GROOTENHUIS, P. D. J. & HAASNOOT, C. A. G. (1993). *Mol. Simul.* **10**, 75–95.
GUNSTEREN, W. F. VAN & BERENDSEN, H. J. C. (1987). *GROMOS. Groningen Molecular Simulation Package*. Univ. of Groningen, The Netherlands.
HAHN, T. (1983). *International Tables for X-ray Crystallography*, Vol. A. Dordrecht: Kluwer Academic Publishers.
HARRIS, K. D. M., TREMAYNE, M., LIGHTFOOT, P. & BRUCE, P. G. (1994). *J. Am. Chem. Soc.* **116**, 3543–3547.
HOLDEN, J. R., DU, Z. & AMMON, H. L. (1993). *J. Comput. Chem.* **14**, 422–437.
JEFFREY, G. A. & SHIONO, R. (1977). *Acta Cryst.* B**33**, 2700–2701.
JÖNSSON, P.-G. (1971). *Acta Cryst.* B**27**, 893–898.
KARFUNKEL, H. R. & GDANITZ, R. J. (1992). *J. Comput. Chem.* **13**, 1171–1183.
KARFUNKEL, H. R., LEUSEN, F. J. J. & GDANITZ, R. J. (1993). *J. Comput.-Aided Mat. Des.* **1**, 177–185.
KOUWIJZER, M. L. C. E., VAN EIJCK, B. P., KOOIJMAN, H. & KROON, J. (1995). *Acta Cryst.* In the press.
KOUWIJZER, M. L. C. E., VAN EIJCK, B. P., KROES, S. J. & KROON, J. (1993). *J. Comp. Chem.* **14**, 1281–1289.
KROON-BATENBURG, L. M. J., VAN DER MSLUIS, P. & KANTERS, J. A. (1984). *Acta Cryst.* C**40**, 1863–1865.
LONGCHAMBON, F., AVENEL, D. & NEUMAN, A. (1976). *Acta Cryst.* B**32**, 1822–1826.
LONGCHAMBON, F., OHANESSIAN, J., AVENEL, D. & NEUMAN, A. (1975). *Acta Cryst.* B**31**, 2623–2627.
MADDOX, J. (1988). *Nature*, **335**, 201.
OHANESSIAN, J., AVENEL, D., KANTERS, J. A. & SMITS, D. (1977). *Acta Cryst.* B**33**, 1063–1066.
PADMAJA, N., RAMAKUMAR, S. & VISWAMITRA, M. A. (1990). *Acta Cryst.* A**46**, 725–730.
PERLSTEIN, J. (1992). *J. Am. Chem. Soc.* **114**, 1955–1963.
PERLSTEIN, J. (1994). *J. Am. Chem. Soc.* **116**, 455–470.
PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A. & VETTERLING, W. T. (1986). *Numerical Recipes, the Art of Scientific Computing*. Cambridge: Cambridge University Press.
STEINER, T. & SAENGER, W. (1992). *Acta Cryst.* B**48**, 819–827.

Acta Cryst. (1995). B**51**, 103–107

A Database Study of Intermolecular NH···O Hydrogen Bonds for Carboxylates, Sulfonates and Monohydrogen Phosphonates

BY B. PIRARD, G. BAUDOUX AND F. DURANT

Laboratoire de Chimie Moléculaire Structurale, Facultés Universitaires Notre-Dame de la Paix, rue de Bruxelles 61, *B*-5000 *Namur, Belgium*

(*Received* 22 *April* 1994; *accepted* 15 *September* 1994)

Abstract

A search of the Cambridge Structural Database (CSD, version 5.05, 1993) was performed in order to compare the geometrical features of the hydrogen bonds involving on the one hand amino groups and on the other hand carboxylates, sulfonates or monohydrogen phosponates. Phosphonates were not considered because only four entries containing amino and phosphonate moieities were located in the CSD. The hydroxylic group of monohydrogen phosphonates primarily acts as a hydrogen-bond donor. The three moieties under study show NH···O hydrogen bonds with similar geometrical features. This statistical analysis has focused on the hydrogen-bond distances and angles and on the distributions of the H atoms around the acceptor O atoms of carboxylates, sulfonates or monohydrogen phosphonates.

