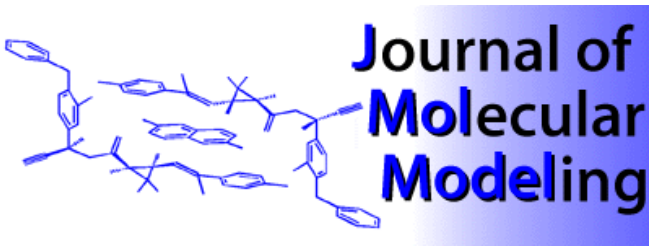
*J. Mol. Model.* **1997**, *3*, 338 – 346






# 3D-QSAR Analysis and Molecular Modeling Investigations of Piritrexim and Analogous[$]

**Romy Fleischer[1,*], Michael Wiese[1], Reinhard Troschütz[2], Mario Zink[2]**

[1]Institute of Pharmaceutical Chemistry, Martin-Luther-University Halle-Wittenberg, Wolfgang-Langenbeckstr.4, 06120 Halle/Saale; Tel: 0345/5525043; Fax: 5527018 (fleischer@medchem5.pharmazie.uni-halle.de)

[2]Institute of Pharmacy and Food Chemistry, University of Erlangen-Nürnberg, Schuhstr.19, 91052 Erlangen

*Received: 9 June 1997 / Accepted: 23 July 1997 / Published: 13 August 1997*

**Abstract**

Quantitative structure-activity relationships for piritrexim and analogues acting as inhibitors of tumour cell growth have been derived. First the Free-Wilson-method was used on a homologous training set of eight derivatives. The selection of variables important for the biological activity of the compounds was carried out with different multivariate methods as multiple linear regression, the partial least squares method und a genetic algorithm. The derivation of three-dimensional structure activity relationships started with a systematic conformational analysis of all compounds. For the conformations having minimal energy and being in agreement with the crystal structure of piritrexim charges were calculated with the AM1 hamiltonian.
For the superimposition of the derivatives two methods were used: maximal similarity of the common substructure or of the molecular fields. A Comparative Molecular Field Analysis with steric and electrostastic fields identified regions important for the activity of the studied compounds independent of the chosen alignment and also correctly predicted the activity of two nonhomologous compounds.

**Keywords:** Free-Wilson-analysis, PLS, genetic algorithm, CoMFA, piritrexim analogous

## Introduction

Piritrexim (ptx) is a lipophilic folate antimetabolite [1] currently undergoing clinical trials as an antineoplastic agent for the treatment of cancer. Several derivatives were synthesized and tested for inhibition of tumour cell growth. It turned out that small structural changes led to large differences in biological activity. To better understand the experimental data classical QSAR, molecular modelling and CoMFA investigations were performed.

## Data and Methods

The compounds were tested for anticancer activity at the NCI. [2] The general framework and the biological activities

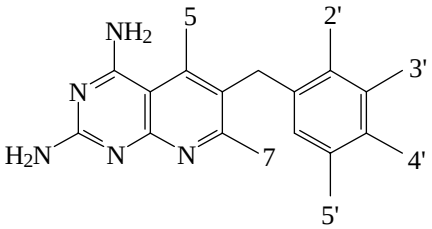
[$] *Presented at the 11. Molecular Modeling Workshop, 6 -7 May 1997, Darmstadt, Germany*

* *To whom correspondence should be addressed*

**Table 1.** *Biological activities and substitution patterns for the homologous set of piritrexim derivatives. A one denotes the presence of the particular substituent and a zero R = H. $GI_{50}$ = growth inhibition by 50 % of cancer cell lines , Me = methyl group, OMe = methoxy group.*

| comp | log $1/GI_{50}$ | Me 5 | Me 7 | OMe 2' | OMe 3' | OMe 4' | OMe 5' |
|---|---|---|---|---|---|---|---|
| ptx | 6.38 | 1 | 0 | 1 | 0 | 0 | 1 |
| p2 | 5.05 | 0 | 1 | 1 | 0 | 0 | 1 |
| p3 | 4.02 | 0 | 1 | 0 | 1 | 1 | 1 |
| p4 | 5.38 | 0 | 0 | 1 | 0 | 1 | 0 |
| p6 | 5.54 | 1 | 1 | 1 | 0 | 0 | 1 |
| p7 | 4.93 | 0 | 0 | 1 | 1 | 1 | 0 |
| p9 | 5.71 | 0 | 0 | 1 | 0 | 0 | 1 |
| p10 | 5.10 | 0 | 0 | 1 | 1 | 0 | 0 |

of the homologous derivatives are shown in Figure 1 and Table 1, respectively, while the structures of the non-homologous derivatives are depicted in Figure 3. Multiple linear regression analysis was performed with STATISTICA 5.0. and Molecular modelling investigations with SYBYL 6.3. [3] The MM3-force field of ALLINGER as implemented in SYBYL and the TRIPOS-force field were used for conformational analysis of the compounds. Charge calculations using the AM1 Hamiltonian were done with MOPAC as implemented in SYBYL. The Advanced CoMFA-module of the SYBYL-program was used for the Comparative Molecular Field Analysis. Standard settings were applied in the CoMFA analysis [4]: the region extended all molecules by at least 4 Å with a grid spacing of 2 Å. A $C_{sp3}$ probe with +1 charge and cutoff values of 30 kcal/mol for both fields with smooth transition were used. The minimum sigma was set to 0.2 kcal/mol. The predictivity of the models was estimated by cross-validation with the leave-one-out method.




