# Solution of the Conformation and Alignment Tensors for the Binding of Trimethoprim and Its Analogs to Dihydrofolate Reductase: 3D-Quantitative Structure−Activity Relationship Study Using Molecular Shape Analysis, 3-Way Partial Least-Squares Regression, and 3-Way Factor Analysis

William J. Dunn III,*[,†,§] Anton J. Hopfinger,[†,‡,§] Cornel Catana,[†] and Chaya Duraiswami[†,‡]

*Department of Medicinal Chemistry and Pharmacognosy and Laboratory of Molecular Modeling and Design, College of Pharmacy, University of Illinois at Chicago, 833 South Wood, m/c 781, Chicago, Illinois 60612, and The Chem21 Group Inc., 1780 Wilson Drive, Lake Forest, Illinois 60045*

*Received July 1, 1996*[⊗]

Molecular recognition is the basis of rational drug design, and for this reason it has been extensively studied. However, the process by which a ligand recognizes and binds to its receptor is complex and not well understood. For the case in which the geometries (conformation and alignment) of the ligand and receptor are known from X-ray crystal structure data, the problem is simplified. The receptor-bound conformation and alignment of the ligand is assumed, and those of additional ligands are inferred. For the general case in which the geometries of the ligand(s) and receptor are unknown, no general treatment or solution is available and receptor−ligand geometries must be obtained indirectly from structure−activity studies or synthesis and evaluation of rigid analogs. A general treatment for solving for the receptor-bound geometry of a series of ligands is presented here. Using molecular shape analysis, for ligand description, tensor analysis of *N*-way arrays by partial least-squares (PLS) regression, and 3-way factor analysis, the receptor-bound geometries of trimethoprim and a series of trimethoprim-like dihydrofolate reductase inhibitors are correctly predicted.

## Introduction

A goal of rational drug design using 3D-quantitative structure−activity relationship (QSAR) is to derive a relationship between some numerical expression of the structures of compounds and their activity. The relationship may be qualitative or quantitative, but no matter which is the case, the objective is to predict the structure(s) of better compounds. Most compounds' effects are receptor mediated. In some cases the structures of the receptor and the drug−receptor complex are known. However, in most cases the structure of the receptor is not known. Thus, a formidable part of rational drug design is to indirectly obtain information about the structure of the drug−receptor complex. This is a daunting and unsolved problem when the compounds are flexible and can assume a number of possible receptor alignments.

The pioneering work of Hansch et al.[1] provided a formalism for deriving QSARs which could be used to predict activity from structure. A major limitation of this approach is that its descriptors are derived from 2-dimensional structures. Comparative molecular field analysis, CoMFA,[2] molecular shape analysis, MSA,[3] distance geometry,[4] and molecular similarity matrices[5] are four methods which explicitly treat the 3-dimensional structure of the ligand and thus yield 3-dimensional quantitative structure−activity relationships, 3D-QSARs.

With the CoMFA method, a receptor-bound conformation and alignment for a reference compound in a series is assumed, and field-related parameters are computed at defined points on a grid about the bioactive compounds. The field probe energies which result are then used as descriptors with PLS regression to derive a 3D-QSAR. This method has been used extensively but is limited in that a receptor-bound conformation and alignment must be initially assumed. It has also been shown that CoMFA models can be very sensitive to the alignment selected[6] as well as conformation, spatial orientation, and grid size.[7]

MSA seeks the active conformation of a molecule, expresses molecular shape similarity in terms of a variety of scalar descriptors, such as common overlap steric volume, COSV, measured relative to some reference compound and conformation, and permits the use of other common 2- and 3-dimensional descriptors in the development of structure−activity models. MSA, like CoMFA, assumes a common receptor-bound conformation and alignment. The active analog approach[8] can generate a hypothetical active site which can geometrically accommodate members of a training set and, thus, provide a graphic model of the receptor. No predictive model results from this method, however. These and other 3D-QSAR methods have been reviewed recently.[9]

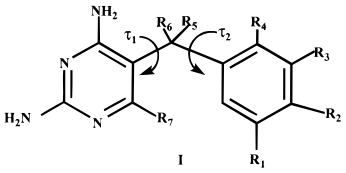
In a recent report, a general solution to the 3D-QSAR problem was proposed[10] and applied to a set of flexible muscarinic receptor ligands in which the receptor alignment was assumed.[11] The method predicted a common receptor-bound conformation for the series which was consistent with other structure−activity data for the series.

Here both the fixed conformation and fixed alignment constraints are relaxed, and the methodology is applied to structure−activity data for trimethoprim (**I**, $R_1 = R_2 = R_3 = OCH_3$, $R_4 = R_5 = R_6 = H$) and a series of trimethoprim-like derivatives which are inhibitors of

† Department of Medicinal Chemistry and Pharmacognosy.
‡ Laboratory of Molecular Modeling and Design.
§ The Chem21 Group Inc.
⊗ Abstract published in *Advance ACS Abstracts,* November 1, 1996.

S0022-2623(96)00491-8 CCC: $12.00 

*Escherichia coli* dihydrofolate reductase, DHFR. The receptor-bound conformation and alignment for the binary enzyme−inhibitor complex are known which makes this an ideal test system on which to test the methodology.




I

## General 3D-QSAR Model

In the development of a general 3D-QSAR model, several assumptions are made. It is assumed that the ligands bind to the receptor in a common conformation and alignment. The common conformation and alignment are necessary in order that functional groups of the ligand, i.e., the pharmacophoric groups, can interact with their complementary groups in the active site of the enzyme. There is considerable, direct evidence supporting this assumption,[12] although some exceptions have been observed.[13] Most notable is the "reverse binding" of methotrexate to DHFR when compared to folic acid. It is further assumed that MSA and physicochemical descriptors can be used to express structural variation within the series of active compounds. There are numerous published accounts supporting this assumption.

The general MSA, 3D-QSAR model[10] is

$$\boldsymbol{Y} = \boldsymbol{T}^*[\boldsymbol{V}(s, n, m), \mathbf{F}(p, r_{i,j,k}, f, n, m), \boldsymbol{H}(h, n, m), \boldsymbol{E}(e, n, m)] \quad (1)$$

Here, **Y** is the biological activity vector or matrix for a series of $N$ compounds. The term in brackets is the composite **VFHE** tensor. **V** is the intrinsic molecular shape tensor composed of shape, $s$, descriptors. These are highly dependent on conformation, $m$, and alignment, $n$, and provide information on molecular shape within the steric contact surface of the molecule. ***F*** is the molecular field tensor where $(p, r_{ijk})$ define the set of field probes, $p$, used to sample space at positions $r_{ijk}$. The $f$ is the field-related molecular features such as dipole moment which are not derived from $(p, r_{ijk})$. The non-**V** and -**F** molecular features are included in **H**. They may or may not be dependent on $m$ and/or $n$ and may be based on the whole molecule or be substituent- or fragment-based. The measured physicochemical features are included in the tensor **E**. These features could be log $P$ and p$K_a$, for example, which form the basis of classical QSAR studies. Generally the set of coefficients, or the transformation tensors, **T**, optimally map the **VFHE** tensor onto **Y**, but in the approach presented here, the **V** and **H** descriptors are applied separately. The data structure for the problem in which a single $\mathbf{V} = \mathbf{X}$ descriptor is used is shown in Figure 1 where **X** is a 3-way array.

