

HQSAR: A New, Highly Predictive QSAR Technique

David R. Lewis

Table of Contents

ABSTRACT	3
INTRODUCTION.....	3
QSAR TECHNIQUES	3
CALCULATION OF MOLECULAR DESCRIPTORS.....	3
STATISTICAL GENERATION OF THE QSAR MODEL.....	4
HQSAR.....	4
HQSAR THEORY	4
MOLECULAR HOLOGRAMS	5
HQSAR PARAMETERS	6
HOLOGRAM LENGTH	7
FRAGMENT SIZE	7
FRAGMENT DISTINCTION	7
FRAGMENT PATTERNS	8
RESULTS AND DISCUSSION	8
DATA SETS.....	9
HQSAR PARAMETERS	9
COMPARISON OF HQSAR AND COMFA	9
FURTHER HQSAR COMPARISONS	10
VALIDITY OF HQSAR MODELS	11
EFFECT OF HOLOGRAM PARAMETERS	12
TIMINGS OF HQSAR RUNS	15
CONCLUSION.....	15
REFERENCES	15

ABSTRACT

QSAR techniques have proven to be extremely valuable in pharmaceutical research, particularly 3D QSAR. However, the necessity of descriptor calculation and selection or conformer generation, and structural alignment makes use of QSAR non-trivial. Demands for analysis of large data sets such as those being generated by combinatorial chemistry and high throughput screening have compounded this problem. Hologram QSAR (HQSAR) is a new technique which employs specialized fragment fingerprints (molecular holograms) as predictive variables of biological activity. By removing the necessity for molecular alignment, HQSAR models can be obtained more rapidly than other techniques which makes HQSAR readily applicable to both small and large data sets. HQSAR models are comparable in predictive ability to those derived from CoMFA studies. In many cases HQSAR can be readily extended to support database searching.

INTRODUCTION

QSAR techniques have proven exceptionally useful in the drug design process^{1,2}. Generation of QSAR models which are able to predict biological activities of new compounds and infer possible activity enhancing modifications are invaluable tools in increasing the efficiency of modern synthesis and screening programs. Hologram QSAR (HQSAR) is a new QSAR method which relates biological activity to structural molecular composition²¹ where molecular composition is described in terms of patterns of substructural fragments.

QSAR TECHNIQUES

QSAR techniques may be divided into two main types: classical Hansch³ type analyses (2D) and 3D techniques, e.g. CoMFA⁴. Both techniques have two basic steps:

- i) Calculation of molecular descriptors
- ii) Statistical generation of the QSAR model

CALCULATION OF MOLECULAR DESCRIPTORS

Classical Hansch type approaches employ molecular descriptors which are most often substituent constants, whole molecule properties such as molar refractivity, logP, orbital energies, or atomic properties such as charge⁵. Descriptors must be selected very carefully⁶ to obtain robust predictive models. This is illustrated by examination of the enormous variety of molecular descriptors employed in traditional QSAR studies. Additionally, values for these descriptor variables must be calculated or determined experimentally. Selection and calculation of descriptor variables for classical QSAR studies can be a time consuming process.

One of the most popular 3D QSAR methods is Comparative Molecular Field Analysis (CoMFA). CoMFA employs variation in field strengths around a set of aligned 3D structures to describe the observed variation in biological activity.

CoMFA has been shown to provide highly predictive models^{7,8,9,10} for a large number of data sets. CoMFA models can be projected back over molecular structures allowing visualization of important 3D information affecting biological activity. Generation of CoMFA models is, however, non-trivial. Decisions regarding molecular conformation and relative alignment can be difficult and complex, especially when data sets contain structurally very diverse molecules where no obvious alignment rule suggests itself. Moreover, the alignment process is not usually amenable to automation which is necessary for the analysis of large data sets.

STATISTICAL GENERATION OF THE QSAR MODEL

Many statistical methods have been employed to generate QSAR models from descriptive variables. The most commonly used techniques are Multiple Linear Regression (MLR) and Partial Least Squares (PLS)¹¹. Both methods have their advantages and disadvantages. Classical QSARs most often use MLR where the ratio of the data points to the number of number of descriptors should not exceed five¹². While PLS analyses are particularly suited to situations where the number of descriptor variables exceeds the number of observations it is often the case that the principal components extracted from the descriptor variables have unclear physical meaning. It should be noted that the CoMFA technique does allow physical interpretation of PLS extracted QSAR model components in terms of 3D contour maps.