5, 7: H, $CH_3$
2', 3', 4', 5': H, $OCH_3$

**Figure 1.** *General framework of the homologous ptx derivatives*

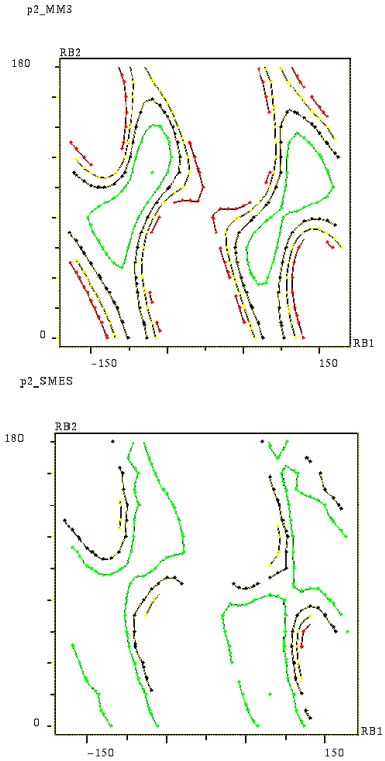



**Figure 2.** *Ramachandran-plot of the conformational energy as function of the two rotatable bonds connecting the 5-deazapterin and the benzene ring with all other rotatable bonds being optimized. Contour maps derived from MM3-calculations are shown on top and those from the Systematic Search method are depicted on bottom. Derivatives p2 (a), p9 (b) and ptx (c). The contour lines are drawn with equal spacing of 1 kcal/mol at one to four kcal above the global minimum in green, white, yellow, and red, respectively.*

**Table 2.** *Effect of parameter elimination on predictivity of the PLS model. The normalized coefficients of the parameters are shown for the starting analysis. Retained parameters are marked by an asterik while empty fields represent eliminated ones.*

| analysis no. | Parameter | | | | | | Fit | cross-validation | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Me5 | Me7 | OMe2' | OMe3' | OMe4' | OMe5' | $r^2$ | $Q^2$ | SDEP | onc |
| *PLS-analysis without compound p3 (n=7)* | | | | | | | | | | |
| 1 | 0.56 | 0.74 | | 0.42 | 0.14 | 0.26 | 0.99 | 0.90 | 0.22 | 3 |
| 2 | * | * | | * | | * | 0.98 | 0.89 | 0.23 | 3 |
| 3 | * | * | | * | * | | 0.99 | 0.83 | 0.29 | 3 |
| 4 | * | * | | * | | | 0.93 | 0.60 | 0.45 | 3 |
| *PLS-analysis with compound p3 (n=8)* | | | | | | | | | | |
| 5 | 0.39 | 0.57 | 0.19 | 0.34 | 0.13 | 0.15 | 1.00 | 0.92 | 0.36 | 5 |
| 6 | * | * | * | * | | * | 0.99 | 0.84 | 0.42 | 4 |
| 7 | * | * | * | * | * | | 0.99 | 0.92 | 0.30 | 4 |
| 8 | * | * | | * | * | * | 0.99 | 0.89 | 0.30 | 3 |
| 9 | * | * | * | | * | * | 0.93 | 0.47 | 0.59 | 2 |
| 10 | * | * | * | * | | | 0.97 | 0.78 | 0.49 | 4 |
| 11 | * | * | | * | | * | 0.95 | 0.78 | 0.38 | 2 |
| 12 | * | * | | * | * | | 0.99 | 0.96 | 0.18 | 3 |