One way in which **T** can be found is by decomposition of the 3-way array by a variation of PLS regression. This decomposition is shown in Figure 2. The 3-way array is decomposed, shown here for one PLS component, into the latent variables $t$ and $u$ which are related through the inner relation $\hat{u} = bt$. **Y** is decomposed into the mean, $y$, and the PLS loadings, $q$, while the 3-way array

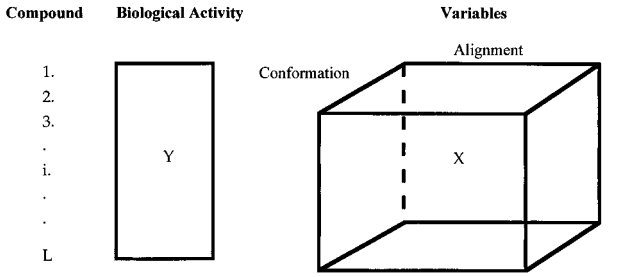



**Figure 1.** Data structure for the MSA conformation/alignment problem with one MSA variable.

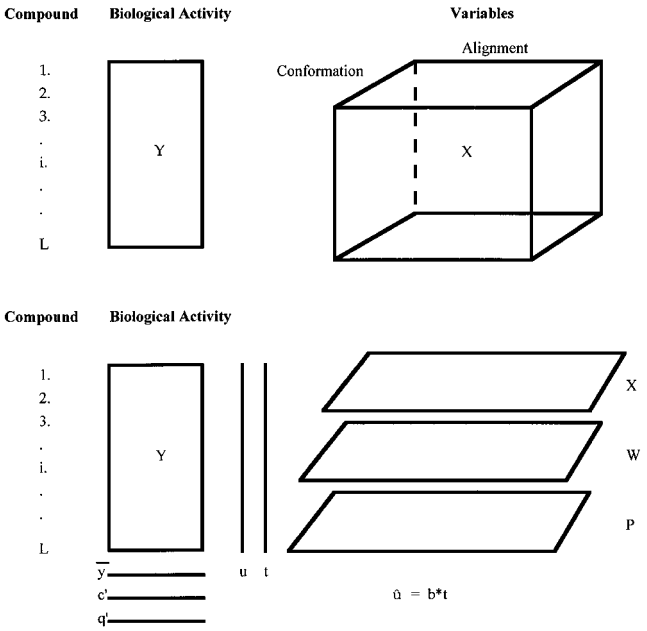



**Figure 2.** PLS decomposition of a 3-way array.

is decomposed into the array of means, $x$, and the array of PLS loadings, $P$.

The PLS models are derived with the usual constraints that the latent variables are those from approximately along the axes of greatest variation in the **X** and **Y** data that are optimally correlated. The algebraic model is shown below.

$$\mathbf{X}_{l,n,m} = \sum_{j=1}^{J}\left[x_{l,n,m} + \sum_{a=1}^{A} t_{l,a} \otimes \mathbf{P}_{a,n,m} + e_{l,n,m}\right] \quad (2)$$

$$Y_{l,i} = y_i + \sum_{a=1}^{A} u_{l,a} q_{a,i} + e_{l,i} \quad (3)$$

$$\hat{u} = bt \quad (4)$$

Equations 2 and 3 represent the general case in which the decomposition is made over $J$ conformation/alignment-dependent variables and $i$ biological activities; $n$ is the alignment index, and $m$ is the conformation index. $A$ is the number of latent variables extracted from the **X** and **Y** data, and $l$ is the compound index. The term $t_{l,a} \otimes P_{a,n,m}$ is the Kronecker product of the vector **t** and the matrix **P**. The elements of **P** are the conformation/alignment weights for the variables, **X**, and are measures of the significance of each conformation/alignment set in explaining the variation in the **Y** data through the inner relation (eq 4).

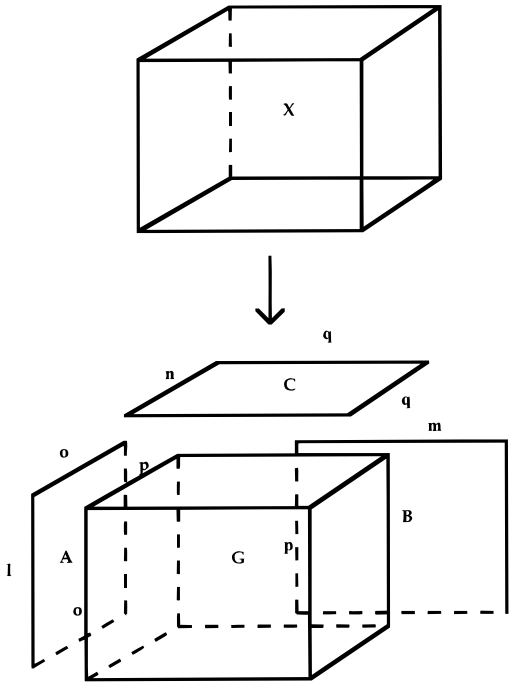



**Figure 3.** Decomposition of a 3-way array with factor analysis.

A complementary method for solution of the desired tensors is 3-way factor analysis. This method has been used mainly in the social sciences[14] but has recently been applied to problems in analytical[15] and environmental chemistry.[16] The decomposition is shown in Figure 3.

The array is decomposed into the three factor-loading matrices, **A**−**C**, and the core 3-way array, **G**. The loading matrices are the compound, alignment, and conformation tensors, respectively. **G** contains the correlation structure for the 3-way **X** array. These are constructed from the eigenvectors of the covariance matrices obtained from unfolding **X** its three ways. The number of components extracted is *o*, *p*, and *q*, respectively. These eigenvectors are used to construct matrices **A**−**C**. Of interest for this application are the elements of **B**, the alignment factor loadings, and **C**, the conformation factor loadings.

The model for decomposition of **X**, composed of a single variable, is shown in eq 5. In the matrix form, the model is shown in eq 6. Here, $\mathbf{X}_{l,m,n}$ is the unfolded 3-way array, and $\mathbf{G}_{o,p,q}$ is the unfolded core matrix. The product $\mathbf{B}_{p,n}$ and $\mathbf{C}_{q,m}$ is the Kronecker product.

$$\mathbf{X}_{l,m,n} = \sum_{o}\sum_{p}\sum_{q} a_{l,o} b_{n,p} c_{m,q} g_{o,p,q} \tag{5}$$

$$\mathbf{X}_{l,m,n} = \mathbf{A}_{l,o}\mathbf{G}_{o,p,q}(\mathbf{B}_{p,n} \otimes \mathbf{C}_{q,m}) \tag{6}$$

An algorithm for the PLS decomposition of *N*-way arrays, where *N* is 3 or greater, has been published.[17] A variation of this algorithm, based on the UNIPALS algorithm,[18] has been developed in our laboratory. The algorithm for factor decomposition of 3-way arrays has also been proposed.[14] Both have been programmed and applied to structure−activity data for trimethoprim and trimethoprim-like DHFR inhibitors for which the crystal structures of *E. coli* DHFR and the binary trimethoprim−enzyme complex are known.[19]

**Table 1.** Structure of DHFR Inhibitors

| no. | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ | $R_6$ | $R_7$ |
|---|---|---|---|---|---|---|---|
| **1**[a] | $-OCH_3$ | $-OCH_3$ | $-OCH_3$ | -H | -H | -H | -H |
| **2**[d] | $-OCH_3$ | -OH | $-OCH_3$ | -H | -H | -H | -H |
| **3**[d] | $-OCH_3$ | $-OCH_3$ | $-OCH_3$ | -H | -H | -H | $-CH_3$ |
| **4**[b] | $-OCH_3$ | -OH | $-OCH_3$ | -H | -H | -H | $-CH_3$ |
| **5**[e] | -H | $-C_6H_5$ | H | -H | -H | -H | -H |
| **6**[e] | $-OCH_3$ | $-OCH_3$ | $-OCH_3$ | -H | $=CH-CH_3$ | | -H |
| **7**[f] | $-OCH_3$ | $-OCH_3$ | $-OCH_3$ | -H | $=CH_2$ | | -H |
| **8**[f] | $-OCH_3$ | -Br | $-OCH_3$ | -H | -H | -H | -H |
| **9**[a] | $-OCH_3$ | -H | $-OCH_3$ | -H | -H | -H | -H |
| **10**[b,c] | $-CH_3$ | -H | $-CH_3$ | -H | -H | -H | -H |
| **11**[a] | -H | -H | $-OCH_3$ | -H | -H | -H | -H |
| **12**[a] | -H | -Br | -H | -H | -H | -H | -H |
| **13**[a] | -H | -H | -Cl | -H | -H | -H | -H |
| **14**[a] | $-CH_2OH$ | -H | $-CH_2OH$ | -H | -H | -H | -H |
| **15**[a] | -H | -H | -H | -H | -H | -H | -H |
| **16**[e] | $-OCH_3$ | $-OCH_3$ | $-OCH_3$ | -H | -H | -OH | -H |
| **17**[e] | $-OCH_3$ | $-OCH_3$ | $-OCH_3$ | -H | -OH | -H | -H |
| **18**[e] | $-OCH_3$ | $-OCH_3$ | $-OCH_3$ | -H | -H | $-CH_3$ | -H |
| **19**[e] | $-OCH_3$ | $-OCH_3$ | $-OCH_3$ | -H | $-CH_3$ | -H | -H |
| **20**[a] | -OH | -H | -OH | -H | -H | -H | -H |