Introduction

The medicinal chemistry literature contains numerous examples of bioisosteric modifications (Davies, Williams & Smith, 1987, and references therein; Lipinski, 1986, and references therein). Bioisosterism is one of the strategies used to develop more potent and less toxic analogues of a lead compound. Within this scope, crystallographic databases such as the Cambridge Structural Database (CSD) (Allen, Kennard & Taylor, 1983) provide an invaluable source of information about potential intermolecular interactions between drug molecules and macromolecules (Klebe, 1994). Crystal packing data enable one to suggest similar patterns of interactions for different functional groups with a common partner (Allen, 1992, and references therein) and hence to consider new bioisosteric replacements.

This strategy has been applied to carboxylic acids and to some of their isosteres. Phosphinic (Howson, Mistry, Broekman & Hills, 1993; Baylis, Campbell & Dingwall, 1984), phosphonic (Chiefari, Galanopoulos, Janowski, Kerr & Prager, 1987; Lipinski, 1986, and references therein), sulfonic acids (Abbenante & Prager, 1992; Lipinski, 1986, and references therein), sulfonamides and tetrazoles (Davies, Williams & Smith, 1987, and references therein; Lipinski, 1986, and references therein) are current surrogates for carboxylic groups. For instance, the carboxylic group of γ-aminobutyric acid, an inhibitory neutrotransmitter, has been replaced by phosphinic, phosphonic or sulfonic groups (Bowery, 1993). In this contribution, we have focused on the

geometrical features of hydrogen bonds involving on the one hand amino groups, and on the other hand carboxylates, sulfonates and monohydrogen phosphonates.

For this purpose, the model fragments shown in Fig. 1 were searched in the CSD. A statistical survey of the geometrical features of NH···O hydrogen bonds was performed for fragments shown in Figs. 1(*a*)–(1*c*). A hydrogen bond is usually described by the hydrogen-bond donor-atom–hydrogen-bond acceptor-atom distance (*D*1), the H atom–hydrogen-bond acceptor-atom distance (*D*2), and the hydrogen-bond angle (*V*1) (Fig. 2) (Taylor & Kennard, 1984, and references therein; Taylor, Kennard & Versichel, 1983, 1984*a*,*b*). Moreover, the question of lone-pair directionality for hydrogen bonding has long been debated. In order to assess lone-pair directionality, the polar spherical coordinates θ and Φ have been defined (Taylor & Kennard, 1984; Taylor, Kennard & Versichel, 1983). The distributions of the variables shown in Fig. 2 have been analyzed for the model fragments given in Figs. 1(*a*)–1(*c*). This study has led us to derive geometrical similarities between NH···O hydrogen bonds involving on the one hand amino moieties and on the other hand carboxylates, sulfonates or monohydrogen phosphonates.

Methods

Version 5.05 of the CSD (1993) was used to perform the searches, which proceeded in three steps. First, the model fragments were searched using the CONN option of the *SEEK* routine. Secondly, the hits were submitted to a geometrical analysis. This analysis primarily aimed to locate all the nitrogen–oxygen distances *D*1 within the range 2.400–3.200 Å, inclusive (Fig. 2), between a target hydrogen-bond acceptor molecule and symmetry-related hydrogen-bond donor molecules. For this purpose, we made use of the CALC INTER option of the *GSTAT* routine. Within the set of crystal structures showing *D*1 values less than 3.200 Å, we further considered only those matching the criteria shown in Figs. 1 and 2. The constraints on the *X*—O (where *X* is a C, S or P atom) bond lengths were introduced to discard the unionized carboxylic, phosphonic and sulfonic moieties. The distance ranges were chosen after a preliminary statistical survey of the *X*—O bond lengths for carboxylates, sulfonates and monohydrogen phosphonates. As heteronuclear hydrogen bonding mostly arises from electrostatic interactions (Gilli, Bertolasi, Ferretti & Gilli, 1994; Mitchell, Thornton, Singh & Price, 1992; Umeyama & Morokuma, 1977), *D*2 (Fig. 2) values as far as 3.000 Å were considered. In accordance with definitions previously reported for hydrogen bonding (Steiner & Saenger, 1992), *V*1 (Fig. 2) was allowed to vary between 90.0 and 180.0°. Moreover, the crystal structures displaying *R* factors larger than 0.070 were discarded. The coordinates of the H atoms were normalized using the HNORM option of the *GSTAT* routine, since normalized X-ray geometries have been found to be in good agreement with neutron diffraction geometries (Taylor & Kennard, 1984). The values of *D*1, *D*2, *V*1, θ and Φ* were computed for the entries matching these crystallographic and geometrical requirements. Thirdly, the KILL option of the *GSTAT* routine enabled one to discard the redundant observations. This corrected data set was submitted to a statistical analysis. Within each data set, the mean, the variance, the minimum and maximum values were computed by the *GSTAT* routine for the distributions of these variables.†