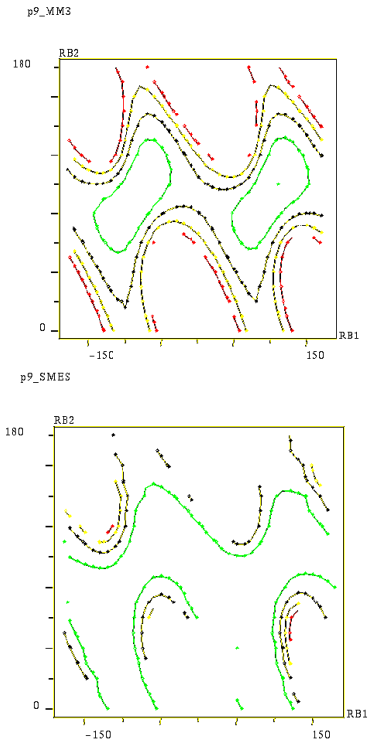



**Figure 2b.**

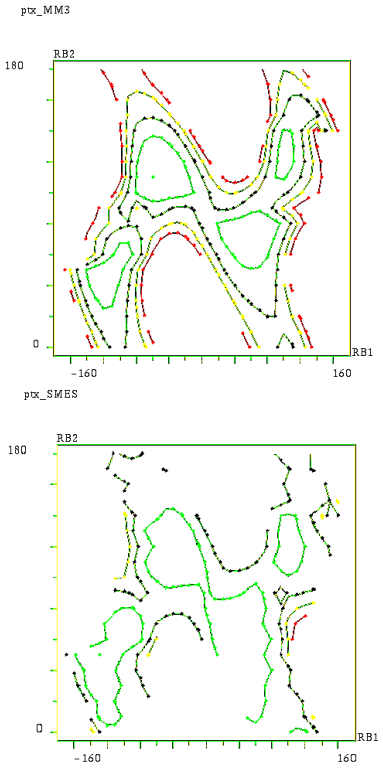



**Figure 2c.**

**Table 3.** *Selected parameters and predictive power of the models found by the genetic algorithm.*

| Parameter | | | | | Fit | Pred |
|---|---|---|---|---|---|---|
| **Me5** | **Me7** | **OMe 3'** | **OMe 4'** | **OMe 5'** | **$r^2$** | **$Q^2$** |
| *without compound p3 (n=7)* | | | | | | |
| * | * | * | * | | 0.99 | 0.84 |
| * | * | * | | | 0.93 | 0.66 |
| *with compound p3 (n=8)* | | | | | | |
| * | * | * | | | 0.95 | 0.84 |
| * | * | * | * | * | 1.00 | 0.85 |
| * | * | * | * | | 0.99 | 0.95 |

## Results and Discussion

*Classical QSAR-analysis*

As a first attempt the Free-Wilson-method [5] was applied to a homologous training set of eight derivatives. Because there is only one compound (p3) without a methoxy group in position 2' (singularity in data) separate analyses with and without p3 were performed. Depending on the data set chosen MLR led to different equations and significant parameters (data not shown) and seems to be not suitable for these data. Therefore the data were analyzed with PLS leading to highly predictive models in case of both data sets. A main problem of a PLS-analysis is the interpretation of the result in structural terms. In order to derive a better interpretable model "non important" parameters were eliminated based on their normalized contribution to the PLS model. The results show that the normalized coefficient is only partly suitable for elimination of parameters. In case of the model with seven compounds the elimination of the least contributing parameter only very slightly decreased predictivity (Table 2, eq. 1 and 2). Whereas for the model derived from eight compounds the picture changed, though no high intercorrelations between the parameters are present. Here three parameters have similar contributions and the predictivity remained constant or dropped considerably depending on the eliminated parameter (Table 2, eq. 5 to 8). Overall the parameter selection procedure yielded an improvement in predictivity and led to a model consistent with the one derived for seven compounds (Table 2, compare eq. 2 and 12).

Also a genetic algorithm [6, 7] (GA) was used for parameter selection. This GA performes a MLR, but the criterion to select the parameters is their predictive power and not the fitting ability as in classical MLR. The GA automatically found the same parameter combination as obtained from the optimal PLS-analysis after manual selection of parameters (Table 3). Therefore, the combination of a genetic algorithm with either MLR or PLS seems promising.

*Conformational Analysis*

The results of a conformational analysis carried out with the TRIPOS- and MM3-force fields are shown in Figure 2 for three derivatives having the same pattern of substitution on the phenyl group, but differing in substitution in position 5 and 7 of the 5-deazapterine ring.

