Analysis of the in vitro antiviral activity of certain ribonucleosides against parainfluenza virus using a novel computer aided receptor modeling procedure

ARTICLE *in* JOURNAL OF MEDICINAL CHEMISTRY · MAY 1989

Impact Factor: 5.45 · DOI: 10.1021/jm00124a005 · Source: PubMed

6 AUTHORS, INCLUDING:

Arup Ghose
Teva Pharmaceutical Industries Ltd.
50 PUBLICATIONS **3,983** CITATIONS
SEE PROFILE

Gordon Crippen
University of Michigan
150 PUBLICATIONS **5,887** CITATIONS
SEE PROFILE

Donald Smee
Utah State University
280 PUBLICATIONS **6,518** CITATIONS
SEE PROFILE

Analysis of the in Vitro Antiviral Activity of Certain Ribonucleosides against Parainfluenza Virus Using a Novel Computer Aided Receptor Modeling Procedure

Arup K. Ghose,* Gordon M. Crippen,† Ganapathi R. Revankar, Patricia A. McKernan, Donald F. Smee, and Roland K. Robins

Nucleic Acid Research Institute, 3300 Hyland Avenue, Costa Mesa, California 92626, and College of Pharmacy, University of Michigan, Ann Arbor, Michigan 48109. *Received July 22, 1988*

The in vitro antiviral activity of 28 nucleosides against the parainfluenza virus type 3 has been analyzed by using a novel computer aided receptor modeling procedure. The method involves an extensive modification of our earlier work (Ghose, A. K.; Crippen, G. M. *J. Med. Chem.* **1985**, *28*, 333). It presents a more straightforward algorithm for the steps that suffered from subjectivity in the earlier method. The method first determines the possible low-energy conformations of the nucleosides, and assigns a priority value for each conformation of each molecule. It then performs the following steps repeatedly, until it finds an acceptable solution. Starting from the conformation of highest priority, the various energetically allowed conformations of the other molecules are superimposed on it. On the basis of the physicochemical property matching (or overlapping), the best superposition is determined. The superimposed molecules are dissected into a minimum number of parts and the local physicochemical properties at different regions are correlated with their binding data (antiviral activity). A modified version of distance geometry has been used for geometric comparison of the structure of the molecules. On the basis of the virus rating (VR) of 28 ribonucleosides, this procedure hypothesized the minimum-energy conformation of 6-(methylthio)-9-β-D-ribofuranosylpurine as a reference conformation and used three physicochemical properties, namely hydrophobicity, molar refractivity, and formal charge density for property matching. The binding-site cavity was divided into seven regions or pockets to differentiate the nature of interaction quantitatively. The model suggests that the 2- and 3-positions of the purine ring and the corresponding atoms of the other rings get some steric repulsion, and nucleosides having a single five-membered heterocyclic ring will better fit this virus. The methylthio group gets a strong attraction from dispersive interaction. Both hydrophilic and dispersive groups are attractive here. Although our calculation supports the previously suggested active conformation of ribavirin, it shows that it is not the global minimum-energy conformation. The difference lies in the orientation of the amide group. The calculated viral rating from this model showed a correlation coefficient of 0.971 with the observed values, and the explained variance and the standard deviation of the fit were 0.880 and 0.125, respectively.

The objective of this work was to develop a complementary model of the binding site of nucleosides to the parainfluenza virus receptor, based upon the virus inhibition data and the chemical structure of the nucleosides. The model could be applied to design novel antiviral agents. The model not only constitutes the geometrical shape of the hypothetical binding site cavity but also gives a quantitative estimation of the interaction of the ligand atoms with the receptor.[1-3] Such a model is very helpful in understanding the ligand–receptor interaction in the absence of the explicit structure of the receptor and the binding site. Not only is the problem of designing such a model very complex[4-7] but one can also question the feasibility of developing any physically realistic model from such limited information. In this introduction we want to consider briefly the various problems of the approach; in the Methods section an approximate solution of the problem will be discussed. We have adopted here an extensively modified procedure from our earlier reports.[1-3] It is very general and can be applied to any comparable problem.

A few important questions associated with this problem can be stated as follows: (1) Does the nucleoside bind at a single particular region of the receptor or is there more than one binding site? If it binds at more than one binding site, are these sites equally responsible for the antiviral activity? (2) Does one nucleoside molecule bind with the receptor in one and only one conformation? (3) Among the enormous number of moderately low energy conformations of the nucleosides, how does one determine the binding conformations of the various nucleosides? (4) How can one determine the relative orientation of the nucleosides at the binding site? (5) How can one formulate the binding-site cavity from the hypothetical binding conformation of the nucleosides and their relative orientation at the binding site? (6) How much can we learn about the receptor from the structure of the nucleoside and its binding data? (7) How can we model the interaction of the nucleosides with the receptor in absence of the explicit structure of the receptor? The present work will not answer all these questions. It may answer only a few questions on the basis of simplifying assumptions. When the antiviral activity is not simply the result of the binding of the nucleosides with the receptor, other factors such as the degree of drug metabolism may be important. These factors are not considered in the pesent study, except that we wanted to make the variation along those lines minimum. For example, many simple heterocyclic compounds without the sugar ring show antiviral activity because they get ribosylated in the cell. We did not include those data since the activity did not come from the heterocyclic ring alone. Although the nucleosides usually get phosphorylated to nucleotide before showing the antiviral activity, we took nucleosides only, for computational convenience and assumed that the factor is constant for all the compounds.

An algorithm for the modeling of the active-site cavity can be developed if we consider the problem in the reverse order. The interaction of the ligand (nucleoside) with the

† University of Michigan.

(1) Ghose, A. K.; Crippen, G. M. *J. Med. Chem.* **1984**, *27*, 901-914.
(2) Ghose, A. K.; Crippen, G. M. *J. Med. Chem.* **1985**, *28*, 333-346.
(3) (a) Ghose, A. K.; Crippen, G. M. *Comprehensive Medical Chemistry*; Ramsden, C. A. Ed.; Pergamon Press, in press; Vol. 4. (b) Ghose, A. K.; Crippen, G. M. *Book of Abstracts*, 193rd National Meeting of the American Chemical Society, Division of Computers in Chemistry, Denver, Colorado, 1987; American Chemical Society: Washington, DC, 1987; Abstract 5.
(4) Marshall, G. R. *Annu. Rev. Pharmacol. Toxicol.* **1987**, *27*, 193-213.
(5) Gund, P.; Halgren, T. A.; Smith, G. M. *Annu. Rep. Med. Chem.* **1987**, *22*, 269-279.
(6) Cohen, N. C. *Adv. Drug Res.* **1985**, *14*, 41-145.
(7) Hopfinger, A. J. *J. Med. Chem.* **1985**, *28*, 1133-1139

0022-2623/89/1832-0746$01.50/0

receptor ($E_{l,r}$) is a complex function of the properties of the ligand as well as the receptor:

$$E_{l,r} = F(p_1^l, p_2^l, ..., p_1^r, p_2^r, ...) \quad (1)$$

where p's are the physicochemical or structural properties, and the superscripts l and r represent the ligand and the receptor, respectively. If the function is such that properties of the ligand and the receptor can be separated as shown in eq 2 or 3, then the part characterized by the

$$E_{l,r} = F_1(p_1^l, p_2^l, ...) \times F_2(p_1^r, p_2^r, ...) \quad (2)$$

$$E_{l,r} = F_1^l(p_1^l) \times F_1^r(p_1^r) + F_2^l(p_2^l) \times F_2^r(p_2^r) + ... \quad (3)$$

receptor may be estimated by optimizing the function representing the goodness of fit of the calculated and observed binding energies. Some requirements for the separation were discussed in our earlier publication.[8] One most important requirement is that the interacting region should be small. If the region is considerably large, it is still possible to represent the interaction numerically with an expression like 2 or 3, by dividing the region into several subregions and keeping different expressions for different regions. This is an important concept behind the present work or any quantitative structure–activity relationships; however, we advise the reader not to be overly concerned with these equations at this point.

The basic forces for nonbonded molecular interactions are of broadly two types: electrostatic and dispersive. However, hydrophobic interaction which is the entropic consequence of these forces also plays a very important role in the biological interaction. These three basic interactions can be modeled with three physicochemical properties of the ligands, namely, octanol–water partition coefficient[9-11] for hydrophobic interaction, molar refractivity[8] for dispersive interaction, and formal charge density or electrostatic potential[12,13] for electrostatic interaction. Since the interaction at various regions of the active receptor site is possibly different, in order to use any optimization program to evaluate the receptor-dependent part we have to know the relative orientation of the ligands at the active site. In the Methods section we have described an algorithm to make an educated guess of the relative mode of binding of the ligands at the active site and modeling the ligand–receptor interaction to explain the binding energy of the ligands quantitatively.

Methods

The various steps in this approach may be summarized as follows: (1) generation of a three-dimensional structure of the ligand molecule, (2) assignment of the physicochemical properties of the atoms in the molecule, (3) evaluation of the low-energy regions of the conformational space for each ligand, (4) determination of the geometrically possible superpositions of the ligands, (5) decision of the best relative superposition of the ligands, (6) generation of the approximate structure of the active-site cavity from the relative superposition of the ligands, (7) dissection of the active-site cavity to differentiate the nature of interaction at different regions, and to explain the binding energy of the ligands quantitatively. These steps are explained in detail below.

