Tautomerism in 5,8-Diaza-7,9-dicarbaguanine ('Alloguanine')[1])

by **Trixie Wagner***[a]), **Bo Han**[b]), **Guido Koch**[a]), **Ramanarayanan Krishnamurthy**[c]), and **Albert Eschenmoser***[b])[c])

[a]) *Novartis* Pharma AG, NIBR Basel, CH-4002 Basel (e-mail: trixie.wagner@novartis.com)
[b]) Laboratory of Organic Chemistry, Swiss Federal Institute of Technology (ETH), Hönggerberg, HCI H309, CH-8093 Zürich (e-mail: eschenmoser@org.chem.ethz.ch)
[c]) The *Skaggs* Institute for Chemical Biology at The *Scripps* Research Institute (TSRI), 10550 North Torrey Pines Road, La Jolla, CA-92037, USA

Rolf Huisgen zum 85. Geburtstag gewidmet

An X-ray structure analysis of the title compound reveals that this purinoid exists in the crystal as two tautomers which interact with each other in the mode of a reverse-*Watson–Crick* base pair.

We recently reported on the synthesis and chemical properties of a family of purinoid heterocycles, 5,8-diaza-7,9-dicarba-purines ('allopurines'[2])), which undergo regioselective *C*-nucleosidations with cyclic iminium salts under mild conditions to afford *C*-nucleosides that are isosteric with N(9)-nucleosides of natural purines [3] (*Fig. 1*). While our interest in the chemistry of this family of purine analogs originated in studies toward an etiology of the structure type of the natural nucleic acids, it became clear that this family of heterocycles with its close structural relationship to the family of the natural purines also deserves attention from a medicinal chemistry point of view. In that context, the guanine member (alloG) was crystallized at *Novarti*s and subjected to an X-ray structure analysis. We report here the result of this analysis; it provides information about the accessibility of a tautomeric form in this guanine-like heterocycle. This is of interest in context with phenomena of self-recognition of (potential) nucleobases through self-pairing between two different tautomeric forms of one and the same heterocycle.

In our work on the properties of homo-DNA [1][5] and pyranosyl-RNA [6][7], we had encountered very robust duplexes of oligonucleotides whose stability is exclusively due to strong purine–purine pairing in the *Watson–Crick* mode between diaminopurine and xanthine, and between guanine and isoguanine (*Fig. 2*). In the pyranosyl-RNA series, it was furthermore demonstrated that purine–purine pairing in this mode requires equal sense of chirality of the two (homochiral) duplex strands. It was also in this series that a clear-cut example of self-recognition through pairing between two

---

[1]) Chemistry of α-Aminonitriles, Part 44; for Part 43, see [1].

[2]) To circumvent confusion when dealing with the analog of isoguanine of the natural series, we refrain from naming the family 'isopurines' and propose to call them 'allopurines' instead. Note the existence of another family of purine isomers [2] that differs from the allopurine family by having interchanged in the purine nucleus only the atoms of positions 5 and 7.

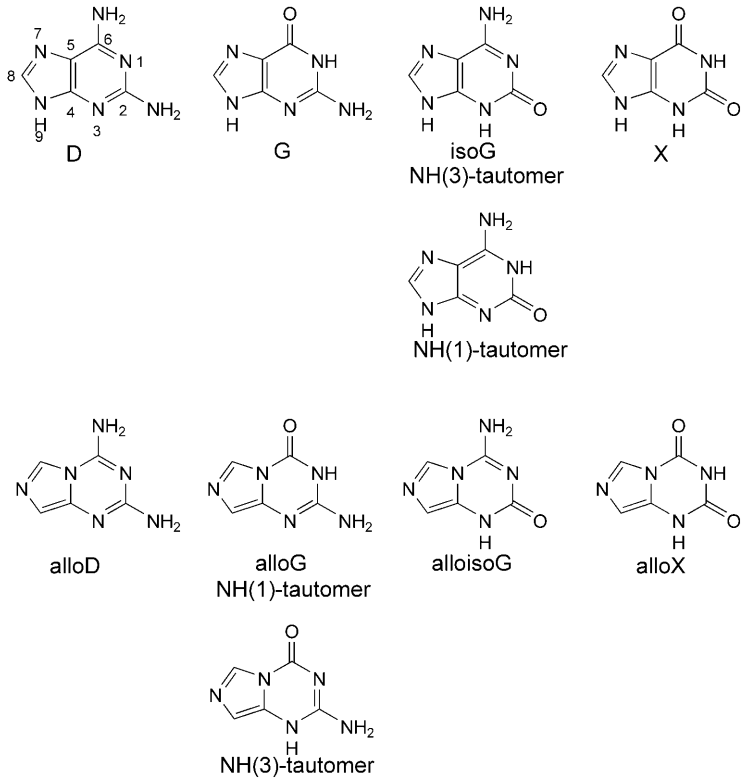


Fig. 1. *Juxtaposition of the chemical formulae of members of the purine family with those of corresponding member of the 'allo-purine' family*