[a] Hansch, C.; Li, R.; Blaney, M. J.; Langridge, R. *J. Med. Chem.* **1982**, *25*, 777. [b] Selassie, D. C; Li, R.; Poe, M.; Hansch, C. *J. Med. Chem.* **1991**, *34*, 46. [c] Li, R.; Poe, M. *J. Med. Chem.* **1988**, *31*, 366. [d] Roth, B.; Sterlitz, Z. J.; Rauckman, S. B. *J. Med. Chem.* **1980**, *23*, 379. [e] Rey-Bellet, G.; et al. *Eur. J. Med. Chem. Chim.-Ther.* **1975**, *10*, 7. [f] Kompis, I.; Then, L. R. *Eur. J. Med. Chem.-Chim. Ther.* **1984**, *19*, 529.

## Molecular Modeling and 3D-QSAR Analysis

The structures of the trimethoprim analogs were taken from the literature and are given in Table 1. The activities are given in Table 2. For some inhibitors, the concentration of the inhibitor that produces 50% of the enzyme inhibition, $IC_{50}$ values, was given, while for other analogs $K_i$ values, the equilibrium constant for the inhibitor binding to the enzyme, were reported. In order to be consistent, since these end points are related by the Cheng−Prusoff equation,[20] the $K_i$ values were converted to $IC_{50}$ values using the procedure previously reported.[21] The composite structure−activity training set has a greater than $10^5$-fold range in *E. coli* DHFR inhibition activity.

In previous MSA QSAR studies, in this laboratory, of DHFR inhibition, the MSA molecular shape variable, COSV, was shown to be a significant descriptor for DHFR inhibition.[22] Therefore, this variable was selected as a QSAR descriptor for this study. The COSVs were computed from structures constructed in the following manner. The ligand analogs were built using the bond lengths and bond angles reported for the crystal structure of trimethoprim as determined by neutron diffraction.[19] The crystal valence geometry was used to compute partial charges using the MNDO semiemperical molecular orbital method.[23] Fixed valence conformational analysis was performed for each of the analogs at 10° resolution for the torsion angles $\tau_1$ and $\tau_2$ defined in **I** using the MMII nonbonded potential, a Coulomb potential with a molecular dielectric of 3.5, and a MMII-scaled hydrogen-bonding potential.[24] This is the same force field used in previous MSA DHFR 3D-QSAR studies.[22]

The availability of the enzyme-bound geometry of trimethoprim permits the evaluation, exploration, and validation of the solutions of the conformation and alignment problem with respect to both ligand energetics and geometry. Ligand conformational analysis permits the construction of molecular shape descriptors