(1) Construction of the molecule satisfying the normal (crystallographic) bond lengths and bond angles was done by using our MOLGEN program[1] and crystallographic fragment library.

(2) The assignment of the atomic physicochemical properties was made by classifying the atoms into different types with our CLASIF program.[8,10,11] The charge density was calculated by using CNDO/2 method[14] on an arbitrary conformation having the most elongated structure.

(3) In the next step the possible low-energy conformations in the entire conformational space were determined. This is one of the slowest steps in this approach, but it is performed only once. We used our fixed valance structure conformational analysis program CONFOR as outlined previously.[1] In this program we used MM2 (1977) parameters for van der Waals interaction and CNDO/2 charge density for electrostatic interaction. The torsional parameters were collected from the literature.[1] The conformational energy was initially minimized by using the pattern-search technique,[15] until the process converged. Next a combinatorial conformational search with an incremental value of 20° for each (important) dihedral angle was performed. The conformations having energy less than or equal to 5 kcal/mol were recorded on a file for each molecule.

(4) Next the various ways of superimposing the molecules were evaluated. However, there are an infinite number of ways of superimposing a molecule upon another. Unless one is conservative, this step becomes extremely lengthy and impractical. We wrote a program, STRUCOMP, which used conformations having energy less than a preassigned value or a specified number of lowest energy conformations. Each acceptable conformation of the various ligands was given a priority value:

$$P = E_{obs} + E_c \quad (4)$$

where E_{obs} is the observed binding free energy (or any biological activity data which is directly proportional to the binding energy) and E_c is the conformational energy compared to the global minimum energy. This function has two important properties; for a particular compound it gives highest priority to the conformation having lowest energy (the lower the value of P, the higher is the priority). Among the global minimum energy conformations of the various molecules, it gives highest priority to the molecule having the lowest binding free energy. When neither of these conditions are maintained, it gives a relative priority of the conformations in different molecules. The priority function was used to pick up a possible active conformation (the reference structure for the comparison purpose). However, its ultimate acceptance was determined by its success in quantitative explanation of the binding data. A small number of conformations having the lowest P values were selected as possible candidates (reference) for the structural comparison. Each conformation ultimately led to a possible solution of the problem. Once a conformation was accepted for comparison, the low-energy conformations of the other molecules were superimposed on

(8) Ghose, A. K.; Crippen, G. M. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 21–35.
(9) Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; Wiley: New York, 1979.
(10) Ghose, A. K.; Crippen, G. M. *J. Comput. Chem.* **1986**, *7*, 565–577.
(11) Ghose, A. K.; Pritchett, A.; Crippen, G. M. *J. Comput. Chem.* **1988**, *9*, 80–90.
(12) *Quantum Chemistry in the Biomedical Sciences*; Weinstein, H., Green, J. P., Eds.; New York Academy of Sciences: New York, 1981.
(13) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. In *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.
(14) Pople, J. A.; Beveridge, D. L. In *Approximate Molecular Orbital Theory*; McGraw-Hill: New York, 1970.
(15) Hooke, R.; Jeeves, T. A. *J. Assoc. Comput. Mach.* **1961**, *8*, 212.

this molecule (see Appendix I for detailed steps). This program gave all geometrically possible superpositions.

(5) At this point we have a large number of geometrically possible superpositions of each molecule on each reference structure (hypothesized active conformation). Determination of the best superposition on a particular reference structure constitutes the next part. The best superposition may be considered as the representation of the relative binding mode at the active site. If such decision can be made, one can get two important pieces of information: (i) the approximate structure of the active-site cavity from the location of the atoms of the ligand molecules, and (ii) the complementary feature of the active site from the correlation of the various physicochemical properties at different regions of the ligands with their binding energies. However, this is an interlocking problem. The same set of ligand molecules may bind to more than one biological receptor, and the relative binding modes of the ligands may not be the same in the two receptors, since the complementary structure of the receptor dictates the binding mode. In other words, we cannot determine the relative binding modes of the ligands on the basis of the physicochemical properties of the ligands alone. The best way to solve the problem may be to use an iterative approach: use the physicochemical properties of the ligand alone to make the first decision about their relative superpositions (binding mode at the active site); correlate the physicochemical properties of the ligands at the various parts of the hypothetical active-site cavity with their binding energy to determine the receptor-dependent properties; use the receptor-dependent properties to reevaluate the relative binding modes; and continue the process until it converges. In the present algorithm, however, we have not adopted the iterative approach and left it for future implementation. In order to measure the goodness of superposition, it is necessary to have a mathematical function. The mathematical function can be formulated in various ways. We used two different functions to evaluate the best superposition (see Appendix II for details). The program OPTSUP evaluated the superposition having best value of these functions for each molecule on a particular reference structure.

(6) The next objective of our calculation was to evaluate the active-site cavity on the basis of the reference conformation and superposition of the ligand molecules as obtained in the previous step. The program SITBLD evaluated the site cavity in terms of the minimum number of site pockets necessary to accommodate all atoms from all ligands. For that, it started with all atoms of the reference structure as the position of site pockets, and then included the atoms of the superimposed molecules only if they were away from the already accepted points by a distance of 2δ or more where δ was the distance limit of acceptance for superposition of the atoms. It then evaluated which atom of the various ligand molecules goes to which site pocket.

(7) The large number of site pockets developed in the previous step should be given different types to differentiate the nature of interaction at different regions. The calculated binding energy (E_{calc}) of a ligand at the active site is given by the eq 5,[2] which is a revised form of eq 3 so as to consider the difference of interaction at different regions. Here E_c is the energy of the conformation under consideration. C's are the site and the physicochemical property dependent adjustable parameters to be determined by some optimization technique, i' is the type of the site i, n_s represents the number of site pockets, n_p represents the number of physicochemical parameters used to model the intermolecular interaction, n_0 represents the number of atoms occupying the ith site pocket, and $P_j(t_k)$ is the jth physicochemical property of the atom type t_k occupying the site pocket. In order to get statistically acceptable values, the total number of adjustable parameters in the fitting study should be much lower than the number of ligand molecules. Each site type and each physicochemical property in eq 5 need one adjustable parameter. Since we prefer to model the interaction in terms of three basic physicochemical properties, namely the octanol–water partition coefficient, molar refractivity, and atomic charge density, the number of site types n_t should be less than (number of molecules/3). In order to give the large number of site pockets the limited number of types n_t, we first picked n_t number of site pockets, hence termed primary site pockets. The rest of the secondary site pockets were given the type of the closest primary site pocket. The guideline for picking the primary site pockets is still not very certain and may need more extensive reasearch. One major mathematical limitation is that it should not lead to linearly dependent parameters for two or more site types. We used two properties of the site pockets to select the primary ones, namely, the number of ligand molecules using the site pocket, or the correlation coefficient of the physicochemical properties of the ligand atoms occupying the site pocket with the overall binding energy. In the first approach the site pockets used by the maximum number of ligands are considered as primary site pockets. In the second approach, the site points giving the maximum sum of the magnitude of correlation coefficients of the physicochemical properties of the ligands with the biological activity were considered as the primary site pockets. Although the second approach seems to be more attractive, in the ultimate correlation study it does not always give the best statistic for the fit since the physicochemical properties are changed extensively when the neighboring secondary site points are included. The program MDLINT not only does all these jobs but it deletes most unnecessary physicochemical parameters and unimportant part of the receptor-site cavity from the expression of the interaction energy (biological activity) using reverse stepwise regression.[16]

$$E_{calc} = -CE_c + \sum_{i=1}^{n_s}\sum_{j=1}^{n_p}[C_{i'j}\sum_{k=1}^{n_0}P_j(t_k)] \qquad (5)$$

Critique of the Method. To construct the active-site pocket in step 5, we took all atoms of all molecules which are not within a preassigned distance from the already accepted points. This approach may suffer from the following difficulties.

(1) In the binding process a part of the ligand molecule may remain outside the binding-site cavity (in the biophase) and inclusion of that part during the construction of the model for the binding-site cavity may not be desirable. However, the evaluated expression for the ligand–receptor interaction gives some answer related to this problem.

(2) The part which was not superimposed on the reference structure may attain several conformations. In our calculation we used the conformation having minimum energy. However, the steric requirements of the binding-site cavity may force it to attain a different conformation.

Since the computer time necessary in the steps between the choice of the reference structure and explanation of the binding energy is lengthy, it may not be possible to

(16) Snedecor, G. W.; Cochran, W. G. *Stastical Methods*; 6th ed.; The Iowa State University Press: Ames, 1967; pp 381–418.

study too many reference structures. We plan to use modified distance geometry[17] to speed the calculation and provide a more educated guess of the reference structure in the near future.

When the biological activity is not a true binding energy of the ligand with the receptor, and therefore has different physical units, the priority values will change, if the unit is changed. Medicinal chemists often have to accept that situation. In such a situation, we recommend to use the same unit of biological activity to compare the results of two different studies.

At the boundary between different types of site pockets in the model, the interaction changes discontinuously, which is far from reality; a continuous function for the interction is more desirable. However, many molecular interactions change sharply with distance, if not discontinuously.

The ionization of acidic ligands or the protonation of the basic compounds may alter the biological activity. Since most of the compounds considered here were neither strongly acidic nor strongly basic, no such consideration was given. However, one can use P^{H}–P^{ka} or simply P^{ka} (for fixed P^{H} binding studies) as an extra parameter in the expression for calculated binding energy (eq 5) to account for that factor.