HQSAR

HQSAR is a new QSAR technique that avoids many of the problems associated with classical or 3D QSAR approaches²¹. Only 2D structures and activity are required as input—no complex descriptor selection process or 3D molecular alignment is required. HQSAR converts the molecules of a data set into counts of their constituent fragments. The patterns of fragment counts from dataset molecules are then related to observed biological activity data using Partial Least Squares analysis. Both steps, fragment counting and PLS analysis, are very fast. Nevertheless, the method is robust and highly predictive for many data sets.

In this paper, the general performance and behavior of HQSAR is examined by performing HQSAR analyses on a number of data sets for which previous QSAR studies have been published.

HQSAR THEORY

HQSAR works by identifying patterns of substructural fragments relevant to biological activity in sets of bioactive molecules. Unlike maximal common subgraph algorithms and the Stigmata algorithm¹³ which seek structural commonalities, HQSAR yields a predictive relationship between substructural features in the data set and biological activity using PLS.

As with most QSAR methods, there are two steps to generate a QSAR model: calculation of the molecular descriptor and subsequent statistical analysis. The following section describes the calculation of molecular holograms and the parameters that affect their nature. No discussion of PLS and its use in statistical analysis is included here as this has been covered in great detail elsewhere²².

Molecular Holograms

The first stage in the HQSAR method is the generation of molecular holograms. A molecular hologram is an array containing counts of molecular fragments and is related to traditional binary 2D fingerprints employed in database searching and molecular diversity techniques. The process of hologram generation is depicted in Figure 1.

The input data set consists of the 2D chemical structures¹⁴ and their associated biological data. The molecular structures are broken down on the fly into all possible linear and branched fragments of connected atoms of size between M and N atoms. Each unique fragment in the data set is assigned a specific large positive integer by means of a cyclic redundancy check (CRC) algorithm. Each of these integers corresponds to a bin in an integer array of fixed length L (L is generally in the range 50 to 500). Bin occupancies are incremented according to the fragments generated. Thus, all generated fragments are hashed¹⁵ into array bins in the range 1 to L. This array is called a molecular hologram, and the bin occupancies are the descriptor variables.

The use of hashing greatly reduces the size of the molecular hologram but leads to a phenomenon called "fragment collision". During fragment generation, identical fragments are always hashed to the same bin, and the corresponding occupancy for that bin is incremented. However, as the hologram length is generally smaller than the total number of unique fragments, different unique fragments can hash to the same bin causing "collisions" between fragments. This is discussed further in the section on hologram length.

Computation of the Molecular Holograms for a data set of structures yields a data matrix of dimension R x L, where R is the number of compounds in the data set and L is the length of the Molecular Hologram. For QSAR purposes, a matrix of target variables (biological activities) is also created. Standard PLS analysis is then applied to identify a set of orthogonal explanatory variables (components) that are linear combinations of the original L variables. Leave-one-out crossvalidation is applied to determine the number of components that yields an optimally predictive model. Once an optimal model is identified, PLS yields a mathematical equation that relates the Molecular Hologram bin values to the corresponding biological activity of each compound in the data set.

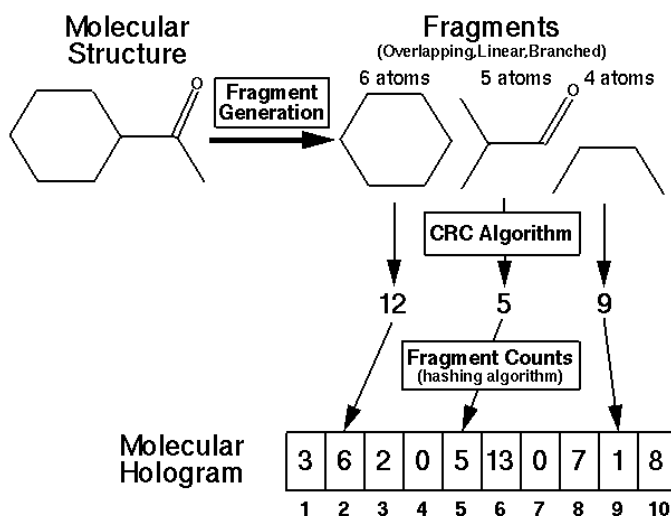


Figure 1: Generation of Molecular Holograms.