**Table 2.** MSA Data and Activities of DHFR Inhibitors

| no. | log (1/$IC_{50}$) | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Alignment 1 | | | | | |
| **1** | 8.23 | 0.4324 | 0.4563 | 0.4351 | 0.4387 | 0.7379 | 0.4408 | 0.4533 | 0.7752 | 0.7931 | 1.0000 |
| **2** | 7.96 | 0.4344 | 0.4512 | 0.4273 | 0.4373 | 0.6708 | 0.4379 | 0.4511 | 0.6212 | 0.7201 | 0.8446 |
| **3** | 7.00 | 0.4369 | 0.4443 | 0.4575 | 0.4529 | 0.4494 | 0.4604 | 0.4788 | 0.5632 | 0.7050 | 0.9055 |
| **4** | 6.52 | 0.4432 | 0.4543 | 0.4396 | 0.4439 | 0.6759 | 0.4565 | 0.4605 | 0.5632 | 0.6943 | 0.8274 |
| **5** | 6.40 | 0.4337 | 0.4533 | 0.4323 | 0.4383 | 0.6668 | 0.4375 | 0.4595 | 0.6918 | 0.7322 | 0.7678 |
| **6** | 6.30 | 0.4393 | 0.5070 | 0.4321 | 0.4762 | 0.6761 | 0.4845 | 0.4996 | 0.7284 | 0.6708 | 0.8905 |
| **7** | 5.60 | 0.4213 | 0.4891 | 0.4170 | 0.4592 | 0.6815 | 0.4570 | 0.4810 | 0.7350 | 0.6396 | 0.8921 |
| **8** | 8.53 | 0.4335 | 0.4521 | 0.4354 | 0.4422 | 0.6997 | 0.4390 | 0.4575 | 0.7357 | 0.7667 | 0.8527 |
| **9** | 7.75 | 0.4332 | 0.4557 | 0.4331 | 0.4399 | 0.6958 | 0.4411 | 0.4530 | 0.7301 | 0.7561 | 0.8872 |
| **10** | 7.45 | 0.4385 | 0.4547 | 0.4317 | 0.4408 | 0.6423 | 0.4379 | 0.4489 | 0.7097 | 0.6902 | 0.7835 |
| **11** | 6.40 | 0.4296 | 0.4555 | 0.4346 | 0.4421 | 0.6395 | 0.4382 | 0.4504 | 0.6804 | 0.7105 | 0.7779 |
| **12** | 6.30 | 0.4319 | 0.4551 | 0.4334 | 0.4433 | 0.6564 | 0.4385 | 0.4517 | 0.6627 | 0.7035 | 0.7622 |
| **13** | 6.14 | 0.4341 | 0.4481 | 0.4337 | 0.4455 | 0.6495 | 0.4391 | 0.4492 | 0.6332 | 0.6591 | 0.7299 |
| **14** | 5.83 | 0.4315 | 0.4449 | 0.4369 | 0.4451 | 0.6574 | 0.4403 | 0.4508 | 0.6677 | 0.6618 | 0.7445 |
| **15** | 5.71 | 0.4312 | 0.4444 | 0.4295 | 0.4495 | 0.6273 | 0.4494 | 0.4598 | 0.6109 | 0.6174 | 0.6753 |
| **16** | 5.35 | 0.4614 | 0.3963 | 0.4570 | 0.3968 | 0.7924 | 0.3878 | 0.4085 | 0.7699 | 0.6596 | 0.6635 |
| **17** | 5.35 | 0.4415 | 0.4996 | 0.4371 | 0.4869 | 0.6835 | 0.4805 | 0.4880 | 0.6959 | 0.6660 | 0.8199 |
| **18** | 4.00 | 0.4295 | 0.4413 | 0.4062 | 0.4361 | 0.7387 | 0.4376 | 0.4523 | 0.7615 | 0.5679 | 0.6249 |
| **19** | 4.00 | 0.4827 | 0.4701 | 0.4899 | 0.4546 | 0.7248 | 0.4592 | 0.4847 | 0.7067 | 0.7230 | 0.7773 |
| **20** | 2.78 | 0.4320 | 0.4462 | 0.4341 | 0.4349 | 0.6414 | 0.4351 | 0.4587 | 0.6362 | 0.6379 | 0.7242 |
| | | | | | | Alignment 2 | | | | | |
| **1** | 8.23 | 0.5865 | 0.5243 | 0.5207 | 0.4831 | 0.7229 | 0.5060 | 0.5889 | 0.7543 | 0.7946 | 1.0000 |
| **2** | 7.96 | 0.6125 | 0.4893 | 0.5638 | 0.5005 | 0.5956 | 0.5482 | 0.5976 | 0.5488 | 0.7017 | 0.8372 |
| **3** | 7.00 | 0.7027 | 0.5011 | 0.5441 | 0.4624 | 0.4624 | 0.5788 | 0.6500 | 0.4587 | 0.6600 | 0.9068 |
| **4** | 6.52 | 0.5925 | 0.4427 | 0.5630 | 0.4416 | 0.6506 | 0.5605 | 0.5900 | 0.4642 | 0.6975 | 0.8256 |
| **5** | 6.40 | 0.5685 | 0.5048 | 0.5207 | 0.4894 | 0.6583 | 0.5125 | 0.5456 | 0.6626 | 0.6935 | 0.7757 |
| **6** | 6.30 | 0.6467 | 0.4833 | 0.5710 | 0.4693 | 0.6434 | 0.5426 | 0.6448 | 0.6908 | 0.7065 | 0.8899 |
| **7** | 5.60 | 0.6396 | 0.4793 | 0.5628 | 0.4630 | 0.6472 | 0.5364 | 0.6366 | 0.6902 | 0.6457 | 0.8913 |
| **8** | 8.53 | 0.6288 | 0.4956 | 0.6009 | 0.4768 | 0.6859 | 0.5721 | 0.6137 | 0.6841 | 0.7258 | 0.8493 |
| **9** | 7.75 | 0.5603 | 0.4786 | 0.4761 | 0.4200 | 0.6797 | 0.4607 | 0.5127 | 0.6725 | 0.7291 | 0.8842 |
| **10** | 7.45 | 0.5901 | 0.4522 | 0.5277 | 0.4266 | 0.5826 | 0.5018 | 0.5672 | 0.6883 | 0.6563 | 0.7785 |
| **11** | 6.40 | 0.4946 | 0.4273 | 0.4672 | 0.4142 | 0.6056 | 0.4519 | 0.4946 | 0.6335 | 0.6727 | 0.7844 |
| **12** | 6.30 | 0.5341 | 0.4663 | 0.4948 | 0.4743 | 0.6164 | 0.4919 | 0.5403 | 0.6496 | 0.6825 | 0.7745 |
| **13** | 6.14 | 0.5398 | 0.4233 | 0.4767 | 0.4156 | 0.5909 | 0.4640 | 0.5125 | 0.5909 | 0.6265 | 0.7340 |
| **14** | 5.83 | 0.5804 | 0.4463 | 0.6092 | 0.4397 | 0.5942 | 0.5695 | 0.4430 | 0.5880 | 0.5938 | 0.7464 |
| **15** | 5.71 | 0.4534 | 0.4067 | 0.4792 | 0.4195 | 0.5857 | 0.4590 | 0.4742 | 0.5577 | 0.5598 | 0.6764 |
| **16** | 5.35 | 0.6196 | 0.5395 | 0.5911 | 0.4720 | 0.6589 | 0.5460 | 0.5909 | 0.6244 | 0.7620 | 0.7686 |
| **17** | 5.35 | 0.7152 | 0.4887 | 0.6283 | 0.4860 | 0.6356 | 0.6160 | 0.7147 | 0.6620 | 0.6439 | 0.8171 |
| **18** | 4.00 | 0.5387 | 0.5000 | 0.5443 | 0.5539 | 0.4441 | 0.5892 | 0.5626 | 0.4573 | 0.4861 | 0.4679 |
| **19** | 4.00 | 0.5491 | 0.4892 | 0.5709 | 0.5235 | 0.4535 | 0.5549 | 0.5387 | 0.4513 | 0.4503 | 0.4593 |
| **20** | 2.78 | 0.5257 | 0.4258 | 0.4673 | 0.4108 | 0.6019 | 0.4621 | 0.5072 | 0.5843 | 0.5903 | 0.7181 |
| | | | | | | Alignment 3 | | | | | |
| **1** | 8.23 | 0.7654 | 0.7051 | 0.7034 | 0.7494 | 0.7401 | 0.7124 | 0.7494 | 0.8226 | 0.8448 | 1.0000 |
| **2** | 7.96 | 0.6091 | 0.5311 | 0.5307 | 0.5554 | 0.5796 | 0.5507 | 0.5876 | 0.5894 | 0.6812 | 0.8434 |
| **3** | 7.00 | 0.6935 | 0.6484 | 0.6257 | 0.6548 | 0.6365 | 0.6045 | 0.6885 | 0.6570 | 0.7098 | 0.9073 |
| **4** | 6.52 | 0.5923 | 0.5522 | 0.5271 | 0.5620 | 0.5590 | 0.5202 | 0.5993 | 0.5691 | 0.6871 | 0.8300 |
| **5** | 6.40 | 0.5462 | 0.4734 | 0.4744 | 0.5206 | 0.5154 | 0.4796 | 0.5225 | 0.5901 | 0.6236 | 0.7761 |
| **6** | 6.30 | 0.6722 | 0.6595 | 0.5981 | 0.6649 | 0.7133 | 0.6064 | 0.6581 | 0.7148 | 0.7442 | 0.8893 |
| **7** | 5.60 | 0.6626 | 0.6388 | 0.5983 | 0.6447 | 0.6473 | 0.5968 | 0.6513 | 0.7164 | 0.6697 | 0.8902 |
| **8** | 8.53 | 0.6183 | 0.5432 | 0.5484 | 0.5916 | 0.5835 | 0.5501 | 0.6001 | 0.6695 | 0.6909 | 0.8534 |
| **9** | 7.75 | 0.6714 | 0.6047 | 0.6021 | 0.6458 | 0.6389 | 0.6079 | 0.6490 | 0.7201 | 0.7249 | 0.8849 |
| **10** | 7.45 | 0.4663 | 0.5274 | 0.5113 | 0.4867 | 0.5319 | 0.5036 | 0.4680 | 0.5374 | 0.5326 | 0.5263 |
| **11** | 6.40 | 0.4937 | 0.5381 | 0.5294 | 0.5030 | 0.5803 | 0.5225 | 0.4910 | 0.5610 | 0.5534 | 0.5493 |
| **12** | 6.30 | 0.5293 | 0.4555 | 0.4575 | 0.5170 | 0.4991 | 0.4778 | 0.5213 | 0.5892 | 0.6158 | 0.7742 |
| **13** | 6.14 | 0.3570 | 0.3893 | 0.3868 | 0.3686 | 0.3674 | 0.3654 | 0.4033 | 0.3492 | 0.3408 | 0.3586 |
| **14** | 5.83 | 0.5340 | 0.4901 | 0.4833 | 0.4959 | 0.5448 | 0.4648 | 0.5363 | 0.5433 | 0.5447 | 0.6606 |
| **15** | 5.71 | 0.4538 | 0.4046 | 0.4023 | 0.4540 | 0.4519 | 0.3811 | 0.4669 | 0.4644 | 0.4636 | 0.5804 |
| **16** | 5.35 | 0.5357 | 0.5353 | 0.5367 | 0.6133 | 0.5350 | 0.5597 | 0.5345 | 0.5364 | 0.5306 | 0.5220 |
| **17** | 5.35 | 0.4077 | 0.3837 | 0.3744 | 0.3839 | 0.3857 | 0.4064 | 0.3912 | 0.3791 | 0.4030 | 0.3915 |
| **18** | 4.00 | 0.6679 | 0.6025 | 0.6059 | 0.6303 | 0.6237 | 0.5855 | 0.6310 | 0.7095 | 0.6955 | 0.7753 |
| **19** | 4.00 | 0.5796 | 0.5091 | 0.5540 | 0.5579 | 0.5798 | 0.5438 | 0.5983 | 0.5889 | 0.6770 | 0.6986 |
| **20** | 2.78 | 0.5214 | 0.4338 | 0.4368 | 0.4978 | 0.4787 | 0.4823 | 0.4494 | 0.4892 | 0.4840 | 0.4763 |
| | | | | | | Alignment 4 | | | | | |
| **1** | 8.23 | 0.5311 | 0.5414 | 0.4826 | 0.4766 | 0.6571 | 0.4719 | 0.5506 | 0.7494 | 0.7832 | 1.0000 |
| **2** | 7.96 | 0.5918 | 0.5046 | 0.5758 | 0.4823 | 0.5606 | 0.5275 | 0.5732 | 0.5515 | 0.6979 | 0.8381 |
| **3** | 7.00 | 0.6530 | 0.4623 | 0.4922 | 0.4601 | 0.4519 | 0.5499 | 0.6161 | 0.4551 | 0.6519 | 0.9056 |
| **4** | 6.52 | 0.5706 | 0.4344 | 0.5220 | 0.4404 | 0.6373 | 0.5459 | 0.5793 | 0.4481 | 0.6877 | 0.8397 |
| **5** | 6.40 | 0.5340 | 0.5033 | 0.5021 | 0.4732 | 0.6192 | 0.4941 | 0.5147 | 0.6466 | 0.6793 | 0.7666 |
| **6** | 6.30 | 0.6054 | 0.5086 | 0.5277 | 0.4614 | 0.5980 | 0.5043 | 0.6136 | 0.6797 | 0.7004 | 0.8912 |
| **7** | 5.60 | 0.5967 | 0.4970 | 0.5204 | 0.4563 | 0.5965 | 0.5060 | 0.6036 | 0.6847 | 0.6418 | 0.8902 |
| **8** | 8.53 | 0.5911 | 0.4901 | 0.5567 | 0.4480 | 0.6355 | 0.5232 | 0.5539 | 0.6463 | 0.6789 | 0.7607 |
| **9** | 7.75 | 0.5025 | 0.5147 | 0.4511 | 0.4292 | 0.6237 | 0.4532 | 0.4841 | 0.6540 | 0.6974 | 0.8808 |
| **10** | 7.45 | 0.5493 | 0.4757 | 0.4912 | 0.4215 | 0.5366 | 0.4867 | 0.5410 | 0.6551 | 0.6303 | 0.7253 |
| **11** | 6.40 | 0.4649 | 0.4609 | 0.4336 | 0.4126 | 0.6015 | 0.4322 | 0.4760 | 0.6015 | 0.6390 | 0.7270 |
| **12** | 6.30 | 0.4916 | 0.4472 | 0.4732 | 0.4523 | 0.5922 | 0.4592 | 0.4864 | 0.5838 | 0.6099 | 0.6433 |

