Graham Richards

STRUCTURE ACTIVITY RELATIONSHIPS FROM MOLECULAR SIMILARITY MATRICES

Andrew C. Good, Sung-Sau So and W. Graham Richards*.

Physical Chemistry Laboratory, Oxford University, South Parks Road, Oxford OX1 3QZ, United Kingdom.

Abstract

An alternative method for determining structure activity correlations is presented. Ligand molecules are described using data matrices derived from the results of N by N (each molecule compared to every other) molecular similarity calculations. The matrices were analysed using a neural network pattern recognition technique and partial least squares statistics, with the results obtained compared to those achieved using comparative molecular field analysis (CoMFA). The molecular series used in the study comprised 31 steroids. The resultant pattern recognition analysis showed clustering of compounds with high, intermediate and low affinity into separate regions of the neuron output plots. The cross-validated correlation coefficients obtained from statistical analyses of the matrices against steroid binding data compared well with those achieved using CoMFA. These results show that data matrices derived from molecular similarity calculations can provide the basis for rapid elucidation of both qualitative and quantitative structure activity relationships .

Introduction

It is generally thought that non covalent forces dominate receptor drug interactions, and that these forces can be described in terms of steric and electrostatic effects. In an attempt to relate these effects to observed biological data, a number of approaches have been presented[1-4]. Among these, Hopfinger[2-3] described a method for comparing net differences in fields and volumes. Cramer et al[4] proposed the now widely applied technique[5] of comparative molecular field analysis (CoMFA), which uses partial least square statistics[6-7] to analyse the steric and electrostatic fields lattices of ligand molecules.

Molecular similarity calculations have become established as a method both for generating parameters in structure activity relationships, and in optimizing structure superposition[8-20]. As originally introduced by Carbo[8-9],

$$R_{AB} = \frac{\int P_A P_B dV}{(\int P_A^2 dV)^{1/2} (\int P_B^2 dV)^{1/2}}$$

molecular similarity R_{AB} is determined from the structural properties P_A and P_B of the two molecules being compared. The numerator measures property overlap while the denominator normalizes the similarity result. Electron density, electrostatic potential and shape have all been used as the structural property P.

The Carbo index has been used in a previously published QSAR study[13] in which electrostatic potential similarity calculations were executed against a lead molecule and the results correlated with activity. This technique is similar to that applied by Hopfinger in that he also used the results from comparisons with a single reference molecule when undertaking QSAR determination.

In this paper we propose an approach somewhere between those of Cramer and Hopfinger. It is clear that, as with Hopfinger, the results obtained from steric and electrostatic comparisons using the Carbo index measure the net effect of similarities between molecules. However, by

comparing each molecule to every other in a series (N by N calculation), a similarity matrix is formed which provides a numerical representation of how all the molecules interrelate. The matrix thus implicitly introduces some of the location dependence of the steric and electrostatic parameters used within CoMFA.

The methodology was tested using the 31 steroids utilized in the original CoMFA study[4] for which testosterone-binding globulin (TBG) and/or corticosteroid-binding globulin (CBG) binding affinities were known[21-23].

Model Building

Steroid models were built from the co-ordinates of related structures in the Cambridge Crystallographic Database[24]. Substituents (hydroxyl groups, acetyl groups etc.) were added using the modelling program CHEM-X[25] as required, ensuring that the starting geometry of any given substituent was identical across all relevant steroids. At this point each steroid was minimized within CHEM-X using the default force field. AM1 MOPAC[26] calculations were then undertaken from which point charges were back calculated to fit the consequent molecular electrostatic potentials[27]. Finally the resultant structures were superimposed by a least squares fit of the 3, 5, 6, 13, 14 and 17 carbon atoms. Binding data for the 31 steroids are shown in Table I. The steroid structures used are shown in Figure 1 (structures available in CSSR format from the authors).

Experimental

In this work we use a modified version of the ASP software[14], utilizing Gaussian functions rather than grids to elucidate electrostatic potential and shape similarity[19-20]. Guassians used to estimate the inverse distance (1/r) dependence of point charge potentials were applied to measure electrostatic similarity. For this study the 3 Gaussian approximation to the 1/r curve was used[19]. Gaussians representing the square of STO3G[28] wave functions were used to measure shape similarity. Gaussians constrained to drop to zero outside the atomic van der

Waals radii[20] (simulating the "hardness" of bonded atoms) were used here, since they are most sensitive to structural differences. These functions permit rapid analytical integral evaluations, greatly enhancing the speed of similarity calculations. Shape and electrostatic potential similarity calculations have been incorporated into the same program and results matrices are generated automatically. 31 by 31 similarity matrices were calculated for steroid shape (SHAPE) and electrostatic potential (ESP). A 31 by 62 matrix was also created containing both shape and electrostatic potential similarity data (SHESP). 178 seconds CPU time were required on a Silicon Graphics Iris 4D-20 to calculate the 31 by 31 matrices for both shape and electrostatic potential (a total of 930 similarity calculations).

While the extra information provided through the use of these similarity matrices is extremely useful, the amount of data present also causes problems. As with CoMFA, the data sets created from N by N similarity calculations are underdetermined (more columns than rows). The dimension reduction techniques of neural networks and partial least squares statistics have therefore been employed to analyse the matrix data sets.

Study 1 - Graphical Analysis Using Neural Networks

For the first investigation the matrix data was analysed qualitatively using a neural network. Neural networks are hardware or software systems attempting to simulate some functions of the brain[29], including the problem of pattern recognition. For these studies we have chosen to use the reversible non-linear dimensionality reduction (ReNDeR) procedure utilized by Livingstone et al[30]. This method uses a symmetric neural network of the kind shown in Figure 2. The input and output layers contain as many neurons as there are parameters. The encoding and decoding layers contain a smaller number of neurons (around a third the number of input neurons works well), while the central parameter layer contains the number of neurons equal to the dimensionality of the ReNDeR plot required (usually 2). The output from the neurons within the parameter layer are used as the co-ordinates of the ReNDeR plot.

For this experiment neural network software developed in house was used. A 31-10-2-10-31 network was used to analyse the shape and electrostatic potential matrices, while a 62-20-2-20-62 system was applied to the combined matrix. Each matrix was analysed for around 2-3 hours on a PC-486 until the decrease in rms fit error was negligible (less time required relative to the original work of Livingstone[30] et al because of the incorporation of a momentum[29] term into the neural network. This greatly increases convergence speed). The results of the neural network analyses are shown in Figures 3a-f. The definitions of high, intermediate and low affinity compounds are shown in Table I.

Study 2 - QSAR Analysis Using Partial Least Squares

In the second investigation PLS statistics were used to analyse the similarity matrices quantitatively. The SYBYL modelling package[31] implementation of PLS[7] was used for the study. The steroid structures 1-21 were used in creating the QSAR equations for both the CBG and TBG studies, as was the case for the original CoMFA study[4]. The equations derived using the CBG affinity data were then used to predict the activities of steroids 22-31, again in the same vein as the original study.