complementary tautomers of the same nucleobase was identified: p-RNA strands (hexamers and octamers) containing isoguanine as the only nucleobase undergo neat duplex formation by virtue of what has to be the reverse-*Watson–Crick* base-pairing mode if, and only if, the backbones of the two homochiral strands have the opposite sense of chirality[3]). This case of nucleobase self-recognition through pairing between two (in the reverse-*Watson–Crick* mode) complementary tautomers is a specific property of isoguanine, which, in contrast to guanine, is known to exist in more than one tautomer [10]. From our observations about the pairing properties of guanine and isoguanine in both the homo-DNA [4][5] and pyranosyl-RNA [6][7] series, it must be inferred that isoguanine as a pairing partner in oligonucleotide sequences can assume both the NH(1)- and NH(3)-tautomeric form. The observed chiroselectivity in the pyranosyl-RNA series of duplex formation that is based on isoG/isoG pairing underscore this inference.

---

[3]) Complex formation based exclusively on self-recognition of the same nucleobases is not that uncommon, yet most often brought about either by one-sided protonation (or deprotonation) of the base (see, *e.g.*, [8]), or on double mode self-pairing (*e.g.*, quadruplexes deriving from guanine-sequences, *e.g.*, [9]).

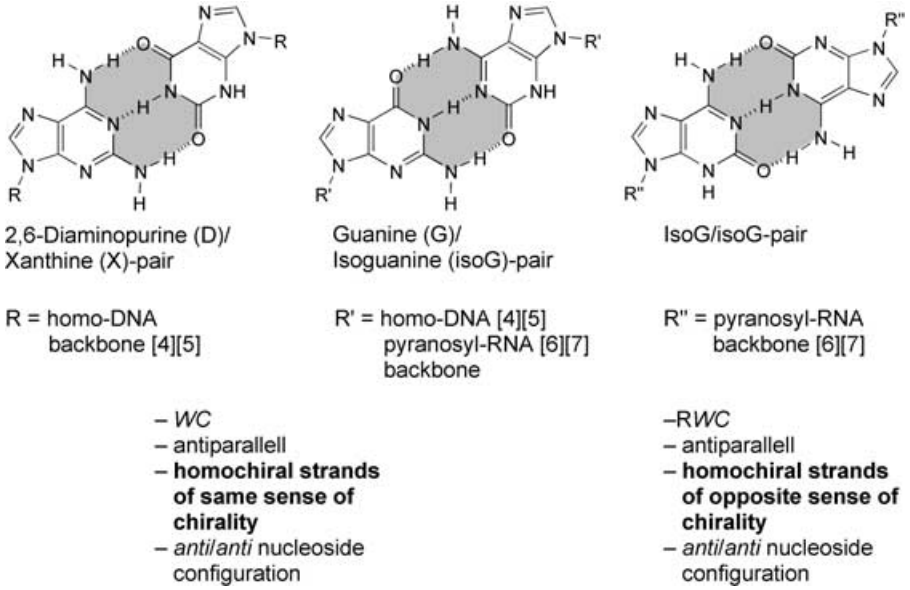


Fig. 2. Watson – Crick (*WC*) *and reverse*-Watson – Crick (*RWC*) *modes of purine – purine base-pairing observed in the homo-DNA and pyranosyl-RNA series*

An interesting constitutional difference between the purine base isoguanine and the purinoid base allo-isoguanine consists in that the latter – in contrast to the former – cannot (be expected to) adopt the form of an NH(1)-tautomer but should exist exclusively as NH(3)-tautomer (*Fig. 1*). On the other hand, the question arises whether in the allo-purine family in which the six-membered part of the heterocycle is a 1,3,5-triazine (and not a pyrimidine) derivative, it might be the guanine member that, besides its conventional NH(1)-form, has access to the NH(3)-tautomeric form. If this were the case, nucleobase self-recognition in the reverse *Watson – Crick* mode due to pairing between different tautomers of the same nucleobase could be expected for guanine in the allopurine series, as observed for isoguanine in the purine series (*Fig. 3*).

Interestingly enough, it is exactly this kind of reverse *Watson – Crick* base pair that the X-ray structure analysis of alloG has uncovered to exist in the crystal structure of this purinoid.

