Complexes of Lysine, Histidine, and Arginine with Sulfonated Azo Dyes: Model Systems for Understanding the Biomolecular Recognition of Glycosaminoglycans by Proteins

William H. Ojala,[†] Elise A. Sudbeck,[‡] Linh Khanh Lu,[†] Timothy I. Richardson,[§] Rex E. Lovrien,[§] and William B. Gleason*[,†,⊥]

Contribution from the Department of Laboratory Medicine & Pathology, Biomedical Engineering Center, and Department of Chemistry, University of Minnesota, Minneapolis, Minnesota 55455, and Department of Biochemistry, University of Minnesota, St. Paul, Minnesota 55108

Received April 6, 1995[⊗]

Abstract: The X-ray crystal structures of six salts composed of amino acids and sulfonated azo dyes have been determined, four of them at low temperature (173 K). The compounds are DL-lysine/4-[(4-hydroxyphenyl)azo]-benzenesulfonate (HABS) monohydrate (**1**), DL-lysine/7-hydroxy-8-(phenylazo)-1,3-naphthalenedisulfonate (Orange G) dihydrate (**2**), L-lysine/Orange G 1.5-hydrate (**3**), DL-histidine/Orange G trihydrate (**4**), L-histidine/Orange G trihydrate (**5**), and tosylarginine methyl ester (TAME)/4-[(2-hydroxy-6-*tert*-butyl-1-naphthalenyl)azo]benzenesulfonate ("Little Rock Orange," LRO) (**6**). By virtue of their basic side chains, these amino acids are the ones most important in the binding interactions between proteins and sulfated macromolecules such as glycosaminoglycans in living systems. The sulfonate salts described here serve as model systems for these interactions. Close intermolecular approaches between the dye sulfonate groups and neighboring amino acids and water molecules are examined, and the graph-set formalism is used to describe packing patterns and to identify corresponding interactions in different crystal structures. The recurrence of certain interactions between sulfonate groups and amino acid functional groups in these small-molecule crystal structures, including numerous interactions mediated by water molecules, suggests specificity that may also be a feature of the interactions between proteins and sulfated biological macromolecules.

Introduction

The diverse biological activities of sulfonated compounds have made them potential pharmaceutical agents and important lead compounds for the development of new drugs.[1] The sulfonated azo dyes Evans Blue and Congo Red have been shown to bind to the HIV protease and reverse transcriptase, thereby inhibiting viral replication.[2] Congo Red is widely used as a histological stain for amyloid deposits[3] implicated in Alzheimer's disease as well as in neurodegenerative diseases in sheep, cattle, and mink. Congo Red inhibits the buildup of protease-resistant plaque in infected cells and inhibits replication of the scrapie infective agent.[4] Suramin, a sulfonated urea derivative, has also been evaluated for anti-HIV activity[2b,5] and is a growth factor receptor poison[6] showing potential as an anti-cancer agent by virtue of its ability to inhibit angiogenesis.[7] Sulfonated polymeric compounds can mimic heparin and exhibit antithrombogenic activity.[8] Thus, a molecular-level understanding of the ways in which sulfonated compounds can interact with proteins should be of great utility for the design of new drugs.

Does the sulfonate group bind to proteins and exert its biological effects by mimicking the sulfate group? Significantly, sulfated compounds are involved in a wide range of biological processes. Glycosaminoglycans (GAGs) are but one example; these are large, usually heterogeneously sulfated polysaccharides which have been implicated in diverse biological processes including embryonic development, blood coagulation, and wound healing.[9] Crystallographic, spectroscopic, and theoretical studies on smaller sulfated carbohydrates can provide information relevant to understanding the properties of the less experimentally tractable GAGs.[10] Sulfated carbohydrates have also been shown to inhibit HIV infection and replication.[11] The key role of the sulfate group is evident since even sulfated polymers that are not carbohydrate derivatives can exhibit anti-

[†]Biomedical Engineering Center.
[‡]Department of Chemistry.
[§] Department of Biochemistry.
[⊥] Department of Laboratory Medicine & Pathology.
[⊗] Abstract published in *Advance ACS Abstracts,* February 15, 1996.

(1) (a) Mohan, P. *Pharm. Res.* **1992**, *9*, 703−714. (b) Baba, M.; Schols, D.; Mohan, P.; De Clercq, E.; Shigeta, S. *Antiviral Chem. Chemother.* **1993**, *4*, 229−234. (c) Weaver, J. L.; Pine, P. S.; Anand, R.; Bell, S.; Aszalos, A. *Antiviral Chem. Chemother.* **1992**, *3*, 147−151. (d) Mohan, P.; Singh, R.; Baba, M. *J. Med. Chem.* **1991**, *34*, 212−217. (e) García-Villalón, D.; Gil-Fernández, C. *Antiviral Chem. Chemother.* **1992**, *3*, 9−14.

(2) (a) Brinkworth, R. I.; Fairlie, D. P. *Biochem. Biophys. Res. Commun.* **1992**, *188*, 624−630. (b) Balzarini, J.; Mitsuya, H.; De Clercq, E.; Broder, S. *Int. J. Cancer* **1986**, *37*, 451−457.

(3) (a) Turnell, W. G.; Finch, J. T. *J. Mol. Biol.* **1992**, *227*, 1205−1223. (b) Klunk, W. E.; Pettegrew, J. W.; Abraham, D. J. *J. Histochem. Cytochem.* **1989**, *37*, 1273−1281. (c) Zhang, S.; Holmes, T.; Lockshin, C.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 3334−3338.

(4) (a) Caughey, B.; Ernst, D.; Race, R. E. *J. Virol.* **1993**, *67*, 6270−6272. (b) Caughey, B.; Race, R. E. *J. Neurochem.* **1992**, *59*, 768−771.

(5) Mohan, P.; Hopfinger, A. J.; Baba, M. *Antiviral Chem. Chemother.* **1991**, *2*, 215−222.

(6) Sachsenmaier, C.; Radler-Pohl, A.; Zinck, R.; Nordheim, A.; Herrlich, P.; Rahmsdorf, H. J. *Cell* **1994**, *78*, 963−972.

(7) Gagliardi, A.; Hadd, H.; Collins, D. C. *Cancer Res.* **1992**, *52*, 5073−5075.

(8) Jozefowicz, M.; Jozefonvicz, J. *Pure Appl. Chem.* **1984**, *56*, 1335−1344.

(9) (a) Hardingham, T. E.; Fosang, A. J. *FASEB J.* **1992**, *6*, 861−870. (b) Sugiura, N.; Sakurai, K.; Hori, Y.; Karasawa, K.; Suzuki, S.; Kimata, K. *J. Biol. Chem.* **1993**, *268*, 15779−15787.

(10) (a) Whitfield, D. M.; Tang, T.-H. *J. Am. Chem. Soc.* **1993**, *115*, 9648−9654. (b) Lamba, D.; Glover, S.; Mackie, W.; Rashid, A.; Sheldrick, B.; Pérez, S. *Glycobiology* **1994**, *4*, 151−163. (c) Zsiška, M.; Meyer, B. *Carbohydr. Res.* **1993**, *243*, 225−258. (d) Polvorinos, A. J.; Contreras, R. R.; Martin-Ramos, D.; Romero, J.; Hidalgo, M. A. *Carbohydr. Res.* **1994**, *257*, 1−10. (e) Lamba, D.; Mackie, W.; Rashid, A.; Sheldrick, B.; Yates, E. A. *Carbohydr. Res.* **1993**, *241*, 89−98. (f) Kanters, J. A.; van Dijk, B.; Kroon, J. *Carbohydr. Res.* **1991**, *212*, 1−11. (g) Lamba, D.; Mackie, W.; Sheldrick, B.; Belton, P.; Tanner, S. *Carbohydr. Res.* **1988**, *180*, 183−193.

0002-7863/96/1518-2131$12.00/0 © 1996 American Chemical Society

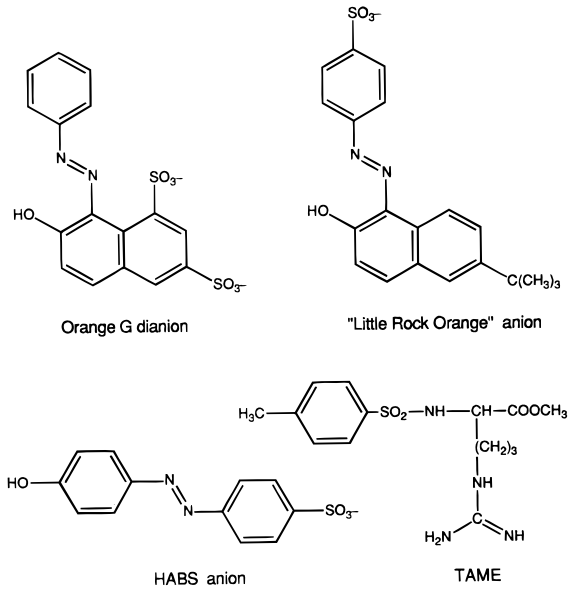
HIV activity; polymers based on poly(vinyl alcohol sulfate) are reported potent inhibitors of the virus.[12] Very recently a report has appeared of the *in vivo* amyloidosis-arresting effects of both sulfated and sulfonated small molecules and the implications of this observation for Alzheimer's disease.[13]

The effectiveness of sulfonated or sulfated compounds as pharmaceutical agents depends on their ability to bind to specific sites on target proteins. On the basis of results reported in this paper and in conjunction with an X-ray crystal structure determination of Congo Red, we have reported a model for the binding of Congo Red to the active site of the HIV-1 protease.[14] Sulfonate groups of this inhibitor can be placed within hydrogen-bonding distance of two arginine side chains which are located at both ends of the protease binding cavity. Evans Blue has been reported to bind with even higher affinity.[2] This can be understood as the result of additional interactions with basic side chains, since both arginine and lysine side chains of the protease are located in positions where tight binding interactions can occur with the four sulfonate groups of Evans Blue. This example shows how a detailed picture of the modes of interaction between sulfonate groups and the side chains of basic residues such as arginine and lysine can be used for understanding how sulfonated materials might bind to proteins.

X-ray crystallography of complexes of sulfonated materials and basic amino acids can provide this detailed picture because the positions and orientations of the sulfonate groups interacting with basic functional groups can be determined. For example, a study using the Cambridge Structural Database has revealed a pattern of preferred hydrogen-bonding geometries assumed by the sulfonyl anion as it interacts with a variety of hydrogen-bond donors.[15] Because binding of sulfonate or sulfate groups to proteins is most likely to occur through the basic side chains of lysine, histidine, and arginine, interactions involving these amino acids are of special interest. Preferred conformations assumed by lysine and arginine side chains in contact with anions (including sulfate) have been the subject of a recent crystallographic report.[16] In other recent work, interactions between basic amino acid side chains and sulfate groups have been shown to be important in the binding of the antiulcer drug sucrose octasulfate (SOS)[17] to acidic fibroblast growth factor (aFGF). In the crystal structure of the SOS−aFGF complex,[18] the SOS molecule is bound to aFGF primarily through interactions between sulfate groups from SOS and lysine and arginine side chains from aFGF. Additional important interactions include sulfate contacts to backbone peptide amide groups and glutamine residues.