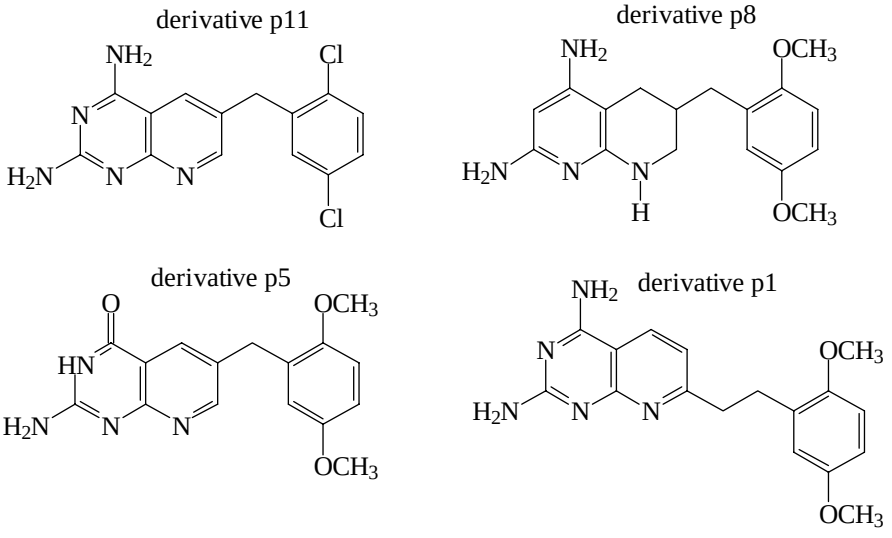



**Figure 3.** *Structures of the non homologous derivatives used to test the predictivity of the CoMFA models*

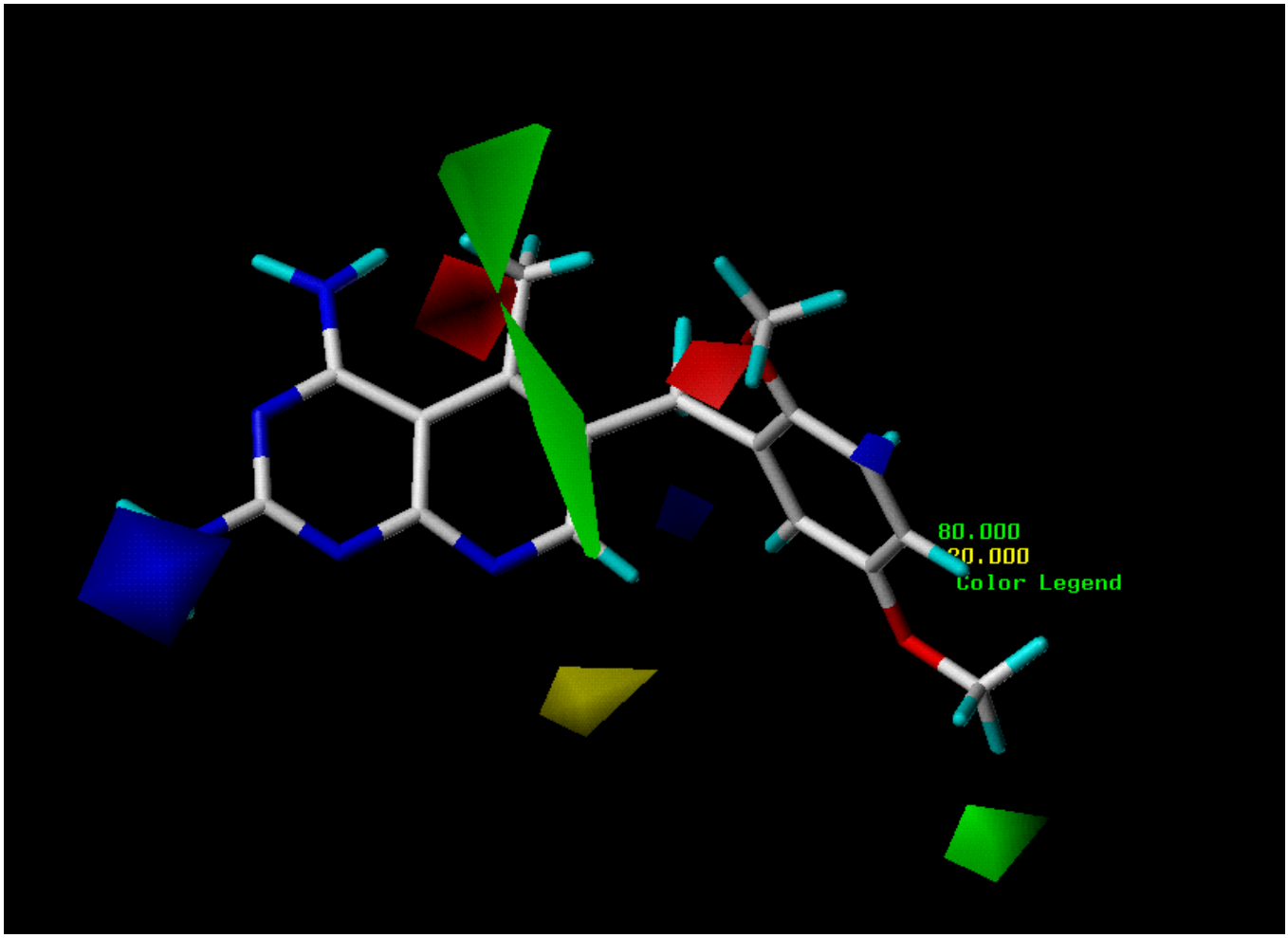



**Figure 4.** *CoMFA model derived from the homologous derivatives using conformations obtained with the Tripos force field after field fit alignment.*
*The steric field is shown in green and yellow. Higher activity is favoured by bulky substituents near green and no or smaller substituents near yellow. The electrostatic field is shown in blue and red. Activity is increased by more positive charge near blue and more negative charge near red.*

In both force fields the bound conformation from the X-ray structure of the ptx-enzyme complex [8,9] is not the energetical minimum, but the energy difference is small for derivatives without a methyl group in position 7 (ptx and p2). A methyl group in position 5 of the pteridine ring additionally restricts the conformational flexibility and favours conformations close to the X-ray structure. If a methyl group in position 7 (p9) is present the bound conformation becomes energetically unfavourable.