PLS runs were undertaken using shape and electrostatic potential similarity data both individually (SHAPE and ESP) and together (SHESP). The similarity matrix containing both shape and electrostatic potential data was automatically loaded into a SYBYL data table using the SYBYL PROGRAMMING LANGUAGE[31]. The matrix was added to the table without data scaling. All cross-validated runs were undertaken using 21 cross-validation groups and 5 components. A cross-validated PLS run of the full (SHESP) similarity matrix took 5 seconds to run on an Iris 4D-20.

CoMFA runs were also performed on the 21 steroids. The CoMFA grid spacing used was 2.0Å for all but two runs when a 1.0Å spacing was applied. An sp^3 carbon atom with a +1 charge was used as the probe atom. The energy cut-off was set to 30 kcal mol^{-1}. The MINIMUM_SIGMA (set to 2.0) and COMFA_STD options were also enabled. All cross-

validated runs were undertaken using 21 cross-validation groups and 5 components. All other variables were left as the SYBYL defaults. PLS runs were initiated for four separate CoMFA runs. These runs were made up as follows: (i) analysis of CoMFA electrostatic field only, (ii) analysis of CoMFA steric field only, (iii) analysis of CoMFA using both steric and electrostatic fields (full field), (iv) analysis of CoMFA using both steric and electrostatic fields with a 1.0Å grid spacing. The four analyses were used for both CBG and TBG correlations. A cross-validated full field CoMFA analysis PLS run took 20 seconds with a 2.0Å grid spacing, and 17 minutes with a 1.0Å spacing (increase by up to 2 orders of magnitude if the MINIMUM_SIGMA option is not used). The results of all the PLS runs are shown in Table II, together with the standard error produced when predicting the CBG affinity of the remaining steroids. The correlations producing the lowest cross-validated standard error were deemed to be the best and are shown here.

Results and discussion

The neuron plots from the electrostatic potential similarity matrix show excellent clustering of the high affinity compounds for both CBG (Figure 3a) and TBG binding (Figure 3b). Grouping of intermediate and low affinity compounds into bands across the plot can also be seen. The CBG plot for the shape similarity matrix (Figure 3c) shows tight clustering for the majority of active compounds, although there is no clear grouping (banding) of the intermediate and low affinity compounds away from the high affinity cluster. The TBG plot for shape (Figure 3d) shows little clustering of compounds according to affinity. The point distributions of the plots created using both shape and electrostatic potential data (Figures 3e-f) are similar to those derived from the electrostatic matrix. For CBG affinity (Figure 3e), the banding of compounds according to affinity appears somewhat more defined, but little discernible improvement can be seen for TBG affinity (Figure 3f).

The quantitative behavior of the PLS analyses closely mirrors the qualitative properties of the neuron plots. Table 2 shows the results of the PLS runs undertaken using the similarity

matrices and CoMFA. For CBG binding shape, both in similarity (SHAPE) and CoMFA form, produced the best cross-validated r^2 values. Correlation of electrostatic similarities (ESP) and CoMFA data was also good. Interestingly, combining the two properties did not produce better correlation, with both the full (SHESP) similarity matrix and CoMFA analysis yielding cross-validated r^2 values less than those achieved by shape alone. The SHESP matrix model description utilizes only 1 component and is close to that produced using just the electrostatic (ESP) matrix. The use of a 1Å grid density when generating CoMFA data marginally increased the cross-validated r^2. The errors and cross-validated r^2 values achieved by CoMFA were slightly better than those attained using similarity.

With regard to the predictions of steroids 22-31, little difference was seen between the two methods, with CoMFA electrostatics giving the lowest standard error of prediction for compounds 22-30. For these compounds the difference in predictive standard error achieved across all similarity and CoMFA PLS runs was only 0.215. None of the QSAR equations were able to closely predict steroid 31, every run overestimating its affinity. The CoMFA runs were found to overestimate the affinity of 31 slightly more than similarity, with a consequent relative increase in overall standard error. Again the full (SHESP) similarity matrix and CoMFA analysis did not produce the best results. It was also found that the use of a 1.0Å density grid decreased prediction accuracy.

For the TBG binding models, both the electrostatics only (ESP) similarity matrix and CoMFA run produced significantly higher cross-validated r^2 values than their shape counterparts. Once more the use of a full CoMFA field analysis, utilizing both a 1.0Å and 2.0Å grid separation, actually decreases cross-validated correlation (in this case lower than both shape only and electrostatics only CoMFA runs). However, the use the full (SHESP) similarity matrix results in an improved cross-validated r^2 while decreasing the number of components required for the model description. The model was still close to that derived using the ESP matrix alone. For the TBG series, similarity matrices produced better overall cross-validated r^2 values than those derived from CoMFA data.

In summary, the results obtained both in the form of a graphical SAR description and QSAR model using similarity matrices correlate well with observed affinity. QSAR models derived from the similarity matrices are comparable with those produced by the CoMFA technique both in terms of overall result and general behavior.

It is of interest that PLS appears unable to derive an improved model when analysing both steric and electrostatic field data from CoMFA. Using the COMFA_STD option and accurate charges leads to rather different results from those obtained in the original CoMFA study[7]. Both the full 2.0Å grid separation CoMFA analyses of CBG and TBG binding have electrostatic and steric factors contributing approximately equally (data obtained from CoMFA PLS listings[31]) to the QSAR models generated. In the original study no correlation is obtained with electrostatic field . This would at first appear to be a good result, since we would expect electrostatics to play some part in binding. However, when we consider this result in relation to those obtained when considering shape and electrostatics separately, PLS would seem a little "confused". This can be seen in the context of the TBG models created. For the TBG system, electrostatic factors are clearly more important (see both the neuron plots and cross-validated QSAR model r^2 values - Figures 3b / 3d and Table II), yet when shape is added to the CoMFA model it makes a near 50 percent contribution and the correlation decreases. The probable explanation for this is that the amount of data produced by CoMFA creates a lot of noise which PLS has trouble separating from signal. This problem becomes particularly acute when both steric and electrostatic factors are considered in tandem, because (a) extra variables increase noise as well as signal and (b) shape parameter variance tends to be larger than electrostatic parameter variance (a similar problem exists for similarity data, although here the shape variance is the lesser of the two). In an attempt to overcome this problem the MINIMUM_SIGMA and COMFA_STD options[31] have been created. The COMFA_STD facility scales all CoMFA columns so they are all treated equally. Unfortunately, because the only variable filtering undertaken is through the somewhat arbitrary MINIMUM_SIGMA option (removal of columns with a standard deviation lower than a user defined threshold),

large numbers of noise variables are still present in the CoMFA matrix. The use of the COMFA_STD option gives the noise equal importance to the signal, making it difficult for the PLS analysis to obtain the best possible result. This hypothesis is supported by the fact the cross-validated r^2 value actually drops when additional data is provided in the form of a 1Å density grid. Also the 1Å density grid analysis of CBG data, while producing a better cross-validated r^2, was found to be the worst of the models for predicting the affinities of compounds 22-31.