**Table 2** (Continued)

| no. | log (1/$IC_{50}$) | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Alignment 4 (continued) | | | | | |
| **13** | 6.14 | 0.5439 | 0.5031 | 0.3518 | 0.3357 | 0.5572 | 0.3500 | 0.4976 | 0.4154 | 0.4400 | 0.5058 |
| **14** | 5.83 | 0.5520 | 0.4473 | 0.5700 | 0.4494 | 0.5802 | 0.5346 | 0.4749 | 0.5572 | 0.5584 | 0.6649 |
| **15** | 5.71 | 0.4368 | 0.4100 | 0.4557 | 0.4423 | 0.5689 | 0.4433 | 0.4699 | 0.5253 | 0.5270 | 0.6095 |
| **16** | 5.35 | 0.5962 | 0.5382 | 0.5620 | 0.4475 | 0.7857 | 0.5132 | 0.5244 | 0.7069 | 0.7291 | 0.7609 |
| **17** | 5.35 | 0.6760 | 0.4955 | 0.6051 | 0.4666 | 0.5876 | 0.5795 | 0.6831 | 0.6535 | 0.6549 | 0.8157 |
| **18** | 4.00 | 0.6528 | 0.4678 | 0.5857 | 0.4431 | 0.6464 | 0.4786 | 0.6090 | 0.7156 | 0.6727 | 0.8209 |
| **19** | 4.00 | 0.5817 | 0.5119 | 0.5639 | 0.4669 | 0.6550 | 0.5316 | 0.6021 | 0.5745 | 0.6706 | 0.7058 |
| **20** | 2.78 | 0.4961 | 0.4542 | 0.4480 | 0.4221 | 0.5648 | 0.4513 | 0.5048 | 0.5528 | 0.5485 | 0.6526 |

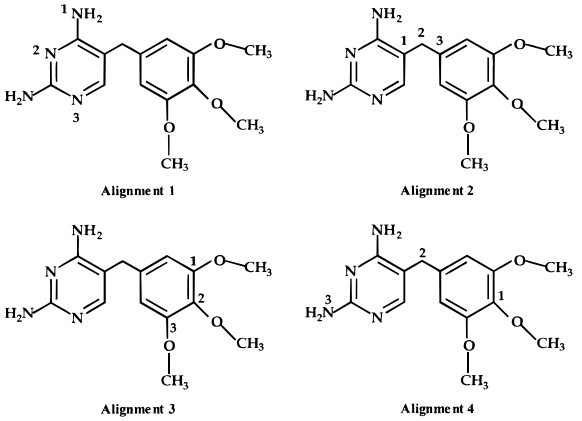



**Figure 4.** Four alignment rules used in this study.

which can be referenced to the enzyme-bound geometry and/or to the intrinsic free space stability of the ligand.

The conformational profiles of the series of analog inhibitors are defined by the torsion angles $\tau_1$ and $\tau_2$. The bound conformation of trimethoprim found in its complex with *E. coli* DHFR is defined by torsion angle states $\tau_1 = 177°$ and $\tau_2 = 76°$,[19] where $\tau_1 = \tau_2 = 0$ corresponds to the reference conformation, in the cis configuration. This conformation is not the free space intramolecular global minimum energy for any of the analogs in the training set. Trimethoprim was used as the shape reference, and 10 trial conformations were considered for each compound. Each of the 10 conformations is operationally equivalent to one another across the set of analogs with respect to bonding topology defining the torsion angles as is discussed below.

Trimethoprim is found to have eight apparent free space minimum energy conformations within 5 kcal/mol of its global intramolecular minimum energy conformation. For each of the other analogs in the data set, the apparent minimum energy conformations within 5 kcal/mol of the respective global minimum energy conformation and nearest in ($\tau_1$, $\tau_2$) torsional angle space to the minimum energy conformations of trimethoprim were considered, that is, the apparent (10° resolution in $\tau_1$ and $\tau_2$) minimum energy conformations within 5 kcal/mol, closest to the $\tau_1$ and $\tau_2$ of the selected eight apparent minima of trimethoprim, were selected. For those compounds that do not have minima for $\tau_1$ and $\tau_2$ values close to those of trimethoprim (compounds **6**, **7**, and **16**−**19**), the $\tau_1$ and $\tau_2$ values were set to those of the trimethoprim minimum.

Compounds **6** and **7** have a $sp^2$ benzylic carbon, while compounds **16**−**19** have substituents at the benzylic carbon, restricting conformational freedom about $\tau_1$ and $\tau_2$. The ($\tau_1$, $\tau_2$) of conformation 9 for each compound corresponds to its secondary minimum energy conformation within ±30° of the crystal-bound conformation of trimethoprim to *E. coli* DHFR. Overall the ($\tau_1$, $\tau_2$) values vary within a range of ±30° of 177° and 76°, respectively. Conformation 10 has the ($\tau_1$, $\tau_2$) values of the crystal-bound conformation ($\tau_1 = 177°$, $\tau_2 = 76°$) of trimethoprim to *E. coli* DHFR. The conformations used in the MSA 3D-QSAR analysis are given in Table 6 for the alignments considered.

The four alignment rules selected in this study are shown in Figure 4. In each test alignment, all compounds in the data set are compared pairwise to trimethoprim using the three alignment atoms defining the alignment rule.

As mentioned above, the COSV was selected as a MSA descriptor to test the methodology reported in this study. The normalized COSV reflects the shape similarity with respect to the shape reference compound and is given in eq 7. In eq 7, $V_{test}$ is the COSV for the conformation of the test compound relative to that of the reference conformation of trimethoprim, $V_{tmp}$. This descriptor was computed for the 20 inhibitors in each of the 10 conformations and for four alignments to give a 20 × 10 × 4 array. The 2-dimensional substituent constants, viz., the hydrophobic constants of the substituents on the phenyl ring, the $\pi$, $\pi_{3,5}$, and the molar refractivity of substituents on the phenyl ring, etc., were taken from the literature.[26] The descriptors are given in Table 4.

$$\mathrm{COSV} = \frac{V_{test} \cap V_{tmp}}{V_{tmp} \cap V_{tmp}} \tag{7}$$

## Results

PLS is sensitive to variance. To avoid dominance of variables with large variation, the data are autoscaled or regularized to give the variables zero mean and unit variance. In this application the data were mean centered but not regularized since the COSV is normalized.