Graphically, the distributions of *D*1 have been displayed as histograms. As *D*2 and *V*1 have been found to be correlated (Steiner & Saenger, 1992, and references therein; Taylor & Kennard, 1984), the variations of *V*1 *versus* *D*2 were investigated for the moieties under study. To clarify the analysis, the polar spherical coordinates θ and Φ were converted into

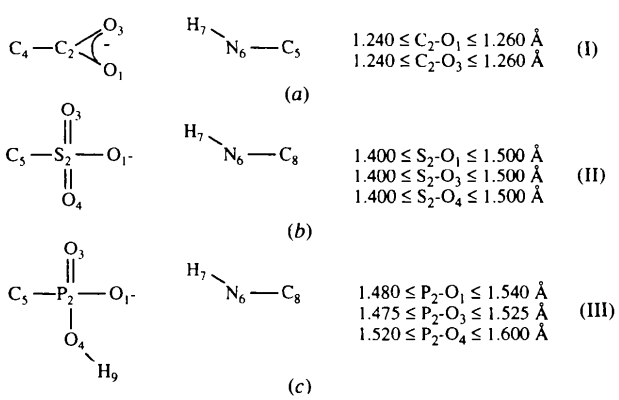


Fig. 1. Fragments under study, amino moieties with (*a*) carboxylates [fragment (I)], (*b*) sulfonates [fragment (II)], (*c*) monohydrogen phosphonates [fragment (III)]; constraints on the *X*—O (*X* is a C, S or P atom) bond lengths.

* The polar spherical coordinates θ and Φ have been defined by the following atoms: O3, C2, O1, H7 for fragment (I); O3, S2, O1, H7 for fragment (II); O3, P2, O1, H7 for fragment (III) (LP2 option of the *GSTAT* routine). The geometrical features of the hydrogen bonds involving the unionized O atom of fragment (III) have not been considered (see *Results and discussion*).

† Lists of refcodes and H-atom representational diagrams have been deposited with the IUCr (Reference: BK0012). Copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

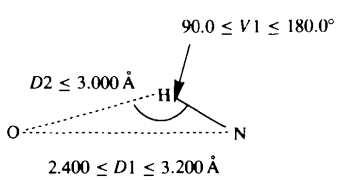


Fig. 2. Geometrical descriptors of hydrogen bonding; constraints on *D*1, *D*2 and *V*1.

Table 1. *D*1 *distribution values*

	Fragment (I)			Fragment (II)			Fragment (III)		
	*D*1 (Å)	*D*2 (Å)	*V*1 (°)	*D*1 (Å)	*D*2 (Å)	*V*1 (°)	*D*1 (Å)	*D*2 (Å)	*V*1 (°)
*N**	878	878	878	320	320	320	65	65	65
Mean	2.836	1.887	157.4	2.946	2.049	150.7	2.810	1.862	155.7
Standard deviation	0.118	0.215	17.6	0.128	0.251	22.3	0.100	0.202	15.5
Minimum	2.542	1.524	90.5	2.574	1.650	91.8	2.611	1.597	90.7
Maximum	3.200	3.000	179.0	3.199	2.975	180.0	3.112	2.980	178.8

* *N* is the number of nonredundant observations.