Given suitable model systems, it should be possible to determine in detail how the sulfonate and sulfate groups interact with functional groups found on the side chains and in the backbone of proteins. We are investigating both sulfonated[14,19] and sulfated compounds[20] in order to understand the biomolecular recognition of sulfated compounds (such as glycosaminoglycans) or sulfonated compounds (such as sulfonated antiviral agents) by proteins. This report focuses on complexes of amino acids and sulfonated compounds, which serve as models for interactions between sulfonate groups and amino acid side chains in biological systems. In addition to possessing important pharmaceutical properties, sulfonated compounds have a long history of use as cocrystallization or coprecipitation agents for biomolecules such as amino acids and nucleic acid bases.[21]

We describe here the structures of six complexes, each formed between one of the chief sulfate/sulfonate-binding amino acids lysine, histidine, and arginine and a sulfonated azo dye. The structures are DL-lysine/4-[(4-hydroxyphenyl)azo]benzenesulfonate (HABS) monohydrate (**1**), DL-lysine/7-hydroxy-8-(phenylazo)-1,3-naphthalenedisulfonate (Orange G) dihydrate (**2**), L-lysine/Orange G 1.5-hydrate (**3**), DL-histidine/Orange G trihydrate (**4**), L-histidine/Orange G trihydrate (**5**), and tosyl arginine methyl ester (TAME)/4-[(2-hydroxy-6-*tert*-butyl-1-naphthalenyl)azo]benzenesulfonate, a dye we have named "Little Rock Orange" (LRO) (**6**). Included are two pairs of salts



differing only in the stereochemistry (DL vs L) of the amino acid, and the salt of a protected amino acid. The arginine derivative TAME is an especially interesting model, since it lacks charged amino and carboxyl termini.

In describing structures **1**−**6**, we focus on the sulfonate interactions. This is done by examination of the solid-state environments of the sulfonate groups in terms of close inter-

(11) (a) Ito, M.; Baba, M.; Shigeta, S.; Wada, S.; Takagi, M.; Kimura, T.; Okuyama, T. *Antiviral Chem. Chemother.* **1991**, *2*, 41−44. (b) McClure, M. O.; Whitby, D.; Patience, C.; Gooderham, N. J.; Bradshaw, A.; Cheingsong-Popov, R.; Weber, J. N.; Davies, D. S.; Cook, G. M. W.; Keynes, R. J.; Weiss, R. A. *Antiviral Chem. Chemother.* **1991**, *2*, 149−156. (c) Witvrouw, M.; Schols, D.; Andrei, G.; Snoeck, R.; Hosoya, M.; Pauwels, R.; Balzarini, J.; De Clercq, E. *Antiviral Chem. Chemother.* **1991**, *2*, 171−179. (d) Schols, D.; De Clercq, E.; Witvrouw, M.; Nakashima, H.; Snoeck, R.; Pauwels, R.; Van Schepdael, A.; Claes, P. *Antiviral Chem. Chemother.* **1991**, *2*, 45−53. (e) Anand, R.; Nayyar, S.; Pitha, J.; Merril, C. R. *Antiviral Chem Chemother.* **1990**, *1*, 41−46. (f) Kozlowski, M. R.; Watson, A. *Antiviral Chem. Chemother.* **1990**, *1*, 175−182. (g) Otake, T.; Schols, D.; Witvrouw, M.; Naesens, L.; Nakashima, H.; Moriya, T.; Kurita, H.; Matsumoto, K.; Ueba, N.; De Clercq, E. *Antiviral Chem. Chemother.* **1994**, *5*, 155−161.

(12) Schols, D.; De Clercq, E.; Balzarini, J.; Baba, M.; Witvrouw, M.; Hosoya, M.; Andrei, G.; Snoeck, R.; Neyts, J.; Pauwels, R.; Nagy, M.; Györgyi-Edelényi, J.; Machovich, R.; Horváth, I.; Löw, M.; Görög, S. *Antiviral Chem. Chemother.* **1990**, *1*, 233−240.

(13) Kisilevsky, R.; Lemieux, L. J.; Fraser, P. E.; Kong, X.; Hultin, P. G.; Szarek, W. A. *Nature Med.* **1995**, *1*, 143−148.

(14) Ojala, W. H.; Ojala, C. R.; Gleason, W. B. *Antiviral Chem. Chemother.* **1995**, *6*, 25−33.

(15) Kanyo, Z. F.; Christianson, D. W. *J. Biol. Chem.* **1991**, *266*, 4264−4268.

(16) Chakrabarti, P. *Int. J. Peptide Protein Res.* **1994**, *43*, 284−291.

(17) Nawata, Y.; Ochi, K.; Shiba, M.; Morita, K., Iitaka, Y. *Acta Crystallogr.* **1981**, *B37*, 246−249.

(18) Zhu, X.; Hsu, B. T.; Rees, D. C. *Structure* **1993**, *1*, 27−34.

(19) (a) Ojala, W. H.; Gleason, W. B.; Richardson, T. I.; Lovrien, R. E. *Acta Crystallogr.* **1994**, *C50*, 1615−1620. (b) Ojala, W. H.; Lu, L. K.; Albers, K. E.; Gleason, W. B.; Richardson, T. I.; Lovrien, R. E.; Sudbeck, E. A. *Acta Crystallogr.* **1994**, *B50*, 684−694.

(20) Ojala, W. H.; Albers, K. E.; Gleason, W. B.; Choo, C. G. *Carbohydr. Res.* **1995**, *275*, 49−65.

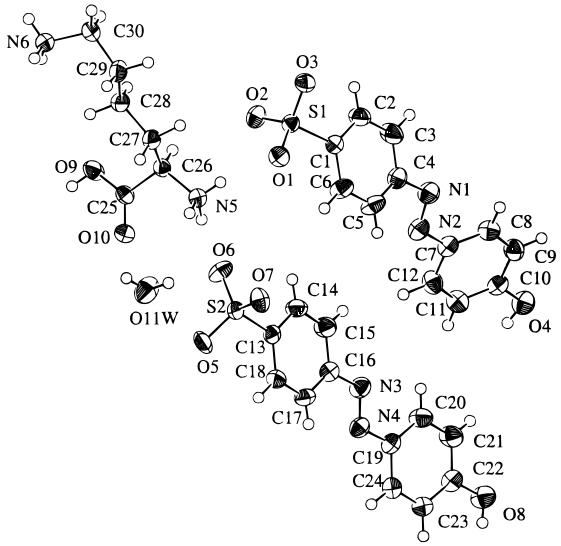


Figure 1. ORTEPII drawing of DL-lysine/HABS (**1**), showing atom numbering. For non-H atoms, 50% probability ellipsoids are shown.

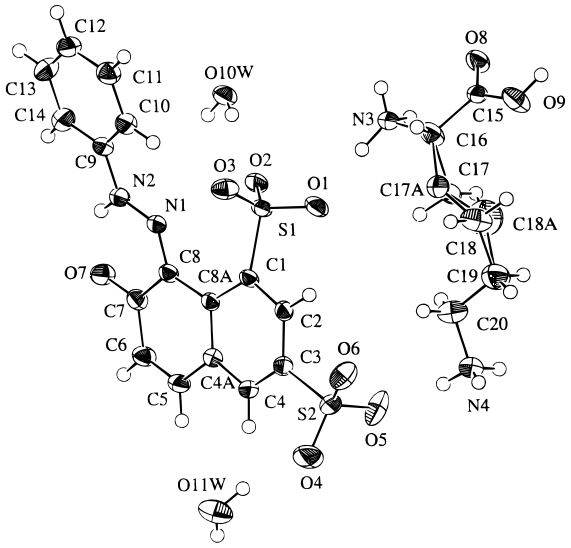


Figure 2. ORTEPII drawing of DL-lysine/Orange G (**2**), showing atom numbering. For non-H atoms, 50% probability ellipsoids are shown.

molecular contacts and use of the graph-set formalism,[22] which provides a complementary view of long-range networks.

Experimental Section

Crystals of **1**−**6** were grown from 2 N HCl. An Enraf-Nonius CAD-4 diffractometer was used for data collection[23] for structures **1**, **4**, **5**, and **6,** and a Rigaku AFC6S diffractometer[24] was used for **2** and

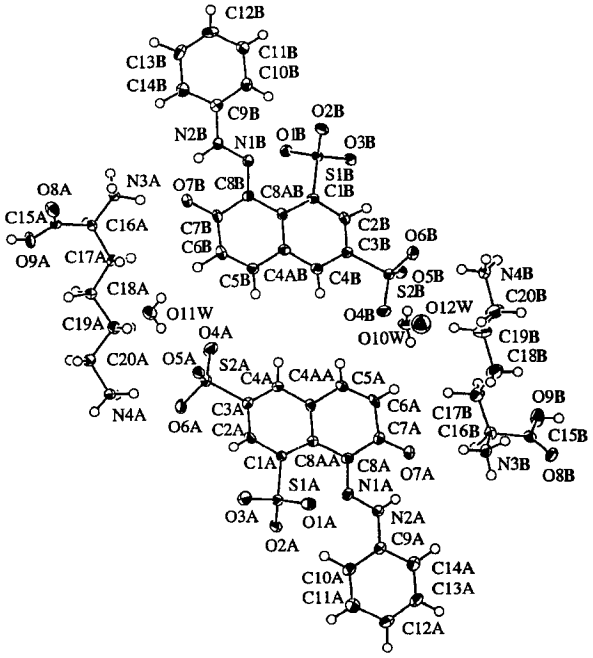


Figure 3. ORTEPII drawing of L-lysine/Orange G (**3**), showing atom numbering. For non-H atoms, 50% probability ellipsoids are shown.

3. Structures **1** and **4** were determined at room temperature; the remaining structures were determined at low temperature (173 K). In all cases Cu Kα radiation was used, and data were collected in the $\theta-2\theta$ scan mode. All structures were solved using SHELXS86[25] and were refined using the TEXSAN software package.[26] Figures were prepared using ORTEPII[27] and PLUTO.[28] Additional details of and results from the crystal structure determinations have been deposited as supporting information.

Results and Discussion

Molecular Geometry. Extensive tables of fractional atomic coordinates, bond lengths, bond angles, and torsion angles have been deposited as supporting information. ORTEPII views of the asymmetric unit of each structure are given in Figures 1−6.

Structural Features of Sulfonated Azo Dyes and Amino Acids in the Complexes. In previous work involving Orange G, we observed the S(1)−C(1) bond to be significantly longer than the S(2)−C(3) bond [for example, 1.795(4) Å vs 1.759(4) Å in the calcium salt[19b]]. This is also the case in the six Orange G molecules in structures **2**−**5**. The S(2)−C(3) bond lengths in these structures are comparable to the S−C bond lengths in HABS and LRO, indicating that the difference found in Orange G is due to an unusually long S(1)−C(1) bond rather than to an unusually short S(2)−C(3) bond. We suggested that lengthening of the S(1)−C(1) bond in Orange G is the result of steric interaction with the bulky phenylazo substituent on C(8),[19b] an interaction not present in HABS or LRO.