The constitution of alloG – one that is not easily deducible from the mode of the purinoid's formation (see the *Scheme* in *Exper. Part*) – was previously determined by X-ray analysis of an *N*-(dimethylamino)formamidine derivative [2c]. It is fully confirmed by the X-ray analysis of the free base itself (*Fig. 4*). What is remarkable in the outcome of this latter analysis is the fact that it discloses a crystal structure in which alloG occurs in two tautomeric forms, and that these two forms interact with each other exactly in the way that corresponds to base pairing in the reverse-*Watson – Crick* mode. This type of pairing interaction between tautomers in the crystalline state is similar to what has been described for isocytosine [11]. The pairing is clearly evidenced by the packing diagram depicted in *Fig. 5*. Each of the two crystallographically independent molecules is located near a center of inversion which places the symmetry-related atoms N(10′),

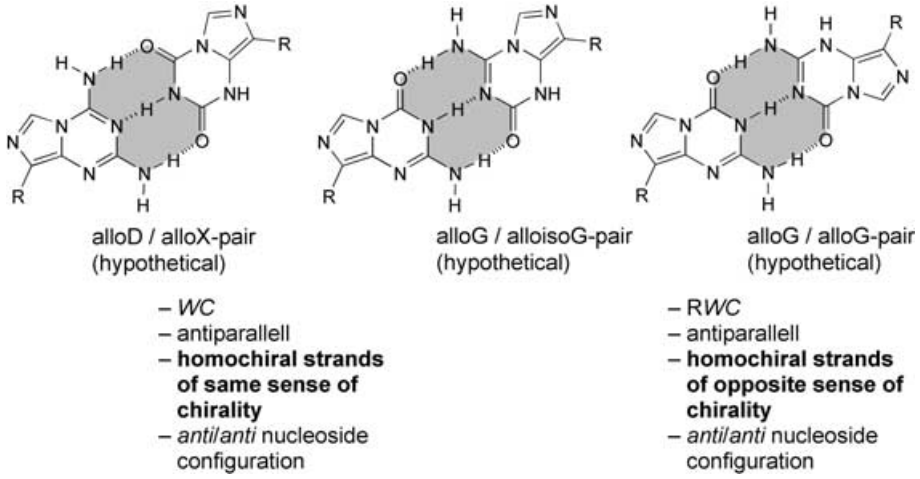


Fig. 3. Watson – Crick (*WC*) *and reverse*-Watson – Crick (*RWC*) *modes of (hypothetical) allopurine – allopurine base-pairing in informational oligomers*

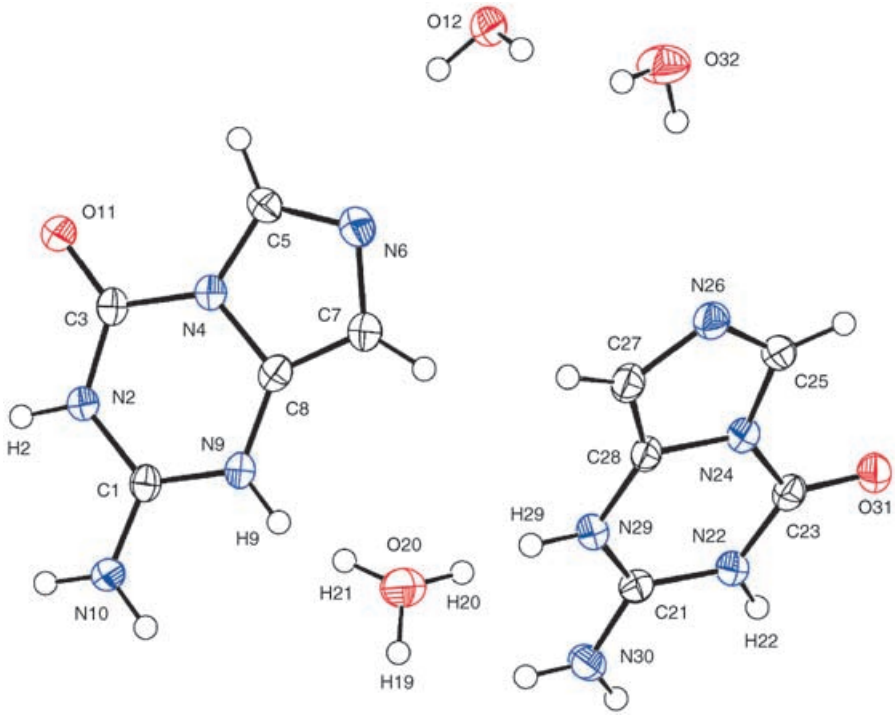


Fig. 4. *Structure of compound* **4** *in the crystal* [13]. The asymmetric unit comprises two independent molecules (I and II) of alloG and three molecules of H_2O. Atomic displacement ellipsoids drawn at the 50% probability level, H-atoms drawn as spheres of arbitrary radius. The H-atoms H(2)/H(22), H(20)/H(21), and H(9)/H(29) are disordered over two sites with a 0.5 : 0.5 distribution.