*CoMFA [4]*

For the conformations that correspond to the bound conformation in accordance with the crystal structure, charges were calculated with the AM1 hamiltonian. To align the compounds two strategies were used: an atom based "database-alignment" and the superimposition of the steric and electrostatic fields "field fit". Both schedules led to very similar orientations, differing only sligthly in the overlap of the common substructure.

The comparison of the $Q^2$ values (Table 4) shows that the field fit alignment is superior for structures optimized with the TRIPOS force field, while for structures optimized with the MM3 force field only negligible differences in the predictive power are observed between the two alignment schedules. The CoMFA contour maps for the field fit alignments are given in Figures 4 and 5 for the TRIPOS and MM3 force field, respectively.

A comparison of the contour maps with the results of the classical statistical analysis reveals similarities and differences. The positive and negative contribution of a methyl group in position 5 and 7, respectively, are clearly reflected by the green and yellow regions surrounding the substituents. However, the detrimental contribution of methoxy groups in position 3' and 4' are not reflected in the CoMFA maps. Instead a positive contribution to activity by an electronegative group in position 5' is indicated by the red contoured region.

**Table 4.** *Summary of the predictivity of the different CoMFA models*

| n | force field | alignment | $r^2$ | $Q^2$ | s | SDEP | onc |
|---|---|---|---|---|---|---|---|
| 8 | TP | DB | 0.996 | 0.55 | 0.07 | 0.70 | 4 |
| 8 | TP | FF | 0.998 | 0.74 | 0.06 | 0.66 | 5 |
| 8 | MM | DB | 0.998 | 0.73 | 0.05 | 0.54 | 4 |
| 8 | MM | FF | 0.999 | 0.72 | 0.04 | 0.68 | 5 |
| 10 | TP | DB | 0.961 | 0.62 | 0.15 | 0.47 | 3 |
| 10 | TP | FF | 0.996 | 0.76 | 0.06 | 0.46 | 5 |
| 10 | MM | DB | 0.998 | 0.69 | 0.04 | 0.52 | 5 |
| 10 | MM | FF | 0.998 | 0.71 | 0.04 | 0.50 | 5 |