The form of this equation for the generated QSAR model is shown by equation 1.

$$\text{Activity}_i = c_0 + \sum_{l=1}^L c_l \text{ xil } c_{il} \quad (1)$$

where xil is the occupancy value of the Molecular Hologram of compound i at position or bin l , c_{il} is the coefficient for that bin derived from the PLS analysis, L is the length of the hologram, Activity_i is the biological activity, and c_0 is a constant.

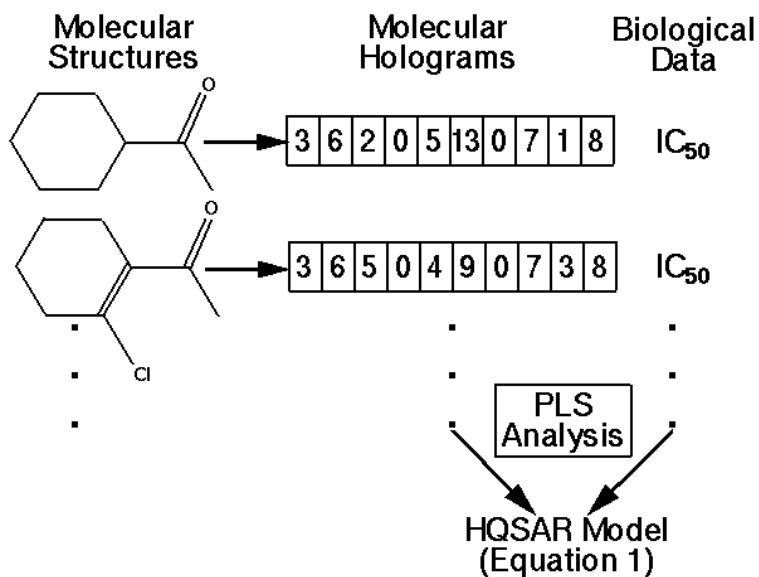


Figure 2: HQSAR Overview

HQSAR Parameters

HQSAR models can be affected by a number of parameters concerning hologram generation: hologram length, fragment size, and fragment distinction.

Hologram Length

The hologram length is a user-definable parameter which controls the number of bins in the hologram fingerprint. Because the hologram length is significantly less than the number of fragments in most compounds, alteration of hologram length, L , will cause the pattern of bin occupancies in data set holograms to change. This means that fragment collisions (where different unique fragments map to the same holographic bin) will be altered. Certain patterns of fragment disposition in the molecular holograms enables PLS to more readily detect the relationship between fragments present in the data set and the variance in biological activity. Twelve default hologram lengths which have been found to yield predictive models on a number of test data sets are provided. These default hologram lengths are prime numbers such that each provides a unique set of fragment collisions.

Fragment Size

Fragment size controls the minimum and maximum length of fragments to be included in the hologram fingerprint. As mentioned previously, molecular holograms are produced by the generation of all linear and branched fragments between M and N atoms in size. The parameters M and N can be changed to include smaller or larger fragments in the holograms. Default fragment lengths ($M = 4$ and $N = 7$) are provided. These values were derived through testing the HQSAR method on a number of different data sets and have proved generally useful.

Fragment Distinction

Depending on the application and data set in question, HQSAR allows fragments to be distinguished based on the atoms, bonds, connections, hydrogens, and chirality parameters which are defined in Table 1 below. Figure 3 shows graphically how these different parameter settings lead to the generation of distinct fragments from the same portion of the original molecule.

Atoms	The <i>atoms</i> parameter enables fragments to be resolved based on elemental atom types, for example, allowing N to be distinguished from P.
Bonds	The <i>bonds</i> parameter enables fragments to be distinguished based on bond orders, for example, in the absence of hydrogen, allowing butane to be distinguished from 2-butene.
Connections	The <i>connections</i> parameter provides a measure of atomic hybridization states within fragments. That is, connections causes HQSAR to keep track of how many connections are made to constituent atoms, and the bond order of those connections.
Hydrogens	By default, HQSAR ignores the hydrogen atoms during fragment generation; the <i>hydrogens</i> parameter overrides this behavior.
Chirality	The <i>chirality</i> parameter enables fragments to be distinguished based on atomic and bond stereo chemistry. Thus, stereochemistry allows cis double bonds to be distinguished from their trans counterparts, and R-enantiomers to be distinguished from S at all chiral centers.