The side chain conformations of the four lysine molecules in structures **1**−**3** are varied. The lysine side chain in **1** is folded

(21) (a) Sudbeck, E. A.; Etter, M. C.; Gleason, W. B. *Chem. Mater.* **1994**, *6*, 1192−1199. (b) Garber, L. T.; Haltiwanger, R. C.; Eggleston, D. S. *Lett. Peptide Sci.* **1994**, *1*, 127−133. (c) Krause, J. A.; Baures, P. W.; Eggleston, D. S. *Acta Crystallogr.* **1993**, *B49*, 123−130. (d) Conroy, M. J.; Lovrien, R. E. *J. Cryst. Growth* **1992**, *122*, 213−222. (e) Kimoto, H.; Saigo, K.; Ohashi, Y.; Hasegawa, M. *Bull. Chem. Soc. Jpn.* **1989**, *62*, 2189−2195. (f) Yamada, S.; Yamamoto, M.; Chibata, I. *J. Org. Chem.* **1973**, *38*, 4408−4412. (g) van der Veen, J. M.; Low, B. W. *Acta Crystallogr.* **1972**, *B28*, 3548−3559. (h) Stein, W. H.; Moore, S.; Bergmann, M. *J. Biol. Chem.* **1944,** *154*, 191−201. (i) Bergmann, M.: Stein, W. H. *J. Biol. Chem.* **1939**, *129*, 609−618. (j) Crosby, B. L.; Kirk, P. L. *Mikrochemie* **1935**, *18*, 137−143. (k) Fischer, E.; Bergell, P. *Ber. Dtsch. Chem. Ges.* **1902**, *35*, 3779−3787.

(22) (a) Etter, M. C. *Acc. Chem. Res.* **1990**, *23*, 120−126. (b) Etter, M. C.; MacDonald, J. C.; Bernstein, J. *Acta Crystallogr.* **1990**, *B46*, 256−262. (c) Bernstein, J.; Etter, M. C.; MacDonald, J. C. *J. Chem. Soc., Perkin Trans. 2* **1990**, 695−698. (d) Etter, M. C. *J. Phys. Chem.* **1991**, *95*, 4601−4610. (e) Bernstein, J.; Davis, R. E.; Shimoni, L.; Chang, N.-L. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 1555−1573.

(23) Enraf-Nonius. *CAD-4 Software,* Version 5.0; Enraf-Nonius: Delft, The Netherlands, 1989.

(24) Molecular Structure Corp. *MSC/AFC Diffractometer Control Software;* MSC, 3200 Research Forest Dr., The Woodlands, TX 77381, 1988.

(25) Sheldrick, G. M. *SHELXS86. Program for the Solution of Crystal Structures;* University of Göttingen: Göttingen, Germany, 1985.

(26) Molecular Structure Corp. *TEXSAN-TEXRAY Structure Analysis Package;* MSC, 3200 Research Forest Dr., The Woodlands, TX 77381, 1985.

(27) Johnson, C. K. ORTEPII. Report ORNL-5138*;* Oak Ridge National Laboratory: Oak Ridge, TN, 1976.

(28) Motherwell, W. D. S.; Clegg, W. *PLUTO. Program for Plotting Molecular and Crystal Structures;* University of Cambridge: Cambridge, England, 1976.

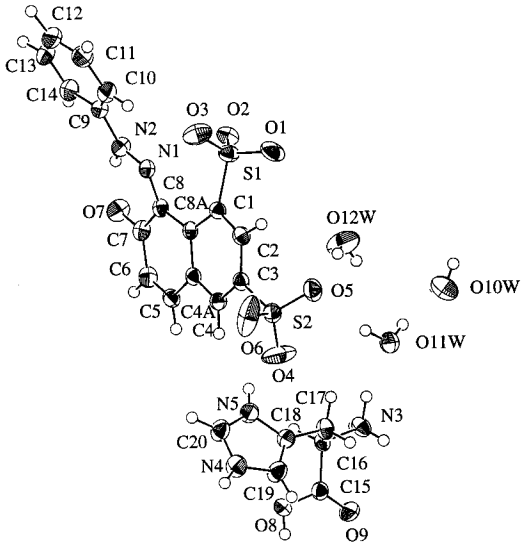


Figure 4. ORTEPII drawing of DL-histidine/Orange G (**4**), showing atom numbering. For non-H atoms, 50% probability ellipsoids are shown.

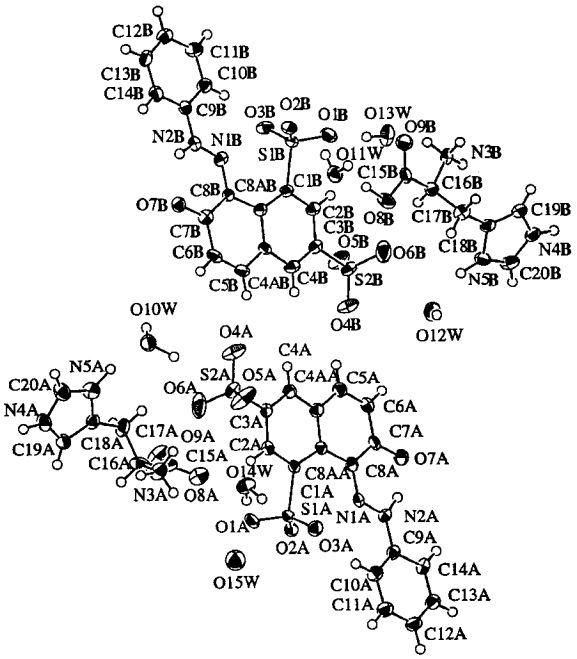


Figure 5. ORTEPII drawing of L-histidine/Orange G (**5**), showing atom numbering. For non-H atoms, 50% probability ellipsoids are shown.

[C(28)−C(29)−C(30)−N(6) torsional angle 61.1(3)°], as is the major orientation of the disordered side chain of **2** [C(17)−C(18)−C(19)−C(20) torsional angle −57.0(5)°]. Folded lysine conformations have been observed in other structures, particularly in those in which both amino groups of lysine make contact with the same anion.[16] Such is not the case here. On the other hand, the folded conformation in **1** allows the terminal amino group to make contact with three different sulfonate oxygen atoms. In contrast, the two crystallographically independent lysine molecules in structure **3** exhibit extended conformations. These two molecules differ from each other chiefly in the torsional angle describing the position of the α-amino group; the orientation of this group is *trans* with respect to the side chain in one molecule and *gauche* in the other.

A *trans* vs *gauche* difference is also observed in the D- and L-histidine molecules in **4** and the L-histidine molecules in **5**. The histidines in **4** are folded, the C(15)−C(16)−C(17)−C(18) torsional angle being *gauche*, but the histidines in **5** are extended, having *trans* orientations about this linkage.

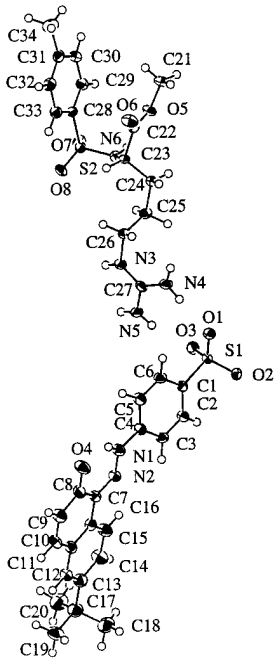


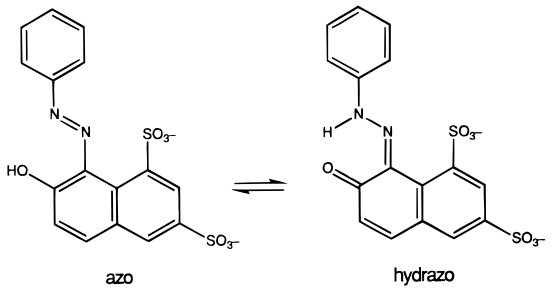
Figure 6. ORTEPII drawing of tosylarginine methyl ester/Little Rock Orange (**6**), showing atom numbering. For non-H atoms, 50% probability ellipsoids are shown.

Although the side chain of tosylarginine methyl ester in **6** is not fully extended, the *gauche* orientation about C(23)−C(24)−C(25)−C(26) is not unusual and has been observed in other small-molecule arginine-containing structures.[16] The guanidinium terminus deviates slightly from planarity by an 8.9(6)° twist about the C(26)−N(3)−C(27)−N(4) linkage. Slight deviations of this type are reported to be common for small-molecule arginine-containing structures and have been ascribed to hydrogen-bonding interactions.[16] Here the twist may be crucial in allowing the guanidinium group to make two simultaneous bidentate contacts with neighboring sulfonate groups as discussed below.

The dye molecules in all six structures can be described as generally planar, but there are some deviations. For example, although one of the two crystallographically independent HABS molecules in structure **1** is almost flat, the other definitely is not, as can be seen from its torsional angles at the azo linkage [C(3)−C(4)−N(1)−N(2) = −149.3(3)°; N(1)−N(2)−C(7)−C(8) = 12.7(4)°]. The LRO molecule in **6** deviates from planarity, its phenyl group being twisted out of the molecular plane [C(3)−C(4)−N(1)−N(2) = −15.7(6)°], but the C(4)−N(1)−N(2)−C(7) and N(1)−N(2)−C(7)−C(8) linkages are sufficiently coplanar to allow formation of an intramolecular hydrogen bond between N(1) and O(4). This high degree of coplanarity in the N(1)−N(2)−C−C−O linkage is also observed in all six Orange G molecules found in structures **2**−**5**, and it allows intramolecular hydrogen-bond formation in these structures as well.

In this intramolecular hydrogen bond, the donor atom is not a naphthol oxygen atom but rather a hydrazo nitrogen atom. An azo compound with a suitably-placed phenolic oxygen can exist as the azo compound, as the hydrazo compound, or as an equilibrium mixture of the two forms in the solid state as well as in solution.[29] As is true of Orange G molecules in salts we

(29) Olivieri, A. C.; Wilson, R. B.; Paul, I. C.; Curtin, D. Y. *J. Am. Chem. Soc.* **1989**, *111*, 5525−5532.



have examined previously,[19] the Orange G molecules in **2**−**5** are better described as hydrazo compounds rather than as azo compounds. The LRO molecule in **6** also assumes the hydrazo form. In each case the hydrogen atom is found on the nitrogen atom instead of on the oxygen atom, and the C−O, C−N, and N−N bond lengths are consistent with a structure that is predominantly hydrazo in character. Certain other sulfonated azo dyes that are of key biomedical interest likewise have the potential to exist as azo compounds, hydrazo compounds, or mixtures of the two. Evans Blue is a notable example. This isomerism may have implications for the conformation of these dyes and for their ability to bind to proteins, so it should be taken into account in the design of structurally related pharmaceutical agents.