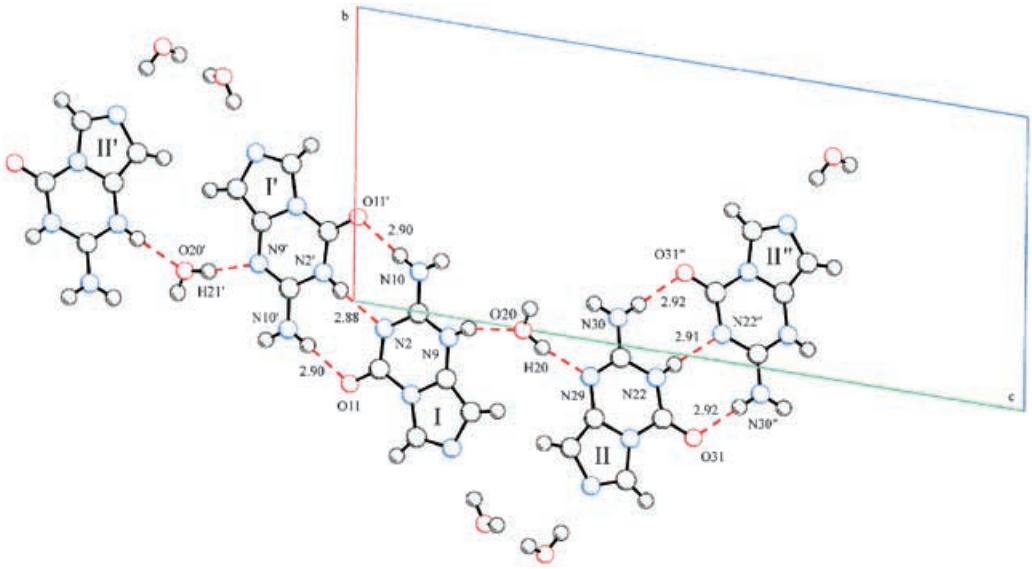


Fig. 5. *Packing of alloG in the crystal* (view down *a*) *visualizing the co-existence of the two tautomers and the reverse*-Watson–Crick *pairing of molecules I and II with their symmetry mates* [14]. Primed/double-primed atoms are related to unprimed atoms by the crystallographic symmetry operations $1-x, -y, -z$ (inversion center between N(2) and N(2′)) and $1-x, -y, 1-z$ (inversion center between N(22) and N(22″)); molecules I and I′ (II and II″) are coplanar. The H-atoms at N(2′) and N(22) are disordered about the centers of symmetry. H-Bond donor/acceptor distances in Å. For each disordered site only one orientation is shown.

N(2′), and O(11′) (N(30′), N(22′), and O(31′)) in reverse-*Watson–Crick* manner within H-bonding distance of N(10), N(2), and O(11) (N(30), N(22), and O(31))[4]). These distances demand that the H-atoms at N(2) and N(22), respectively, have to be disordered about those centers of symmetry, and that the molecules must be present in a 0.5:0.5 ratio in both tautomeric forms. This disorder is propagated through the adjacent atom O20[5]). After the final least-squares cycles with a model that is complete except for the H-atoms in question, the first four peaks in the resulting DF map are located at the positions that correspond to N(1) and N(3) of the allopurine ring with approximately equal peak heights (*Fig. 6*). The average number of electrons per Å^3 for the top six peaks (including the H-atoms at O(20)) is half of the number one obtains when omitting a C-bonded H-atom from the calculations (0.79). A similar distribution of the residual electron density is observed even if the space group symmetry is lowered to *P*1. Here, too, the structure is non-ordered with half-occupied H-positions and, therefore, we believe that it is best described in the centrosymmetric space group *P*-1. *Fig. 7* gives the transcription of the structural observation of the alloG/alloG interaction into the corresponding chemical formula.

---

[4]) Note that the atom numbering used in describing the crystal structure does not correspond to the conventional purine numbering in the chemical formula of alloG.

[5]) A similar behavior has been described, *e.g.*, for isocytosine carboxylic acids [12].