Table 1: Fragment Distinction

Fragment Patterns

It has been asserted that HQSAR works by identifying patterns of substructural fragments relevant to biological activity in sets of bioactive molecules. The reason for employing fragment patterns is that, unless only one atom fragments are considered, each atom in a molecule will be found in a number of different fragments, incrementing a number of holographic bins. Use of fragment patterns also distinguishes HQSAR from the Free-Wilson analysis which considers only independent fragments for which PLS is generally inappropriate. The use of fragment patterns is also analogous to the situation for CoMFA where correlations between lattice point energies play an important role—it is the pattern of energies which is important, not the individual values.

Results and Discussion

In this study, HQSAR analysis was performed on seven different sets of compounds, examining nine types of biological activity. Four of these data sets had previously been studied with CoMFA and three with other QSAR methods. HQSAR runs were performed over twelve different hologram lengths with four different hologram construction parameters settings. This resulted in 48 individual HQSAR experiments for each activity type. The data sets are described below along with the relevant publications.

DATA SETS

IA	42 Benzodiazepines, binding affinities at Diazepam-sensitive (DS) benzodiazepine receptor ^{7,16,17} .
IB	42 Benzodiazepines, binding affinities at Diazepam-insensitive (DI) benzodiazepine receptor ^{7,16,17} .
IC	42 Benzodiazepines, selectivity between Diazepam insensitive and sensitive benzodiazepine receptors (DI/DS) ^{7,16,17} .
II	50 Estrogen related steroids, inhibition of aromatase ⁸ .
III	28 Non-peptide angiotensin II receptor antagonists ⁹ .
IV	21 Steroids, binding affinity at corticosteroid binding protein ⁴ .
V	40 1-[(2-hydroxyethoxy)-methyl]-6-[phenylthio]thymine derivatives, inhibition of HIV-I reverse transcriptase ¹⁸ .
VI	13 Peptide and non-peptides leukotriene antagonists ¹⁹ .
VII	37 bisamidines; activity against Leishmania mexicana amazonensis ²⁰ .

HQSAR Parameters

HQSAR analysis was performed over the default series of hologram lengths of 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, 401 bins. Only fragments between 4 and 7 were considered.

Comparison of HQSAR and CoMFA

A number of data sets were selected from the literature, and HQSAR analyses performed on the 2D structures and their associated biological data as published. The CoMFA study results were not reproduced but rather are listed below as from their respective publications.

HQSAR is able to produce models of comparable predictive quality to those from CoMFA as clearly shown by the results in Figure 4 and Table I. In general, CoMFA tends to give better r^2 values. Model predictivity, in terms of q^2 and crossvalidated standard error of estimate (Std Err(cv)), is, however, quite comparable.

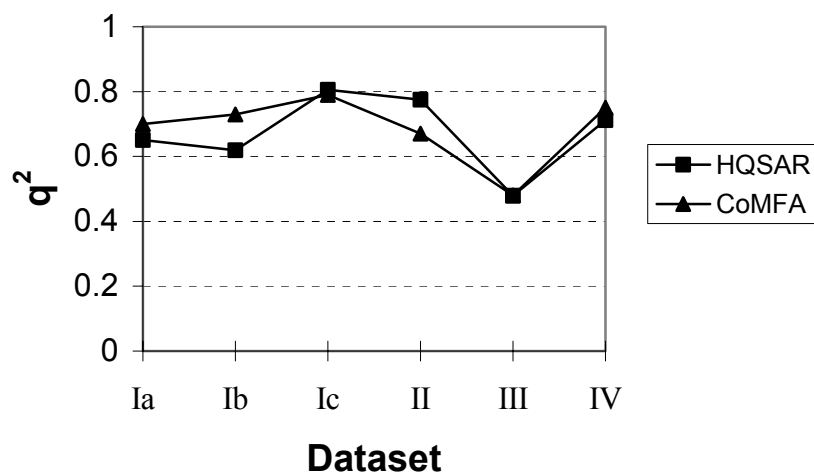


Figure 4: Comparison of q^2 values from HQSAR and CoMFA. The best results from the full set of HQSAR runs was reported.