Sulfonate groups on the dye molecules can assume either staggered or eclipsed orientations, and both orientations are observed in structures **1**−**6**. There appears to be no clear preference for the staggered sulfonate conformation (one S−O bond oriented perpendicular to the naphthalene plane) over the eclipsed conformation (one S−O bond oriented parallel to the naphthalene plane).[37] In these and other structures we have examined,[19] sulfonate shows a lesser tendency than sulfate to assume a fixed, staggered orientation with respect to the moiety to which it is bonded. This torsional variability is exemplified by structure **3**, in which one sulfonate on each dye molecule is eclipsed while the other one lies on the borderline between eclipsed and staggered. In contrast, in crystal structures of sulfated sugars the torsional orientation assumed by the sulfate group is found to be similar from structure to structure and appears to be independent of its environment.[10b,20] If the barrier to rotation of the sulfonate groups in these structures is in fact lower than that for sulfate groups in the sulfated sugars, perhaps it is because the bond connecting the sulfonate SO_3 group to the rest of the molecule is an S−C bond approximately 1.77−1.80 Å long instead of the shorter S−O bond (in sulfate) approximately 1.60 Å long. In sulfate the SO_3 group is always adjacent to the same atom (the linking oxygen atom), so the barrier to rotation for the SO_3 in sulfate is essentially the same from molecule to molecule. In contrast, the SO_3 of sulfonate is connected directly to the rest of the molecule and may be more sensitive to differences in its molecular surroundings. Interestingly, the greater torsional variability shown by sulfonate may actually facilitate its mimicking of sulfate; it may allow the sulfonate group to rotate into an orientation permitting it to bind to a normally sulfate-recognizing protein.

Intermolecular Interactions: **General Features.** It is particularly interesting that in none of the packing arrangements in **1**−**5** (in which the carboxyl groups are not protected as they are in **6**) can dimers formed between carboxyl groups be found. Such hydrogen-bonded dimers are extremely common for protonated carboxylic acids,[30] and their absence in these crystal structures is striking. Close contacts to the carboxyl groups are instead made by sulfonate oxygen atoms or by water molecules that are themselves in contact with one or more sulfonate oxygens. Whether direct or water-mediated, this preference for sulfonate−carboxyl interactions over carboxyl−carboxyl interactions probably facilitates cocrystal formation.

(30) Leiserowitz, L. *Acta Crystallogr.* **1976**, *B32*, 775−802.

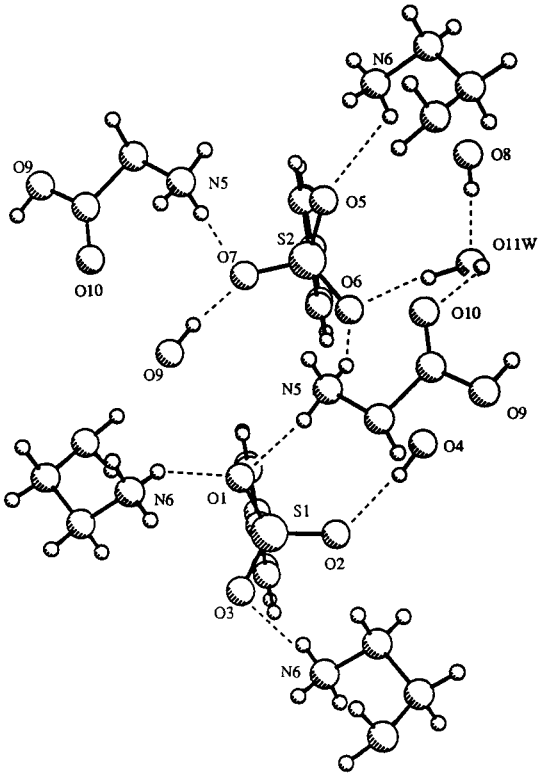


Figure 7. PLUTO view of the crystalline environment of the sulfonate groups in DL-lysine/HABS (**1**), showing close intermolecular contacts. See Table 1 for distances. To improve clarity, only selected portions of the molecules are shown.

Table 1. Intermolecular Contacts (Å) in DL-Lysine/HABS (**1**)[a]

Contacts to the Sulfonate Groups			
O(1)···N(5)[a]	2.875(3)	O(6)···O(11)W[a]	2.815(3)
O(1)···N(6)[b]	2.908(3)	O(6)···N(5)[a]	2.834(3)
O(2)···O(4)[c]	2.810(3)	O(7)···O(9)[b]	2.595(2)
O(3)···N(6)[d]	2.853(3)	O(7)···N(5)[f]	2.867(3)
O(5)···N(6)[e]	2.742(3)		
Contacts to Other Atoms			
O(8)···O(11W)[g]	2.641(3)	O(10)···O(11W)[a]	2.928(3)
O(8)···N(6)[h]	2.942(3)		

[a] See Table 9 for the meaning of the superscript letters.

Sulfonate Contacts in DL-Lysine/HABS (1). The environment of the sulfonate groups in **1** is shown in Figure 7. Distances are given in Table 1. The sulfonate−carboxyl interaction is clearly important here, the approach between sulfonate oxygen atom O(7) and the carboxyl OH group O(9) being the closest hydrogen-bonding contact in this structure. Three of the six sulfonate oxygen atoms participate in two close (<3 Å) intermolecular contacts each, and the remaining three oxygens participate in only one contact each. Close approaches from the sulfonates to the protonated amino groups of the lysine are equally divided (three each) between the α-amino and side-chain amino groups. Only one sulfonate oxygen, O(2), is not in contact with an amino group; this oxygen makes direct contact only with the phenolic hydroxyl group O(4) of another HABS molecule. Only one sulfonate oxygen, O(6), comes into direct contact with the water molecule. That water molecule is also in contact with the carbonyl oxygen atom O(10) of the lysine and with the phenolic oxygen O(8) of one of the HABS molecules. The other phenolic hydroxyl group is in direct

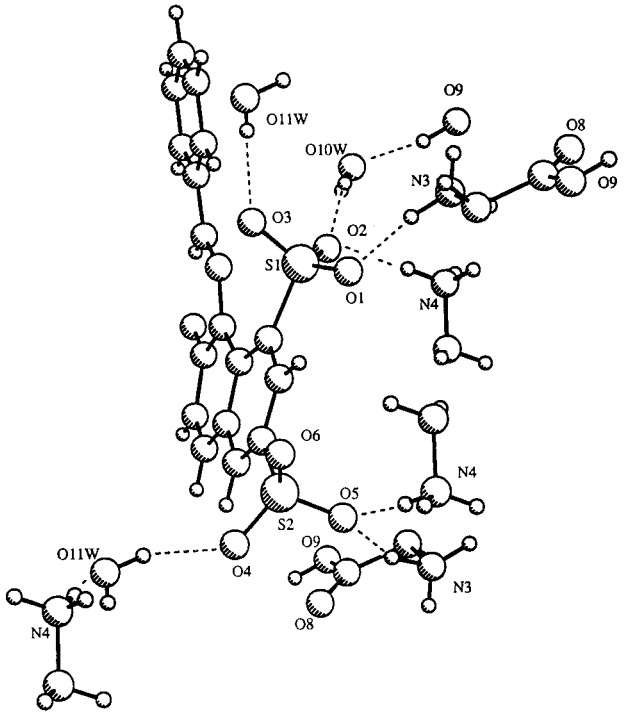


Figure 8. PLUTO view of the crystalline environment of the sulfonate groups in DL-lysine/Orange G (**2**), showing close intermolecular contacts. See Table 2 for distances. To improve clarity, only selected portions of the molecules are shown.

Table 2. Intermolecular Contacts (Å) in DL-Lysine/Orange G (**2**)[a]

Contacts to the Sulfonate Groups			
O(1)···N(3)[a]	2.790(3)	O(4)···O(11W)[a]	2.802(4)
O(2)···O(10W)[a]	2.799(3)	O(5)···N(4)[a]	2.845(4)
O(2)···N(4)[i]	2.844(4)	O(5)···N(3)[i]	2.872(4)
O(3)···O(11W)[j]	2.806(3)	O(5)···O(8)[b]	3.001(4)
Contacts to Other Atoms			
O(7)···O(10W)[c]	2.744(3)	O(9)···O(10W)[k]	2.582(3)
O(8)···N(3)[k]	2.933(4)	O(11W)···N(4)[l]	2.774(4)

[a] See Table 9 for the meaning of the superscript letters.

contact with a sulfonate oxygen atom. These contacts between the sulfonate oxygens and the phenolic hydroxyl groups indicate possible interaction modes for both direct and water-mediated interactions between sulfate or sulfonate groups and tyrosine side chains in proteins.

Sulfonate Contacts in DL-Lysine/Orange G (2). The environment of the sulfonate groups in **2** is shown in Figure 8. Distances are given in Table 2. Two of the sulfonate oxygens make two close hydrogen-bonding contacts, three make a single contact, and one of them, O(6), makes no contacts. The nearly eclipsed position of the S(2)−O(6) bond with respect to the naphthalene ring may make O(6) less accessible than the other sulfonate oxygen atoms. The closest hydrogen-bonding contact in **2** occurs between the carboxyl OH O(9) and water molecule O(10W), which itself is in close contact with sulfonate oxygen O(2). Thus, in **2** the sulfonate−carboxyl interaction that apparently substitutes for carboxyl dimer formation is water-mediated. Structure **2**, like structure **1**, has two sulfonate groups and two amino groups in its asymmetric unit, but **2** has an additional water molecule. Perhaps this is why water-mediated contacts between the sulfonate oxygens and the protonated amino groups of lysine can be found in **2** but not in **1**. Both O(3) and O(4) make water-mediated contacts with the terminal amino groups N(4) of lysine molecules via O(11W) water molecules. O(1) directly contacts the α-amino group N(3), O(2) directly contacts the terminal amino group N(4), and O(5)

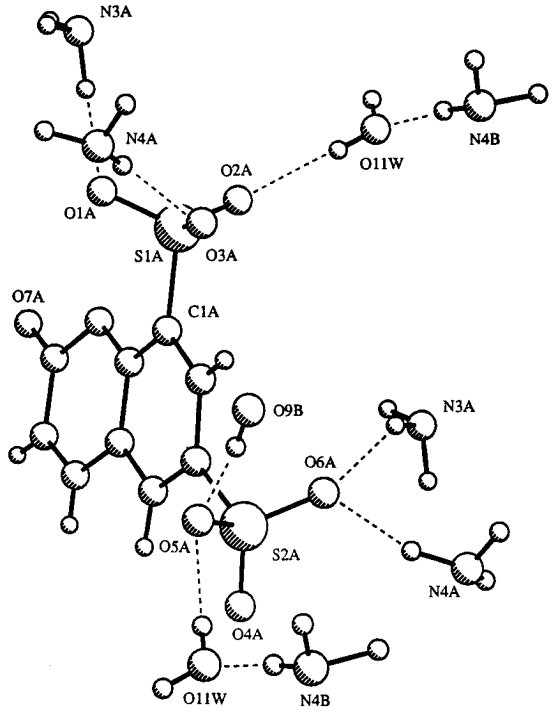


Figure 9. PLUTO view of the crystalline environment of the sulfonate groups of dye molecule A in L-lysine/Orange G (**3**), showing close intermolecular contacts. N3A/B = α-amino; N4A/B = side-chain terminal amino. See Table 3 for distances. To improve clarity, only selected portions of the molecules are shown.

directly contacts both kinds of amino groups. Of the six intermolecular interactions between sulfonate oxygens and amino groups, four of them are to the side-chain terminus (two direct and two water-mediated) and two of them are to the α-amino group (both direct).