As a consequence of these problems, it would be useful to have access to more sophisticated statistical techniques for filtering the signal data from the noise and creating the QSAR model. To this end a test release of the statistics package Golpe[32] was used to analyse the similarity data and CBG affinity. Signal variables were filtered from the noise using progressive exclusion[32], with shape and electrostatic similarity data considered separately. The resulting signal variable data was combined and auto scaled. A preliminary PLS run was undertaken and the 1st component scores plotted against y (affinity) values. The resultant plot suggested a non linear relationship between the data and the CBG affinity. Quadratic PLS was thus used to fully analyse the data. The resultant cross-validated r^2 was 0.828 described in 3 components derived from just 14 shape and 9 electrostatic similarity variables. It was not possible to undertake similar work on the CoMFA data, since the version of Golpe used could not handle the missing values of the electrostatic data columns (points within the van der Waals radii). Nevertheless, the importance of variable filtering (and the availability of quadratic PLS) is evident, since notable improvements in model performance are possible even when analysing relatively small similarity matrices. It is almost certain therefore, that large CoMFA data sets could benefit greatly from this form of analysis.

Similarity matrices are thus shown to provide an interesting alternative method for generating SAR and QSAR models. The matrices are generated quickly, and the amount and nature of the data they contain allows rapid analysis with a relatively low noise to signal ratio. This is extremely useful in areas such as PLS, where analysis of large data sets can be a time

consuming process, and leads to possible extensions in the application of QSAR studies. For example cross-validated r^2 values could be treated as optimizable variables, with molecules used to derive the QSAR model (especially out-liers) shifting in order to maximize correlation.

Conclusions

N by N similarity matrices have been shown to provide a powerful numerical representation of the steric and electrostatic relationships between a series of drug molecules. Matrix generation and analysis are rapid, with excellent qualitative and quantitative correlation found between similarity data and affinities for the steroid set used.

Acknowledgement

This work is supported through an SERC CASE award with British Biotechnology Ltd, Oxford, England.

REFERENCES

[1] Ghose, A.K.; Crippen, G.M.; Revankar, G.R.; Mckernan, P.A.; Smee, D.F.; Robins, R.K. Analysis of the Invitro Antiviral Activity of Certain Ribonucleosides Against Parainfluenza Virus Using a Novel Computer Aided Receptor Modeling Procedure. *J. Med. Chem.* **1989**, 32, 746-756.

[2] Hopfinger, A.J. A QSAR Investigation of DHFR Inhibition by Bakers Triazines Based Upon Molecular Shape Analysis. *J. Am. Chem. Soc.* **1980**, 102, 7196-7206.

[3] Hopfinger, A.J. Theory and Analysis of Molecular Potential Energy Fields in Molecular Shape Analysis: A QSAR Study of 2,4-diamino5-benzylpyrimidines as DHFR Inhibitors. *J. Med. Chem.* **1983**, 26, 990-996.

[4] Cramer, R.D. III.; Patterson, D.E.; Bunce, J.D. Comparative Molecular Field Analysis (CoMFA). Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, 110, 5959-5967.

[5] Mcfarland, J.W. Comparative Molecular Field Analysis of Antococcidial Triazines. *J. Med. Chem.* **1992**, 35, 2543-2550. This paper also contains a number of references to other interesting articles which utilize the CoMFA technique.

[6] Dunn, W.J. III; Wold, S.; Edlund, U.; Hellberg, S.; Gasteiger J. Multivariate Structure Activity Relationships Between Data from Battery of Biological Tests and an Ensemble of Structure Descriptors. The PLS method. *Quant. Struct. Act. Relat.* **1984**, 3, 131-137.

[7] Cramer, R.D. III.; Bunce, J.D.; Patterson, D.E.; Frank, I.E. Crossvalidation, Bootstrapping, and Partial Least Squares Compared with Multiple Linear Regression in Conventional QSAR Studies. *Quant. Struct. Act. Relat.* 1988, 7, 18-25.

[8] Carbo, R.; Leyda, L; Arnau, M. An Electron Density Measure of the Similarity Between Two Compounds. *Int. J. Quantum Chem.* **1980**, 17, 1185-1189.

[9] Carbo, R; Domingo, L. LCAO-MO Similarity Measures and Taxonomy. *Int. J. Quantum Chem.* **1987**, 32, 517-545.

[10] Hodgkin, E.E; Richards, W.G. Molecular Similarity Based on Electrostatic Potential and Electric Field. *Int. J. Quantum Chem. Quantum Biol. Symp.* **1987**, 14, 105-110.

[11] Arteca, G.A.; Jammal, V.B.; Mezey, P.G. Shape Group Studies of Molecular Similarity and Regioselectivity in Chemical Reactions. *J. Comp. Chem.* **1988**, *9*, 608-619.

[12] Burt, C.; Richards, W.G. Molecular Similarity: The Introduction of Flexible Fitting. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 231-238.

[13] Burt, C.; Huxley, P.; Richards, W.G. The Application of Molecular Similarity Calculations. *J. Comp. Chem.* **1990**, 11, 1139-1146.

[14] Automated Similarity Package, Oxford Molecular Ltd, The Magdalen Centre, Oxford Science Park, Sandford on Thames, Oxford OX4 4GA, United Kingdom.

[15] Richard, A.M. Quantitative Comparison of Molecular Electrostatic Potentials for Structure Activity Studies. *J. Comp. Chem.* **1991**, 12 (8), 959-969.

[16] Manaut, M.; Sanz, F.; Jose, J.; Milesi, M. Automatic Search for Maximum Similarity Between Molecular Electrostatic Potential Distributions. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 371-380.

[17] Cioslowski, J.; Fleischmann E.D. Assessing Molecular Similarity from Results of ab Initio Electronic Structure Calculations. *J. Am. Chem. Soc.* **1991**, 113, 64-67.

[18] Meyer A. M.; Richards W.G. Similarity of Molecular Shape. *J. Comput.-Aided Mol. Design*, **1991**, 5, 426-439.

[19] Good A.C.; Hodgkin E.E.; Richards, W.G. The Utilization of Gaussian Functions for the Rapid Evaluation of Molecular Similarity. *J. Chem. Inf. Comput. Sci.* **1992**, 32, 188-191.

[20] Good A.C.; Richards, W.G. Rapid Evaluation of Shape Similarity Using Gaussian Functions. *J. Chem. Inf. Comput. Sci.*. in press.

