First International Electronic Conference on Synthetic Organic Chemistry (ECSOC-1), www.mdpi.org/ecsoc/, September 1-30, 1997

[F0003]

HQSAR

A New, Highly Predictive QSAR Technique

Thorsten Naumann*, David Lowis**

Tripos

* Tripos GmbH, Martin Kollar Str. 13, 81829 München, Germany.
E-mail: tnaumann@tripos.com

** Tripos Inc., 1699 S. Hanley Road, St. Louis, MO 63144-29143, USA

Received: 15 July 1997 / Uploaded: 29 August 1997

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 specialised fragment fingerprints (molecular holograms) as predictive variables of biological activity or other structural related data. By removing the necessity for molecular alignment, models by HQSAR can be obtained more rapidly

than other by 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 and allow extension to database searching capability.

HQSAR

HQSAR is a new QSAR technique that avoids many of the problems associated with classical or 3D QSAR approaches[1]. Only 2D structures and activity are required as input — no selection of descriptors or 3D molecular alignment is needed. HQSAR converts the molecules of a data set into counts of their constituent fragments. These fragment counts are then related to biological data using Partial Least Squares (PLS) analysis. Both steps, fragment counting and PLS analysis, are very fast. Nevertheless, the method is robust and highly predictive for many data sets. The details of the method will be published elsewhere.

In this study[2], 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. The data sets are described below along with the relevant publications. We also calculated CLOGP values for a data set of 1152 compounds, some of the compounds could not be calculated by CLOGP. We used a QSAR model derived by HQSAR to predict the remaining compounds.

Data Sets

Ia	42 Benzodiazepines, binding affinities at Diazepam-sensitive (DS) benzodiazepine receptor[3,4,5] ,,.
Ib	42 Benzodiazepines, binding affinities at Diazepam-insensitive (DI) benzodiazepine receptor[3,4,5] .
Ic	42 Benzodiazepines, selectivity between Diazepam insensitive and sensitive benzodiazepine receptors[3,4,5] (DI/DS)
II	50 Estrogen related steroids, inhibition of aromatase[6] .
III	28 Non-peptide angiotensin II receptor antagonists[7].
IV	21 Steroids, binding affinity at corticosteroid binding protein[8].
V	40 1-[(2-hydroxyethoxy)-methyl]-6-[phenylthio]thymine derivatives, inhibition of HIV-I reverse transcriptase[9].
VI	13 Peptide and non-peptides leukotriene antagonists[10].
VII	37 bisamidines, activity against *Leishmania mexicana amazonesis[11]*.

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.

Table 1: Comparison of HQSAR and CoMFA QSAR models.

	q2		**r2**		**Std Err(cv)**		**Std Err**	
Data set	**HQSAR**	CoMFA	**HQSAR**	CoMFA	**HQSAR**	CoMFA	**HQSAR**	CoMFA
Ia	**0.65**	0.70	**0.90**	0.99	**0.62**	0.59	**0.34**	0.09
Ib	**0.62**	0.73	**0.92**	0.96	**0.25**	0.18	**0.55**	0.45
Ic	**0.81**	0.79	**0.97**	0.98	**0.56**	0.56	**0.24**	0.18
II	**0.78**	0.67	**0.89**	0.89	**0.70**	NA	**0.42**	NA
III	**0.48**	0.48	**0.85**	0.93	**0.81**	0.75	**0.43**	0.28
IV	**0.71**	0.75	**0.85**	0.96	**0.81**	0.66	**0.43**	0.27