Sulfonate Contacts in L-Lysine/Orange G (3). The sulfonate environment in **3** is shown in Figures 9 and 10. Distances are given in Table 3. The view of the asymmetric unit given in Figure 3 shows that this structure is pseudocentrosymmetric and that several similar contacts are made by the two dye molecules. The two closest intermolecular contacts are those between sulfonate oxygens and, again, the carboxyl OH groups. Four sulfonate oxygens make two intermolecular close contacts, seven make single contacts, and one [O(4A), the S(2) sulfonate oxygen which occupies the nearly eclipsed position with respect to the naphthalene ring] makes no contacts. For molecule A, O(1A) makes direct contact with the α-amino group N(3A), O(3A) makes direct contact with the side-chain amino group N(4A), and O(6A) makes direct contact with both amino groups. Sulfonate oxygens from molecule A are also involved in water-bridged interactions with a terminal amino nitrogen: O(11W) connects both O(2A) and O(5A) to N(4B). The surroundings of molecule B in this noncentrosymmetric structure do differ from those of the crystallographically independent molecule A, with only three direct contacts involving sulfonate oxygens (one to the α-amino group and two to the side-chain terminus) but four water-bridged contacts: O(10W) connects O(2B) to both N(4A) and N(3B), and O(5B) to both N(4A) and N(3B) as well. Although dye molecules A and B are not required by crystallographic symmetry to make similar intermolecular contacts, there are nonetheless many, and both molecules resemble the dye molecule in **2** with respect to the types of intermolecular contacts they make (Tables 2 and 4).

Although only direct interactions between sulfonate oxygens and the α-amino group are found in structures **1** and **2**, both

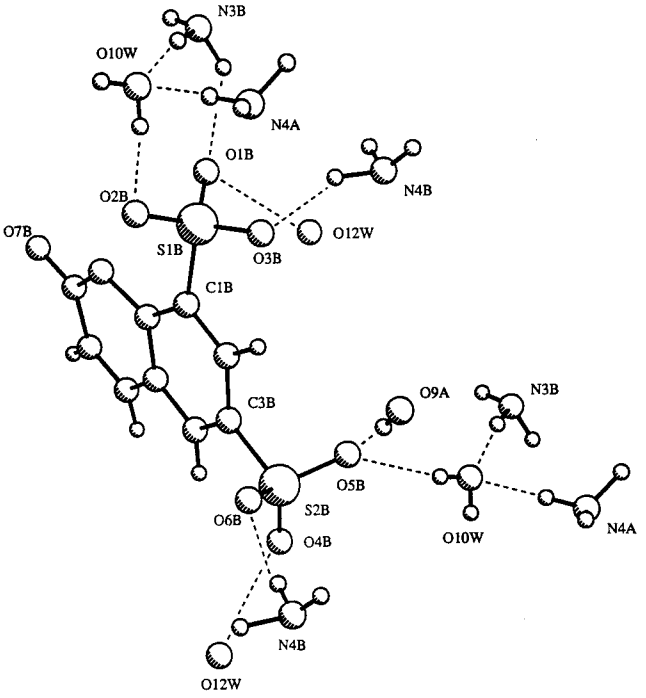


Figure 10. PLUTO view of the crystalline environment of the sulfonate groups of dye molecule B in L-lysine/Orange G (**3**), showing close intermolecular contacts. N3A/B = α-amino; N4A/B = side-chain terminal amino. See Table 3 for distances. To improve clarity, only selected portions of the molecules are shown.

Table 3. Intermolecular Contacts (Å) in L-Lysine/Orange G (**3**)[a]

Contacts to the Sulfonate Groups			
O(1A)···N(3A)[m]	2.740(5)	O(5A)···O(9B)[q]	2.582(4)
O(1B)···N(3B)[n]	2.859(5)	O(5A)···O(11W)[a]	2.958(4)
O(1B)···O(12W)[b]	2.91(1)	O(5B)···O(9A)[r]	2.637(4)
O(2A)···O(11W)[o]	2.735(5)	O(5B)···O(10W)[a]	2.870(4)
O(2B)···O(10W)[p]	2.746(4)	O(6A)···N(4A)[a]	2.839(5)
O(3A)···N(4A)[j]	2.820(5)	O(6A)···N(3A)[o]	2.983(5)
O(3B)···N(4B)[b]	2.788(5)	O(6B)···N(4B)[a]	2.820(5)
O(4B)···O(12W)[a]	2.72(1)		
Contacts to Other Atoms			
O(7A)···N(3B)[a]	2.843(5)	O(10W)···N(3B)[b]	3.009(6)
O(7B)···N(3A)[a]	2.748(5)	O(11W)···N(4B)[t]	2.762(5)
O(10W)···N(4A)[s]	2.892(5)		

[a] See Table 9 for the meaning of the superscript letters.

direct and water-mediated interactions with it are found in structure **3**. In structures **1**−**3** more sulfonate contacts of both kinds are made to the side-chain amino group (15) than to the α-amino group (10), and a higher proportion of those to the side-chain amino group are water-mediated (6 of the 15 for the side-chain amino group compared to 2 of the 10 for the α-amino group). The direct contacts involving sulfonate oxygens in **1**−**3** are divided almost evenly between the two kinds of lysine amino groups. The greater tendency for those contacts made to the side-chain terminal amino group to be water-mediated may result from the chain's flexibility, which allows the terminal amino group to hydrogen bond to the maximum possible number of acceptor molecules, particularly water molecules.

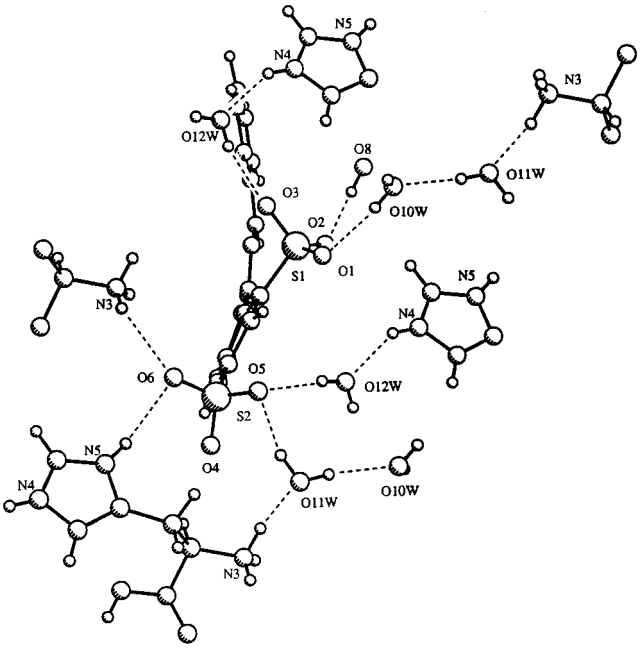


Figure 11. PLUTO view of the crystalline environment of the sulfonate groups in DL-histidine/Orange G (**4**), showing close intermolecular contacts. See Table 5 for distances. To improve clarity, only selected portions of the molecules are shown.

Sulfonate Contacts in DL-Histidine/Orange G (4). The immediate surroundings of the sulfonate groups in **4** are shown in Figure 11. Distances are given in Table 5. There are only two direct short contacts (<3 Å) between the sulfonate oxygens and the amino acid molecules: O(2) to O(8) (a contact with the acidic OH) and O(6) to N(3) (a contact with the protonated α-amino group). However, there is also an approach of 3.090(5) Å between sulfonate oxygen O(6) and N(5) of the histidine ring. Four water-mediated contacts occur between the sulfonate oxygen atoms and the amino acid nitrogen atoms. Sulfonate oxygen O(1) is connected to N(3) by a bridge consisting of *two* water molecules, O(10W) and O(11W). O(3) contacts the protonated ring nitrogen N(4) by way of bridging water molecule O(12W). O(5) contacts the α-amino group N(3) through O(11W) and the ring nitrogen N(4) through O(12W). In this structure only one sulfonate oxygen engages in two close (<3 Å) contacts, four engage in only one contact, and one oxygen atom [O(4), the S(2) sulfonate oxygen in the eclipsed position] engages in none.

Sulfonate Contacts in L-Histidine/Orange G (5). The environments of the sulfonate groups in **5** are shown in Figures 12 and 13. Distances are given in Table 6. Like the L-lysine/Orange G structure (**3**), the L-histidine/Orange G structure (**5**) is pseudosymmetric, even in the placement of the water molecules. In this structure the sulfonate oxygens come into close (<3 Å) direct contact with the α-amino groups but only

Table 4. Types of Contacts (<3 Å) to Sulfonate Oxygens in DL-Lysine/Orange G (**2**) and L-Lysine/Orange G (**3**)

2		**3**, molecule A		**3**, molecule B	
O(1)	α-amino	O(1)	α-amino	O(1)	α-amino; water
O(2)	water bridge to carboxyl OH; terminal amino	O(2)	water bridge to terminal amino	O(2)	water bridge to α-amino and terminal amino
O(3)	water bridge to terminal amino	O(3)	terminal amino	O(3)	terminal amino
O(4)	water bridge to terminal amino	O(5)	water bridge to terminal amino; carboxyl OH	O(5)	water bridge to α-amino and terminal amino; carboxyl OH
O(5)	α-amino; terminal amino	O(6)	α-amino; terminal amino	O(6)	terminal amino
O(6)	nothing	O(4)	nothing	O(4)	water

Table 5. Intermolecular Contacts (Å) in DL-Histidine/Orange G (**4**)[a]

Contacts to the Sulfonate Groups			
O(1)···O(10W)[u]	2.878(4)	O(5)···O(12W)[a]	2.861(4)
O(2)···O(8)[v]	2.635(3)	O(5)···O(11W)[a]	2.889(4)
O(3)···O(12W)[b]	2.876(4)	O(6)···N(3)[b]	2.967(4)
Contacts to Other Atoms			
O(7)···N(3)[w]	2.757(4)	O(11W)···N(3)[a]	2.819(4)
O(10W)···O(11W)[a]	2.753(4)	O(12W)···N(4)[aa]	2.703(4)

[a] See Table 9 for the meaning of the superscript letters.