[21] Dunn, J.F.; Nisula, B.C.; Rodbard, D. Transport of Steroid Hormones: Binding of 21 Endogenous Steroids to Both Testosterone-Binding Globulin and Corticosteroid-Binding Globulin in Human Plasma. *J. Clin. Endocrin. Metab.* **1981**, 53, 58-68.

[22] Mickelson, K.E.; Forsthoefel, J.; Westphal, U. Steroid Protein Interactions. Human Corticosteroid Binding Globulin: Physiochemical proprties and Binding Specificity. *Biochemistry*. **1981**, 20, 6211-6218.

[23] Westphal, U. *Steroid-Protein Interactions II*. Springer Verlag: Berlin **1986**.

[24] Allen, F.N.; Kennard, O; Taylor, R. Systematic Analysis of Structural Data as a Research Technique in Organic Chemistry. *Acc. Chem. Res.* **1983**, 146-153.

[25] CHEM-X, Chemical Design Ltd, Unit 12, 7 West Way, Oxford OX2 0JB, United Kingdom.

[26] MOPAC 5, Stewart, J.J.P. *Q.P.C.E. 455*.

[27] Ferenzcy, G., Reynolds, C.A. and Richards, W.G. Semi Empirical AM1 Electrostatic Potential and AM1 Electrostatic Potential Derived Charges, a Comparison with Ab initio Values. *J. Comp. Chem.* **1990**, 11, 159-159.

[28] Gaussian 88. Frisch, M.J.; Head, Gordon M.; Schlegel, H.B.; Raghavachari, K.; Binkley, J.S.; Gonzalez, C.; Defrees, D.J.; Fox, D.J.; Whiteside, R.A.; Seeger, R.; Melius C.F.; Baker, J.; Martin, R.; Kahn, L.R.; Stewart, J.J.P.; Fluder, E.M.; Topiol, S.; Pople, J.A.; Gaussian. Inc., Pittsburgh, PA, **1988**.

[29] Hertz, J.; Krogh, A.; Palmer, R.G. *Introduction to the Theory of Neural Computation*. Addison-Wesley Publishing Company, **1991**.

[30] Livingstone, D.J.; Hesketh, G.; Clayworth, D. Novel Method for the Display of Multivariate Data Using Neural Networks. *J. Mol. Graph.* **1991**, 9, 115-118.

[31] SYBYL, Tripos Associates Inc, 1699 S. Hanley Rd., Suite 303, St Louis, Missouri 63144.

[32] Baroni, M.; Constantino, G.; Cruciani, G.; Riganelli, R.; Valigi R.; Clementi, S. GOLPE: an Advanced Chemometric Tool for 3D-QSAR Problems. *Quant. Struct. Act. Relat.* in press.

Figure Legends

Figure 1 - Steroid structures used to determine SARs and QSARs.

Figure 2 - Schematic representation of ReNDeR neural network.

Figure 3 - Neuron plots produced by ReNDeR neural network:

(a) from electrostatic potential similarity matrix for CBG affinity.

(b) from electrostatic potential similarity matrix for TBG affinity.

(c) from shape similarity matrix for CBG affinity.

(d) from shape similarity matrix for TBG affinity.

(e) from combined similarity matrix for CBG affinity.

(f) from combined similarity matrix for TBG affinity.

Legend for Figure 3 plots -

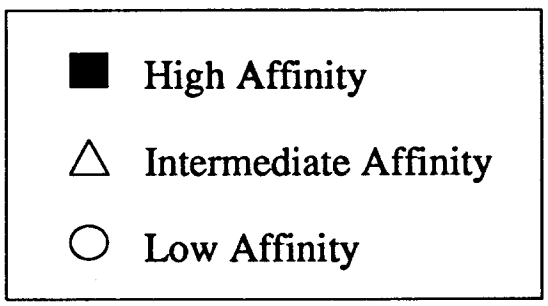


Table Legends

Table I - Steroid binding affinity data.

Footnotes:

(a) Binding affinity data from references 21-23 shown as log 1/K.

(b) Defined as high affinity compound for the neuron plots.

(c) Defined as intermediate affinity compound for the neuron plots.

(d) Defined as low affinity compound for the neuron plots.

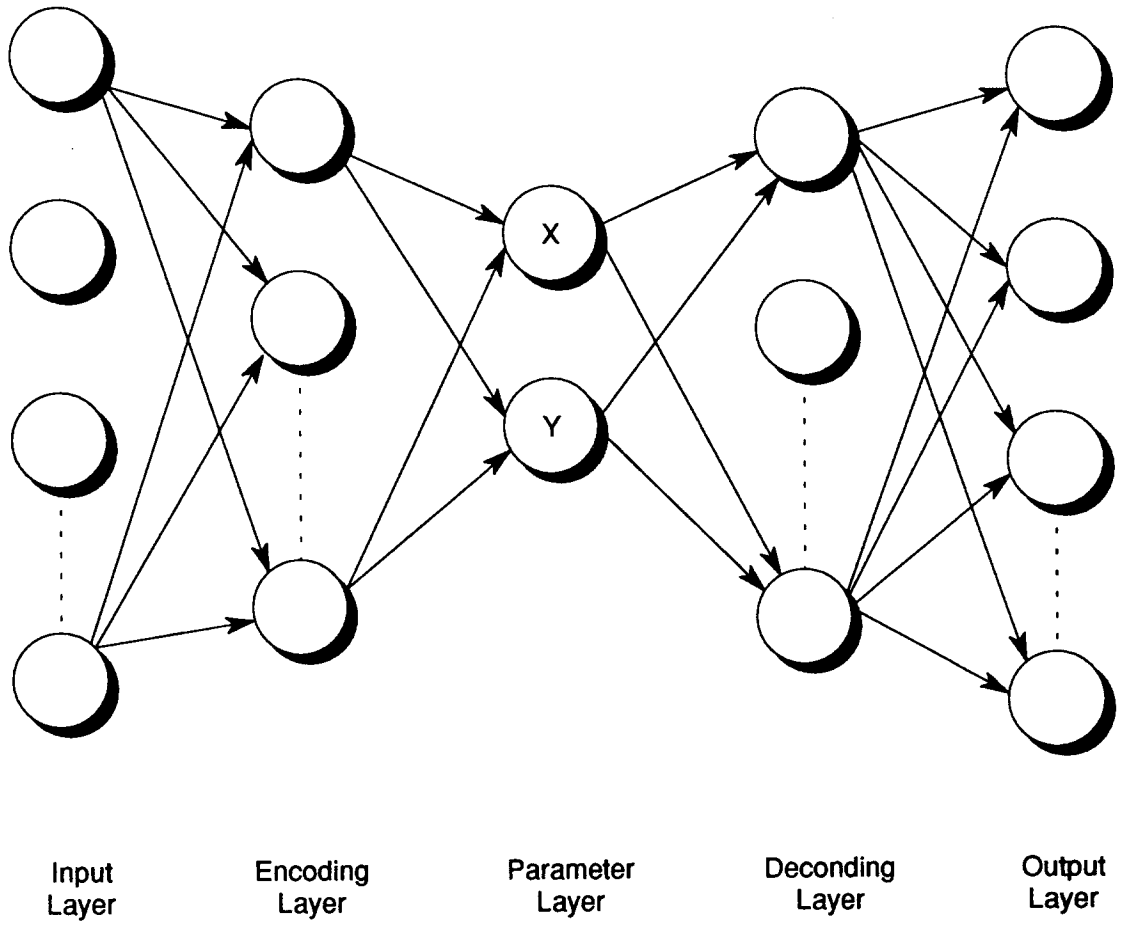
Table II - Results of QSAR studies.

