.7	*126.3	119.7	*123.3	115.1
N2–C3–N4	116.0	120.1	115.6	115.2	121.8	121.9	128.2
C3–N4–C5	*121.5	*121.3	*123.6	*122.8	*120.2	115.9	115.2
N4–C5–C6	123.0	120.6	121.5	121.8	120.5	125.9	122.7
C5–C6–C1	118.5	117.1	116.4	118.2	119.1	118.2	116.3



*sign indicates C–NH–C type of bond angles while no sign indicates C=N–C type of bond angles.

for obtain the positions of all the hydrogen atoms, the peaks for the H1 and H3 involved in the hydrogen bond are observed rather clearly. Each of the nitrogen atoms, N2 and N7, obviously possesses a hydrogen atom, while the carboxyl oxygens, O10 and O11, do not. From these results, it may be concluded that the complex consists of a cytosinyl cation and an *N*-benzoylglycyl anion as a result of proton transfer from the carboxyl group to the amidine group; the complex can be well represented by the formula of Fig. 2(b) rather than by that of Fig. 2(a). The NH···O

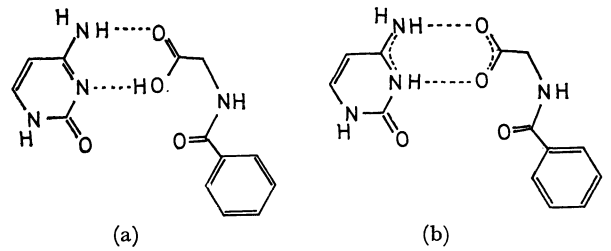


Fig. 2. Hydrogen bond system between cytosine and *N*-benzoylglycine.

hydrogen bond distances are 2.697 and 2.652 Å, fairly shorter than the normal value,[18] probably because of the enhancement of some ionic force. The intermolecular hydrogen bond between the carboxyl and guanido groups in an arginine crystal[19] is analogous to that in the present complex. The least-squares planes of cytosine, the carboxylic group, and a benzene ring in *N*-benzoylglycine were calculated; the displacements of the atoms from each plane are shown in Table 6. The maximum deviation is 0.04 Å for O8 in the cytosine ring. The dihedral angle between cytosine and the carboxyl group is 3.0°.

As may be seen from Fig. 3, the component molecules have several sites, (A), (B),···(F), which are capable

TABLE 6. LEAST SQUARES PLANES AND DEVIATIONS

Atom	Displacement from the planes
C 1	−0.00Å
N 2	−0.01
C 3	0.02
N 4	−0.02
C 5	0.00
C 6	0.00
N 7	0.00
O 8	0.04
C 9	−0.00
O10	0.00
O11	0.00
C12	0.00
C16	0.02
C17	−0.01
C18	−0.00
C19	0.01
C20	0.00
C21	−0.02

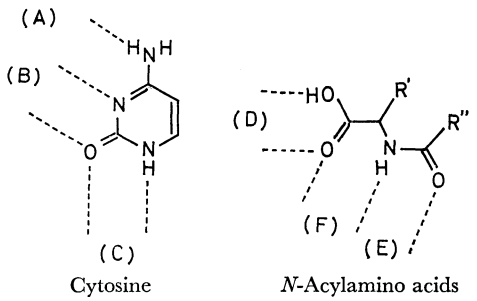


Fig. 3. Hydrogen bonding sites in cytosine and *N*-acylamino acids.

of forming an intermolecular hydrogen bond. The favourable condition of the pairing might be affected not only by their characteristic atomic arrangements, but also by their crystalline fields. Therefore, the hydrogen-bonded complexes found in crystal structures might be rather limited. We expect that the hydrogen-bond systems in some of the complexes

18) W. Fuller, *J. Phys. Chem.*, **63**, 1705 (1959).
19) I. L. Karle and J. Karle, *Acta Crystallogr.*, **17**, 835 (1964).
25) G. S. Parry, *ibid.*, **7**, 313 (1954).
26) R. Gerdil, *ibid.*, **14**, 333 (1961).
27) B. D. Sharma and J. F. McConnell, *ibid.*, **19**, 797 (1965).
28) P. J. Wheatley, *ibid.*, **13**, 80 (1960).