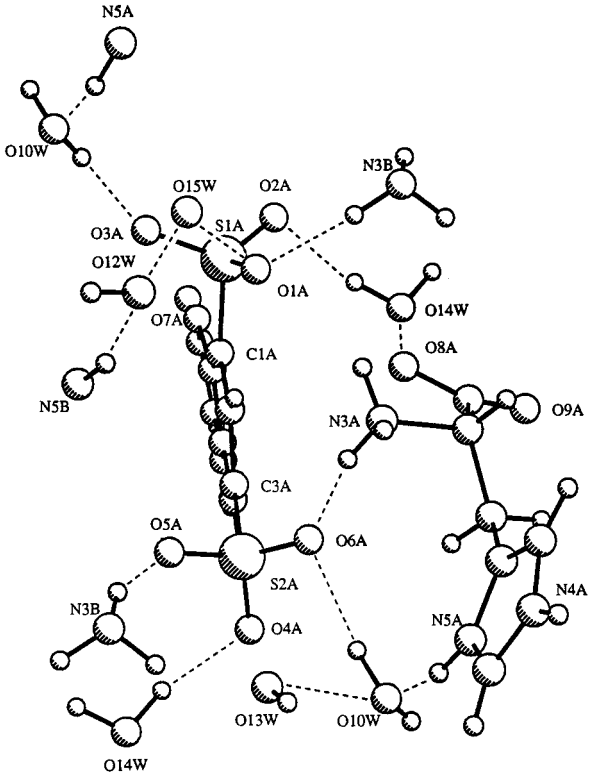


Figure 12. PLUTO view of the crystalline environment of the sulfonate groups of dye molecule A in L-histidine/Orange G (**5**), showing close intermolecular contacts. See Table 6 for distances. To improve clarity, only selected portions of the molecules are shown.

into water-mediated contact with the ring nitrogen atoms, and the close interactions are only with N(4). Each sulfonate oxygen is involved in at least one close intermolecular contact. Here O(4A) and O(4B) come into close contact with water molecules that allow them to make water-mediated contact with the carboxyl OH groups. Two sulfonate oxygens from molecule A participate in two close contacts each; the rest of the sulfonate oxygens all make single contacts. The closest intermolecular approaches in the structure are those between the carboxyl OH groups and water molecules which are themselves in contact with the sulfonate oxygens O(2A) and O(2B).

As is true of the DL- and L-lysine structures **2** and **3**, there are correspondences between close sulfonate−dye approaches in the DL- and L-histidine structures **4** and **5** (Tables 5 and 7). Although in each pair the crystal structures are different, there are nevertheless similar contacts between the amino acid and sulfonated azo dye in spite of the differences in crystal structures, sometimes through the intervention of bridging water molecules. This implies a significant degree of specificity in the interactions between amino acids and sulfonated dye molecules, a fact with important implications for biomolecular recognition of sulfonated/sulfated compounds by proteins.

It is especially interesting that, at least in the two histidine-containing structures examined here, the sulfonate contacts to

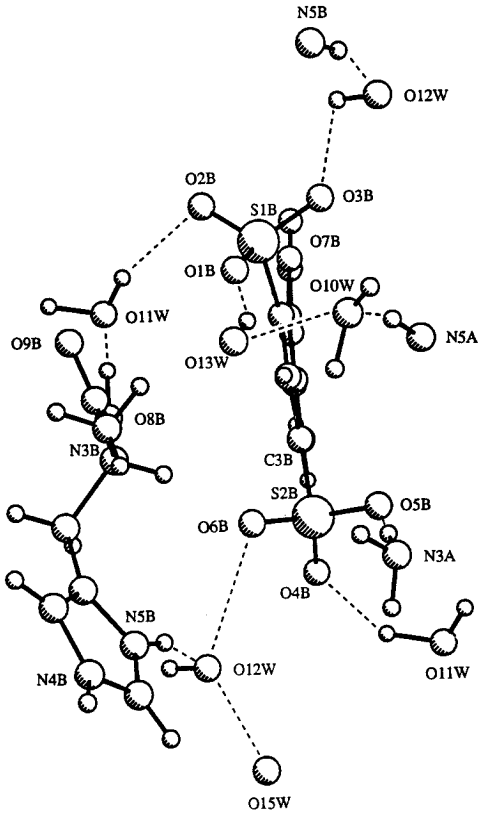


Figure 13. PLUTO view of the crystalline environment of the sulfonate groups of dye molecule B in L-histidine/Orange G (**5**), showing close intermolecular contacts. See Table 6 for distances. To improve clarity, only selected portions of the molecules are shown.

Table 6. Intermolecular Contacts (Å) in L-Histidine/Orange G (**5**)[a]

Contacts to the Sulfonate Groups			
O(1A)···O(15W)[a]	2.699(9)	O(4A)···O(14W)[j]	2.851(6)
O(1A)···N(3B)[n]	2.893(5)	O(4B)···O(11W)[b]	2.798(5)
O(1B)···O(13W)[a]	2.670(5)	O(5A)···N(3B)[p]	2.731(5)
O(2A)···O(14W)[a]	2.841(5)	O(5B)···N(3A)[o]	2.879(6)
O(2B)···O(11W)[a]	2.923(6)	O(6A)···N(3A)[a]	2.879(8)
O(3A)···O(10W)[bb]	2.706(5)	O(6A)···O(10W)[a]	2.995(7)
O(3B)···O(12W)[cc]	2.718(6)	O(6B)···O(12W)[a]	2.875(7)
Contacts to Other Atoms			
O(7A)···N(4A)[dd]	2.627(6)	O(10W)···O(13W)[p]	2.889(5)
O(7B)···N(4B)[ee]	2.718(6)	O(12W)···N(5B)[a]	2.704(6)
O(8A)···O(14W)[a]	2.550(5)	O(12W)···O(15W)[o]	2.788(9)
O(8B)···O(11W)[a]	2.539(5)	O(13W)···N(3B)[a]	2.735(4)
O(9B)···N(3A)[m]	2.802(5)	O(15W)···N(5A)[bb]	2.919(9)
O(10W)···N(5A)[a]	2.763(6)		

[a] See Table 9 for the meaning of the superscript letters.

the histidine ring are water-mediated. This result suggests that histidine side chains may participate as sulfonate acceptor groups in macromolecular complexes via bridging water molecules. As an example, the large movement of histidine 124 in FGF upon binding SOS does not result in any contacts close enough to be directly mediated by hydrogen bonding.[18] At the present stage of refinement of the FGF/SOS complex there are no water molecules in the model. However, a water-mediated contact between histidine and SOS may explain the histidine movement upon complexation.

Sulfonate Contacts in TAME/LRO (6). Interactions involving the sulfonate group in **6** are shown in detail in Figure 14. Distances are given in Table 8. This is the only structure in which the amino acid is protected at the amino and carboxyl groups. Consequently, there are fewer hydrogen-bonding donors and acceptors to interact with water molecules, and in fact no water molecules are found in this structure (in contrast to

Table 7. Types of Contacts (<3 Å) to Sulfonate Oxygens in DL-Histidine/Orange G (**4**) and L-Histidine/Orange G (**5**)

4		**5**, molecule A		**5**, molecule B	
O(1)	two-water bridge to α-amino	O(1)	two-water bridge to ring N	O(1)	two-water bridge to ring N
O(2)	carboxyl OH	O(2)	water bridge to carboxyl OH	O(2)	water bridge to carboxyl OH
O(3)	water bridge to ring N	O(3)	water bridge to ring N	O(3)	water bridge to ring N
O(4)	nothing	O(4)	water bridge to carboxyl OH	O(4)	water bridge to carboxyl OH
O(5)	water bridges to α-amino and ring N	O(6)	α-amino; water bridge to ring N	O(6)	water bridge to ring N
O(6)	α-amino	O(5)	α-amino	O(5)	α-amino

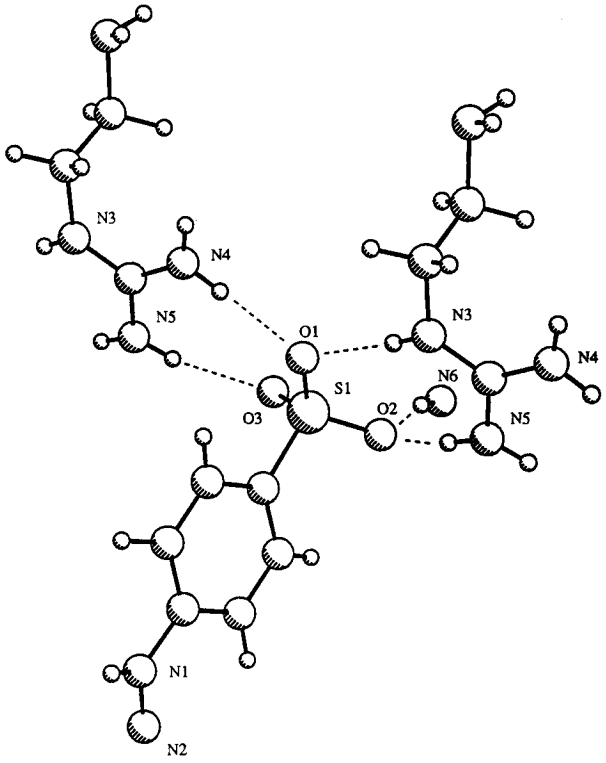


Figure 14. PLUTO view of the crystalline environment of the sulfonate group in tosylarginine methyl ester/Little Rock Orange (**6**), showing close intermolecular contacts. See Table 8 for distances. To improve clarity, only selected portions of the molecules are shown.

Table 8. Intermolecular Contacts (Å) in Tosylarginine Methyl Ester/Little Rock Orange (**6**)[a]

Contacts to the Sulfonate Group			
O(1)···N(3)[b]	2.898(4)	O(2)···N(5)[b]	2.936(4)
O(1)···N(4)[a]	2.937(4)	O(3)···N(5)[a]	2.913(5)
O(2)···N(6)[ff]	2.915(4)		
Contacts to Other Atoms			
O(8)···N(4)[j]	2.934(4)		

[a] See Table 9 for the meaning of the superscript letters.

structures **1**−**5**). Two of the three sulfonate oxygens make two close (<3 Å) intermolecular contacts; the third makes one contact. Interactions between the sulfonate group and the guanidinium group in **6** are *bridging* interactions, in which two oxygen atoms from a single sulfonate group contact two nitrogen atoms of the guanidinium, forming a hydrogen-bonded ring. In structure **6** one of these interactions is "head-on" with respect to the guanidinium: O(1) and O(3) contact the terminal nitrogens N(4) and N(5) of the amino acid. The other bridging interaction is "sideways," in which O(1) and O(2) of the sulfonate group contact an internal and terminal guanidine nitrogen [N(3) and N(5)]. Bridging interactions are in fact a favored mode of contact between sulfonates/sulfates and guanidinium groups, being found in guanidinium alkanesulfonates and arenesulfonates[31] and in other structures as well. Sulfonate-bridging interactions occur between the sulfonate groups of

Table 9. Symmetry Code for Intermolecular Contacts

a	x, y, z	p	$x, y, 1+z$
b	$-1+x, y, z$	q	$1-x, -1/2+y, 1-z$
c	$-x, 1-y, 1-z$	r	$-x, 1/2+y, 2-z$
d	$2-x, 2-y, -z$	s	$-x, 1/2+y, 1-z$
e	$2-x, 1-y, -z$	t	$1-x, -1/2+y, 2-z$
f	$1-x, 1-y, -z$	u	$1-x, 1-y, 1-z$
g	$-x, -y, 1-z$	v	$-1+x, 1+y, z$
h	$-2+x, -1+y, 1+z$	w	$1-x, -y, -z$
i	$-x, 2-y, 1-z$	aa	$x, 1+y, z$
j	$1+x, y, z$	bb	$-1+x, -1+y, z$
k	$1-x, 2-y, 1-z$	cc	$1+x, 1+y, z$
l	$-1-x, 2-y, -z$	dd	$-1+x, -1+y, -1+z$
m	$1+x, y, -1+z$	ee	$1+x, 1+y, 1+z$
n	$-1+x, y, 1+z$	ff	$-x, -1/2+y, 1/2-z$
o	$x, y, -1+z$		

Orange G and the protonated forms of the nucleic acid bases adenine and cytosine,[19a] and between the sulfonate groups and water molecules held in suitable orientations by metal ions in the lithium, magnesium, and calcium salts of Orange G.[19b] An analogous sulfate-bridging interaction is found in the crystal structure of glucosamine 2*N*-sulfate, in which the sulfate group is bridged by the sulfonamide NH and adjacent OH of a neighboring sugar molecule.[20] Their frequent occurrence indicates that bridging interactions are common and favorable for both sulfate and sulfonate and that they are likely to occur not only among small molecules but also where large sulfated biomolecules interact with proteins.