Footnotes:

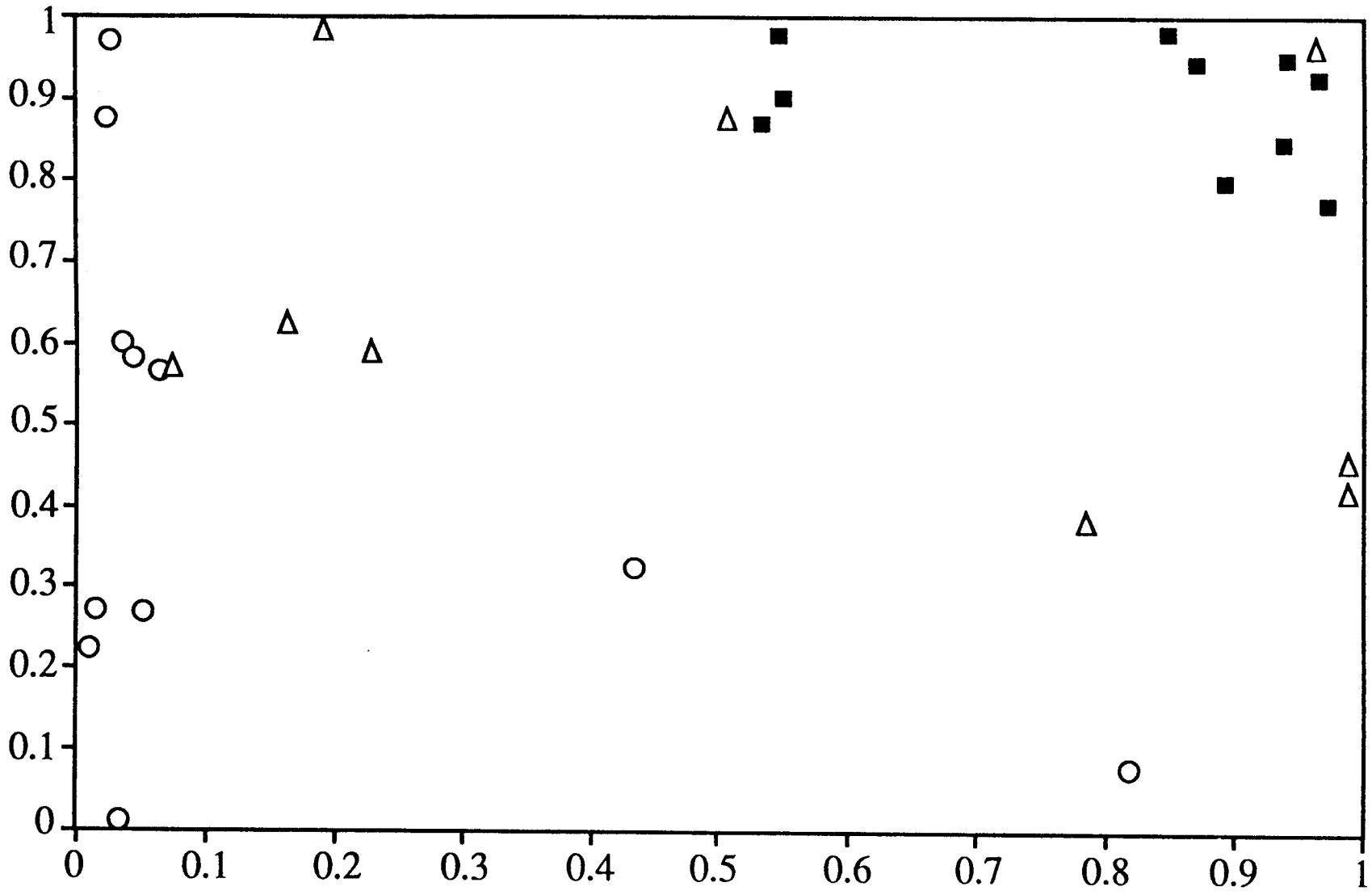
(a) 2.0Å grid density.