Cartesian ones by *GSTATCON*, a Fortran program (Baudoux & Pirard, 1994). The O atoms labelled O(1) in Figs. 1(*a*)–1(*c*) were chosen as the origin for the three Cartesian coordinate systems. The Cartesian coordinates served as input for *KEMIT*, an in-house molecular graphics program (Vanderveken & Vercauteren, 1992). The histograms were drawn using *DELTAGRAPH*® (DeltaPoint, 1990). *GSTATCON* and *KEMIT* have been running on an IBM RISC6000 of the Scientific Computing facility Center of the University of Namur.

Results and discussion

The data sets analyzed

After the redundant observations for the carboxylate and sulfonate groups were discarded, respectively, 878 and 320 hits have been found to match the criteria shown in Fig. 1. Crystal structures with amino and phosphonate moieties are scarce (only four hits) within the CSD. That is why the statistical survey was limited to monohydrogen phosphonate groups. For this latter moiety, we have located within the CSD 65 NH···O [O is either O(1) or O(3), Fig. 1(*c*)] hydrogen bonds. Further, we note that in only 29% of these monohydrogen phosphonate structures does the hydroxylic group participate in hydrogen bonding and then solely as a donor in one of two geometrical configurations (Fig. 3). Within this contribution, we have not further investigated the geometrical features of hydrogen bonds involving the hydroxylic O atoms of monohydrogen phosphonates.

*The distributions of D*1

Fig. 4 shows the distributions of *D*1 for carboxylates (*a*), sulfonates (*b*) and monohydrogen phosphonates (*c*). For each distribution the mean, the standard deviation of the sample, the minimum, and the maximum values are given in Table 1. For the three moieties under study, the mean of *D*1 amounts to 2.836, 2.946 and 2.810 Å, respectively (Table 1). These mean values are similar to those previously reported: 2.755 (12) Å for NH···O=C bonds (Taylor, Kennard & Versichel, 1984*b*) and 2.77 (11) Å for N/OH···$^{-}$OOC bonds (Klebe, 1994). Within the data sets analyzed, most of the *D*1 values (78.0 and 88.0%, respectively) for carboxylates and monohydrogen phosphonates are either near the short end of the range or equal to the sum of the van der Waals radii of nitrogen and oxygen, *i.e.* 2.900 Å (Weast, 1986), while only 47.0% of the *D*1 values for sulfonates belong

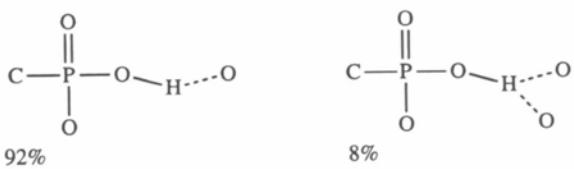


Fig. 3. Hydrogen-bond patterns and occurrence for the hydroxylic O Atom of monohydogen phosphonates.

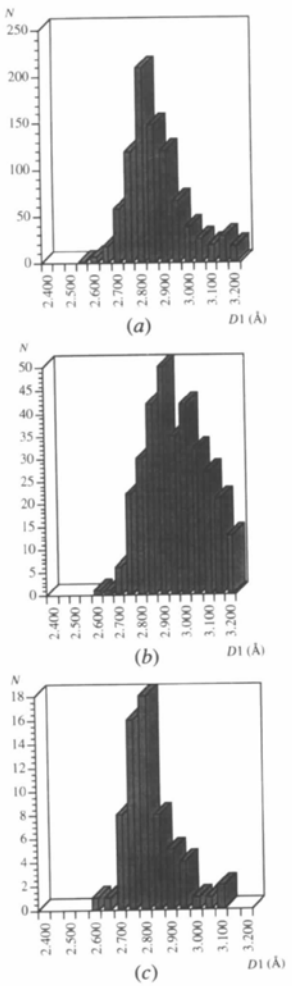


Fig. 4. Distributions of *D*1 for (*a*) carboxylates, (*b*) sulfonates and (*c*) monohydrogen phosphonates; *N* is the number of nonredundant observations.

to this distance range. These discrepancies could arise from different electronic densities on the O atoms of carboxylates, monohydrogen phosphonates and sulfonates, in accordance with the results of electrostatic potential derived charges (computed at the *ab initio* HF STO-3G or STO-3G* level of sophistication) computations (Pirard, Baudoux & Durant, 1994). The charges of the sulfonate O atoms are significantly lower than those computed for carboxylates and monohydrogen phosphonates. However, the minimum values of $D1$ (2.542, 2.574 and 2.611 Å for carboxylates, sulfonates and monohydrogen phosphonates, respectively) remain similar.