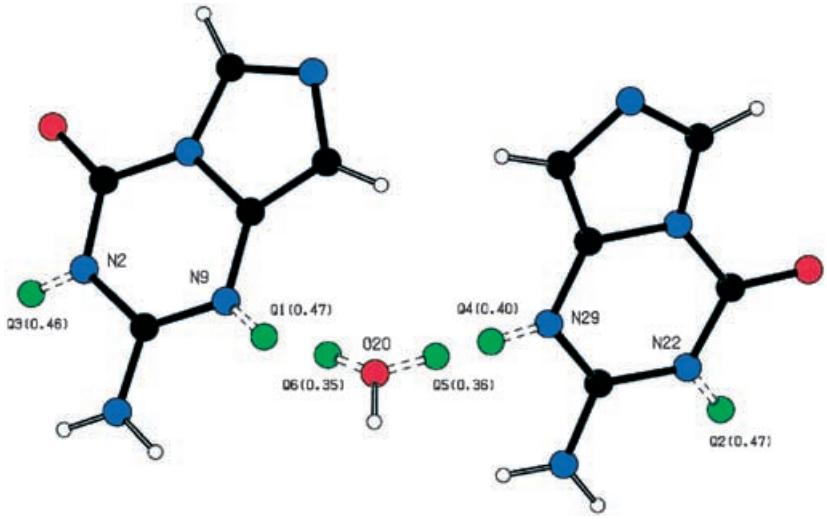


Fig. 6. *Residual electron-density peaks* (green) *in the final refinement stages demonstrate the co-existence of two tautomeric forms* (numbers in parentheses in electrons per Å^3). In this crystal structure, a fully occupied H-atom can be expected to appear with a peak height of *ca.* 0.8 e$^-$/Å^3 in the DF map.

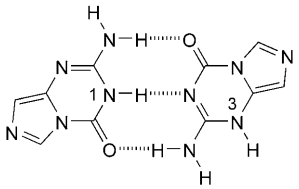


Fig. 7. *Triply H-bonded motif observed for both independent molecules of alloG in the crystal.* One half of the molecules are present in the NH(1) tautomeric form and the other half in the NH(3) form (allopurine numbering).

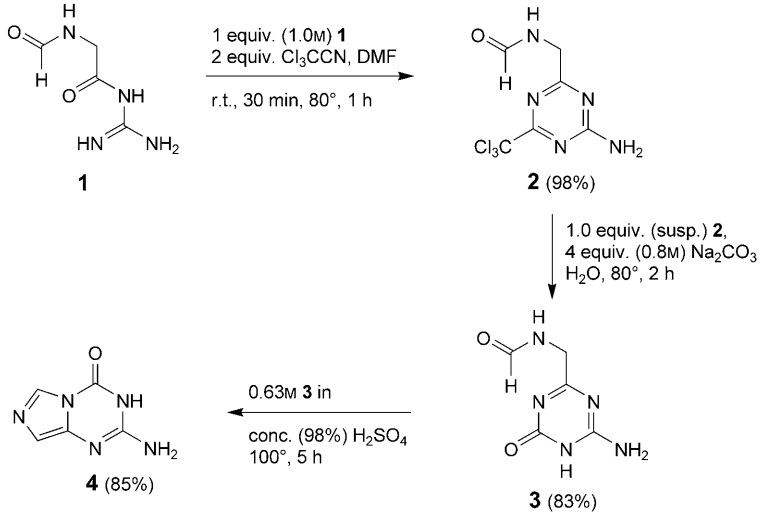
We have not performed any spectroscopic and/or chemical studies that would reveal the position of the equilibrium between the two alloG tautomers in solution. Since an essential part of the stabilization of the observed reverse-*Watson – Crick* pair in the crystalline state is presumably due to the crystal lattice, no conclusions are to be drawn with regard to the equilibrium between the tautomers in solution. However, from a qualitative organo-chemical point of view, it would seem not unreasonable that in allo-guanine the NH(3) tautomer is more easily accessible relative to the (NH-(1)) tautomer than it is in the natural guanine. The chemical formulae of G and alloG tautomers may be considered to illustrate this reasoning: in the NH(1) tautomer of G (*Fig. 1*), the C(6)=O group enjoys linear conjugation with the exocyclic NH_2 group at C(2), whereas in the corresponding tautomer of alloG it does not. On the other hand, there is linear conjugation between these two functional groups in the NH(3) tautomers of both G (not shown) and alloG (shown). The question to which this qualitative reasoning is referring is clearly one that should be subjected to a genuinely theoretical treatment.

This work was supported by *Novartis AG*, Basel, and the *Skaggs* Foundation (TSRI).