Results and Discussion

The present method has been applied to a set of nucleosides active against parainfluenza virus type 3. This is the first effort to study antiviral activity of nucleosides by using any computer aided receptor modeling technique. Antiviral nucleosides are becoming very important in the treatment of viral infections.[18-20] Unfortunately, the mechanism of action of these agents is usually not very well understood.[21,22] In the present work we used the virus rating[23,24] as the antiviral activity parameter. The 28 ribonucleosides active against parainfluenza virus type 3 in Vero cells are shown in Table I. Although the synthesis and antiviral activity of most of these compounds have previously been published,[24-27] the antiviral activity used in the present work is based on only the virus rating against parainfluenza virus in Vero cells (Table I) to lessen the biological variation from different host cell interactions.

The structures of the various molecules and their corresponding atom numberings are shown in Figure 1. The initial valence structures of the molecules were determined from a crystallographic fragment library containing the 2′-endo 3′-exo (type S) ribose ring[28] and various heterocyclic rings as obtained from the literature.[29-34] To discuss the conformational results we used the terminology suggested by Sundaralingam et al.[35] During the conformational analysis the furanoside ring was kept rigid at its crystallographic structure. Although all its substituents were rotated during the pattern search[15] energy minimization, during the combinatorial search only three to four bonds were rotated. These bonds were the glycosyl $C_{1'}$–X bond, where X represents the attached atom of the heterocyclic ring; the backbone $C_{4'}$–$C_{5'}$ and $C_{5'}$–O bonds; and any one important dihedral angle involving the substituent of the heterocyclic ring. The $C_{5'}$–O bond was rotated from 60° to 300° at an increment of 120°. All other bonds were rotated from 0° to 340° at an interval of 20°. The reason for the limited rotation of this group was the result of compromise between two opposing factors. The OH group often forms hydrogen bond with the heterocyclic ring atoms to form a stable structure, and therefore needs rotation to explore such possibility. However, in the latter part of the study, it was necessary to focus on a limited number of lowest energy conformations. Since the rotation of the OH group usually does not take much energy, it makes each low-energy region as a collection of many conformations arising from the rotation of the OH group. Structurally such conformations are not very different and should not be considered as different structures, otherwise one may miss some important conformations in the process of selecting a limited number of conformations. The conformational analysis of nucleosides and nucleotides is a well-studied subject.[35-40] It is also beyond the scope of this paper to discuss the details of our conformational results. Certain important aspects of our conformational calculation are summarized below.

(i) Both the pattern search and the combinatorial search led to the same minimum-energy conformation in 24 out of 28 molecules. In the four molecules the new minimum-energy conformations were only a few tenth of kcal/mol energy less than the old (pattern search) minimum. The minimum-energy conformations of the various molecules as obtained from the combinatorial search, are

(17) Ghose, A. K.; Crippen, G. M. *J. Comput. Chem.* **1985**, *6*, 350–359.
(18) Robins, R. K. *Chem. Eng. News* **1986**, *64*, January 27, 28–40.
(19) Shugar, D. In *Viral Chemotherapy, Vol. II, International Encyclopedia of Pharmacology and Therapeutics*; Pergamon Press: New York, 1986.
(20) Glasso, G. J.; Merigan, T. C.; Buchanan, R. A. In *Antiviral Agents and Viral Diseases of Man*, 2nd ed.; Raven Press: New York, 1984.
(21) Miller, J. P.; Kigwana, L. J.; Streeter, D. G.; Robins, R. K.; Simon, L. N.; Roboz, J. *Ann. N.Y. Acad. Sci.* **1974**, *284*, 211–229.
(22) Smith, R. A. In *Ribavirin: A Broad Spectrum Antiviral Agent*; Smith, R. A., Kirkpatric, W., Eds.; Academic Press: New York, 1980; pp 99–118.
(23) Sidwell, R. W.; Huffman, J. H. *Appl. Microbiol.* **1971**, *22*, 797–801.
(24) Sidwell, R. W.; Allen, L. B.; Huffman, J. H.; Witkowski, J. T.; Cook, P. D.; Tolman, R. L.; Revankar, G. R.; Simon, L. N.; Robins, R. K. *Chemotherapy* **1976**, *6*, 279–294.
(25) Kirsi, J. J.; North, J. A.; Mckernan, P. A.; Murray, B. K.; Canonico, P. G.; Huggins, J. W.; Srivastava, P. C.; Robins, R. K. *Antimicrob. Agents Chemother.* **1983**, *24*, 353–361.
(26) Srivastava, P. C.; Pickering, M. V.; Allen, L. B.; Streeter, D. G.; Campbell, M. T.; Witkowski, J. T.; Sidwell, R. W.; Robins, R. K. *J. Med. Chem.* **1977**, *20*, 257–262.
(27) Petrie, C. R., III; Revankar, G. R.; Dalley, N. K.; George, R. D.; McKernan, P. A.; Hamill, R. L.; Robins, R. K. *J. Med. Chem.* **1986**, *29*, 268–278.
(28) Low, J. N. *Acta Crystallogr.* **1983**, *C39*, 796–798.
(29) Jeffrey, G. A.; Sundaralingam, M. *Adv. Carbohydr. Chem. Biochem.* **1985**, *43*, 203–421.
(30) Prusiner, P.; Sundaralingam, M. *Acta Crystallogr.* **1976**, *B32*, 419–426.
(31) Goldstein, B. M.; Takusagawa, F.; Berman, H. M; Srivastava, P. C.; Robins, R. K. *J. Am. Chem. Soc.* **1983**, 7416–7422.
(32) Gadret, P. M.; Goursolle, M.; Leger, J. M. *Acta Crystallogr.* **1974**, *B30*, 1598–1602.
(33) Larson, S. B. Nucleic Acid Research Institute, personal communication.
(34) Cambridge Structural Database, Cambridge Crystallographic Data Center, University Chemical Laboratory, Cambridge, England.
(35) Altona, C.; Sundaralingam, M. *J. Am. Chem. Soc.* **1972**, *94*, 8205–8212.
(36) Miles, D. L.; Miles, D. W.; Eyring, H. *Biochim. Biophys. Acta* **1978**, *518*, 17–30.
(37) Ponnuswamy, P. K.; Anukanth, A. *J. Theor. Biol.* **1982**, *96*, 233–251.
(38) Miles, D. W.; Revankar, G. R.; Robins, R. K. *J. Phys. Chem.* **1983**, *87*, 2444–2450.
(39) Masojidkova, M.; Zajicek, J.; Budesinsky, M.; Rosenberg, I.; Holy, A. *Collect. Czech. Chem. Commun.* **1985**, *50*, 1899–1905.
(40) Pearlman, D. A.; Kim, S. H. *J. Biomol. Struct. Dyn.* **1985**, *3*, 85–98.