The variations of $V1$ *versus* $D2$

The variations of $V1$ *versus* $D2$ are shown in Figs. 5(*a*)–5(*c*) for the three moieties under study. In Figs. 5(*a*)–5(*c*) three distinct regions are observed.

(1) The region corresponding roughly to the diagonal is populated. But within each data set, the density of the population along the diagonal varies. Most of the observations (79.0, 66.3 and 78.3% for carboxylates, sulfonates and monohydrogen phosphonates, respectively) cluster together in the following area, *i.e.* $1.550 < D2 < 2.300$ Å and $150.0 < V1 < 180.0°$. This corresponds to the area of two-center and major components of three-center hydrogen bonds (Steiner & Saenger, 1992). The rest of the data spread in an elongated cluster, *i.e.* $1.850 < D2 < 3.000$ Å, $90.0 < V1 < 150.0°$. These points represent minor components of three-center hydrogen bonds (Steiner & Saenger, 1992).

(2) The lower triangular region is unpopulated. This region would correspond either to long and almost linear ($D2 > 2.300$ Å and $V1 > 150°$) hydrogen bonds or to long and bent hydrogen bonds ($D2 > 2.600$ Å and $V1 < 150°$).

(3) The upper triangular region is forbidden by mere limitation of the nitrogen–oxygen approach.

The distributions of the H *atoms around the acceptor* O *atoms*

The distributions of the H atoms around the O atom labelled O(1) [Figs. 1(*a*)–1(*c*)] of carboxylates, sulfonates and monohydrogen phosphonates have been considered in order to assess the lone-pair directionality for NH···O hydrogen bonds. The conversion of the polar spherical coordinates θ and Φ into Cartesian ones has enabled us to visualize these distributions. Within each data set, the H atoms form a crown around O(1) [Figs. 6(*a*)–6(*c*)]. Recently, a database survey has also revealed similar distributions of the H atoms around the acceptor O atoms for carboxylates, tetravalent sulfates and tetravalent phosphates (Klebe, 1994). Moreover, some preference for hydrogen bonding in the direction reported as (IV) rather than (V) (Fig. 7) is observed. Similar observations were previously reported for nitro groups (Taylor, Mullaley & Mullier, 1990). The observed