Dataset	q^2		r^2 ^{Std}		Err(cv)		Std Err	
	HQSAR	CoMFA	HQSAR	CoMFA	HQSAR	CoMFA	HQSAR	CoMFA
Ia	0.65	0.70 ⁷	0.90	0.99	0.62	0.59	0.34	0.09
Ib	0.62	0.73 ⁷	0.92	0.96	0.55	0.45	0.55	0.45
Ic	0.81	0.79 ⁷	0.97	0.98	0.56	0.56	0.24	0.18
II	0.78	0.67 ⁸	0.89	0.89	0.70	NA	0.42	NA
III	0.48	0.48 ⁹	0.85	0.93	0.81	0.75	0.43	0.28
IV	0.71	0.75 ¹⁰	0.85	0.96	0.81	0.66	0.43	0.27

Table 2: Comparison of HQSAR and other QSAR models. The best results from the full set of HQSAR runs was reported.

Further HQSAR Comparisons

HQSAR results were also compared to results from other QSAR methods. The number of tests sets is too small to represent a survey of the literature, but these were chosen at random from a literature search. In general, HQSAR outperforms most other QSAR methods.

Data set V: Hansch analysis employing MR, hydration energy, and atomic charge

Data set VI: Apex 3D

Data set VII: Flexible fitting and use of molecular similarity indices

Data displayed in figure 5 and table 2 demonstrate that, at least for the data sets considered, HQSAR gives more predictive models (q^2 , Std Err(cv)) with better correlation with biological data (r^2).

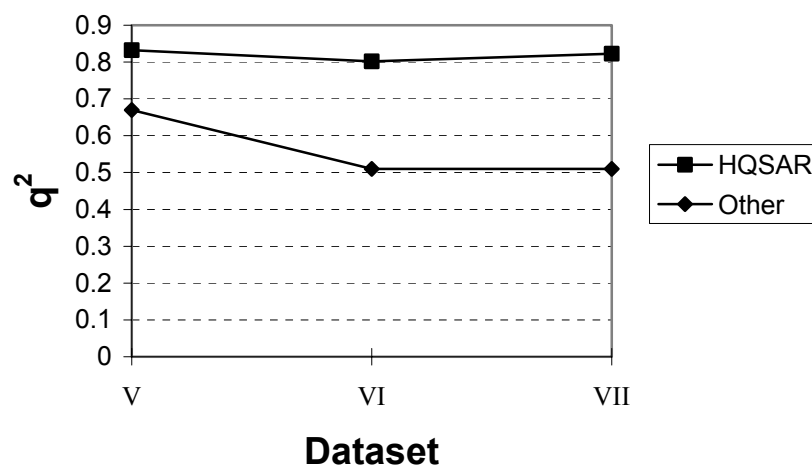


Figure 5: Comparison of q^2 values from HQSAR and other QSAR methods. The best results from the full set of HQSAR runs was reported.

Dataset	q^2		r^2 Std		Err(cv)		Std Err	
	HQSAR	Lit	HQSAR	Lit	HQSAR	Lit	HQSAR	Lit
V	0.83	0.67 ¹⁸	0.99	0.87	0.79	0.74	0.18	NA
VI	0.80	0.51 ¹⁹	0.99	0.87	0.64	0.51	0.17	0.44
VII	0.82	0.51 ²⁰	0.94	0.69	0.25	0.32	0.14	NA

Table 2: Comparison of HQSAR and other QSAR models. The best results from the full set of HQSAR runs was reported.

Validity of HQSAR Models

Figure 6 shows the results of randomization testing on the benzodiazepine selectivity data set (Ic). 1000 HQSAR runs were performed in which the biological data was randomized with respect to the training set compounds. Randomization of data and subsequent model evaluation is performed to assess the statistical validity of the QSAR model, eliminating the prospect of a chance correlation between descriptor variables and biological activity. Randomization tests were performed for all data sets in this study and all gave results qualitatively similar to that shown in Figure 6.