Application of 3-way PLS to the array gave, by cross-validation, a model with two significant PLS components. The model accounted for 61% (37% + 24%) of the variance in the **Y** data. The parameters of interest from the decomposition are the scaled **X** loadings, *W*. The *W* values are given in Table 3.

In order to rank the set of conformations/alignments considered, a scoring function incorporating the conformation and alignment weights, CAW (eq 8), similar to

$$\mathrm{CAW}_{m,n} = \sum_{a=1}^{A} \mathrm{Var}_a W^2_{a,m,n} \tag{8}$$

that used in our previous 3D-QSAR study of muscarinic agents,[11] was computed. $\mathrm{Var}_a$ is the **Y** variance ex-

**Table 3.** Conformation and Alignment Loadings (Normalized to Length 1) for Common Overlap Steric Volume Data

| | conformation | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| alignment | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | −0.0183 | 0.0031 | −0.0118 | −0.0019 | −0.0530 | −0.0004 | −0.0126 | 0.0005 | 0.1835 | 0.2890 |
| 2 | 0.0842 | 0.0344 | 0.0058 | −0.0284 | 0.1733 | −0.0159 | 0.0644 | 0.2012 | 0.3057 | 0.4628 |
| 3 | 0.1284 | 0.1710 | 0.1430 | 0.1226 | 0.1542 | 0.1278 | 0.1577 | 0.1933 | 0.2331 | 0.4149 |
| 4 | −0.0215 | 0.0392 | −0.0366 | 0.0179 | −0.0576 | 0.0110 | −0.0289 | 0.0284 | 0.1354 | 0.2250 |

**Table 4.** Two-Dimensional Descriptors

| no. | $S$ | $\pi_{3,5}$ | $\pi_{3,4,5}$ | $\pi$ | log $P$ | MR |
|---|---|---|---|---|---|---|
| **1** | 1.34 | 0.02 | −0.52 | −1.196 | 0.484 | 2.36 |
| **2** | 1.34 | 0.02 | −0.65 | −1.133 | 0.547 | 1.86 |
| **3** | 1.06 | 0.02 | −0.52 | −1.196 | 0.933 | 2.36 |
| **4** | 1.06 | 0.02 | −0.65 | −1.133 | 0.996 | 1.86 |
| **5** | 1.34 | 0.02 | 1.98 | 1.799 | 3.479 | 2.74 |
| **6** | 0.51 | 0.02 | −0.52 | −1.196 | 1.158 | 2.36 |
| **7** | 0.69 | 0.02 | −0.52 | −1.196 | 0.629 | 2.36 |
| **8** | 1.34 | 0.02 | 0.46 | 0.385 | 2.065 | 2.46 |
| **9** | 1.34 | 0.02 | 0.02 | −0.081 | 1.599 | 1.68 |
| **10** | 1.34 | 1.12 | 1.12 | 0.909 | 2.589 | 1.23 |
| **11** | 1.34 | 0.11 | 0.11 | −0.170 | 1.510 | 0.99 |
| **12** | 1.34 | 0.00 | 0.86 | 0.724 | 2.454 | 1.09 |
| **13** | 1.34 | 0.67 | 0.67 | 0.624 | 2.304 | 0.81 |
| **14** | 1.34 | −2.06 | −2.06 | −2.163 | −0.485 | 1.54 |
| **15** | 1.34 | 0.00 | 0.00 | 0.000 | 1.591 | 0.31 |
| **16** | 0.92 | 0.02 | −0.52 | −1.196 | 0.883 | 2.36 |
| **17** | 0.92 | 0.02 | −0.52 | −1.196 | 0.883 | 2.36 |
| **18** | 0.86 | 0.02 | −0.52 | −1.196 | −0.874 | 2.36 |
| **19** | 0.86 | 0.02 | −0.52 | −1.196 | −0.874 | 2.36 |
| **20** | 1.34 | −1.34 | −1.34 | −1.423 | 0.257 | 2.78 |

**Table 5.** Factor Loadings for Compounds

| | component | | | component | |
|---|---|---|---|---|---|
| no. | 1 | 2 | no. | 1 | 2 |
| **1** | −0.48 | −0.26 | **11** | 0.11 | −0.11 |
| **2** | −0.10 | 0.01 | **12** | 0.07 | −0.06 |
| **3** | −0.22 | 0.10 | **13** | 0.47 | −0.16 |
| **4** | −0.07 | 0.03 | **14** | 0.12 | 0.07 |
| **5** | 0.01 | −0.08 | **15** | 0.30 | 0.07 |
| **6** | −0.27 | −0.08 | **16** | 0.05 | −0.05 |
| **7** | −0.23 | −0.07 | **17** | 0.26 | −0.22 |
| **8** | −0.14 | −0.11 | **18** | −0.09 | 0.66 |
| **9** | −0.22 | −0.16 | **19** | 0.03 | 0.55 |
| **10** | 0.13 | −0.12 | **20** | 0.25 | −0.01 |

**Table 6.** Factor Loadings for Alignment

| | factor | | | factor | |
|---|---|---|---|---|---|
| alignment | 1 | 2 | alignment | 1 | 2 |
| 1 | −0.53 | −0.65 | 3 | 0.82 | −0.24 |
| 2 | −0.18 | 0.69 | 4 | −0.10 | 0.21 |

plained by the **X** data in component $a$. The $CAW_{10,2} = 0.10$, $CAW_{10,3} = 0.07$, and $CAW_{9,2} = 0.05$. Conformation 10, the receptor-bound conformation for trimethoprim, is the highest scored conformation in conjunction with alignments 2 and 3, the highest scored of the four alignments.

The application of 3-way factor analysis to the data gave results consistent with those of PLS. Two eigenvalues were extracted from each of the **M**, **P**, and **Q** matrices for construction of **A**−**C**. The two factors explain 71% (58% + 13%), 90% (61% + 29%), and 93% (84% + 9%) of the variance in **X**, respectively. The factor-loading matrices are given in Tables 5−7. As used here, this formalism does not yield a QSAR in the sense of a classical Hansch analysis but is used to provide a means to weigh and statistically evaluate the possible geometries which are significant for receptor recognition of the ligand. This approach can also be thought of as a filtering process to generate "simple" 3D-QSARs.

**Table 7.** Factor Loadings for Conformation

| | factor | | | factor | |
|---|---|---|---|---|---|
| conformation | 1 | 2 | conformation | 1 | 2 |
| 1 | −0.14 | −0.40 | 6 | −0.26 | −0.09 |
| 2 | −0.28 | 0.14 | 7 | −0.14 | −0.31 |
| 3 | −0.24 | −0.10 | 8 | 0.21 | 0.50 |
| 4 | −0.32 | 0.11 | 9 | 0.31 | 0.06 |
| 5 | 0.17 | 0.52 | 10 | 0.69 | −0.42 |

The highest weighted MSA tensors were used to generate a 3D-QSAR model, using the COSV, and its square term only. The resulting 3D-QSARs for conformation 10/alignment 2, conformation 10/alignment 3, and conformation 9/alignment 2 are eqs 9−11, respectively.

$$\log(1/IC_{50}) = 5.51COSV^2 + 2.74 \qquad (9)$$

$$n = 20,\ R^2 = 0.50,\ \text{XV-}R^2 = 0.42,\ F = 17.839$$

$$\log(1/IC_{50}) = 2.9COSV^2 + 4.62 \qquad (10)$$

$$n = 20,\ R^2 = 0.25,\ \text{XV-}R^2 = 0.09,\ F = 6.617$$

$$\log(1/IC_{50}) = 12.38COSV - 1.91 \qquad (11)$$

$$n = 20,\ R^2 = 0.52,\ \text{XV-}R^2 = 0.42,\ F = 19.473$$

Here, $n$ is the number of compounds, $R^2$ is the correlation coefficient, and XV-$R^2$ is the cross-validated $R^2$, or sometimes referred to as $Q^2$, computed by the leave-one-out method. Statistically, these results are similar for conformation 10/alignment 2 and conformation 9/alignment 2. The result for conformation 10/alignment 3 (eq 10) is not significant. The second-order term in eq 9 suggests a nonlinear relationship and that factors other than COSV may also be important. These are explored below.