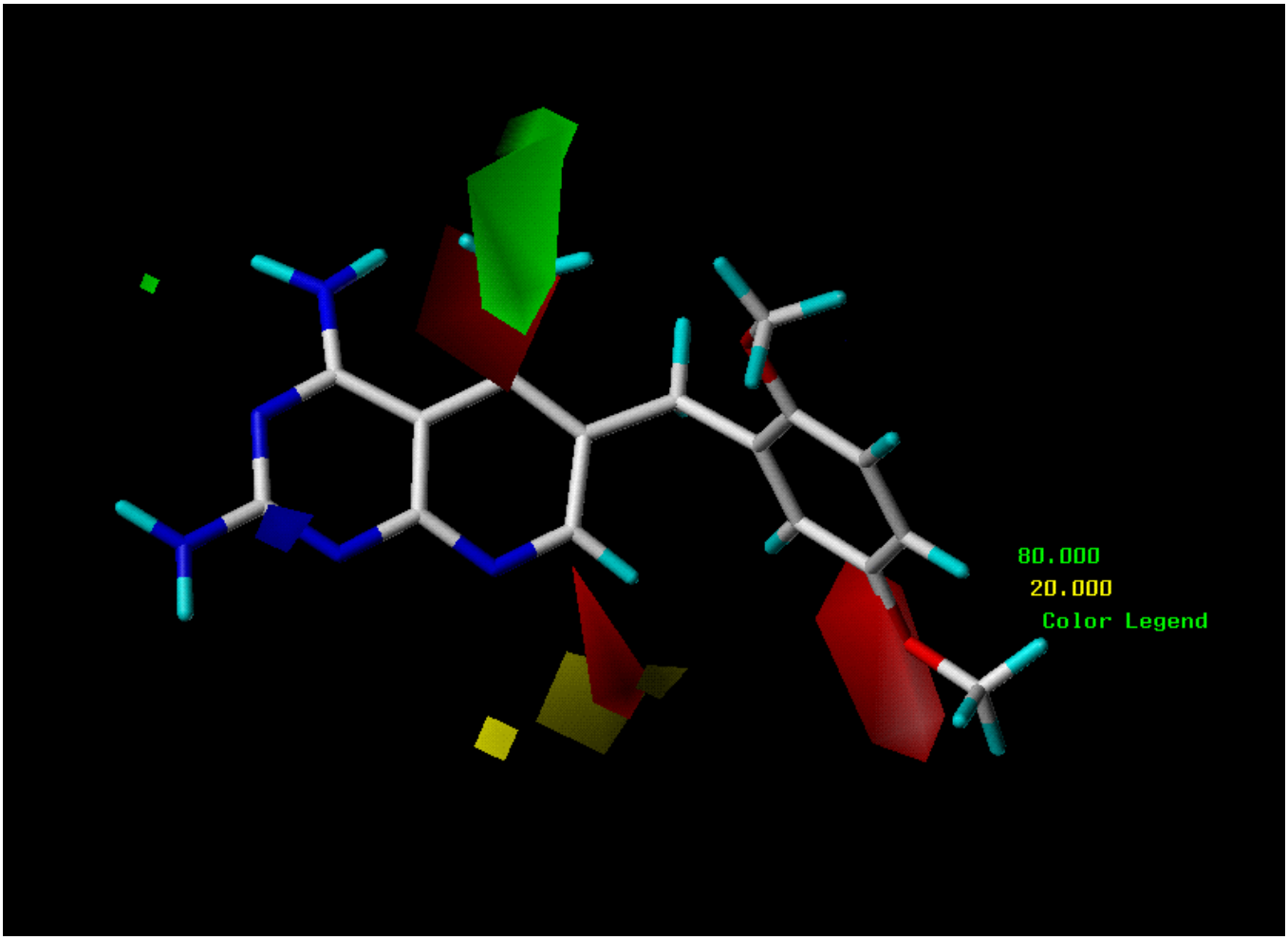



**Figure 5.** *CoMFA model derived from the homologous derivatives using conformations obtained with the MM3 force field after field fit alignment. For designation of the coloured regions see Figure 4.*

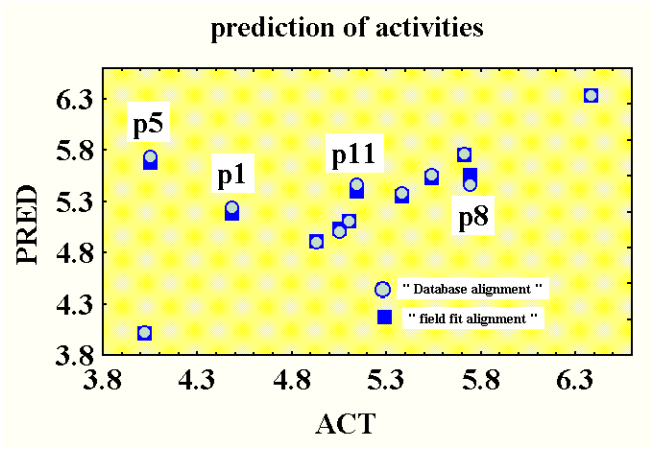



**Figure 6.** *Activity prediction for non homologous derivatives.*

The predictive power of the CoMFA-models was tested with four additional non homologous compounds which are depicted in Figure 3.

The models can predict the activity of p8 and p11, but not of p1 and p5 (Figure 6). Therefore p8 and p11 were included in the CoMFA data set. The derived CoMFA-models for the Tripos and the MM3 force field using the field fit alignment are shown in Figures 7 and 8, respectively.

The contour maps of the models derived for 8 and 10 derivatives are similar, indicating the stability of the derived model. The fact that the CoMFA model derived from the homologous set of compounds cannot predict the activities of compounds p1 and p5 can be understood as they represent structural singularities within the data set. Compound p1 is structurally different and bears an ethylene instead of a methylene bridge, thus the benzene ring occupies space outside the volume of all other derivatives not covered by the CoMFA