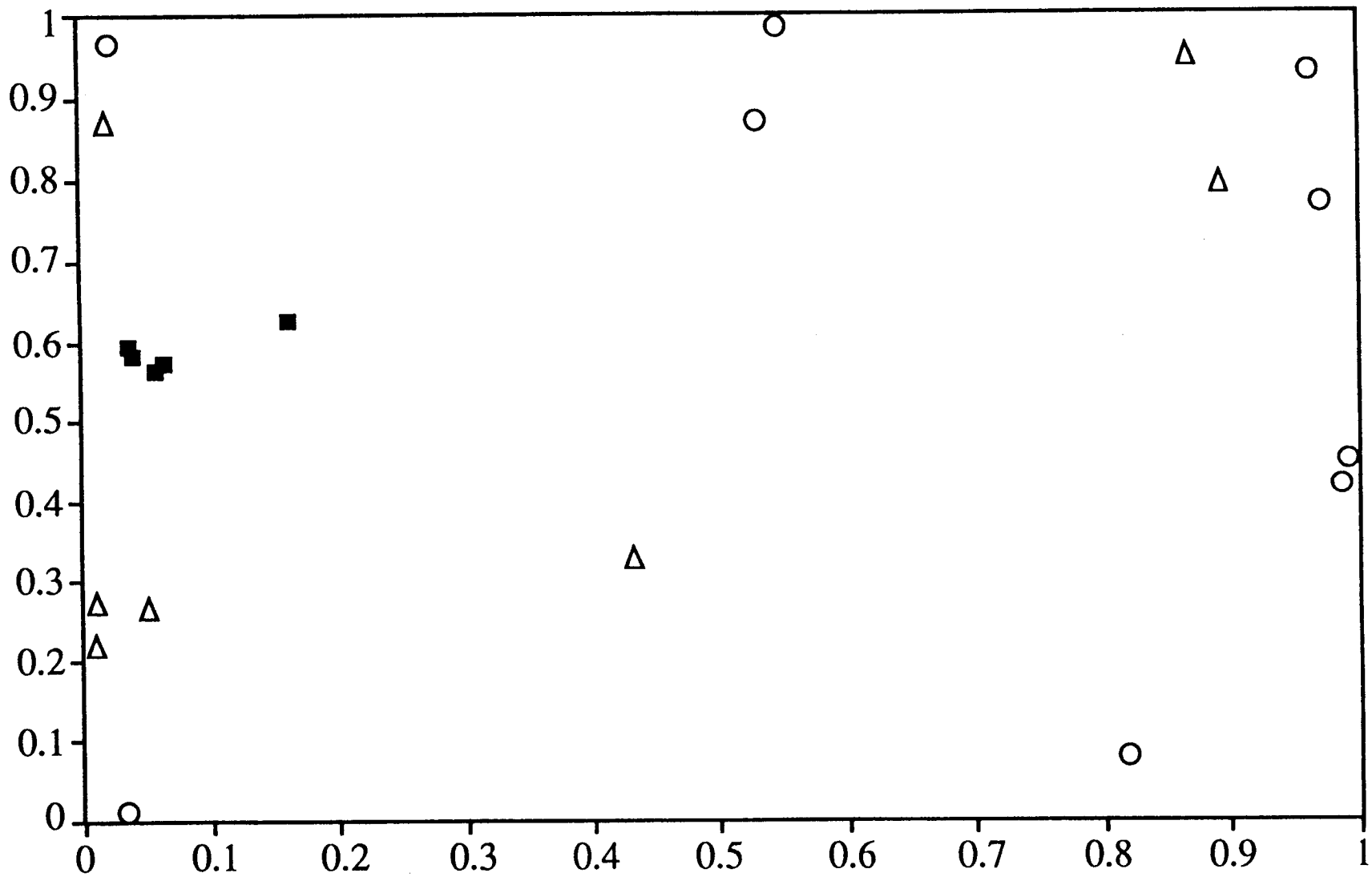
(b) 1.0Å grid density.

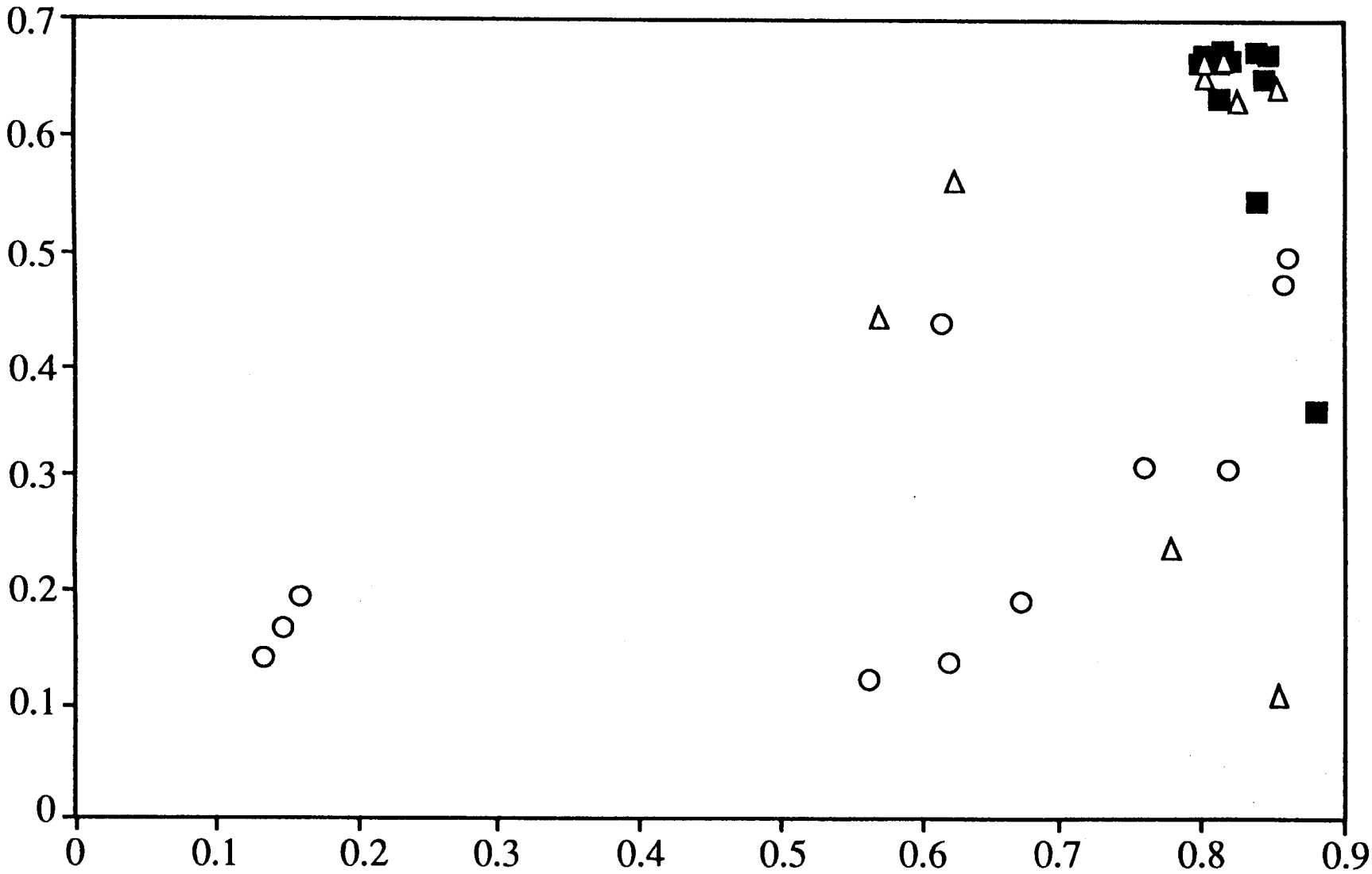
(c) Error including compound 31.