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

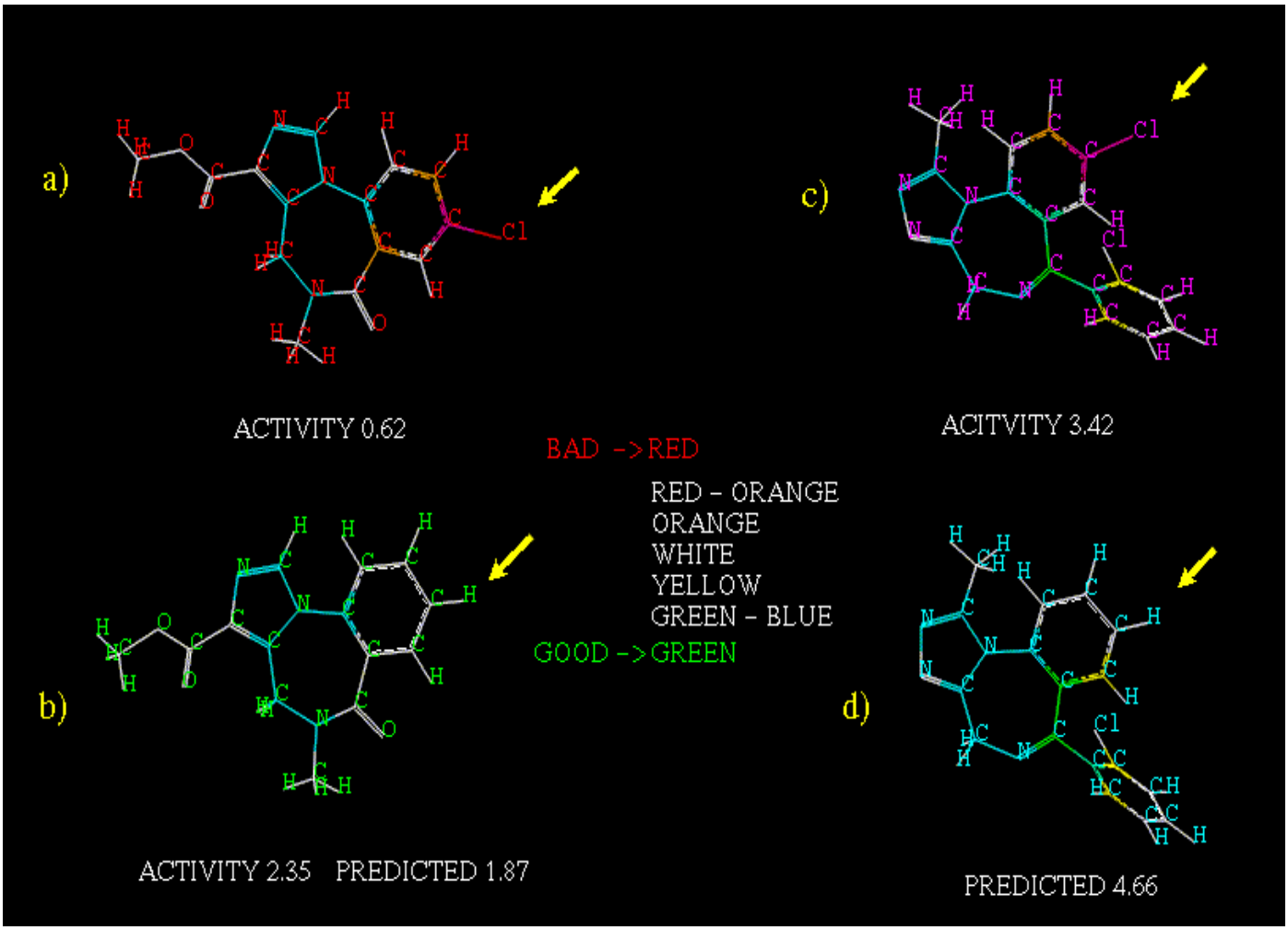
Data set VI: Apex 3D

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

Table 2: Comparison of HQSAR and other QSAR models.

	q_2		**r_2**		**Std Err(cv)**		**Std Err**	
Data set	**HQSAR**	Lit	**HQSAR**	Lit	**HQSAR**	Lit	**HQSAR**	Lit
V	**0.83**	0.67	**0.99**	0.87	**0.70**	0.74	**0.18**	NA
VI	**0.80**	0.51	**0.99**	0.87	**0.64**	0.51	**0.17**	0.44
VII	**0.82**	0.51	**0.94**	0.69	**0.25**	0.32	**0.14**	NA