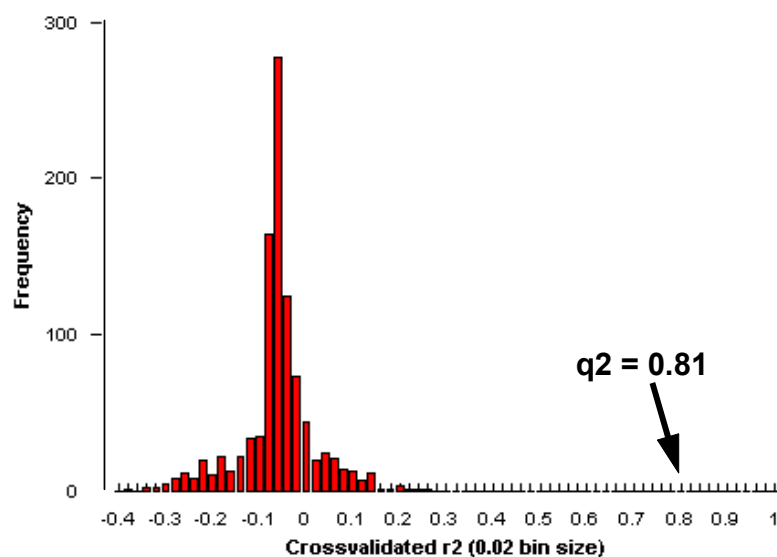


Figure 6: Histogram of q^2 vs. frequency of occurrence for 1000 HQSAR runs with randomized biological data.

Effect of Hologram Parameters

It was asserted in the theory section that employing different hologram lengths would affect HQSAR models through alteration of the fragment disposition in the holograms (changing the hologram lengths also changes the pattern of fragment collisions). Variation in q^2 values for a variety of hologram lengths is shown in Figure 7. Hologram construction parameters were set at atoms and bonds only for the generation of this data.

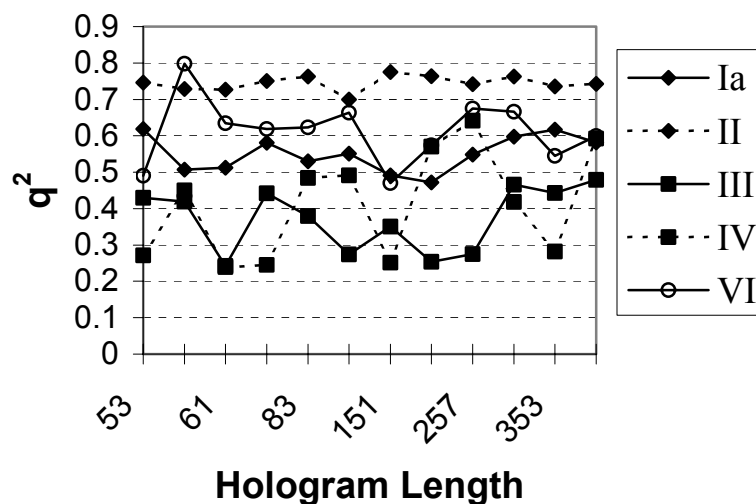


Figure 7: Variation in HQSAR q^2 as a function of Hologram Length.

It appears from the plot in figure 7 that there is no direct correlation between hologram length and the predictive quality of the HQSAR model produced. This situation was expected as the pattern of bin occupancies in the molecular hologram will change in a non-trivial manner with respect to alteration of hologram length. It is evident, however, that for some hologram lengths, PLS is able to obtain much stronger correlations between the fragment types present in the data set and the biological activity. It should be noted that for all hologram lengths some kind of predictive model results. Additionally, figure 7 suggests that careful consideration should be given when reporting HQSAR q^2 as values can vary in a significant manner depending on HQSAR parameter settings. For the datasets shown in Figure 7, the following ranges of q^2 have been observed. Data set I $q^2 = 0.47-0.62$, Data set II $q^2 = 0.70-0.76$, Data set III $q^2 = 0.24-0.48$, Data set IV $q^2 = 0.24-0.64$ and Data set VI $q^2 = 0.47-0.80$. It is suggested that the most reasonable HQSAR q^2 value to report is the median value. q^2 's reported here earlier have been the maximum values to provide a direct comparison with reported q^2 's from other QSAR techniques since q^2 from other QSAR techniques under comparison are reported either for the best model from various ones published or the only reported value (authors are assumed to have reported the highest q^2 obtained for the best model).

The effect of the information content of the molecular holograms on HQSAR models was also investigated by altering the fragment distinction parameters (atoms, bonds, connections, hydrogens, and chirality) prior to hologram generation (figure 8).