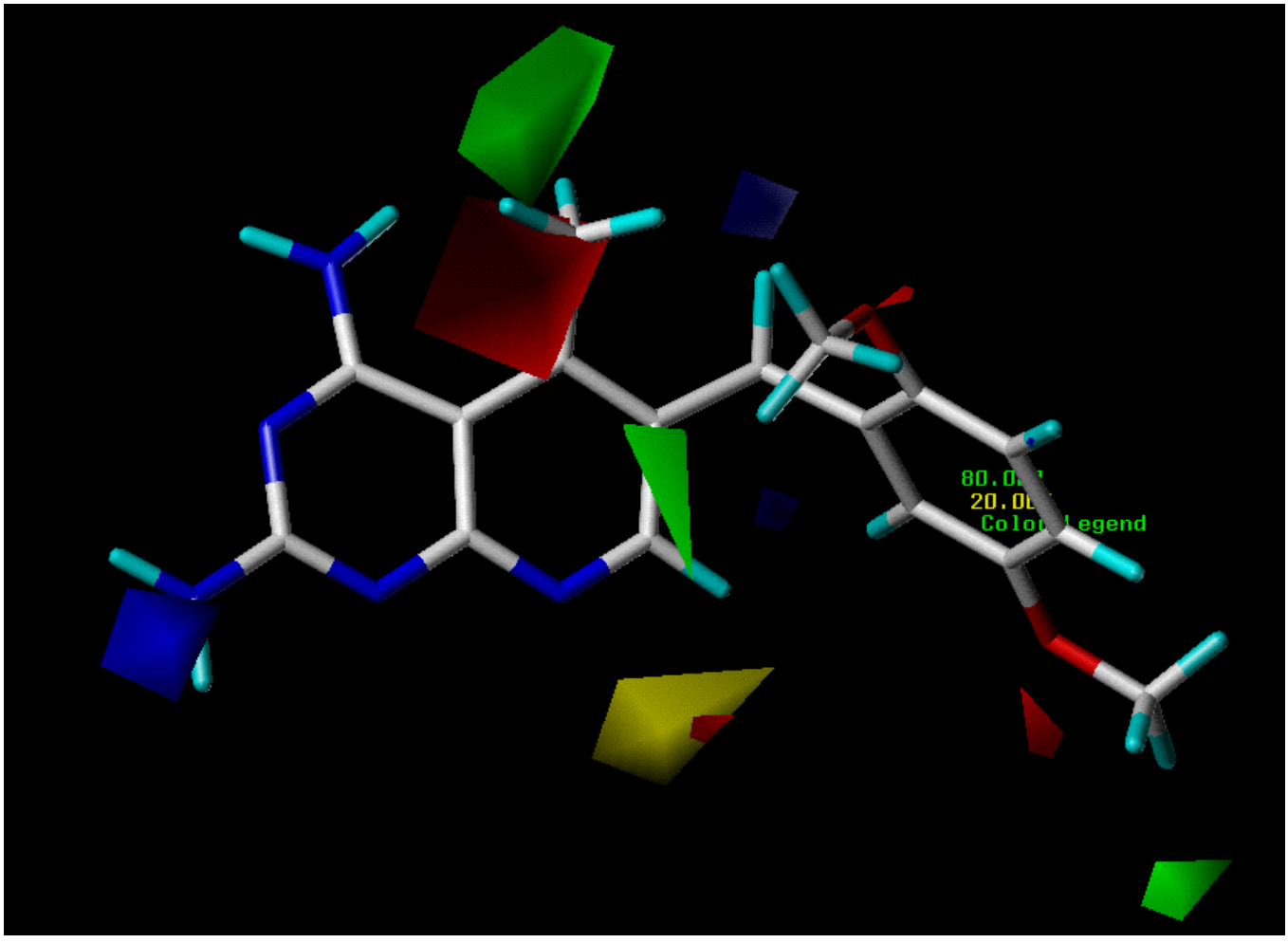


**Figure 7.** *CoMFA model derived from the conformations obtained with the Tripos force field after field fit alignment. For designation of the coloured regions see Figure 4.*