Experimental Part[6])

Procedure Referring to the Scheme: N-*Formyl-glycyl-guanidine* (**1**). This compound was prepared according to *Hoffmann*'s procedure for the synthesis of *N*-formyl-L-valyl-guanidine [15]. Guanidine hydrochloride (15.00 g, 157 mmol) was added to a turbid suspension of MeONa (8.48 g, 157 mmol) in 200 ml of anh. MeOH. The resulting suspension was stirred at r.t. for 40 min. The formed precipitate, NaCl, was filtered off *via Celite* and washed with 50 ml of MeOH. To the colorless filtrate thus obtained was added *N*-formylglycine ethyl ester [16] (41.14 g, 314 mmol, 2.0 equiv.) in 20 ml of MeOH. The mixture was stirred at r.t. for 6 h, during which a white precipitate appeared within min. After cooling at 4° for 2 h, the white solid was filtered, washed with 50 ml of cold MeOH, dried under h.v. (r.t., overnight) to give 15.60 g (69%) of **1** as white solid. ^{1}H-NMR (300 MHz, (D_6)DMSO): 3.67 (*d*, $J = 5.1$, NHCH_2); 6.20–8.40 (5*m*, NH); 8.03 (*s*, CHO). ^{13}C-NMR (75 MHz, (D_6)DMSO): 44.05; 161.07; 162.23; 165.54; 179.40. ESI-MS (pos.): 145.3 (28, $[M+H]^+$), 167.2 (20, $[M+Na]^+$), 289.2 (52, $[2M+H]^+$), 311.2 (100, $[2M+Na]^+$).

Scheme. *Synthesis of 5,8-Diaza-7,9-dicarbapurine* ('Alloguanine'; **4**)



2-Amino-4-[(formylamino)methyl]-6-(trichloromethyl)-1,3,5-triazine (**2**). This compound as well as its hydrolysis product **3** were prepared according to methodology described by *Kelarev et al.* [17]. To a suspension of **1** (14.4 g, 0.1 mol) in 100 ml of DMF was added dropwise of Cl_3CCN (28.8 g, 0.2 mol, 2 equiv.). After adding, the mixture was stirred at r.t. for 30 min until the soln. became clear. After heating at 80° (bath temp.) for 1 h, the light brown soln. was evaporated to dryness. The white solid thus obtained was washed with H_2O (200 ml), dried under h.v. (P_2O_5) to give 26.5 g (98%) of **2** as white solid[7]). ^{1}H-NMR (300 MHz, (D_6)DMSO): 4.29 (*d*, $J =$ 6.0, CH_2); 8.15 (*d*, $J = 1.2$, CHO); 8.23 (br., 1 H, NH_2); 8.27 (br., 1 H, NH_2); 8.51 (*t*, $J = 5.7$, N*H*CHO). ^{13}C-NMR (75 MHz, (D_6)DMSO): 42.89; 95.99; 161.28; 166.87; 172.06; 176.65. ESI-MS (neg.): 268.0 (100, $[M-H]^-$), 270.0 (96), 272.0 (30).

*4-Amino-6-[(formylamino)methyl]-1*H*-[1,3,5]triazin-2-one* (**3**). Compound **2** (13.5 g, 50 mmol) was suspended in 250 ml of 0.8M aq. Na_2CO_3 soln. The mixture was stirred at 80° for 2 h, at which time **2** was

---

6) For general remarks, see, *e.g.*, [7]. See also [3c] and supplemental information referring to it.

7) 2-Amino-4-formylaminomethyl-6-trichloromethyl-s-triazine **2** can be dissolved well in MeOH. It is sparingly soluble in H_2O.

gradually dissolved, and a clear light yellow soln. was obtained. TLC ($CHCl_3$/MeOH, 25% aq. NH_3) showed the consumption of **2** (R_f 0.86) and the formation of a major product (R_f 0.22). The mixture was cooled to 0° and neutralized with conc. aq. HCl to pH 7. The white precipitate was filtered, washed with H_2O (2 × 100 ml), MeOH (2 × 100 ml), and Et_2O (2 × 100 ml) to give 7.1 g (83%) of **3** as white solid[8]). ^{1}H-NMR (300 MHz, (D_6)DMSO): 4.05 (*d*, *J* = 6.0, CH_2); 7.39 (br., 1 H, NH_2); 7.51 (br., 1 H, NH_2); 8.10 (*s*, CHO); 8.36 (*t*, *J* = 5.7, N*H*CHO); 11.54 (br., NH). ^{13}C-NMR (75 MHz, (D_6)DMSO, 60°): 40.48; 156.00; 161.61; 165.06; 168.85. ESI-MS(pos.): 142.3 (48, $[M - CO + H]^+$), 170.2 (36, $[M + H]^+$), 339.2 (100, $[2\,M + H]^+$), 361.1 (38, $[2\,M + Na]^+$), 530.0 (34, $[3\,M + Na]^+$), 698.9 (14, $[4\,M + Na]^+$), 867.9 (5, $[5\,M + Na]^+$).