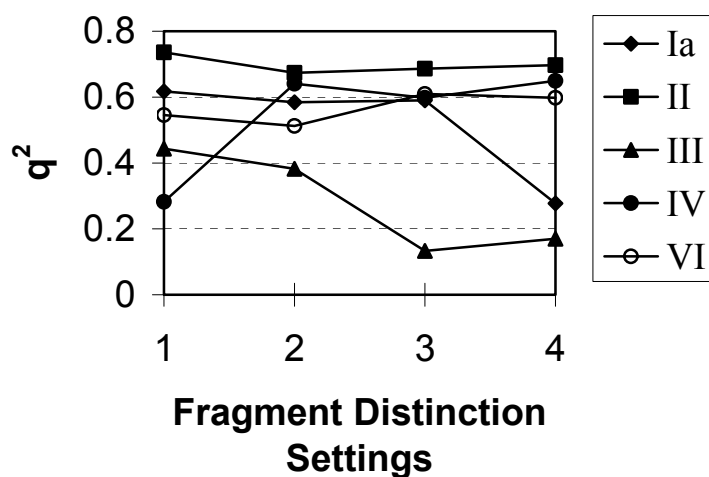


Figure 8: Variation in HQSAR q^2 vs. Fragment Distinction parameters.

Hologram Construction Parameters for Figure 8:

1 = Atoms & Bonds

2 = Atoms, Bonds & Connections

3 = Atoms, Bonds, Connections & Hydrogens

4 = Atoms, Bonds, Connections, Hydrogens & Chirality.

As was the case with variation in hologram length, the predictive quality of HQSAR models does not appear to depend on the information content of the holograms in any simple way. Alteration of the manner in which molecular fragments are defined will, as in the case of changing hologram length, disperse fragments differently over the holograms and also alter the information content. In most cases the setting of only atoms and bonds parameters appears sufficient for a predictive model to result. Depending, however, on the data set under study, it may be pertinent to use connections, hydrogens and chirality parameters to involve more molecular information. Further investigation into the incorporation of hybridization and chirality information is necessary.

Data Sets	TIMINGS (Sec)			
	(Parameters A=atoms, B=bonds, C=connections, H=Hydrogens, Ch=Chirality)			
	A-B	A-B-C	A-B-C-H	A-B-C-H-Ch
Ia	44	65	125	180
II	67	103	365	581
III	19	32	77	118
IV	14	31	75	208
V	37	48	82	115
VI	7	12	36	62
VII	30	39	85	136

Table 3: Timings for HQSAR runs.

Timings of HQSAR Runs

Table 3 shows the speed with which HQSAR can obtain QSAR models, the results of which are shown in Tables 1 and 2 and Figures 5 and 6. The only overhead in generating HQSAR models is the time required to enter the biological and molecular data.

All runs were made over the 12 default hologram lengths on an SGI O2 R10000 with 128Mb of memory. The starting point for each run was a SYBYL table containing the training set molecules and biological data, the finish point was the selection of the final model on the basis of standard error of estimate (Std Err (cv)) from the 12 possibilities.

CONCLUSION

HQSAR is a rapid, highly predictive QSAR technique. Results described earlier show that HQSAR can readily produce highly predictive QSAR models over a wide variety of data sets. In terms of q^2 values, the predictivity of HQSAR models is comparable to those derived from CoMFA and better than the other methods under comparison such as Apex3D and Hansch analysis for the data sets considered. Most importantly, HQSAR allows very rapid generation of QSAR models making it applicable to both small and large data sets.

Work is on-going to further develop HQSAR for virtual screening of databases. HQSAR's reliance on molecular holograms allows ready extension to database searching to find structurally similar highly active molecules.

REFERENCES

1. Van de Waterbeemd H., "The History of Drug Research: from Hansch to the Present", *Quant. Struct. Act. Relat.*, 11, 200, 1992.
2. Van de Waterbeemd H., "Recent Progress in QSAR Technology". *Drug Des. Discov.*, 9, 277, 1993.