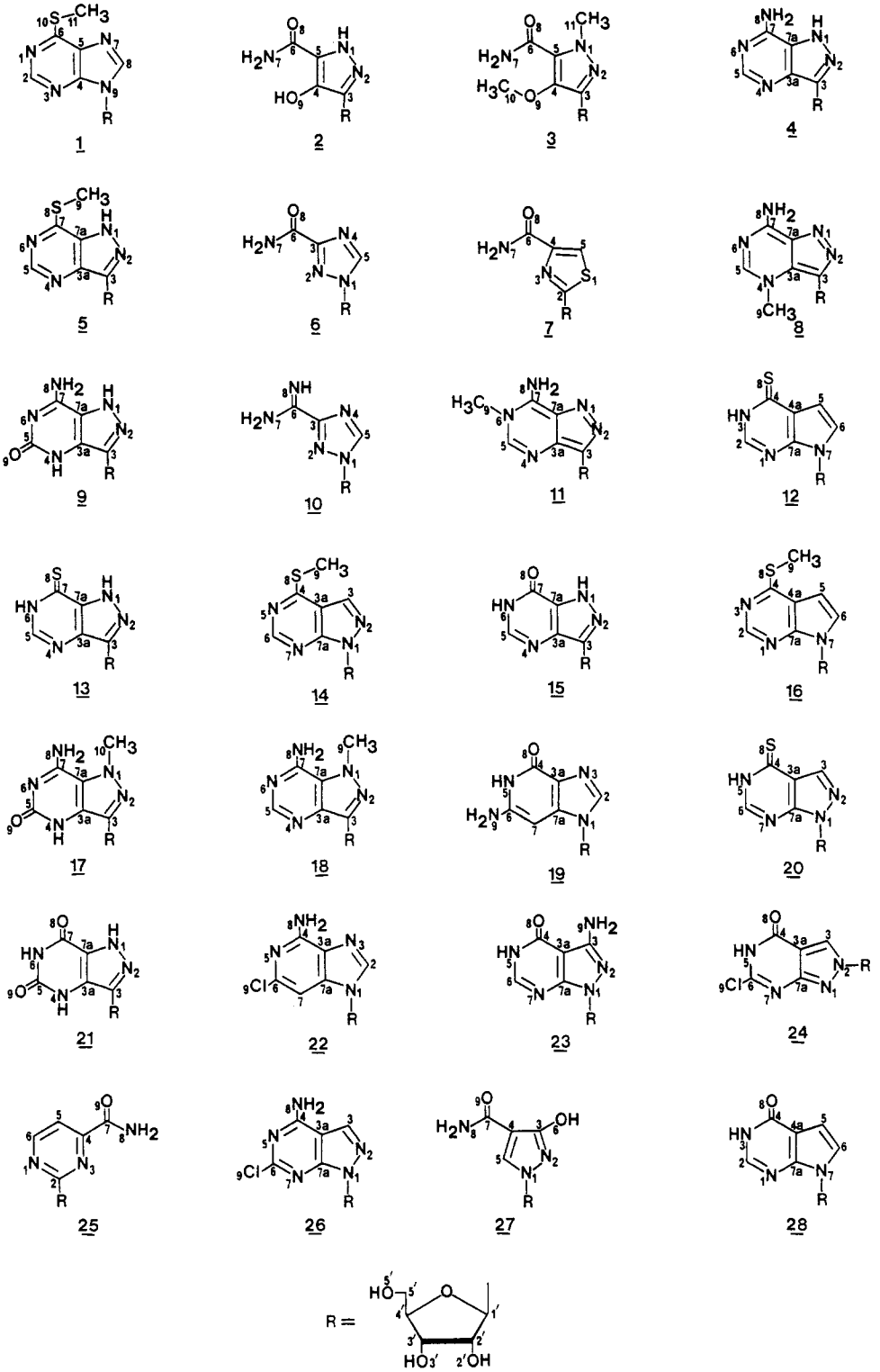

Figure 1. The structure and atom numbering of the various nucleosides used in the present study. For non-hydrogen atoms the subscripted digits belong to the atom label; for hydrogen atoms the subscript indicates the number of atoms.

given in Table II. The binding (active) conformations are summarized in Table III.

(ii) The molecules having a single five-membered heterocyclic base ring (e.g., **2**, **3**, **6**, **7**, **10**, **27**) follow similar conformational patterns. A bulky atom (such as S) or negatively charged atom (such as N) at the 2-position relative to the ribofuranosyl group is stabilized at the anti conformation relative to the O_1'–C_1'. The conformational effect may be reversed if a bulky group like CH_3O is introduced at the 5-position of the heterocyclic ring. The effect of an OH group is not sufficient to reverse this effect. Although the rotational barrier of most of these heterocyclic rings is less than 5 kcal/mol, for compound **3**, with a methoxyl group at the 5-position, the rotation seems to be partly restricted at the high syn conformations.

(iii) When a six-membered ring is fused at the 2,3-position of the five-membered heterocyclic ring, most of the molecules exist in the anti to high anti conformation. The rotational barrier comes from the high syn conformations and varies considerably with the substituent of the 3-position of the purine ring or its equivalent position in other heterocyclic ring, e.g., the molecule **8** is almost frozen

Table I. Various Azole Nucleoside Inhibitors of Parainfluenza Virus Type 3 Used in the Modeling Study

no.	nucleosides[b]	virus rating[a] obs	cal	dif
1	6-(methylthio)-9-β-D-ribofuranosylpurine	2.20	2.16	−0.04
2	4-hydroxy-3-β-D-ribofuranosyl-1*H*-pyrazole-5-carboxamide (pyrazofurin)	1.65	1.77	0.12
3	1-methyl-4-methoxy-3-β-D-ribofuranosyl-1*H*-pyrazole-5-carboxamide	1.65	1.58	−0.07
4	7-amino-3-β-D-ribofuranosyl-1*H*-pyrazolo[4,3-*d*]pyrimidine (formycin A)	1.40	1.22	−0.18
5	7-(methylthio)-3-β-D-ribofuranosyl-1*H*-pyrazolo[4,3-*d*]pyrimidine	1.35	1.39	0.04
6	1-β-D-ribofuranosyl-1*H*-1,2,4-triazole-3-carboxamide (ribavirin)	1.30	1.39	0.09
7	2-β-D-ribofuranosylthiazole-4-carboxamide (tiazofurin)	1.30	1.16	−0.14
8	7-amino-4-methyl-3-β-D-ribofuranosylpyrazolo[4,3-*d*]pyrimidine (4-methylformycin A)	1.28	1.32	0.04
9	7-amino-3-β-D-ribofuranosyl-1*H*-pyrazolo[4,3-*d*]pyrimidin-5(4*H*)-one (oxoformycin A)	1.22	1.01	−0.21
10	1-β-D-ribofuranosyl-1*H*-1,2,4-triazole-3-carboxamidine	1.20	1.17	−0.03
11	7-amino-6-methyl-3-β-D-ribofuranosylpyrazolo[4,3-*d*]pyrimidine (6-methylformycin A)	1.16	1.17	0.01
12	7-β-D-ribofuranosylpyrrolo[2,3-*d*]pyrimidine-4(3*H*)-thione	1.15	0.92	−0.23
13	1,6-dihydro-3-β-D-ribofuranosyl-7*H*-pyrazolo[4,3-*d*]pyrimidine-7-thione (thioformycin B)	1.10	1.26	0.16
14	4-(methylthio)-1-β-D-ribofuranosylpyrazolo[3,4-*d*]pyrimidine	1.05	0.88	−0.17
15	3-β-D-ribofuranosyl-1*H*-pyrazolo[4,3-*d*]pyrimidin-7(6*H*)-one (formycin B)	0.94	0.84	−0.10
16	4-(methylthio)-7-β-D-ribofuranosylpyrrolo[2,3-*d*]pyrimidine	0.90	1.05	0.15
17	7-amino-1-methyl-3-β-D-ribofuranosylpyrazolo[4,3-*d*]pyrimidin-5(4*H*)-one	0.88	0.94	0.06
18	7-amino-1-methyl-3-β-D-ribofuranosylpyrazolo[4,3-*d*]pyrimidine (1-methylformycin A)	0.86	1.08	0.22
19	6-Amino-1-β-D-ribofuranosylimidazo[4,5-*c*]pyrimidin-4-one (3-deazaguanosine)	0.80	0.74	−0.06
20	1-β-D-ribofuranosylpyrazolo[3,4-*d*]pyrimidine-4(5*H*)-thione	0.70	0.77	0.07
21	3-β-D-ribofuranosyl-1*H*-pyrazolo[4,3-*d*]pyrimidine-5,7(4*H*,6*H*)-dione (oxoformycin B)	0.68	0.75	0.07
22	4-amino-6-chloro-1-β-D-ribofuranosylimidazo[4,5-*c*]pyridine (2-chloro-3-deazaadenosine)	0.35	0.27	−0.08
23	3-amino-1-β-D-ribofuranosylpyrazolo[3,4-*d*]pyrimidin-4(5*H*)-one	0.28	0.22	−0.06
24	6-chloro-2-β-D-ribofuranosylpyrazolo[3,4-*d*]pyrimidin-4(5*H*)-one	0.25	0.24	−0.01
25	2-β-D-ribofuranosylpyrimidine-4-carboxamide	0.25	0.24	−0.01
26	4-amino-6-chloro-1-β-D-ribofuranosylpyrazolo[3,4-*d*]pyrimidine	0.15	0.24	0.09
27	3-hydroxy-1-β-D-ribofuranosylpyrazole-4-carboxamide	0.15	0.12	−0.03
28	7-β-D-ribofuranosylpyrrolo[2,3-*d*]pyrimidin-4(3*H*)-one	0.08	0.36	0.28

[a] See ref 23 for the definition of virus rating. obs stands for the observed virus rating, cal stands for the calculated one, and dif represents the difference between the observed and calculated values. [b] See Figure 1 for the structure of the compounds.