Analysis of Intermolecular Interactions Using the Graph-Set Formalism. General Principles. Qualitative discussions of hydrogen-bonding patterns, such as the discussion above, have proven very useful in the past but often leave unanswered the question of whether *all* the patterns, particular the most general ones, have in fact been identified. An attractive method for investigating this topological question is by the use of graph-set analysis. Following earlier work along these lines by Wells,[32] Hamilton and Ibers,[33] and Kuleshova and Zorky,[34] Etter and co-workers[22] developed a hydrogen-bond classification scheme based on graph sets. This method has been used by several groups in recent years to identify and classify hydrogen-bonding patterns in the solid state.[35] A recent, particularly relevant example is a study by Shimoni and Glusker[36] of

(31) (a) Russell, V. A.; Etter, M. C.; Ward, M. D. *J. Am. Chem. Soc*. **1994**, *116*, 1941−1952. (b) Russell, V. A.; Etter, M. C.; Ward, M. D. *Chem. Mater*. **1994**, *6*, 1206−1217.

(32) Wells, A. F. *Structural Inorganic Chemistry*; Clarendon Press: Oxford, 1962; pp 294−315.

(33) Hamilton, W. C.; Ibers, J. A. *Hydrogen Bonding in Solids*; W. A. Benjamin: New York, 1968; pp 19−21.

(34) Kuleshova, L. N.; Zorky, P. M. *Acta Crystallogr*. **1980**, *B36*, 2113−2115.

(35) (a) Bernstein, J. *Acta Crystallogr*. **1991**, *B47*, 1004−1010. (b) Shimoni, L.; Carrell, H. L.; Glusker, J. P.; Coombs, M. M. *J. Am. Chem. Soc*. **1994**, *116*, 8162−8168. (c) Aakeröy, C. B.; Nieuwenhuyzen, M. *J. Am. Chem. Soc*. **1994**, *116*, 10983−10991. (d) Several other examples can be found in *Chem. Mater*. **1994**, *6*.

(36) Shimoni, L.; Glusker, J. P. *Protein Sci*. **1995**, *4*, 65−74.

(37) Our criterion for an "eclipsed" sulfonate conformation is that the S−O bond most nearly coplanar with the naphthalene ring be located within a torsional angle of 15° with respect to the ring.

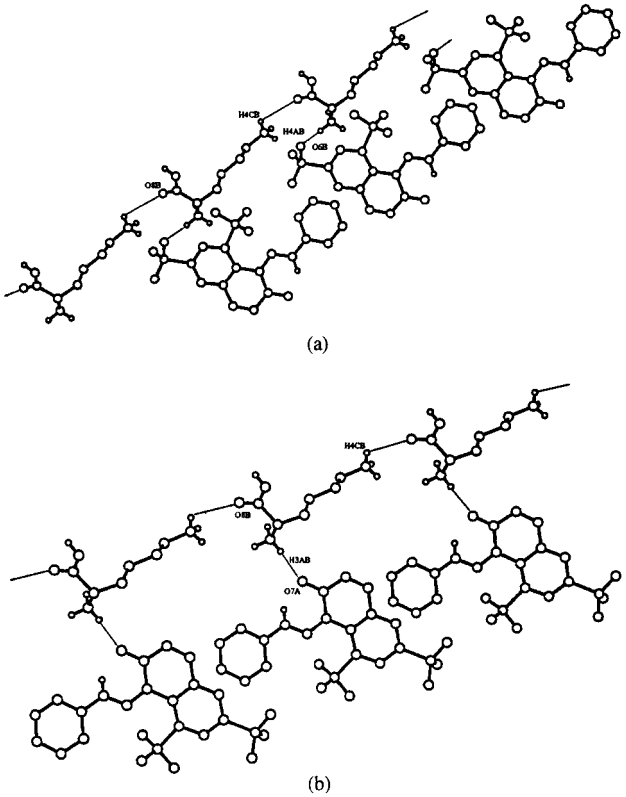


Figure 15. Branched-chain motifs illustrating some second-level hydrogen-bond interactions in **3**. (a) $D_2^2(7)C_1^1(9)$ branched-chain motif composed of two interactions: N4B−H4AB···O6B and N4B−H4CB···O8B. (b) $D_2^2(7)C_1^1(9)$ branched chain from N3B−H3AB···O7A and N4B−H4CB···O8B interactions. Hydrogen bonds are shown as narrow lines.

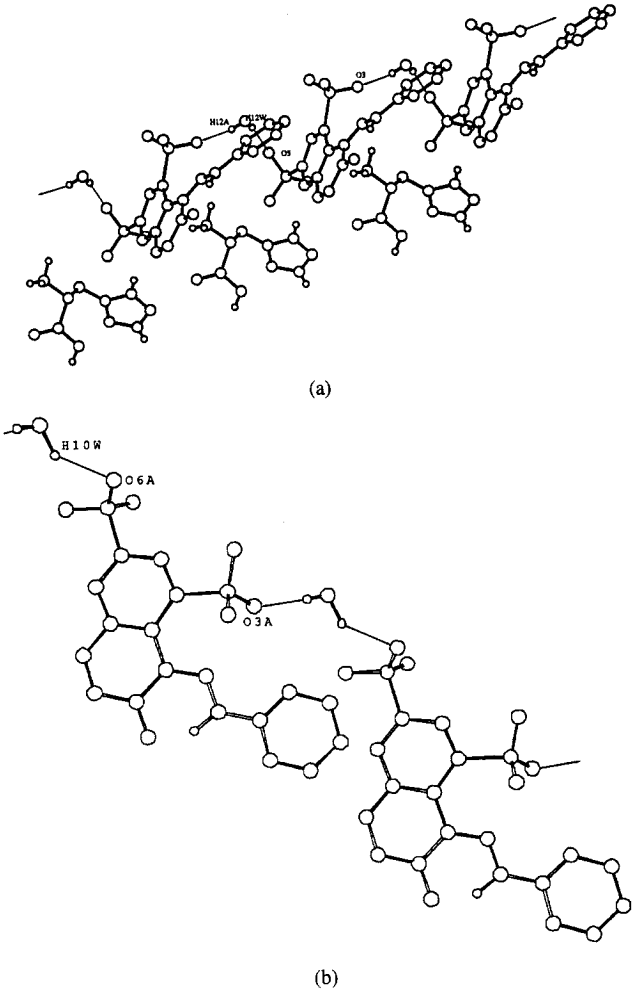


Figure 16. Second-level graph set motifs $C_2^2(10)$ for **4** (a) and **5** (b). The $C_2^2(10)$ motif is characteristic of a combination of two hydrogen bonds in four amino acid/Orange G hydrate salts (**2**−**5**): two different water protons hydrogen-bonded to oxygen atoms of two symmetry-unrelated sulfonate groups. Hydrogen bonds are shown as narrow lines.

hydrogen-bonding patterns involving arginine, asparagine, and glutamine aimed at elucidating specific binding motifs defined by these side chains in protein−nucleic acid interactions. As an aid in examining protein−glycosaminoglycan interactions, we present here selected results of the graph-set analysis of hydrogen-bonding patterns in our dye−amino acid complexes. Extensive tables and diagrams giving more complete details of the analysis are included as supporting information.

The method utilizes the following descriptors: the pattern designator *G*, the degree *r*, the number of hydrogen bond donors *d*, and the number of hydrogen bond acceptors *a*. These descriptors are used to form the general graph-set expression $G_d^a(r)$. The pattern designator *G* indicates the type of pattern made by a specific hydrogen bond or combination of hydrogen bonds throughout a crystal. G takes on the symbolic values S (designating an intramolecular hydrogen bond, a bond to the molecule's *self*), D (designating a finite group of molecules, such as a *dimeric* pair), C (designating an infinite *chain* of molecules), and R (designating a *ring* formed by hydrogen-bonded molecules). The superscript *a* and subscript *d* terms represent the number of hydrogen bond acceptors and donors, respectively, involved in the repeat unit of the specified pattern. The degree *r* is determined by counting the number of atoms comprising the repeat unit of the hydrogen bond motif. The number of different types of hydrogen bonds present in the pattern is called the *level* of the hydrogen bond graph set; thus, "first level" describes a pattern formed by only one type of hydrogen bond, and "second level" describes a pattern formed by two types. Bernstein, Davis, and co-workers[22e] provide a detailed discussion of hydrogen-bond graph sets and numerous examples of applications.

Graph-Set Comparison of Related Structures. DL-Lysine Salts: DL-Lysine/HABS (1) and DL-Lysine/Orange G (2). The first-level graph sets for both **1** and **2** consist mainly of finite complex interactions. The exceptions are a ring motif in **1**, an $R_2^2(28)$ motif defined by phenolic OH···^-O_3S interactions, and a ring motif in **2**, an $R_2^2(10)$ motif defined by α-NH_3^+···acid carbonyl interactions. Each of these motifs when combined with other hydrogen-bonding interactions in the respective structures generates a finite centrosymmetric spiral ring pattern as a second-level motif.

Lysine/Orange G Salts: DL-Lysine/Orange G (2) and L-Lysine/Orange G (3). First-level graph sets for both **2** and **3** consist of finite complex motifs except for one infinite motif in each structure. In **2** this motif is an $R_2^2(10)$ ring formed by α-NH_3^+···acid carbonyl interactions, and in **3** (Figure 15) this motif is a $C_1^1(9)$ chain formed by side-chain NH_3^+···acid carbonyl interactions. In centrosymmetric **2**, all ring motifs in the second-level graph set are located about crystallographic inversion centers. In noncentrosymmetric **3**, *no* ring motifs occur in either the first- or second-level graph sets. The chain motif found in **3** generates eight branched-chain patterns in the second-level graph set, two of which are shown in Figure 15. In the second-level graph sets of both **2** and **3**, $D_2^2(5)$ and $D_2^2(9)$ motifs defined by interactions between a single sulfonate and two water molecules are found.