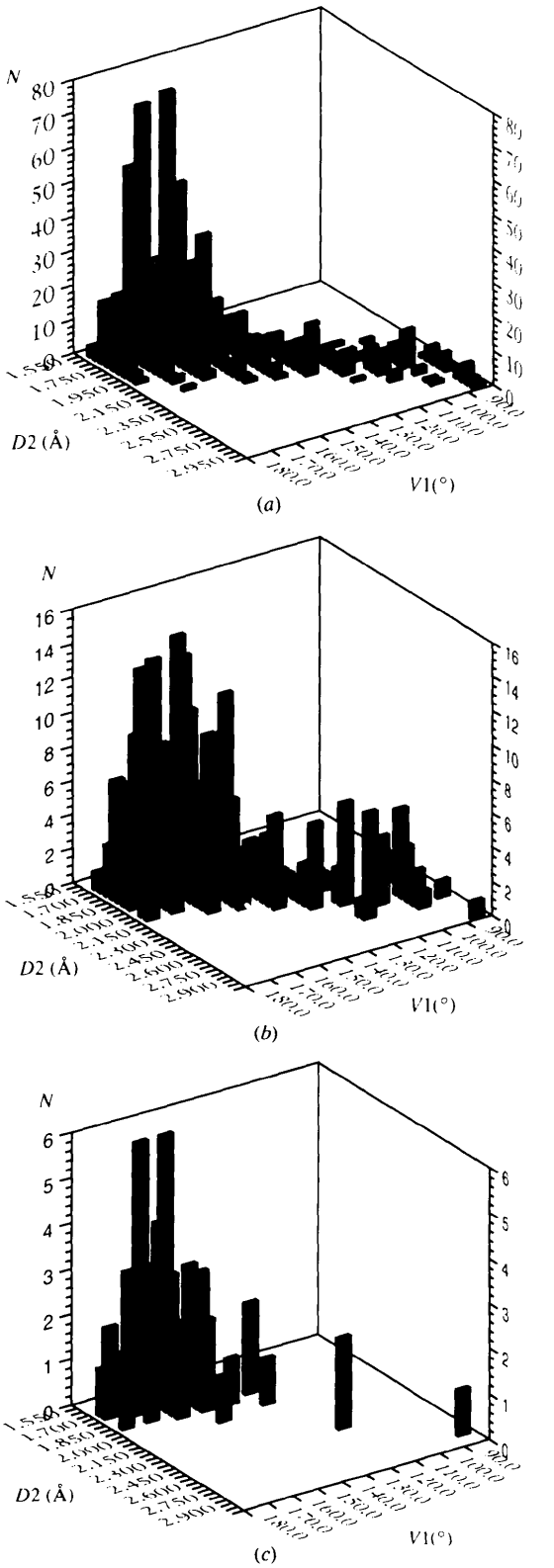


Fig. 5. Histograms of N *versus* $D2$ and $V1$ for (*a*) carboxylates, (*b*) sulfonates and (*c*) monohydrogen phosphonates. N is the number of nonredundant observations.

distributions are consistent with the results of *ab initio* HF STO-3G or STO-3G* molecular electrostatic potential computations for γ-aminobutyric acid, its sulfonic, and phosphonic analogues (Pirard, Baudoux & Durant, 1994).

Summary

A database survey of the geometrical features of the NH···O hydrogen bonds for carboxylates, sulfonates and monohydrogen phosphonates has revealed similarities. These three moieties mostly form short linear hydrogen bonds ($1.550 < D2 < 2.300$ Å and $150.0 < V1 < 180°$). Moreover, the H atoms tend to form a crown around the acceptor O atoms [Figs. 6(*a*)–6(*c*)]. However, the sulfonates tend to form longer hydrogen bonds than the carboxylates and monohydrogen phosphonates (Fig. 4). This could arise from differences in electronic densities on the O atoms of carboxylates, sulfonates and monohydrogen phosphonates.

BP is indebted to the ADIR society for financial support. The authors also thank the National Belgian Foundation for Scientific Research (FNRS), IBM-Belgium, and the Facultés Universitaires Notre-Dame de la Paix (FUNDP) for the use of the Namur Scientific Computing Facility.

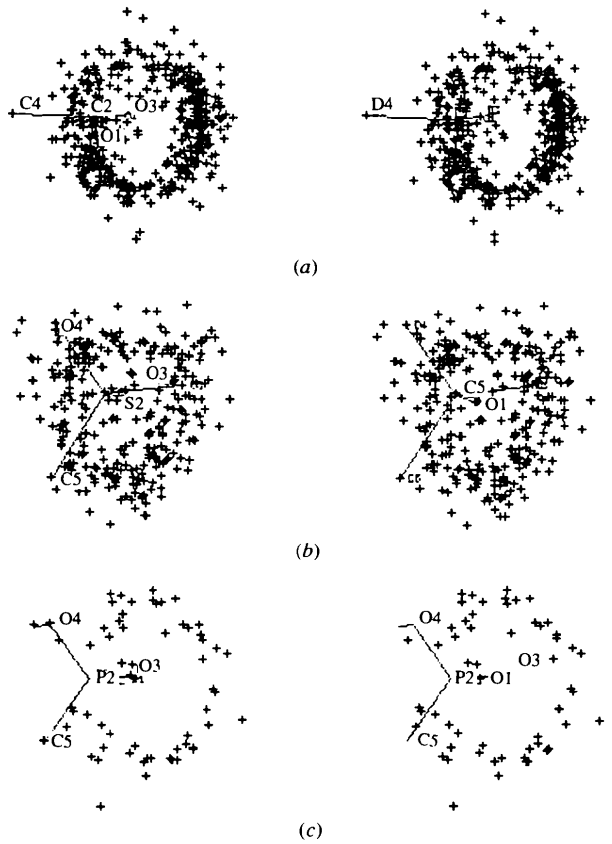
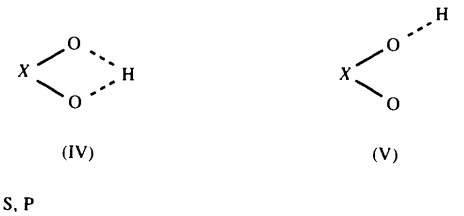


Fig. 6. Stereo diagrams for the distributions of H atoms around (*a*) carboxylates, (*b*) sulfonates and (*c*) monohydrogen phosphonates.