Table II. Description of the Minimum-Energy Conformations of the Molecules

compd no.	torsion angles,[a] deg
1	$\omega_1 = 60$, $\omega_2 = 60$, $\omega_3 = 180$, $\omega_4 = 180$, ω(4–N9–C1′–O1′) = 200, ω(5–6–S10–C11) = 80, ω(6–S10–C11–H) = 60
2	$\omega_1 = 300$, $\omega_2 = 60$, $\omega_3 = 180$, $\omega_4 = 60$, ω(4–3–C1′–O1′) = 60, ω(3–4–O9–H) = 180, ω(5–6–N7–H) = 0, ω(4–5–7–O8) = 0
3	$\omega_1 = 60$, $\omega_2 = 60$, $\omega_3 = 180$, $\omega_4 = 60$, ω(4–3–C1′–O1′) = 240, ω(N2–N1–C11–H) = 0, ω(3–4–O9–C10) = 80, ω(4–O9–C10–H) = 180, ω(5–6–*n*7–H) = 0, ω(4–5–6–O8) = 180
4	$\omega_1 = 60$, $\omega_2 = 300$, $\omega_3 = 60$, $\omega_4 = 180$, ω(3a–3–C1′–O1′) = 240, ω(7a–7–N8–H) = 160
5	$\omega_1 = 60$, $\omega_2 = 60$, $\omega_3 = 180$, $\omega_4 = 60$, ω(3a–3–C1′–O1′) = 240, ω(7a–7–S8–C9) = 80, ω(7–S8–C9–H) = 60
6	$\omega_1 = 60$, $\omega_2 = 60$, $\omega_3 = 180$, $\omega_4 = 60$, ω(N2–N1–C1′–O1′) = 200, ω(3–6–N7–H) = 0, ω(O8–6–3–N2) = 0
7	$\omega_1 = 60$, $\omega_2 = 60$, $\omega_3 = 180$, $\omega_4 = 60$, ω(N3–2–C1′–O1′) = 60, ω(4–6–N7–H) = O, ω(N3–4–6–O8) = 180
8	$\omega_1 = 60$, $\omega_2 = 60$, $\omega_3 = 60$, $\omega_4 = 60$, ω(3a–3–C1′–O1′) = 280, ω(5–N4–C9–H) = 0, ω(7a–7–N8–H) = 160
9	$\omega_1 = 60$, $\omega_2 = 60$, $\omega_3 = 180$, $\omega_4 = 180$, ω(3a–3–C1′–O1′) = 240, ω(7a–7–N8–H) = 160
10	$\omega_1 = 60$, $\omega_2 = 60$, $\omega_3 = 180$, $\omega_4 = 60$, ω(N2–N1–C1′–O1′) = 180, ω(N2–3–6–N8) = 0, ω(3–6–N7–H) = 160
11	$\omega_1 = 60$, $\omega_2 = 300$, $\omega_3 = 60$, $\omega_4 = 180$, ω(3a–3–C1′–O1′) = 240, ω(7–N6–C9–H) = 60, ω(7a–7–N8–H) = 160
12	$\omega_1 = 60$, $\omega_2 = 60$, $\omega_3 = 180$, $\omega_4 = 180$, ω(7a–N7–C1′–O1′) = 200
13	$\omega_1 = 60$, $\omega_2 = 60$, $\omega_3 = 180$, $\omega_4 = 180$, ω(3a–3–C1′–O1′) = 240
14	$\omega_1 = 180$, $\omega_2 = 60$, $\omega_3 = 300$, $\omega_4 = 60$, ω(7a–N1–C1′–O1′) = 40, ω(3a–4–S8–C9) = 80, ω(4–S8–C9–H) = 60
15	$\omega_1 = 60$, $\omega_2 = 300$, $\omega_3 = 60$, $\omega_4 = 180$, ω(3a–3–C1′–O1′) = 240
16	$\omega_1 = 60$, $\omega_2 = 60$, $\omega_3 = 180$, $\omega_4 = 180$, ω(7a–N7–C1′–O1′) = 200, ω(4a–4–S8–C9) = 80, ω(4–S8–C9–H) = 60
17	$\omega_1 = 60$, $\omega_2 = 60$, $\omega_3 = 180$, $\omega_4 = 180$, ω(3a–3–C1′–O1′) = 240, ω(7a–7–N8–H) = 160, ω(N2–N1–10–H) = 0
18	$\omega_1 = 60$, $\omega_2 = 60$, $\omega_3 = 60$, $\omega_4 = 60$, ω(3a–3–C1′–O1′) = 240, ω(7a–7–N8–H) = 160, ω(N2–N1–C9–H) = 0
19	$\omega_1 = 60$, $\omega_2 = 60$, $\omega_3 = 180$, $\omega_4 = 180$, ω(7a–N1–C1′–O1′) = 200, ω(N5–6–N9–H) = 160
20	$\omega_1 = 60$, $\omega_2 = 60$, $\omega_3 = 180$, $\omega_4 = 180$, ω(7a–N1–C1′–O1′) = 220
21	$\omega_1 = 60$, $\omega_2 = 60$, $\omega_3 = 180$, $\omega_4 = 180$, ω(3a–3–C1′–O1′) = 240
22	$\omega_1 = 60$, $\omega_2 = 60$, $\omega_3 = 180$, $\omega_4 = 180$, ω(7a–N1–C1′–O1′) = 200, ω(3a–4–N8–H) = 160
23	$\omega_1 = 180$, $\omega_2 = 60$, $\omega_3 = 300$, $\omega_4 = 60$, ω(7a–1–C1′–O1′) = 40, ω(N2–3–N9–H) = 160
24	$\omega_1 = 60$, $\omega_2 = 60$, $\omega_3 = 180$, $\omega_4 = 60$, ω(N1–N2–C1′–O1′) = 220
25	$\omega_1 = 60$, $\omega_2 = 60$, $\omega_3 = 180$, $\omega_4 = 60$, ω(N1–2–C1′–O1′) = 60, ω(4–7–N8–H) = 0, ω(N3–C4–C7–O9) = 180
26	$\omega_1 = 300$, $\omega_2 = 180$, $\omega_3 = 180$, $\omega_4 = 180$, ω(7a–N1–C1′–O1′) = 40, ω(3a–4–N8–H) = 160
27	$\omega_1 = 60$, $\omega_2 = 60$, $\omega_3 = 180$, $\omega_4 = 60$, ω(N2–N1–C1′–O1′) = 200, ω(N2–3–O6–H) = 180, ω(4–7–N8–H) = 0, ω(O9–7–4–3) = 0
28	$\omega_1 = 60$, $\omega_2 = 60$, $\omega_3 = 180$, $\omega_4 = 180$, ω(7a–N1–C1′–O1′) = 200

[a] For the sign of the torsion angle we followed the IUPAC convention: Klyne, W.; Prelog, V. *Experientia* **1960**, *16*, 621; only difference is that our angle varies from 0 to 360°, but in the IUPAC convention, an angle greater than 180 is considered as negative. $\omega_1 = \omega$(C3′–C4′–C5′–O5′), $\omega_2 = \omega$(C2′–C3′–O3′–H), $\omega_3 = \omega$(C_4′–C5′–O5′–H), and $\omega_4 = \omega$(C1′–C2′–O2′–H). For other dihedral angles the numbers within the parentheses represent the atom number as shown in Figure 1.

at the high anti conformation due to the methyl group at the 4-position.

(iv) The rotational barrier of the $C_{4'}$–$C_{5'}$ bond lies between 6.0 and 7.0 kcal/mol. The g+ conformation is usually energetically favored over the g− conformation. Approximate values of the rotational barrier can be obtained in both pattern search and combinatorial search energy minimization. However, the value may be over-

Table III. Description of the Binding Conformations of the Molecules

compd no.	torsion angles,[a] deg	energy, kcal/mol
1	ω_1 = 60, ω_3 = 180, ω(4–N9–C1′–O1′) = 200	0.00
2	ω_1 = 60, ω_3 = 180, ω(4–3–C1′–O1′) = 200, ω(4–5–6–O8) = 0	1.17
3	ω_1 = 60, ω_3 = 180, ω(4–3–C1′–O1′) = 220, ω(4–5–6–O8) = 180	0.60
4	ω_1 = 60, ω_3 = 180, ω(3a–3–C1′–O1′) = 200	0.59
5	ω_1 = 60, ω_3 = 180, ω(3a–3–C1′–O1′) = 200	0.66
6	ω_1 = 60, ω_3 = 180, ω(N2–N1–C1′–O1′) = 200, ω(O8–6–3–N2) = 180	0.73
7	ω_1 = 60, ω_3 = 180, ω(N3–2–C1′–O1′) = 180, ω(N3–4–6–O8) = 180	0.63
8	ω_1 = 60, ω_3 = 60, ω(3a–3–C1′–O1′) = 240	4.91
9	ω_1 = 60, ω_3 = 180, ω(3a–3–C1′–O1′) = 200	0.35
10	ω_1 = 60, ω_3 = 180, ω(N2–N1–C1′–O1′) = 200, ω(N2–3–6–N8) = 180	0.11
11	ω_1 = 60, ω_3 = 180, ω(3a–3–C1′–O1′) = 200	0.77
12	ω_1 = 60, ω_3 = 180, ω(7a–N7–C1′–O1′) = 200	0.00
13	ω_1 = 60, ω_3 = 180, ω(3a–3–C1′–O1′) = 240	0.53
14	ω_1 = 60, ω_3 = 180, ω(7a–N1–C1′–O1′) = 200	0.49
15	ω_1 = 60, ω_3 = 180, ω(3a–3–C1′–O1′) = 200	0.58
16	ω_1 = 60, ω_3 = 180, ω(7a–N7–C1′–O1′) = 200	0.00
17	ω_1 = 60, ω_3 = 180, ω(3a–3–C1′–O1′) = 200	0.37
18	ω_1 = 60, ω_3 = 60, ω(3a–3–C1′–O1′) = 200	0.64
19	ω_1 = 60, ω_3 = 180, ω(7a–N1–C1′–O1′) = 200	0.00
20	ω_1 = 60, ω_3 = 180, ω(7a–N1–C1′–O1′) = 200	0.07
21	ω_1 = 60, ω_3 = 180, ω(3a–3–C1′–O1′) = 200	0.34
22	ω_1 = 60, ω_3 = 180, ω(7a–N1–C1′–O1′) = 200	0.00
23	ω_1 = 60, ω_3 = 180, ω(7a–1–C1′–O1′) = 200	0.81
24	ω_1 = 60, ω_3 = 180, ω(N1–N2–C1′–O1′) = 200	0.11
25	ω_1 = 60, ω_3 = 180, ω(N1–2–C1′–O1′) = 200, ω(N3–4–7–O9) = 180	1.20
26	ω_1 = 60, ω_3 = 180, ω(7a–N1–C1′–O1′) = 200	1.02
27	ω_1 = 60, ω_3 = 180, ω(N2–N1–C1′–O1′) = 200, ω(O9–7–4–3) = 0	0.00
28	ω_1 = 60, ω_3 = 180, ω(7a–N1–C1′–O1′) = 200	0.00

[a] Torison angles, not indicated, were kept fixed at their minimum-energy conformation; see Table II.

estimated if the maximum-energy conformation has van der Waals penetration between movable atoms.