2-Amino-6-oxo-5,8-diaza-7,9-dicarbapurine (**4**). This compound was prepared following a protocol described by *Kim et al.* [18]. Compound **3** (4.23 g, 25 mmol) was dissolved in 40 ml of 98% H_2SO_4, and the resulting light brown soln. was stirred at 100° for 5 h. After cooling to r.t., the acid soln. was carefully poured into 500 ml of Et_2O with rapid stirring. The Et_2O was decanted and fresh Et_2O was added; this sequence was repeated twice. The precipitated salt was suspended in 100 ml of H_2O, and the soln. was made basic (pH 9) by adding 25% aq. NH_3. To the clear light brown soln. thus obtained, AcOH was added dropwise until pH 6.5. The light brown precipitate formed was filtered, washed with 50 ml of cold H_2O, and dried under h.v. (P_2O_5) to give 3.21 g (85%) of **4** as light brown solid[9]). R_f (CH_2Cl_2/MeOH, 4 : 1) 0.33. UV ($c = 2.50 \times 10^{-4}$ M in pH 7, 0.1M phosphate buffer): 220, (10200), 261 (7300). ^{1}H-NMR (300 MHz, (D_6)DMSO): 6.46 (*s*, NH_2); 6.52 (*d*, *J* = 0.9, H–C(9)); 7.93 (*d*, *J* = 0.9, H–C(7)); 11.17 (br., NH). ^{13}C-NMR (75 MHz, (D_6)DMSO): 111.66 C(9); 125.37 C(7); 137.79; 145.08; 149.45. ESI-MS (neg.): 150.0 (100, $[M - H]^-$), 301.0 (32, $[2\,M - H]^-$), 452.0 (4, $[3\,M - H]^-$).

A sample of 1.80 g (11.9 mmol) of **4** obtained in an analogous experiment was crystallized by dissolving the sample in 150 ml of boiling H_2O and collecting 1.18 g (66%) of colorless crystals at *ca.* 0°. M.p. > 150°.

X-Ray Analysis of **4**. Unit-cell determination and intensity-data collection were performed on a *Bruker AXS* three-circle goniometer with a *SMART 6000 CCD* and graphite-monochromatized CuK_α radiation (λ 1.54178 Å) from a sealed tube generator[10]). A semi-empirical absorption correction was applied, based on the intensities of symmetry-related reflections measured at different angular settings ($T_{min} = 0.8093$, $T_{max} = 0.9468$) [19]. The structure was solved by direct methods and refined by full-matrix least-squares on F^2 [20]. Compound **4** crystallizes with two independent molecules and three molecules of H_2O in the asymmetric unit in the centrosymmetric space group *P*-1. Non-H-atoms were refined with anisotropic displacement parameters, H-atoms at C-atoms were calculated in idealized positions and refined using a riding model. All H-atoms at N- and O-atoms were located in DF maps and refined isotropically with suitable SIMU and SADI restraints. For both independent molecules, two tautomeric forms are present in the crystal lattice, the NH(1) tautomer and the NH(3) tautomer, with a 0.5 : 0.5 distribution. Therefore, the multiplicity of H(2), H(9), H(22), H(29), H(20), and H(21) is 0.5. For not fully occupied H-atoms the displacement parameter was set to be $U_{iso}(H) = 1.2\ U_{iso}(N/O)$.

Crystal Data of **4**. Colorless rod from H_2O, size 0.20 · 0.05 · 0.05 mm^3, $C_5H_8N_5O_{2.5}$ ($C_5H_5N_5O \cdot 1.5\ H_2O$), $M_r = 178.16$ (151.14 + 1.5 · 18.01), triclinic, space group *P*-1 (No. 2) with $a = 3.672(2)$, $b = 9.475(3)$, $c = 21.963(6)$ Å, $\alpha = 98.296(9)$, $\beta = 92.669(9)$, $\gamma = 94.987(9)^0$, $V = 753.2(5)$ Å^3, $Z = 4$, $D_c = 1.570$ g · cm^{-3}, $\mu = 1.105$ mm^{-1}, 15793 reflections measured (ω-scans), 2754 independent ($R_{int} = 0.0376$), $2.04^0 < \theta < 69.14^0$, $T = 100$ K, 280 parameters, 41 restraints, $R_1 = 0.0379$, $wR_2 = 0.0850$ for 2327 reflections with $I > 2\sigma(I)$, $R_1 = 0.0477$, $wR_2 = 0.0901$ for all 2754 data, GoF = 1.072, restrained GoF = 1.065, res. el. dens. = +0.19/−0.24 e · Å^{-3}.

REFERENCES

[1] K. Groebke, J. Hunziker, W. Fraser, L. Peng, U. Diederichsen, K. Zimmermann, A. Holzner, C. Leumann, A. Eschenmoser, *Helv. Chim. Acta* **1998**, *81*, 375.