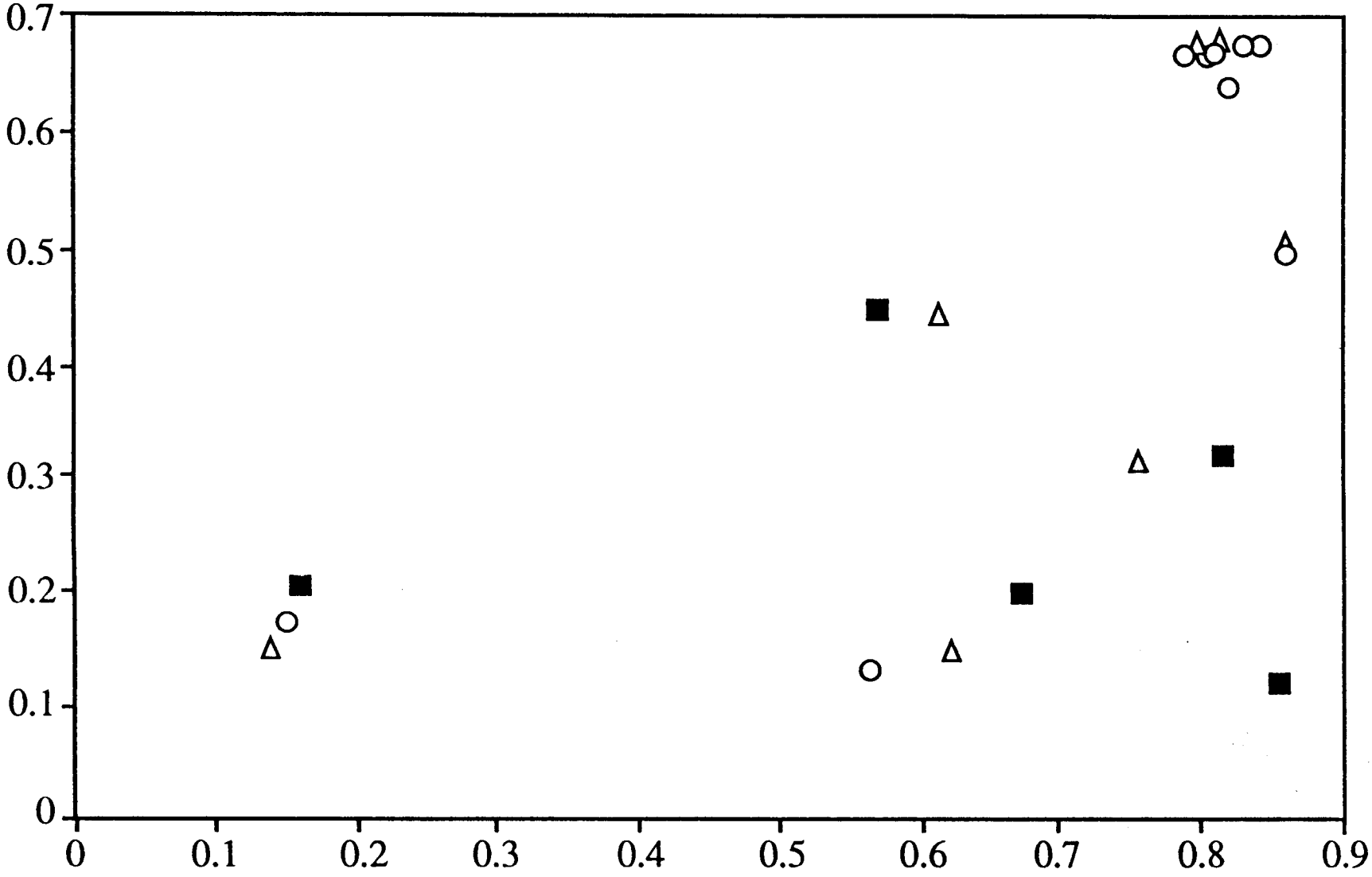

X
Y
Input Layer
Encoding Layer
Parameter Layer
Deconding Layer
Output Layer

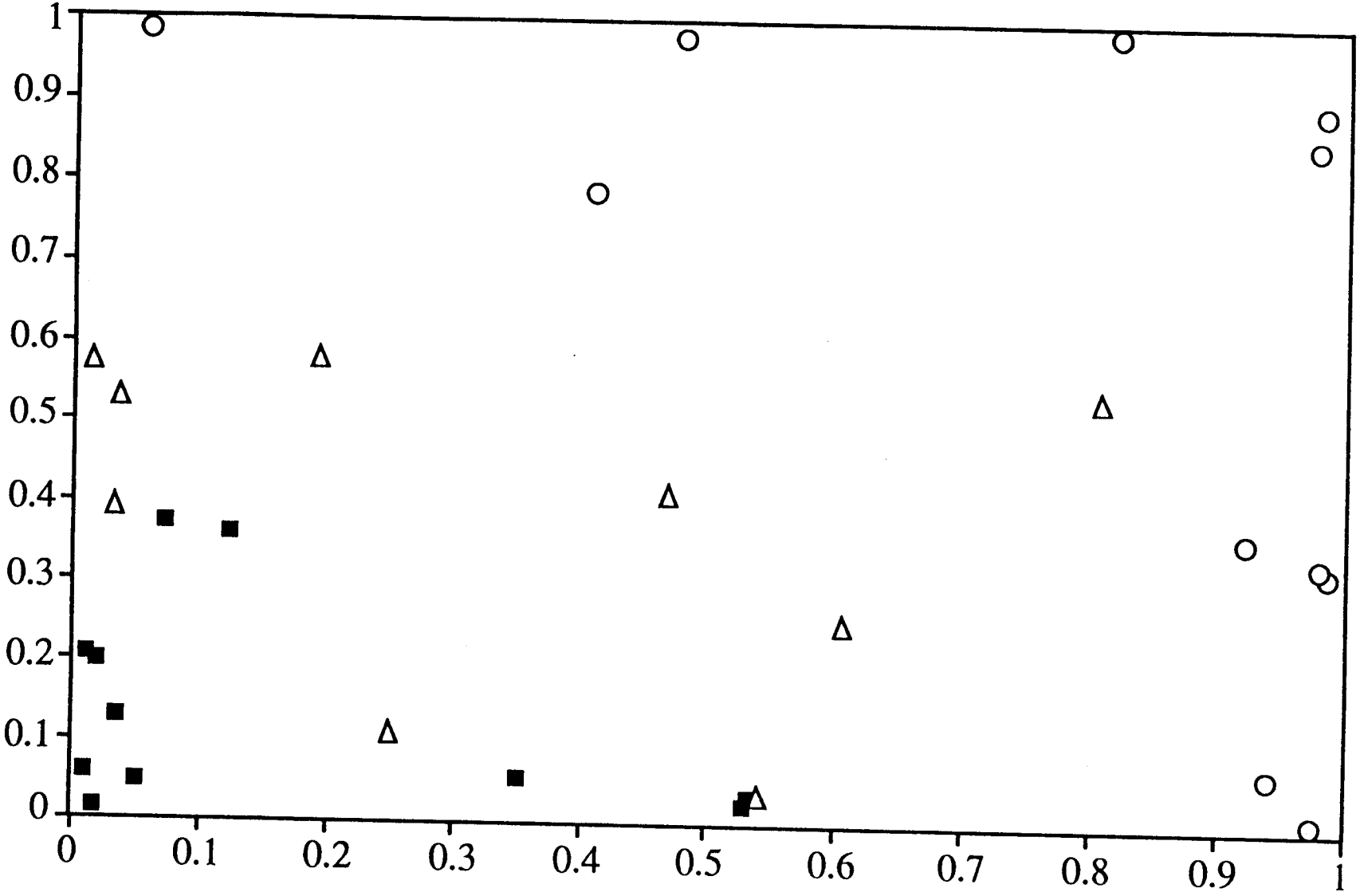
Compound	CBG affinity[a]	TeBG affinity[a]
1	-6.279[c]	-5.322[d]
2	-5.000[d]	-9.114[b]
3	-5.000[d]	-9.176[b]
4	-5.763[d]	-7.462[c]
5	-5.613[d]	-7.146[c]
6	-7.881[b]	-6.342[d]
7	-7.881[b]	-6.204[d]
8	-6.892[c]	-6.431[d]
9	-5.000[d]	-7.819[c]
10	-7.653[b]	-7.380[c]
11	-7.881[b]	-7.204[c]
12	-5.919[c]	-9.740[b]
13	-5.000[d]	-8.833[b]
14	-5.000[d]	-6.633[d]
15	-5.000[d]	-8.176[c]
16	-5.225[d]	-6.146[d]
17	-5.225[d]	-7.146[c]
18	-5.000[d]	-6.362[d]
19	-7.380[b]	-6.944[c]
20	-7.740[b]	-6.996[c]
21	-6.724[c]	-9.204[b]
22	-7.512[b]	-
23	-7.553[b]	-
24	-6.779[c]	-
25	-7.200[b]	-
26	-6.144[c]	-
27	-6.247[c]	-
28	-7.120[b]	-
29	-6.817[c]	-
30	-7.688[b]	-
31	-5.797[c]	-