(v) Most of these observations satisfy the existing experimental and theoretical results. Some of the discrepancies resulted from the difference in the sugar ring conformation. The sugar ring of the nucleosides and nucleotides exists in two distinct conformations, $C_{3'}$ endo $C_{2'}$ exo and $C_{3'}$ exo $C_{2'}$ endo. The present conformational results are based on the second type. Apart from that, some interesting differences of our results from the existing X-ray studies are given below. (1) The conformation of the amide group in ribavirin suggests a better hydrogen bonding interaction of the amide group with the N_4 atom rather than the N_2 atom due to the higher electron charge density on the N_4 atom. (2) For tiazofurin the syn conformation is found to be more stable than the anti conformation. We used AMPAC (AM1) molecular orbital package[41,42] to check these results. The molecular orbital calculation supported our molecular mechanics results in both the cases. During this calculation we noticed that some of the bond distances involving the hydrogen atoms were very short in the X-ray data.[31]

In the model-building process the minimum-energy conformation of molecule **1** was found to be the best reference structure in the ultimate fitting of the antiviral activity. Four statistical parameters,[16] viz., standard deviation, correlation coefficient, explained variance, and *f*-test, were used to measure the goodness of fit. This structure is of course the most plausible candidate but not the only possibility. We studied altogether eight reference structures. None of the others were as good as the present model using nucleoside **1** as the reference. This structure was comparable to the suggested active conformation of ribavirin[36] (compound **6**); the only difference was that it possesses an anti conformation very close to the high anti region. The best model was obtained by superposing the molecules on the basis of their property matching using function F_1 (see Appendix II). In the data set there were six nucleosides having single five-membered heterocyclic ring (compounds **2**, **3**, **6**, **7**, **10**, and **27**). Except **27**, the rest of these compounds showed reasonably good antiviral activity in the present system. The superimposed structures suggest that compounds **3**, **6**, **7**, and **10** all satisfy the suggested active conformation of ribavirin.[30,36] However, for compound **2** the carboxamide group has flipped over. The conformation having the usual orientation had the energy 2.5 kcal/mol above the global minimum. Since only 100 lowest energy conformations were studied during superpositions, this conformation was not considered. Several nucleosides containing fused heterocyclic rings with an amino group at the 6-position of the purine ring or equivalent position showed considerable activity, suggesting that the orientation of the amide group may not be essential for this antiviral activity. Unlike most other five-membered heterocyclic ring compounds, compound **27** is barely active. Although this compound had the necessary orientation for amide group, to get a better fit of the physicochemical parameters, the heterocyclic ring played the key role, putting the amide group in a different place. The superposition of all these compounds are shown in Figure 2. The important conclusions here are (i) the crystallographic orientation of the amide group may be the preferred conformation for the antiviral activity, but it is not necessarily the minimum energy conformation of ribavirin, and (ii) in order to increase the antiviral activity we have to stabilize the active conformation with minimum alteration of the physicochemical property distribution within the ring.

For nucleosides **4**, **5**, **9**, **11**–**23**, **26**, and **28**, containing a fused six-membered heterocyclic ring, the superposition resulted from either the minimum-energy conformation or the low-energy conformation, in the expected anti conformation. The superposition of some of these molecules (**1**, **4**, **8**, **19**, **24**, and **25**) are shown in Figure 3. In the superposition process compound **8** exhibited confor-

(41) QCPE Program No. 506, Department of Chemistry, Indiana University, Bloomington, IN.
(42) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.

Table IV. Description of the Site Pockets

site pocket	coefficients of phys chem param: hydrophobicity[a]	dispersive[b]	electrostatic[c]	description (geometrical[d] and physicochemical[e])
1		−0.0827		Binds the lower part of the six membered fused heterocyclic ring and also the 2′-OH group. Negative contribution from molar refractivity, indicating steric repulsion.
2	−0.6197	−0.2219		Binds N_1 atom of the purine ring. Small hydrophilic group is preferable.
3	−1.2185	0.5197	6.3771	Binds the C_6 atom of the purine ring. Hydrophilic atom with positive charge is preferable. Since molar refractivity makes a positive contribution, larger groups like sulfinamide may be worth trying, although such a group is not planar.
4	−0.1895	−0.0470	5.7077	Binds N_7 and C_8H of the purine ring. Small hydrophilic atoms with positive charge is preferable.
5	−0.8911	0.0919		Binds the ribose ring, including the N_9 atom of the purine ring. Hydrophilic group is preferred here.
6	0.7012	−0.7665		Binds the C_4 and C_5 atoms of the purine ring. Small hydrophobic atoms are prefered.
7	−0.8582	0.2088		Binds the thiomethyl group in molecule 1 and the amide oxygen in molecule 3 and the similar atoms in other molecules. Hydrophilic and dispersive forces are operating in this region. Hydrophilic atoms and groups with high atomic refractivity should be tried.

[a] Octanol–water partition coefficient was used to model the hydrophobic interaction. [b] Molar refractivity was used to model the dispersive and steric interactions. [c] CNDO/2 atomic charge density was used to model the electrostatic interaction. [d] Geometrical description is provided with respect to the reference molecule, which is generally applicable for many other molecules. [e] The physicochemical properties of the preferred atoms suggested here does not consider the effect on the conformational properties or the physicochemical properties of the other part of the molecule. Those effects can be estimated by the actual generation of the molecule and docking the low-energy conformations of the molecule at the hypothetical active site in the desired orientation.

Table V. Statistics of the Study

no. of compounds	no. of site pkt	no. of parameter	SD	correl coeff	explained variance	*F*-test, %
28	7	15	0.125	0.971	0.880	99.99

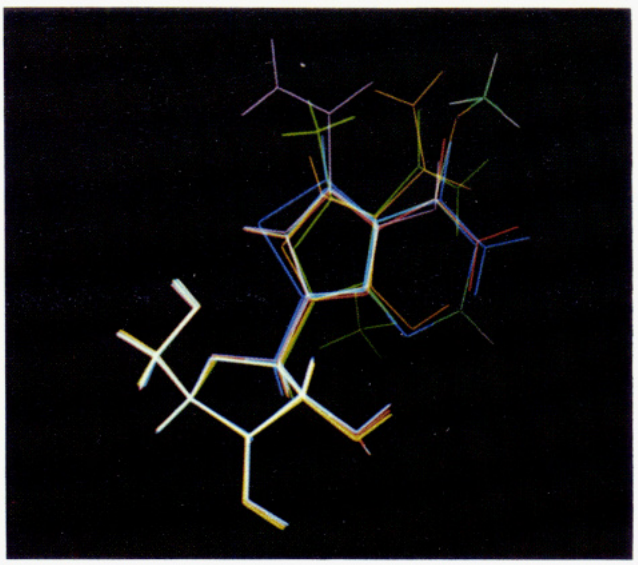

Figure 2. Five-membered heterocyclic molecules superimposed over the reference molecule 1. Molecule 1 in "half bond" color, molecule 2 in yellow, molecule 3 in green, molecule 6 in red, molecule 7 in blue, and molecule 27 in magenta.

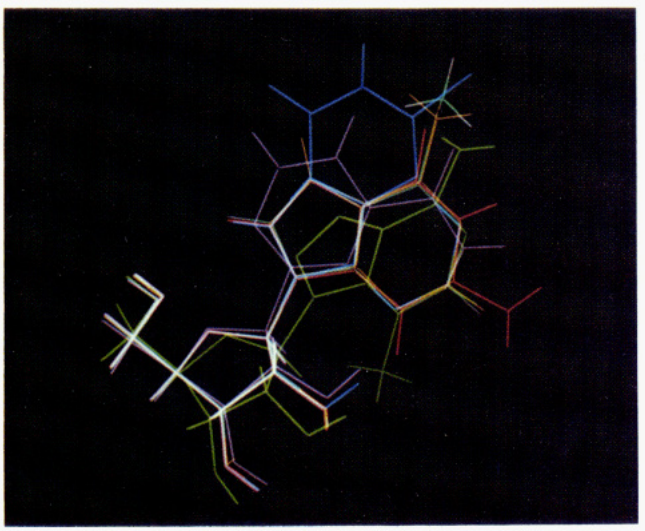

Figure 3. The fused heterocyclic compounds superimposed over the reference molecule 1. Molecule 1 in "half bond" color, molecule 4 in yellow, molecule 8 in green, molecule 19 in red, molecule 24 in blue, and molecule 25 in violet. The conformational problem did not allow a better superposition for molecule 8.

mational problem. This molecule is stable only at high anti conformation. Attaining the anti conformation at 240° needed nearly 5 kcal/mol energy. The high antiviral activity of this compound was still accounted for, because the conformational energy was not used during the correlation study. At this point it is hard to decide whether the binding site has such flexibility to accommodate it, or that some other factors are complicating its antiviral activity.