3. Hansch C. and Fujita T., "Pi-Sigma-Rho analysis. A Method for the Correlation of Biological Activity and Chemical Structure.", *J. Am. Chem. Soc.*, 86, 1616, 1964.
4. Cramer R. D. III Patterson D. E. and Bunce J. D., "Comparative Molecular Field Analysis (CoMFA). 1. Effect of shape on the Binding of Steroids to Carrier Proteins.", *J. Am. Chem. Soc.*, 110, 5959, 1988).
5. Livingstone D. J., "Structure Property Correlations in Drug Research, Chapter 4", 81 Ed. Han van de Waterbeemd, R. G. Landes Company, Austin, 1996.
6. Topliss J. G. and Edwards R. P., "Chance Factors in Studies of Quantitative Structure-Activity Relationships.", *J. Med. Chem.*, 22, 1238 1979.
7. Wong G., Koehler K. F., Skolnick P., Gu Z., Ananthan S., Schonholzer P., Humkeler W., Zhang W. and Cook J. M., "Synthetic and Computer-Assisted Analysis of the Structural Requirements for Selective, High Affinity Ligand Binding to Diazepam Insensitive Benzodiazepine Receptors." *J. Med. Chem.*, 36, 1820 1993.
8. Oprea T. I., and Garcia A. E., "Three-Dimensional Quantitative Structure-Activity Relationships of Steroid Aromatase Inhibitors.", *J. Comp. Aid. Mol. Des.*, 10, 186, 1996.
9. Belvisi L., Bravi G., Catalane G., Mabilia M., Salimbeni A. and Scolastico C., "A 3D QSAR CoMFA Study of Non-peptide Angiotensin II Receptor Antagonists.", *J. Comp. Aid. Mol. Des.*, 10, 567, 1996.
10. Kellogg G. E., Kier L. B., Gaillard P. and Hall L. H., "E-state Fields: Applications to 3D QSAR." *J. Comp. Aid. Mol. Des.*, 10, 513, 1996.
11. Wold S., Abano C. and Dunn W. J., "Multivariate Data Analysis in Chemistry", in *Chemometrics: Mathematics and Statistics in Chemistry.*, Ed. B. Kowalski, Reidel, Netherlands, 1984.
12. van de Waterbeemd H., "Structure Property Correlations in Drug Research, Chapter 3", 81 Ed. Han van de Waterbeemd, R. G. Landes Company, Austin, 1996.
13. Shemetulskis H. E., Weininger D., Blankley C. J., Yang J. J. and Humblet C., "Stigmata: an Algorithm to Determine Structural Commonalities in Diverse Data sets.", *J. Chem. Inf. Comp. Sci.*, 36, 862, 1996.
14. Ash S., Cline M. A. Homer R. W., Hurst T. and Smith G. B., "SYBYL Line Notation (SLN): A Versatile Language for Chemical Structure Representation.", *J. Chem. Inf. Comp. Sci.*, 37, 71 1997.
15. Knuth D. E., "Sorting and Searching.", Addison-Wesley, reading, MA, 1973.
16. Wong G. and Skolnick P., "High Affinity Ligands for Diazepam-Insensitive Benzodiazepine Receptors.", *Eur. J. Pharm. Sec.*, 225, 63, 1992.
17. Gu Z., Wong G., Dominguez C. de Costa B. R., Rice K.C. and Skolnick P., "Synthesis and Evaluation of Imidazo[1,5-a][1,4]-benzodiazepine Esters with High Affinities and Selectivities at Diazepam Insensitive (DI) Benzodiazepine Receptors.", *J. Med. Chem.*, 36 479, 1993.
18. Hannongbua S., Lawtrakul L. and Limtrakul J., "Structure-Activity Correlation Study of HIV-I Inhibitors: Electronic and Molecular Parameters.", *J. Comp. Aid. Mol. Des.*, 10, 145, 1996.

19. Hariprasad V. and Kulkarni V. M., "A Proposed Common Spatial Pharmacophore and the Corresponding Active Site Conformations of Some Peptide Leukotriene Receptor Antagonists.", *J. Comp. Aid. Mol. Des.*, 10, 284, 1996.
20. Montanari C. A., Tute M. S., Beezer A. E. and Mitchell J. C., "Determination of Receptor-Bound Drug Conformations by QSAR Using flexible Fitting to Derive a Molecular Similarity Index.", *J. Comp. Aid. Mol. Des.*, 10, 67, 1996.
21. Hurst T., and Heritage T. "HQSAR - A Highly Predictive QSAR Technique Based on Molecular Holograms", 213th ACS Natl. Meeting, San Francisco, CA, 1997, CINF 019.
22. Wold S., Albano C., Dunn W.J. III, Edlund U., Esbensen K., Geladi P., Hellberg S., Johansson E., Lindberg W., Sjostrom M., "Multivariate Data Analysis in Chemistry", in *CHEMOMETRICS: Mathematics and Statistics in Chemistry*, B. Kowalski, Reidel, Dordrecht, Netherlands, (1984).