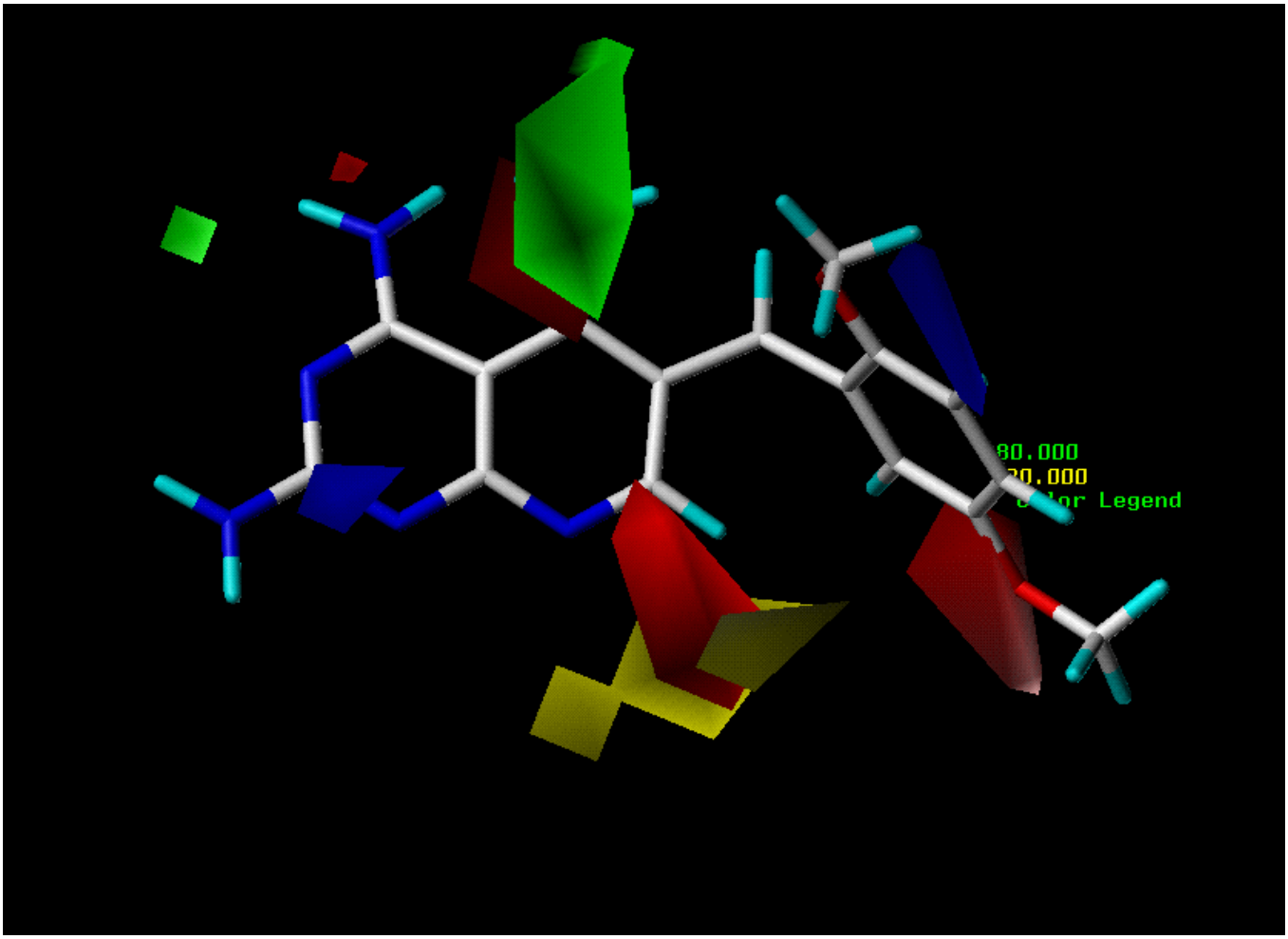



**Figure 8.** *CoMFA model derived from the conformations obtained with the MM3 force field after field fit alignment. For designation of the coloured regions see Figure 4.*

model. Compound p5 has no 2,4-diamino-partial-structure but a 2-keto moiety instead. It might bind to the enzyme in a different way, as observed in the example of dihydrofolate and methotrexate. To improve the derived CoMFA model data on more structurally divers derivatives are needed.

## Conclusion

The 2D-QSAR analysis using either PLS or a GA yielded equations with very high cross-validated predictivity. In case of PLS a parameter selection further improved the results. The genetic algorithm gave nearly identical results and performed the parameter selection automatically. A CoMFA model was developed that could also correctly predict the activity of two non-homologous derivatives.

*Acknowledgment*: We thank the Fonds der Chemischen Industrie for financial support.

## Abbreviations

| | |
|---|---|
| CoMFA | Comparative Molecular Field Analysis |
| MLR | Multiple Linear Regression |
| PLS | Partial Least Squares |
| onc | optimal number of components |
| SDEP | standard error of prediction |
| $r^2$ | explained variance |
| $Q^2$ | explained variance by cross-validation |
| s | standard error of estimate |
| F | ratio of explained variance / unexplained variance |
| DB | database-alignment |
| FF | field-fit-alignment |
| n | number of compounds |
| TP | systematic search with the Tripos force field without electrostatics |
| MM | molecular mechanics optimization with the MM3 force field and the dihedral-driver-option |

## References

1. Troschütz, R.; Zink, M.; Dennstedt, T. Synthese und Antitumorwirkung von Iso-Piritrexim, einem lipophilen Folatantagonisten. *Arch. Pharm.* **1995**, 328, 535-540.

2. Boyd, M. R. Status of the preclinical antitumor drug discovery screen. *Principles and Practises of oncology*. **1989**, 3, 1-12.
3. Tripos Assoc. *SYBYL Theory Manual.* St. Louis, USA Version 6.3, 1996.
4. Cramer, R. D. III; De Priest, S. A.; Patterson, D. E.; Hecht, P. The developing practise of comparative molecular field analysis. In *3D-QSAR in Drug Design-Theory Methods and Applications*, Kubinyi, H., Ed.; ESCOM Leiden, 1993; pp 443-485.
5. Seydel, J. K.; Schaper, K. J. Chemische Struktur und biologische Aktivität von Wirkstoffen; VCH Weinheim 1979.
6. Leardi, R.; Boggia, R.; Terrile, M. Genetic algorithm as a strategy for feature selection. *J. of Chemometrics.* **1992**, 6, 267-281.
7. Leardi, R. Application of a genetic algorithm to feature selection under full validation conditions and outlier detection. *J. of Chemometrics*. **1994**, 8, 65-79.
8. Cody, V.; Sutton, P. A. Conformational analysis of lipophilic antifolates, crystal structure of piritrexim and a theoretical evaluation of its binding to dihydrofolate reducatse. *J. Am. Chem. Soc.* **1988**, 110, 6219-6224.
9. Davies, J. F.; Delcamp, T. J.; Prendergast, N. J.; Ashford, V. A.; Freisheim, J. H.; Kraut, J. Crystal structures of recombinant human dihydrofolate reductase complexed with folate and 5-deazafolate. *Biochemistry*. **1990**, 29, 9467-9479.