The modeling of the interaction of the ligand with the hypothetical site cavity showed some interesting results. It divided the site cavity into seven different pockets, a schematic representation of which is shown in Figure 4 using the reference molecule. Such dissection is valid for most of the molecules. The quantitative and qualitative description of these site pockets are given in Table IV. A negative coefficient for hydrophobic parameters indicates the region is hydrophilic in nature. Since molar refractivity is directly proportional to the molecular volume, the negative coefficient for molar refractivity is considered as steric susceptibility. The negative value for electrostatic interaction indicates that a negative charge is preferred. Although such a description alone can qualitatively be used for the future design of possible antiviral compounds, the best way to use this model is to superimpose the low-energy conformations of the molecule in the desired orientation. If any part of a molecule goes far beyond the site cavity, its fate is uncertain because that part may be sterically

Table VI. Illustration of the Evaluation of the Antiviral Activity Using the Model Compound 3: Atom Types and Their Physicochemical Properties

atom	atom type[a]	Oc–H_2O part. coeff	molar refrac	formal chrg density
$O_{1'}$	O-59	0.1017	1.2000	−0.2365
$C_{1'}$	C-8	−0.5894	3.0137	0.1561
$C_{2'}$	C-8	−0.5894	3.0137	0.1285
$C_{3'}$	C-8	−0.5894	3.0137	0.1330
$C_{4'}$	C-8	−0.5894	3.0137	0.1267
$C_{5'}$	C–6	−0.8062	3.2624	0.1327
$O_{2'}$	O-56	−0.0114	1.4430	−0.2515
$O_{3'}$	O-56	−0.0114	1.4430	−0.2460
$O_{5'}$	O-56	−0.0114	1.4430	−0.2472
N_1	N-73	0.3682	2.6295	0.0318
N_2	N-75	−0.0179	4.5123	−0.1541
C_3	C-28	0.1199	2.5000	0.0378
C_4	C-26	0.0281	3.8182	0.1026
C_5	C-34	−0.2682	3.4494	−0.0192
C_6	C-40	0.0689	2.7938	0.3586
N_7	N-72	−0.1964	3.0059	−0.2348
O_8	O-58	−0.2487	1.6506	−0.3639
O_9	O-60	0.2443	1.8434	−0.2157
C_{10}	C-5	−1.0237	3.4006	0.1201
C_{11}	C-5	−1.0237	3.4006	0.0432
H-$C_{1'}$	H-51	0.1591	1.0026	−0.0114
H-$C_{2'}$	H-47	0.3383	0.8000	−0.0115
H-$C_{3'}$	H-47	0.3383	0.8000	−0.0296
H-$C_{4'}$	H-47	0.3383	0.8000	−0.0001
H-$C_{5'}$	H–47	0.3383	0.8000	−0.0090
H-$C_{5'}$	H–47	0.3383	0.8000	−0.0136
H-$O_{2'}$	H-50	−0.2757	0.8000	0.1270
H-$O_{3'}$	H-50	−0.2757	0.8000	0.1243
H-$O_{5'}$	H-50	−0.2757	0.8000	0.1289
H-N_7	H-50	−0.2757	0.8000	0.1255
H-N_7	H-50	−0.2757	0.8000	0.1180
H-C_{10}	H-47	0.3383	0.8000	−0.0015
H-C_{10}	H-47	0.3383	0.8000	0.0002
H-C_{10}	H-47	0.3383	0.8000	−0.0051
H-C_{11}	H-51	0.1591	1.0026	0.0267
H-C_{11}	H-51	0.1591	1.0026	0.0098
H-C_{11}	H-51	0.1591	1.0026	0.0192

[a] See ref 11 for the atom classification and their molar refractivity. The charge densities were obtained from the CNDO/2 calculation. For hydrophobicity, some recently modified values, which are yet unpublished, were used. These values are fairly comparable to the values reported in ref 11 and will be supplied on request.

inaccessible. The qualitative suggestion for structural modification often fails since such modification changes the electronic as well as physicochemical properties of the other parts. Since the sugar ring was not modified in the present data set, due to the linear dependency of the parameters the method could not give any definite picture about the location of the major interaction point in the sugar moiety. The color coded surface of the present site model is shown in Figure 5.

The analysis of the contribution of the various physicochemical properties of the ligand at different regions toward antiviral activity gives some interesting information. The dispersive interaction of the sulfur atoms in the thioethers or thioamides may be a major source of attractive interaction for the receptor. However, the steric repulsion of the six-membered heterocyclic ring in site pocket 1 did not allow the activity to attain the desired level in these compounds. A single five-membered ring is desirable in that respect. It may also be a good idea to fuse the amide group of the five-membered heterocyclic ring with another five-membered ring. In such a system the conformation will be frozen to the active conformation but the steric repulsion will be less compared to the six membered ring fusion.

The statistics of the correlation study is shown in Table V. The evaluation of the calculated binding energy of a

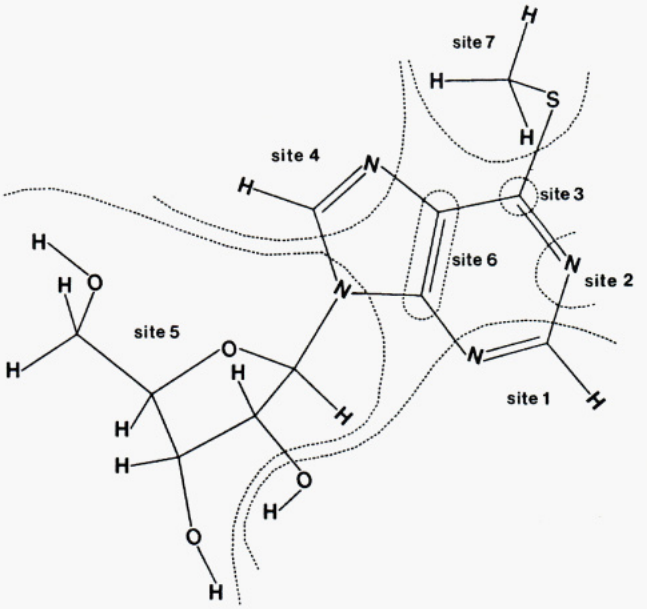

Figure 4. An schematic representation of the dissection of the site cavity into different types. The wall is not a physical barrier but indicates the change of the nature of interaction.

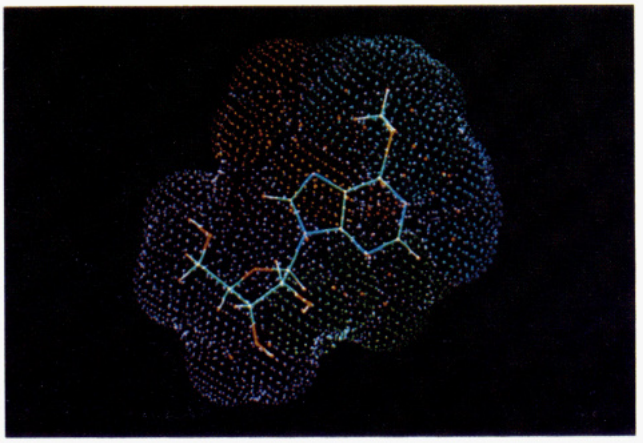

Figure 5. Color coded representation of the surface of the active-site cavity. The view from the side of the $C_{2'}$ and $C_{3'}$ atoms of the sugar ring. Color settings are as follows: site type 1 green, site type 2 blue, site type 3 red, site type 4 orange, site type 5 magenta, site type 6 yellow, and site type 7 cyan.

model compound (molecule **3**) is illustrated in Tables VI and VII. Table VI shows how the atoms of the molecule occupy the various site pockets, the types of the atoms, and their atomic physicochemical properties. The local property at any site type is the sum of the properties of the atoms occupying it. The interaction or contribution toward the antiviral activity is simply the multiple of the physicochemical property and the corresponding coefficient as given in Table IV. The calculated antiviral activity is the sum of all these products.

Any one doing this type of empirical modeling might be interested in a work where a particular set of compounds has been tested for more than one virus. Unfortunately, most of the virus systems are highly specific in their selectivity (see, for example, Table VIII), very few compounds have such broad antiviral activity, which indirectly indicates the difference of the binding site among these systems. In order to compare their nature we are working on an elaborate project by using a different set of active compounds against different viruses and comparing the different models thus obtained. This will be published in the future.

Table VII. Illustration of the Evaluation of the Biological Activity: Local Physicochemical Properties in Different Site Types and the Corresponding Interaction Energy[a] for Molecule **3**

site	Oc–H_2O part. coeff (hydrophobic contribution)	molar refrac (dispersive contribution)	charge density (elec stat contribution)	atom occupying the site pocket[b]
1	c	9.0870		$O_{2'}$, O_9, C_{10}, H–C_{10} (2),
		(–0.752)		H–$O_{2'}$
2	–0.2757	0.8000		H–N_7
	(0.171)	(–0.178)		
3	–0.1275	5.7997	0.1238	C_6, H_7
	0.155)	(3.014)	(0.789)	
4	–0.1961	13.5502	–0.0234	N_1, N_2, C_{11}, H–C_{11} (3)
	(0.037)	(0.637)	(0.134	
5	–1.6520	28.5058		$O_{1'}$, $C_{1'}$, $C_{2'}$, $C_{3'}$,
	(1.472)	(2.620		$C_{4'}$, $C_{5'}$, $O_{3'}$, $O_{5'}$,
				C_3, H–$C_{1'}$, H–$C_{2'}$,
				H–$C_{3'}$, H–$C_{4'}$, H–$C_{5'}$ (2)
				H–$O_{3'}$, H–$O_{4'}$
6	0.0982	8.0676		C_4, C_5, H–C_{10}
	(0.069)	(–6.183)		
7	–0.5244	2.4506		O_8, H–N_7
	(0.450)	(0.688)		

[a] The interaction energy is simply the product of physicochemical property and the corresponding site-dependent coefficient as shown in Table V. [b] The number within the parentheses represents the number of hydrogen atoms. [c] Unused property.