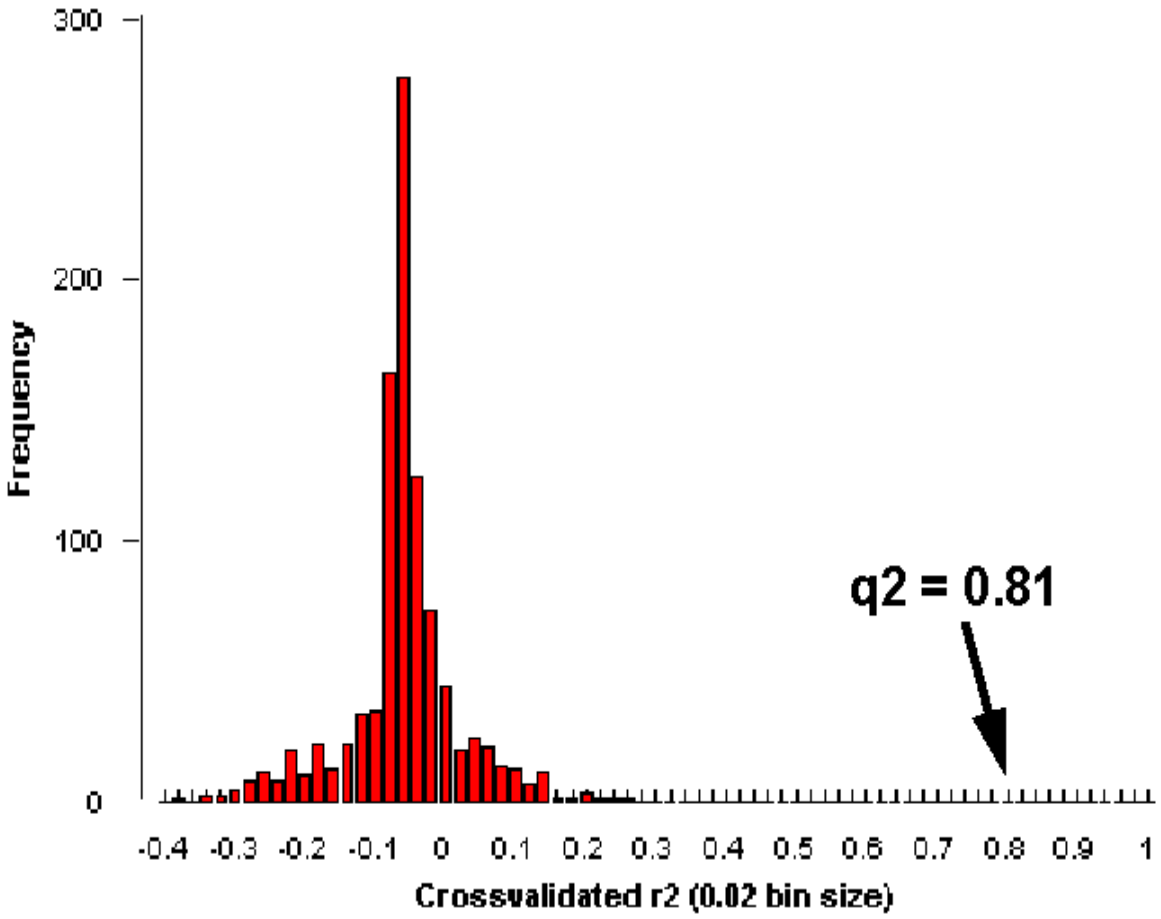
Figure 1 shows the contribution graph of compounds of the Ic data set. As one can see on the colorcoding the Cl in the compounds a) and c) has a negative contribution. Compound a), b) and c) are included in the dataset. We changed the Cl to a Hydrogen and predict the new activity with the model derived by HQSAR. The compound d) is not included in the data set if one changes the negative fragment to a neutral or positv the activity increases



Validity of HQSAR Models

Figure 2 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 2.

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



Prediction of non biological data

HQSAR can also be used to predict other data related to the 2D structure of a molecule. We used a data set of 1152 molecules. This data set represents biologically active compounds (77 biological classes) of the IndexChemicus 93 database (IC93)[12]. We used the data set to calculate CLOGP.The program was able to calculate values for 77 % of the structures in data set. In order to complete the data information we used HQSAR to predict the LOGP values of all not calculated compounds.

Results

884 compounds out of 1152 were used to calculate a QSAR model to predict LOGP values. We used the HQSAR standard settings to perform the calculation.

Hologram Lengths : 53, 59, 61, 71, 83, 97, 151, 199

Atom Counts in Fragments : Min = 4 ; Max = 7

Atoms: on

Bonds: on

Connections: on

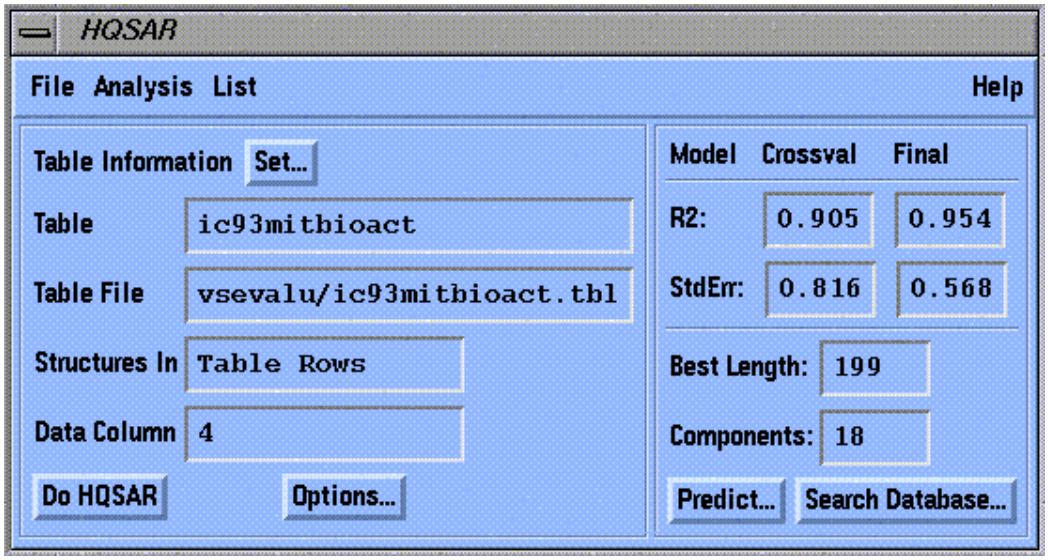


Figure 3. HQSAR interface; shows actual input data and the results of the calculation.