---

8) Compound **3** is soluble in DMSO, and insoluble in CH_2Cl_2, MeOH, and H_2O.

9) Compound **4** can be dissolved in DMF, DMSO, hot H_2O and hot MeOH. It is not soluble in ether, CH_2Cl_2. It is thermally stable for 10 h at 190°, undergoing slow sublimation.

10) CCDC-267004 contains the supplementary crystallographic data for this paper (excluding structure factors). These data can be obtained free of charge *via* http://www.ccdc.cam.ac.uk/data_request/cif (or from the *Cambridge Crystallographic Data Centre*, 12 Union Road, Cambridge CB2 1EZ, UK; fax: +44-(0)1223-336033; e-mail: deposit@ccdc.cam.ac.uk).

[2] S. H. Kim, D. G. Bartholomew, L. B. Allen, R. K. Robins, G. R. Revankar, P. Dea, *J. Med. Chem.* **1978**, 21, 883.
[3] a) Z. Wang, H. K. Huynh, B. Han, R. Krishnamurthy, A. Eschenmoser, *Org. Lett*. **2003**, *5*, 2067; b) B. Han, Z. Wang, R. Krishnamurthy, B. Jaun, A. Eschenmoser, *Org. Lett*. **2003**, *5*, 2071; c) B. Han, B. Jaun, R. Krishnamurthy, A. Eschenmoser, *Org. Lett*. **2004**, *6*, 3691
[4] B. Han, V. Rajwanshi, J. Nandy, R. Krishnamurthy, A. Eschenmoser, *Synlett* **2005**, 744.
[5] K. Groebke, 'Über Purin-Purin-Paarungen bei Hexopyranose-Nukleinsäuren', Dissertation No. 10149, ETH Zürich, 1993.
[6] R. Krishnamurthy, S. Pitsch, M. Minton, C. Miculka, N. Windhab, A. Eschenmoser, *Angew. Chem.* **1996**, *108*, 1619; *Angew. Chem., Int. Ed.* **1996**, *35*, 1537.
[7] S. Pitsch, S. Wendeborn, R. Krishnamurthy, A. Holzner, M. Minton, M. Bolli, C. Miculka, N. Windhab, R. Micura, M. Stanek, B. Jaun, A. Eschenmoser, *Helv. Chim. Acta* **2003**, *86*, 4270.
[8] A. Rich, D. R. Davies, F. H. C. Crick, J. D. Watson, *J. Mol. Biol.* **1961**, *3*, 71; D. B. Lerner, D. R. Kearns, *Biopolymers* **1981**, *20*, 803; K. A. Hartman Jr., A. Rich, *J. Am. Chem. Soc*. **1965**, *87*, 2033; K. Gehring, J-L. Leroy, M. Guéron, *Nature* **1993**, *363*, 561.
[9] S. Arnott, R. Chandrasekaran, C. M. Marttila, *Biochem. J.* **1974**, *141*, 537; S. B. Zimmerman, G. H. Cohen, D. R. Davis, *J. Mol. Biol*. **1975**, *92*, 181.
[10] J. Sepiol, Z. Kazimierczuk, D. Shugar, *Z. Naturforsch. C* **1976**, *31*, 361.
[11] B. D. Sharma, J. F. McConnell, *Acta Crystallogr*. **1965**, *19*, 797.
[12] L. M. Toledo, K. Musa, J. W. Lauher, F. W. Fowler, *Chem. Mater.* **1995**, *7*, 1639.
[13] A. L. Spek, PLATON, Utrecht University, Utrecht, The Netherlands, 2000.
[14] E. Keller, SCHAKAL, University of Freiburg, Freiburg, Germany, 1999.
[15] E. Hoffmann, D. Diller, *Can. J. Chem.* **1965**, *43*, 3103.
[16] P. K. Martin, H. R. Matthews, H. Rapoport, G. Thyagarajan, *J. Org. Chem*. **1968**, *33*, 3758.
[17] V. I. Kelarev, Ammar Dibi, A. F. Lunin, *Chem. Heterocycl. Compd. (New York)*, **1985**, *21*, 1284 (engl. transl. from *Khim. Geterotsikl. Soedin.* (Russian) **1985**, *21*, 1557).
[18] S. H. Kim, D. G. Bartholomew, L. B. Allen, R. K. Robins, G. R. Revankar, P. Dea, *J. Med. Chem.* **1978**, *21*, 883.
[19] G. M. Sheldrick, SADABS, version 2004/1, University of Göttingen, Göttingen, Germany.
[20] G. M. Sheldrick, SHELXTL, version 6.10, Bruker AXS, Inc., Madison, WI, USA, 2000.

Received April 1, 2005