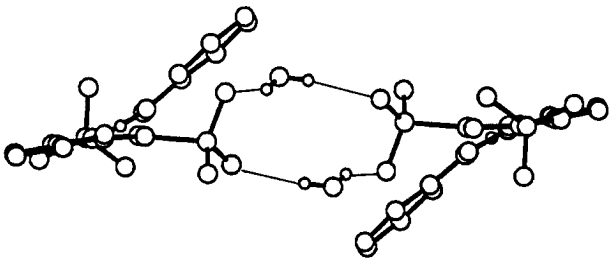


Figure 17. $R_4^4(12)$ motif found in the second-level hydrogen bond graph set of **4**. The motif is composed of two pairs of *symmetry-related* hydrogen bonds. Hydrogen bonds are shown as narrow lines.

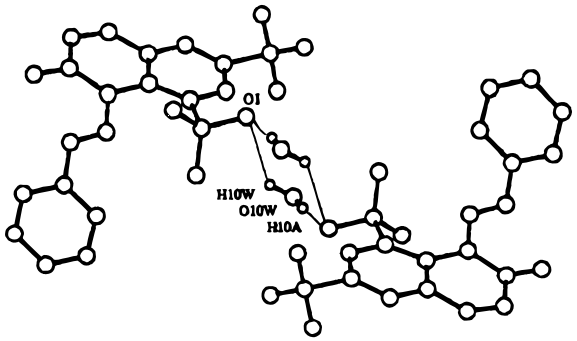


Figure 18. $R_4^2(8)$ motif found in the second-level graph set of **4**. The pattern is composed of two different water protons hydrogen-bonded to symmetry-related sulfonate oxygen atoms. Hydrogen bonds are shown as narrow lines.

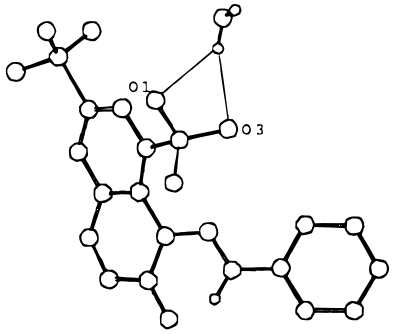


Figure 19. $R_1^2(4)$ motif found in the second-level hydrogen bond graph sets of **4**. A bifurcated water proton hydrogen-bonded to two different oxygen atoms of the same sulfonate group defines the motif. Hydrogen bonds are represented by narrow lines. A similar ring motif is present in **5**.

Histidine/Orange G Salts: DL-Histidine/Orange G (4) and L-Histidine/Orange G (5). The first-level graph sets of both **4** and **5** are composed entirely of finite complex motifs (no chains or rings). The second-level graph sets include infinite chains and rings as well as finite complex interactions, although the finite motifs predominate in both salts. Examination of the hydrogen-bonding interactions in **4** and **5** reveals that the number of water···sulfonate interactions in these structures is sufficiently large to allow trends in the graph sets to be identified. These trends may provide a basis for predicting the occurrence of similar hydrogen-bonding motifs in compounds related to **4** and **5,** compounds bearing similar hydrogen-bonding functional groups. These water···sulfonate interactions can be classified into four main types.

Type I. *The two hydrogen atoms of a given water molecule are hydrogen-bonded to oxygen atoms of symmetry-unrelated sulfonate groups.* In spite of stereochemical differences between the amino acids in **4** and **5**, this interaction specifically forms a $C_2^2(10)$ chain motif in *both* salts (Figure 16). Structure **4** incorporates a single chain of this type while **5,** due to the larger

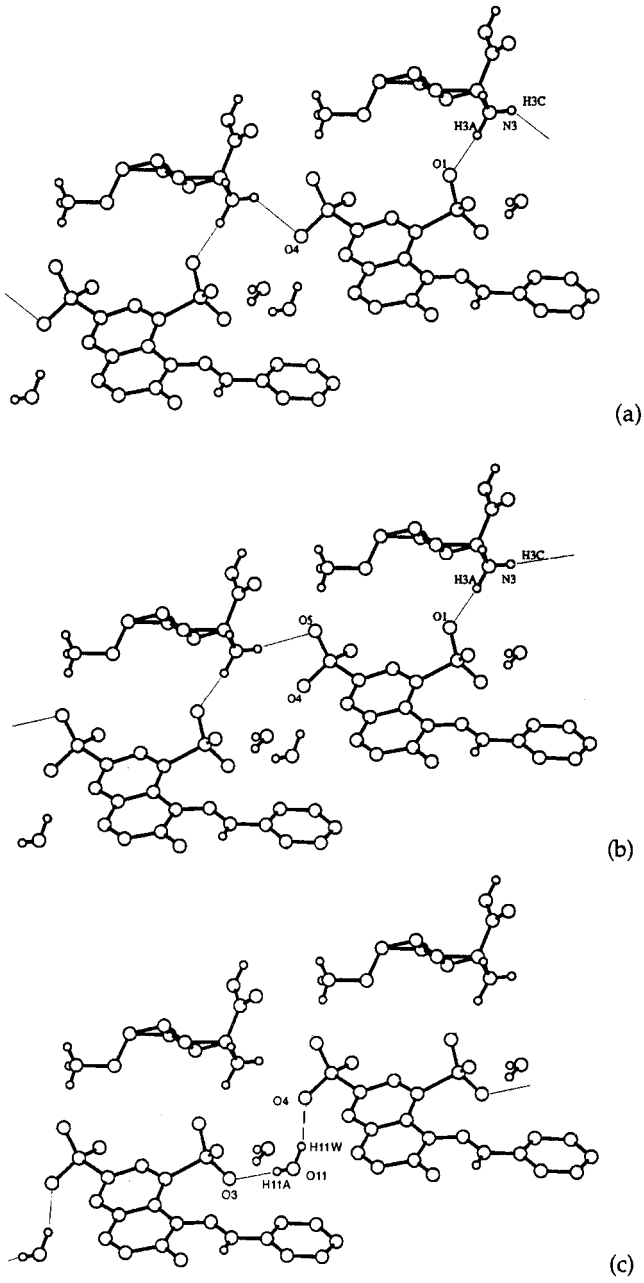


Figure 20. Three different $C_2^2(10)$ graph-set motifs (a−c) can be defined in the molecular packing arrangement of **2**. Narrow lines represent hydrogen bonds. (a, b) $C_2^2(10)$ chains formed by different amino−sulfonate hydrogen bonds. (c) $C_2^2(10)$ chain formed by two different water−sulfonate hydrogen bonds. Water molecules and amino groups in this structure link dye molecules together in a similar manner.

number of water molecules present, incorporates six of these chains. This chain motif is the *only* type of motif found among the second-level interactions in **4** and **5** that involves two water···sulfonate hydrogen bonds. The same interaction is found in all four Orange G salts presented here (structures **2**−**5**) and is also found in the crystal structure of the diammonium salt of Orange G.[19b]

Type II. *The two hydrogen atoms of a given water molecule are hydrogen-bonded to two different oxygen atoms of symmetry-related sulfonate groups.* One interaction of this type is found in **4** (racemic histidine/Orange G) but not in **5** (enantiomerically pure L-histidine/Orange G). This centrosymmetric $R_4^4(12)$ motif is shown in Figure 17.

Type III. *Two hydrogen atoms from a given water molecule are hydrogen-bonded to symmetry-related sulfonate oxygen*

atoms. This type of interaction is present in **4** and is described by the centrosymmetric $R_4^2(8)$ motif shown in Figure 18.

Type IV. *One water proton is hydrogen-bonded to two different oxygen atoms of the same sulfonate group (a bifurcated hydrogen bond).* This interaction is found in both **4** and **5** and forms a characteristic $R_1^2(4)$ motif. An example is shown in Figure 19.

Orange G Salts: DL-Lysine/Orange G (2), L-Lysine/Orange G (3), DL-Histidine/Orange G (4), and L-Histidine Orange G (5). As noted earlier, similarities among the second-level graph sets of the Orange G salts **2**−**5** can be found in spite of these salts' varied compositions and stereochemistries. The most noteworthy similarity involves the formation of two hydrogen bonds between a single water molecule and one or more sulfonate groups. In every instance (a total of eleven from **2**−**5**) in which a single water molecule is hydrogen-bonded to two different sulfonate groups of symmetry-related dye molecules, the result is the same motif, $C_2^2(10)$. Although the number of crystal structures presented here is too low to allow definitive predictions, in these and other salts of Orange G both the $C_2^2(10)$ and $R_1^2(4)$ motifs are unquestionably common. The $C_2^2(10)$ motif in **2** is of particular interest because it occurs not only between water and sulfonate but also between the protonated *amino* group and sulfonate in this same structure. Here the same hydrogen-bonding chain pattern is formed by two very different functional groups (Figure 20). This fact suggests that under certain circumstances water molecules can mimic amino groups with respect to hydrogen-bonding behavior toward sulfonate groups. Although the numbers of hydrogen-bond donors and acceptors from the dye and amino acid molecules in **2** are unequal, the pattern is maintained through the incorporation of water molecules that effectively substitute for amino groups in these interactions.

Arginine Structure (6). The "head-on" interaction between the guanidinium and sulfonate, which involves both primary amino groups, and the "side-on" interaction, which involves one primary and the secondary amino group, are both examples of the $R_2^2(8)$ motif. This same motif was observed in other arginine-containing small-molecule structures by Shimoni and Glusker,[36] who also found structures in which carboxylate or sulfite serves as the bidentate H-bond acceptor. It seems likely that sulfate could serve as well.

Conclusion. These crystal structures provide a detailed view of interactions between sulfonated molecules and basic amino acid functional groups important in binding sulfated biological macromolecules to proteins. They reveal numerous examples of water-mediated contacts, particularly between sulfonate oxygens and amino acid nitrogens, strongly suggesting that water molecules play a significant role in mediating the interactions between sulfate/sulfonate groups of biologically-important macromolecules and their target proteins. From both graph-set analysis and examination of close intermolecular contacts, there is evidence of specific patterns of interaction between sulfonate groups and acceptor sites from structure to structure, facilitated by mediating water molecules. These results demonstrate that X-ray analysis of suitable model compounds can yield information which is useful in understanding the biomolecular recognition of sulfonated and sulfated compounds by proteins.

Acknowledgment. The support of the Minnesota Medical Foundation, the American Cancer Society, and the Graduate School and Biomedical Engineering Center of the University of Minnesota are gratefully acknowledged. Support for L.K.L. was provided by the Summer Undergraduate Fellowship Program of the Biomedical Engineering Center.

Supporting Information Available: Graph-set analysis of the individual structures, including tables and figures, details of the data collection and structure refinement, and tables giving the fractional atomic coordinates, bond lengths and angles, torsional angles, intermolecular distances, and least-squares planes for **1**−**6** (224 pages); and structure factor lists for **1**−**6** (227 pages). This material is contained in many libraries on microfiche, immediately follows this article in the microfilm version of the journal, can be ordered from the ACS, and can be downloaded from the Internet; see any current masthead page for ordering information and Internet access instructions.

JA951121F