Additionally the highest weighted MSA tensors can be used with the **H** and **E** tensors and simple regression methods to derive MSA 3D-QSARs. Further QSAR development was carried out with genetic function analysis[27] in conjunction with a one-component PLS regression analysis, using substituent descriptors in combination with COSV. From X-ray crystallography, the values of $\tau_1$ and $\tau_2$ of trimethoprim bound to *E. coli* DHFR are 177° and 76°, respectively. This is the geometry of conformation 10. NMR studies also show that in the binary complex of trimethoprim with the *Lactobacillus casei* enzyme, the benzyl group is in a dynamic state in which $\tau_1 = 190 \pm 10°$ and $\tau_2 = 73 \pm 10°$. The same conformational flexibility can be expected for the *E. coli* enzyme in solution. Thus the torsion angle entropy for $\tau_1$ and $\tau_2$ was computed by using the TAU theory[25] for polymers and used as a

descriptor to generate 3D-QSAR. The ($\tau_1$ and $\tau_2$) conformational entropy, *s*, for the compounds is defined in eq 12, where $n_0$ is the number of torsion angle units and $s_i(\tau)$ is the *i*th torsion angle unit entropy. The *S* values are listed in Table 4.

$$S = \frac{1}{n_0}\sum_{i=1}^{n_0} S_i(\tau) \qquad (12)$$

The descriptor COSV was used with 2-dimensional descriptors, viz., $\pi$, MR, etc.,[26] a to construct QSAR models. The QSARs were derived using the genetic function analysis, GFA, approximation.[27] Modeling with variable selection was carried out using genetic PLS, or G/PLS, as recently described.[28] Using standard substituent descriptors for conformation 10/alignment 2, in combination with COSV, a one-component PLS model was derived and rotated into original descriptor space yielding eq 13.

$$\log(1/\mathrm{IC}_{50}) = 0.36\pi - 0.35\mathrm{MR}^2 + 17.55\mathrm{COSV} + 0.05\mathrm{NOV} + 1.89S^2 - 10.28 \qquad (13)$$

$$n = 20,\ R^2 = 0.913,\ \text{XV-}R^2 = 0.816,\ F = 29.325$$

The QSAR is optimal for prediction rather than being optimal for mechanistic interpretation. This results from the use of G/PLS and the choice of a one-component model. G/PLS selects variables for the single latent variable which are most highly correlated so the variables in eq 13 are to some extent redundant.

Random scrambling of the **y** vector was used to validate eq 13. The biological activity measures of all 20 compounds in the data set were repetitively randomized 30 times to generate 30 arrays of reordered variables, while the **X** data (independent variable data) were left intact. A random number generator was used to allocate the integers between 1 and 20 to sequences of 20 numbers each time. In each cycle, the resulting arrangement of random integers was employed in order to reorder the **Y** data. Full data analysis was carried out on these scrambled data. In each of the 30 cases, the scrambled data yielded a much lower $R^2$ value than the original data which indicates that the two 3D-QSAR equations did not result from chance.[29] The results are given in Table 8. Thus, conformation 10 and alignment 2 or 3 lead to significant and reliable MSA 3D-QSAR models.

**Table 8.** Validation Results from Scrambling

| scrambled BA | C10A2-$R^2$ | XV-$R^2$ C10A2 |
|---|---|---|
| 1 | 0.148 | −4.43 |
| 2 | 0.358 | −2.37 |
| 3 | 0.157 | −0.55 |
| 4 | 0.245 | −0.42 |
| 5 | 0.4 | −2.36 |
| 6 | 0.296 | −1.00 |
| 7 | 0.358 | −0.50 |
| 8 | 0.213 | −0.59 |
| 9 | 0.431 | −1.17 |
| 10 | 0.23 | −2.40 |
| 11 | 0.339 | −0.69 |
| 12 | 0.222 | −0.27 |
| 13 | 0.285 | −1.11 |
| 14 | 0.575 | −1.50 |
| 15 | 0.421 | −0.17 |
| 16 | 0.197 | −0.91 |
| 17 | 0.316 | −0.70 |
| 18 | 0.316 | −0.83 |
| 19 | 0.277 | −0.71 |
| 20 | 0.308 | −0.27 |
| 21 | 0.27 | −2.34 |
| 22 | 0.328 | −1.03 |
| 23 | 0.25 | −2.16 |
| 24 | 0.515 | −0.23 |
| 25 | 0.383 | −0.55 |
| 26 | 0.519 | −1.22 |
| 27 | 0.125 | −0.57 |
| 28 | 0.227 | −1.90 |
| 29 | 0.303 | −0.34 |
| 30 | 0.216 | −0.71 |
| average | 0.182 | −2.57 |
| range | 0.125−0.575 | −4.43−0.17 |

## Discussion

The results of this investigation are consistent with previously published QSAR studies of the inhibition of DHFR and also with results from X-ray crystal structure studies of *E. coli* DHFR and its binary complex with trimethoprim. From the crystal structure, the active site of DHFR is a cavity 15 Å deep, lined with hydrophobic amino acids that make van der Waals contacts with the *m*-methoxy groups of trimethoprim.[30] This explains the importance of the descriptor $\pi$ in the model. This descriptor has also been found by others[31] to be significant in DHFR QSARs.

The role of alignment in the receptor recognition process has not been studied as extensively as that of ligand conformation. The alignments used here were chosen so that they would represent overlap of common substructural elements located in very different regions of the ligands. Alignment 1 fixes the ligands at atoms of the 2,4-diaminopyrimidine group, alignment 2 fixes the series at the methylene bridge, alignment 3 fixes the ligands at the trimethoxybenzyl group, and alignment 4 is the most rigid and fixes the ligands over the entire structure. Thus, the alignments do explore realizing shape commonality over different representations of the ligands.

The MSA/tensor methodology selects alignments 2 and 3 as the most significant. Of these, alignment 2 is preferred as it allows the ligands flexibility at their structural extremes so that they can orient their different functional groups for interaction with their complements in the active site. One might expect that alignment 1 should also be significant since the 2,4-diaminopyrimidine group anchors the ligands to the receptor as shown from $^{15}$N NMR spectroscopic studies of the binary complex in solution. The protonated N1 of trimethoprim forms a hydrogen-bonding/charge transfer interaction with the carboxylate group of Asp-26 in the *E. coli* enzyme.[32] However, this alignment does not allow ligand flexibility as discussed earlier.

The work presented here carries 3D-QSAR to a new level. This is the first report of a general methodology for treating structure−activity data for flexible compounds that can assume different conformations and receptor alignments. The methodology correctly predicts the receptor-bound conformation of trimethoprim and its analogs and gives results which are consistent with published QSAR studies. It provides a framework within classical QSAR for exploring higher dimensional conformation and alignment descriptor space while simultaneously considering activity. Now that this step has been taken, the formalism can easily be extended

to higher dimensions, e.g., time. This is currently being explored in our laboratory.

The tensor representation of flexible molecular interactions and their solution by the PLS factor analysis procedures are applicable to a variety of chemical recognition problems and not just receptor-independent 3D-QSAR. The same conformation and alignment problems exist when the receptor geometry is known but not the ligand−receptor binding complex geometry. The modeling of crystallization requires solving for conformation and alignment. In fact, all molecular assemblies have embedded in them conformation/alignment multiplicity which can be treated by the approach presented here. Thus, the tensor representation/PLS factor analysis solution described offers hope to treat a universe of previously dimensionally forbidden problems in chemistry.