Table 7. Short intermolecular distances

N 7–O10	2.652	1	N4 –O15	2.742	4
N 2–O11	2.697	1	C 5–O11	3.601	4
C 5–C16	3.687	2	C 5–O12	3.719	4
C 5–C21	3.664	2	N 5–O 8	3.516	4
C 6–C14	3.766	2	C 9–C18	3.762	5
C 6–O15	3.443	2	C 1–O 8	3.636	6
O 8–C 1	3.636	3	C18–C12	3.568	6
O 8–N 7	2.758	3	N 4–C 6	3.674	7
C12–C18	3.568	3	O15–C 6	3.774	8
C 1–O 8	3.694	4	C20–O 8	3.421	8
N 2–N 2	3.597	4	C21–N 2	3.569	8
N 2–C 3	3.705	4	C21–C 3	3.496	8
N 2–O 8	3.575	4	O22–C 9	3.606	8
N 2–O11	3.473	4	O22–O11	2.884	8
C 3–C 9	3.347	4	O22–C12	3.472	8
C 3–C10	3.681	4	C16–C19	3.528	9
C 3–O11	3.090	4	C17–C19	3.720	9
C 3–O15	3.613	4	C21–C18	3.753	9
N 4–C 9	3.206	4	C21–C19	3.575	9
N 4–O10	3.759	4	O22–O10	2.790	10
N 4–O11	3.190	4	O22–N13	3.637	10
N 4–C12	3.405	4	O22–C14	3.870	10
N 4–C14	3.734	4	O22–C17	3.397	10

1..	x,	y,	z,
2..	$-x$,	$-y$,	$1-z$,
3..	x,	$1+y$,	z,
4..	$-x$,	$1-y$,	$1-z$,
5..	$1-x$,	$-y$,	$2-z$,
6..	x,	$-1+y$,	z,
7..	$-x$,	$1/2+y$,	$1/2-z$,
8..	x,	$1/2-y$,	$1/2+z$,
9..	$1-x$,	$1/2+y$,	$3/2-z$,
10..	x,	$-1/2-y$,	$-1/2+z$.

listed in Table 3 are similar to that in cytosine—*N*-benzoylglycine.

The principal architecture of the present complex crystal is as follows; the complexes are linked to form an infinite column in the direction of the *b*-axis by hydrogen bonds between N7 and O8 in the cytosine molecules, with a distance of 2.758 Å. In this column, water molecules are located between O11 and O10 with distances of 2.789 and 2.884 Å, respectively so as to be hydrogen-bonded along the column. Two of these columns are joined through the hydrogen bonds of N4 with O15 at 2.742 Å. The intermolecular contacts shorter than 3.9 Å are given in Table 7.

Implications concerning Biological Aspects. The details of the interaction between *N*-acylamino acids and nucleotide bases in crystalline complexes can now be presented on the basis of our results. Similar types of molecular interaction may exist *in vivo*, but it is as yet impossible to predict such complex formation in a particular biological process. Here we will only point out some features of the present results which may be of biological significance.

At the terminal position of *t*-RNA molecules, the ususal sequence is CCA. Cytosine and adenine, not uracil nor thymine, are the bases with which *N*-acylamino acid from the complexes. This coincidence seems very attractive in considering the interaction of *t*-RNA and amino acids.

Another point is *N*-acylation, which is necessary for complex formation between amino acids and nucleotide bases. It is widely known that the hydrogen atom of a terminal amino group in many peptides is replaced by formyl or acetyl groups.[20–23] Marcker[24] discovered that methionine-*t*-RNA, which can tranfer a formylated amino acid into polypeptide, is involved in the initiation of protein synthesis in *E. Coli*. Thus, we assume that the binding of *N*-acylamino acid and a nucleotide base, similar to that of the present complex, has occurred in several steps of the protein synthesis.

In conclusion, the complex formation of adenine and cytosine with *N*-acylamino acid occurs generally. The molecular arrangement in the present complexes may be thought of as a candidate in considering possible interactions in biological systems.

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

20) E. Margoliash, E. L. Smith, G. Kreil, and H. Tuppy, *Nature*, **192**, 1121 (1961).
21) K. Narita, *Biochem. Biophys. Research Comms.*, **5**, 160 (1961).
22) K. Hofmann and H. Yajima, *J. Amer. Chem. Soc.*, **83**, 2289 (1961).
23) K. Hofmann, H. Yajima, and E. T. Schwartz, *ibid.*, **82**, 3732 (1960).
24) K. A. Marcker and F. Sanger, *J. Mol. Biol.*, **8**, 835 (1964).