Table VIII. Comparison of the Virus Rating of the Compounds in Different Test Systems[a]

compd	RNA virus Rhino1A	DNA virus HSV2	DNA virus Adeno2
1	0.80	1.25	0.00
2	0.60	1.00	0.00
3	0.00	0.30	1.00
4	0.10	0.20	0.00
5	0.00	0.30	0.80
6	0.70	1.40	0.30
7	0.00	1.19	0.20
8	0.00	0.00	0.00
9	0.00	0.75	0.00
10	0.20	0.00	0.00
11	0.00	0.88	0.00
13	0.90	0.00	1.00
14	0.60	0.00	0.00
15	0.30	0.00	0.00
16	0.60	0.00	0.00
18	0.00	0.20	0.00
19	1.00	1.06	0.20
20	0.00	0.00	0.00
21	0.00	0.12	0.00
22	0.00	0.00	0.00
23	0.00	0.00	0.00
24	0.00	0.00	0.00
25	0.00	0.00	0.00
26	0.00	0.00	0.00
27	0.00	0.00	0.00
28	0.00	0.10	0.00

[a] Rhino1A = rhinovirus type 1A, HSV2 = herpes simplex virus type 2, Adeno2 = adenovirus type 2.

Conclusion

We have presented a very straightforward procedure for undertaking conformational and physicochemical properties of certain nucleosides to rationalize their antiviral activity against parainfluenza virus and a model for the active site has been developed. To our knowledge, it is for the first time any such method has been applied to rationalize the antiviral activity of the nucleosides. The method is very general and can be used in any comparable receptor modeling problem.

The modeling of a receptor from the structure of nucleosides alone is risky, since it suffers from various uncertainties. However, in absence of the detailed knowledge of the mechanism of action and the structure of the biological receptor, three dimensional structure directed quantitative structure–activity relationships is a reasonable approach to rationalize antiviral activity. The present approach, using the conformational properties, molecular orbital charge density, hydrophobicity, and molar refractivity of 28 nucleosides, gave the three-dimensional structure of a hypothetical binding-site cavity, together with a quantitative expression for the antiviral activity with the electronic and physicochemical properties of the ligand molecules. The model[46] can be used qualitatively for the direction of alteration of the physicochemical properties in the ligand during the future drug design. More appropriately, it can be used quantitatively by actual generation of the molecule in the computer, and placing the low-energy conformations in the hypothetical receptor site cavity in the energetically best orientation. Some synthetic work to verify the predictive power of the model has been under taken, and will be reported in the future.

The success of this type of computer aided empirical model building approach for drug design depends on very close collaboration between the experimental and computational scientists. One should be aware of the limitation of this approach: (i) the outcome of the method is totally dependent on the antiviral (biological) activity data used, and it is important that precise biological activity data be used for this method; (ii) if the antiviral activity does not represent the binding data to a purified receptor, one should be critical about the various simplifying assumptions made during the model building process and should try to keep the metabolic factors constant for all the compounds studied (see introduction section for detailed consideration).

Acknowledgment. We thank Dr. S. Larson for supplying the X-ray data of a few heterocyclic rings. Support in part from NSF (Grant No. DMB-8705006) and NIH (Grant No. GM 37123) is acknowledged.

Appendix I

Given the reference structure and the various low-energy conformations of the molecules, the problem was to de-

(43) Kenknight, C. E. *Acta Crystallogr.* **1984**, *A40*, 708–712.
(44) Thornber, C. W. *Rev. Chem. Soc.* **1979**, *8*, 563.
(45) Varmuza, K. In *Pattern Recognition in Chemistry, Lecture Notes in Chemistry*; Springer-Verlag: Berlin, 1980; Vol. 21.
(46) The model consisted of 67 small spheres of seven types. The coordinates of the center of the spheres and their radii and types may be obtained from the authors on request.

termine the various ways of superposing a molecule on the reference structure. By the term reference structure, it is meant a particular conformation of a particular ligand. The choice of the reference structure is based on the priority function as discussed in the method section (step 4). The program has two options: complete search and selective search. We used a modified distance geometry[17] (more appropriately distance matching) approach to superpose the molecules. The complete search consists of the following steps.

(1) The interatomic distance matrices of the reference structure and of the conformation of the molecule to be compared are generated first.

(2) A basic superposition consists of taking all combinations of three atoms from the reference structure and all permutations of three atoms from the second structure. If the distances between the superimposed atoms match, the basic superposition holds. The condition of distance matching is

$$d_t + \delta \geq d_r \leq d_t - \delta \qquad \text{(i)}$$

where d_r represents the distance in the reference molecule and d_t represents that in the trial molecule.

(3) When a basic superposition is accepted, the contacts between the other atoms are mere consequences. These consequences are determined by checking the distances between the nonsuperimposed atoms from the superimposed atoms in the two molecules. If the matching criterion, as given above, is satisfied, the atoms are assumed to be superimposed. If more than one atom of the trial molecule satisfy the condition, we took the one having the minimum absolute deviation per distance from the superimposed atoms. The superposition based on only distance checking cannot distinguish between enantiomers. We solved the problem by actual superposition of the two structures in the evaluated mode using rigid translation and rotation.[43] The superposition was represented by a vector, each element of which gives the atom superimposed on the reference structure. The superposition criterion may be described as *all or none*. An atom of the test molecule was assumed to be totally superimposed on the nearest atom of the reference molecule, if the distance was less than a preassigned value. The merits and demerits of such simplification have already been discussed.[1-3]

Due to the enormity of the combinatorial problem, we simplified the search by two different ways: (1) taking only the "important atoms" for the basic superposition, (2) taking only those basic superpositions in which "similar atoms" are superposed. By the term "important atoms" we mean the heteroatoms, carbon multiply bonded to heteroatoms, and hydrogen attached to heteroatoms. These are the atoms responsible for strong regiospecific electrostatic and dispersive interactions and constitutes the pharmacophore (the groups responsible for the biological activity). Defining "similar atoms" is the problem of bioisosterism.[44] We simplified the situation by classifying the atoms very broadly, as follows: (i) all heteroatoms which are not multiply bonded to electronegative heteroatoms, (ii) all other heteroatoms, (iii) carbon multiply bonded to heteroatoms, (iv) carbon otherwise, (v) hydrogen attached to heteroatoms, and (vi) hydrogen otherwise.

All these features were written in the STRUCOMP program using fortran77.

Appendix II

The problem here was to select the best superposition of a ligand from the various geometrically feasible ones. In absence of the explicit structure of the binding site, we used a few physicochemical properties of the ligand which are representative of the various molecular interactions. Three physicochemical properties, viz. octanol–water partition coefficient, molar refractivity, and formal charge density, and two different functions were used for this purpose. The first function assumes that the hypothesized reference structure is the best possible structure, and any deviation from it will incur a penalty:

$$F_1 = \sum_k |x_k| - |x_k - x_{j(k)}| \qquad \text{(ii)}$$

where x_k represents the physicochemical property fo the kth atom of the reference structure, $j(k)$ is the atom superimposed on the kth atom of the reference structure. The corresponding physicochemical property will be zero if no atom is superimposed. The above function assumes that the interaction of the ligand atoms with the receptor site is quadratic in nature, and the refererence structure lies on the peak; therefore it will decrease if the physicochemical property changes on either side of the "ideal" value. The interactions in many places, however, are linear with the physicochemical property. Since the coefficients of the physicochemicl properties in the expression of ligand–receptor interaction (eq 5) are not known, we assumed that the reference structure experiences attraction from all its atoms. In other words, the sign of the physicochemical property for attractive interaction is the sign of the corresponding physicochemical property of the reference structure. This gives a second function to express the goodness of a superposition:

$$F_2 = \sum_k [x_k/|x_k|]x_{j(k)} \qquad \text{(iii)}$$

If one is interested in using more than one physicochemical property, he can easily take sum over the desired physicochemical properties. However, that brings the problem of scaling the parameters.[45] We wanted to use the conventional "autoscaling":

$$x_{i,\text{new}} = [x_{i,\text{old}} - \mu_{\text{old}}]/s_{\text{old}} \qquad \text{(iv)}$$

where μ_{old} is the mean of the old parameter, and s_{old} is its variance. The new parameter has unit variance and zero mean. However, it is expected that the scaling should be such that it will make a linear transformation of the function expressing the goodness of superposition while considering any particular parameter. Otherwise the ordering of the binding modes using unscaled and scaled parameters will differ, making the calculation physically unrealistic. The function F_1 in general will maintain the condition only if all the atoms of the reference structure are occupied. The function F_2 in general cannot maintain the condition. In both the cases the problem is caused by the μ_{old} part of the parameter transformation. Furthermore, autoscaling puts the parameters on either sides of zero. For charge density and hydrophobicity, positive and negative values have a physical significance. However, molar refractivity is closely related to molar volume and is always positive. Making molar refractivity negative in any scaling is unwanted. All these problem can be avoided, if we used a modified "autoscaling"

$$x_{i,\text{new}} = x_{i,\text{old}}/s_{\text{old}} \qquad \text{(v)}$$

The new parameters here have one common property: their variances are unity.

All these aspects were written in OPTSUP program using fortran77.