As one can see the cross validated R2 value is 0.905, which means that the model is 90 % predictive. The standard error of 0.816 (cross val.) is also very good. In figure 3 the comparison between the predicted and the actual value (calculated LogP) is shown.

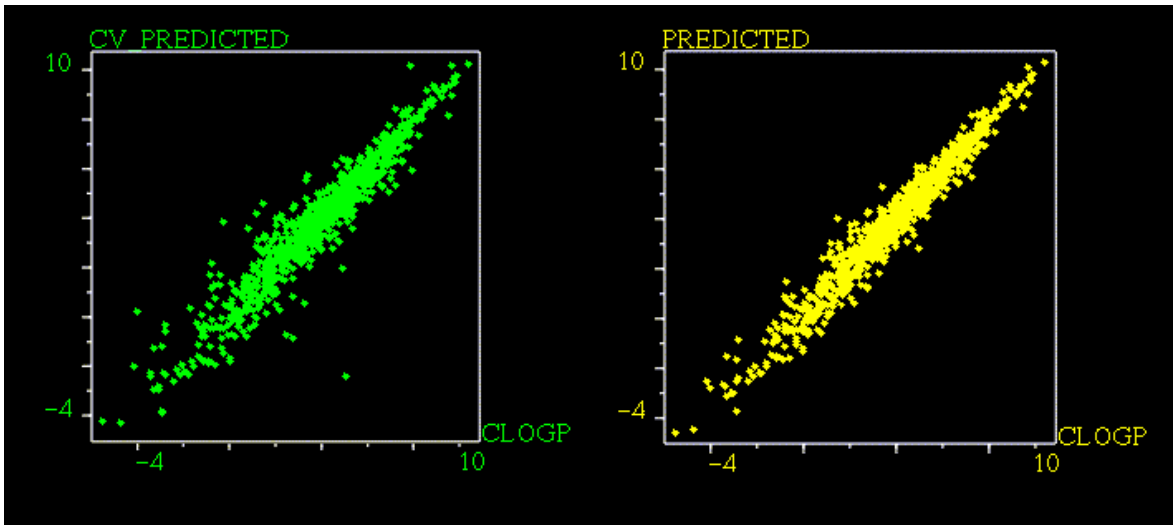
Figure 4 List of the actual and the predicted values (cross validated and for the whole data set)

HQSAR Predicted Values

Row	Actual	Predicted	Crossval
2	2.08000	0.83802	0.50738
4	0.01000	-1.35446	-1.68896
5	1.61000	0.51798	0.14370
11	2.07000	1.59955	1.54833
17	5.55000	5.59631	4.96676
18	5.09000	5.13686	5.11266
19	4.65000	5.40350	6.16716
20	6.15000	5.50864	5.49572
21	3.61000	3.78106	3.81914
22	4.17000	4.49248	4.55339
23	4.73000	5.00834	5.04123
24	4.69000	5.05883	5.18425
25	4.69000	5.02686	5.11516
26	4.69000	4.93514	4.95052
27	5.21000	5.20052	5.22224
28	5.21000	5.10127	5.10681
29	3.56000	3.68380	3.69351
30	5.48000	5.53054	5.53905

Close

Figure 5. Graph of the predicted vs. Actual (cross val. and whole data set)



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 q2 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.

Refrences

[1] 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.

[2] David Lowis Tripos Inc "Tripos Technical Notes"1997

[3] 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.

[4] Wong G. and Skolnick P., "High Affinity Ligands for Diazepam-Insensitive Benzodiazepine Receptors.", *Eur. J. Pharm. Sec.*, **225**, 63, 1992.

[5] 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.

[6] Oprea T. I., and Garcia A. E., "Three-Dimensional Quantitive Structure-Activity Relationships of Steroid Aromatase Inhibitors.", *J. Comp. Aid. Mol. Des.*, **10**, 186, 1996

[7] 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

[8] 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, 188).

[9] 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.

[10] 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.

[11] 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.

[12] Index Chemicus – Subset from 1993; Institute of Scientific Information, Inc. (ISI), 3501 Marked St., Philadelphia PA 19104

Comments

During 1-30 September 1997, all comments on this poster should be sent by e-mail to ecsoc@listserv.arizona.edu with **F0003** as the message subject of your e-mail. After the conference, please send all the comments and reprints requests to the author(s).