**Acknowledgment.** Resources of the Laboratory of Molecular Modeling and Design were used for part of this work. The Chem21 Group gratefully acknowledges support from a NSF SBIR grant.

## References

(1) Hansch, C. A Quantitative Approach to Biochemical Structure-Activity Relationships. *Accts. Chem. Res.* **1969**, *2*, 232.
(2) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
(3) Hopfinger, A. J.; Burke, B. J. Molecular Shape Analysis: A Formalism to Quantitatively Establish Spatial Molecular Similarity. In *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Eds.; Wiley and Sons: New York, 1990; p 173.
(4) Srivastava, S.; Richardson, W. W.; Bradley, M. P.; Crippen, G. M. In *3D-QSAR in Drug Design*: *Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993; p 409.
(5) Good, A. C.; Peterson, S. J.; Richards, W. G. QSAR's from Similarity Matrices. Technique Validation and Application in the Comparison of Different Similarity Evaluation Methods. *J. Med. Chem.* **1993**, *36*, 2929.
(6) Folkers, G.; Merz, A.; Rognan, D. In *3D-QSAR in Drug Design*: *Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993; p 583.
(7) Cho, S. J.; Tropsha, A. Cross-Validated $R^2$-Guided Region Selection for Comparative Molecular Field Analysis: A Simple Method to Achieve Consistent Results. *J. Med. Chem.* **1995**, *38*, 1060.
(8) Marshall, G. R. In *3D-QSAR in Drug Design*: *Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993; p 80.
(9) Green, S. M.; Marshall, G. R. 3D-QSAR: A Current Perspective. *TiPS* **1995**, *16*, 285.
(10) Hopfinger, A. J.; Burke, B. J.; Dunn, W. J., III. A Generalized Formalism for Three-Dimensional Quantitative Structure-Property Relationship Analysis for Flexible Molecules Using Tensor Representation. *J. Med. Chem.* **1994**, *37*, 3768.
(11) Burke, B. J.; Dunn, W. J., III; Hopfinger, A. J. Construction of a Molecular Shape Analysis-Three-Dimensional Quantitative Structure-Analysis Relationship for an Analog Series of Pyridobenzodiazepinone Inhibitors of Muscarinic 2 and 3 Receptors. *J. Med. Chem.* **1994**, *37*, 3775.
(12) Baker, D. J.; Beddell, C. R.; Champness, J. N.; Goodford, P. J.; Norrington, F. E. A.; Smith, D. R.; Stammers, D. K. The Binding of Trimethoprim to Bacterial Dihydrofolate Reductase. *FEBS. Lett.* **1981**, *126*, 49.
(13) Dean, P. M. In *3D-QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993.
(14) Levin, J. Three-mode Factor Analysis. *Psychol. Bull.* **1965**, *64*, 442−452.
(15) Apellof, C. J.; Davidson, E. R. Three Dimensional Rank Annihilation for Multicomponent Determinations. *Anal. Chim. Acta* **1983**, *146*, 9−14. Sanchez, E.; Kowalski, B. R. Generalized Rank Annihilation Factor Analysis. *Anal. Chem.* **1986**, *58*, 496−499.
(16) Zeng, Y.; Hopke, P. K. The Application of Three-mode Factor Analysis (TMFA) to Receptor Modeling of Scenes Particle Data. *Atmosph. Environ.* **1992**, *26A*, 1701−1711.
(17) Lohmöller, J.-B.; Wold, H. Three-Mode Path Models with Latent Variables and Partial Least Squares (PLS) Parameter Estimation. *Proceedings of the European Meeting of the Psychometric Society*; University of Gronigen: The Netherlands, 1980; p 50.
(18) Glen, W. G.; Dunn, W. J., III; Scott, D. R. Principal Components Analysis and Partial Least Squares Regression. *Tetrahedron Comput. Method* **1989**, *2*, 349.
(19) Koetzle, T. F.; Williams, G. J. B. The Crystal and Molecular Structure of the Antifolate Drug Trimethoprim (2,4-Diamino-5-(3,4,5-triemthoxybenzyl)pryrimidine). A Neutron Diffraction Study. *J. Am. Chem. Soc.* **1976**, *98*, 2074. Mathews, D. A.; Bolin, J. T.; Burridge, J. M.; Filmen, D. J.; Volz, K. W.; Kaufman, B. T.; Beddell, C. R.; Champness, J. N.; Stammers, D. K.; Kraut, J. Refined Crystal Structures of *Escherichia coli* and Chicken Liver Dihydrofolate Reductase Containing Bound Trimethoprim. *J. Biol. Chem.* **1985**, *260*, 381.
(20) Cheng, Y.; Prusoff, W. H. Relationship between the inhibition constant ($K_i$) and the concentration of inhibitor which causes 50 percent inhibition ($I_{50}$) of enzyme reaction. *Biochem. Pharmacol.* **1973**, *22*, 3099−3108.
(21) (a) Burke, B. J.; Hopfinger, A. J. 1-(Substituted-benzyl)imidazole-2(3H)-thione Inhibitors of Dopamine-Hydroxylase. *J. Med. Chem.* **1990**, *33*, 274. (b) Koehler, M. G.; Rowberg-Schaefer, K.; Hopfinger, A. J. A molecular shape analysis and quantitative structure-activity relationship investigation of some triazine-antifolate inhibitors of *leishmania* dihydrofolate reductase. *Arch. Biochem. Biophys.* **1988**, 266, 152−161.
(22) Mabilia, M.; Pearlstein, R. A.; Hopfinger, A. J. Molecular shape analysis and energetics-based intermolecular modeling of benzylpyrimidine dihydrofolate reductase inhibitors. *Eur. J. Med. Chem. - Chim. Ther.* **1985**, *20*, 163−174.
(23) Dewar, M. J. S.; Thiel, W. Ground States of Molecules. 38. The MNDO method, Approximations and Parameters. *J. Am. Chem. Soc.* **1977**, *99*, 4899. Dewar, M. J. S.; Thiel, W. Ground States of Molecules. 39. MNDO Results for Molecules Containing Hydrogen, Carbon, Nitrogen and Oxygen. *J. Am. Chem. Soc.* **1977**, *99*, 4907.
(24) Hopfinger, A. J.; Perlstein, R. A. Molecular Mechanics Force-Field Parameterization Procedures. *J. Comput. Chem.* **1985**, *5*, 486.
(25) Koehler, M. G.; Hopfinger, A. J. Molecular Modeling of Polymers: 5. Inclusion of intermolecular energetics in estimating glass and crystal-melt transition temperatures. *Polymer* **1989**, *30*, 116−126.
(26) Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; Wiley-Interscience: New York, 1979.
(27) Rogers, D.; Hopfinger, A. J. Application of Genetic Function Approximation to Quantitative Structure-Activity Relationships and Quantitative Structure-Property Relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854−866.
(28) Rogers, D.; Dunn, W. J., III. Genetic Partial Least Squares. *J. Comput.-Aided Mol. Des.* **1996**, in press.
(29) Wold, S.; Eriksson, L. Statistical Validation of QSAR Results. In *Chemometric Methods in Molecular Design*; Waterbeemd, H., Ed.; VCH Publishers, Inc.: New York, 1995; p 311.
(30) Li, R.-L.; Poe, M. Quantitative Structure-Activity Relationships for the Inhibition of *Escherichia coli* Dihydrofolate Reductase by 5-(Substituted benzyl)-2,4-diaminopyrimidines. *J. Med. Chem.* **1988**, *31*, 366.
(31) Selassie, S. D.; Li, R.-L.; Poe, M.; Hansch, C. On the Optimization of Hydrophobic and Hydrophilic Substituent Interactions of 2,4-Diamino-5-(substituted-benzyl)pyrimidines with Dihydrofolate Reductase. *J. Med. Chem.* **1991**, *34*, 46.
(32) Huang, F.; Yang, Q.-X.; Huang, T.; Gelbaum, L.; Kuyper, L. F. The conformations of trimethoprim/*E. coli* dihydrofolate reductase complexes. A $^{15}N$ and $^{31}P$ NMR study. *FEBS Lett.* **1991**, *38*, 44.

JM960491R