X = C, S, P

Fig. 7. Preferred hydrogen-bonding directions for the three moieties under study.

References

ABBENANTE, G. & PRAGER, R. H. (1992). *Aust. J. Chem.* **45**, 1791–1800.

ALLEN, F. H. (1992). *Accurate Molecular Structure Determination*, edited by A. DOMENICANO & I. HARGITTAI, pp. 355–378. New York: Oxford Univ. Press.

ALLEN, F. H., KENNARD, O. & TAYLOR, R. (1983). *Acc. Chem. Res.* **16**, 146–153.

BAUDOUX, G. & PIRARD, B. (1994). *GSTATCON*. Facultés Univ. Notre-Dame de la Paix, Namur, Belgium.

BAYLIS, E. K., CAMPBELL, C. D. & DINGWALL, J. G. (1984). *J. Chem. Soc. Perkin Trans.* 1, pp. 2845–2853.

BOWERY, N. G. (1993). *Annu. Rev. Toxicol.* **33**, 109–147.

CHIEFARI, J., GALANOPOULOS, S., JANOWSKI, W. K., KERR, D. I. B. & PRAGER, R. H. (1987). *Aust. J. Chem.* **40**, 1511–1518.

DAVIES, R. H., WILLIAMS, H. & SMITH, H. J. (1987). *Introduction to the Principles of Drug Design*, 2nd ed., edited by H. J. SMITH, pp. 82–95. London: Wright.

DeltaPoint (1990) *DELTAGRAPH*® Rel. 1.5.

GILLI, P., BERTOLASI, V., FERRETTI, V. & GILLI, G. (1994). *J. Am. Chem. Soc.* **116**, 909–915.

HOWSON, W., MISTRY, J., BROEKMAN, M. & HILLS, J. M. (1993). *Biorg. Med. Chem. Lett.* **3**, 515–518.

KLEBE, G. (1994). *J. Mol. Biol.* **237**, 212–235.

LIPINSKI, C. A. (1986). *Annual Reports in Medicinal Chemistry*, Vol. 21, edited by D. M. BAILEY & R. C. ALLEN, pp. 283–291. Orlando: Academic Press.

MITCHELL, J. B., THORNTON, J. M., SINGH, J. & PRICE, S. L. (1992). *J. Mol. Biol.* **226**, 251–262.

PIRARD, B., BAUDOUX, G. & DURANT, F. (1994). *Eur. J. Med. Chem.* **29**, 513–517.

STEINER, TH. & SAENGER, W. (1992). *Acta Cryst.* B**48**, 819–827.

TAYLOR, R. & KENNARD, O. (1984). *Acc. Chem. Res.* **17**, 320–326.

TAYLOR, R., KENNARD, O. & VERSICHEL, W. (1983). *J. Am. Chem. Soc.* **105**, 5761–5766.

TAYLOR, R., KENNARD, O. & VERSICHEL, W. (1984*a*). *J. Am. Chem. Soc.* **106**, 244–248.

TAYLOR, R., KENNARD, O. & VERSICHEL, W. (1984*b*). *Acta Cryst.* B**40**, 280–288.

TAYLOR, R., MULLALEY, A. & MULLIER, G. W. (1990). *Pestic. Sci.* **29**, 197–213.

UMEYAMA, H. & MOROKUMA, K. (1977). *J. Am. Chem. Soc.* **99**, 1316–1332.

VANDERVEKEN, D. J. & VERCAUTEREN, D. P. (1992). *KEMIT*. Facultés Univ. Notre-Dame de la Paix, Namur, Belgium.

WEAST, R. C. (1986). *Handbook of Chemistry and Physics*, 67th ed., D-188. Boca Raton: CRC Press.