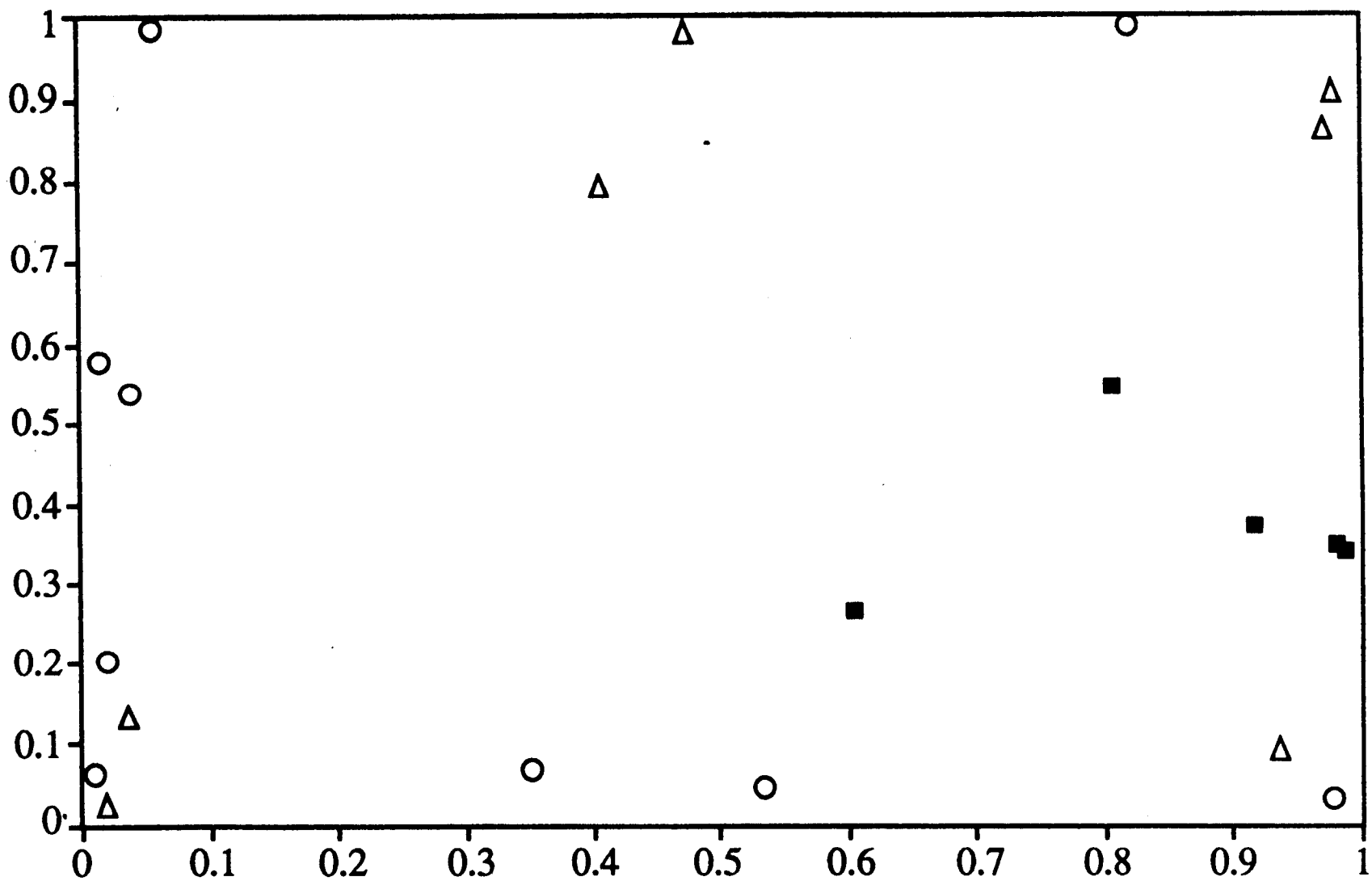

1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9


0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Independent Variables	Cross Validated Standard Error	Cross Validated R^2	No. of Components (Degrees of Freedom)	Standard Error of Predictions
		Corticosteroid-Binding Globulins		
ESP matrix	0.850	0.501	1	0.411 (0.595[c])
SHAPE matrix	0.821	0.633	5	0.404 (0.710[c])
SHESP matrix	0.820	0.533	1	0.514 (0.646[c])
CoMFA - electrostatics	0.718	0.644	1	0.352 (0.619[c])
CoMFA - shape	0.604	0.761	2	0.421 (0.760[c])
CoMFA - full field	0.690[a]/ 0.678[b]	0.689[a]/ 0.716[b]	2[a]/ 3[b]	0.396[a]/ 0.567[b] (0.746[ac]) / (0.835[bc])
		Testosterone-Binding Globulins		
ESP matrix	0.699	0.733	4	-
SHAPE matrix	1.146	0.244	3	-
SHESP matrix	0.665	0.743	3	-
CoMFA - electrostatics	0.826	0.603	3	-
CoMFA - shape	0.972	0.483	4	-
CoMFA - full field	1.010[a]/ 1.004[b]	0.441[a]/ 0.414[b]	4[a]/ 3[b]	-