Journal of Computer-Aided Molecular Design, 7 (1993) 253–262
ESCOM

J-CAMD 197

Inhibitors of prolyl endopeptidase: Characterization of the pharmacophoric pattern using conformational analysis and 3D-QSAR

Thierry Langer* and Camille G. Wermuth**

Laboratoire de Pharmacochimie Moléculaire, Centre de Neurochimie, 5 rue Blaise Pascal, F-67084 Strasbourg Cedex, France

Received 27 October 1992
Accepted 10 December 1992

Key words: Prolyl endopeptidase inhibitors; Computer-aided molecular design; Conformational analysis; Molecular dynamics; CoMFA; 3D-QSAR

SUMMARY

A structure–activity study has been carried out on several compounds known as inhibitors of the serine protease prolyl endopeptidase. Conformational analysis has been done using different molecular mechanics methods such as molecular dynamics, or a randomized conformational search method. The conformers obtained were classified using geometric and energetic criteria. A pattern recognition analysis was done in order to divide conformers according to families. The resulting dominant families, for all compounds investigated, showed very similar geometric features. Based on the lowest energy conformers obtained after randomized conformational analysis, a 3D-QSAR model was established using the CoMFA approach. The validity of this model was verified by prediciting correctly the activity of other molecules not used in the construction of this model.

INTRODUCTION

Prolyl endopeptidase (PEP; EC 3.4.21.26) is a serine protease that is widely distributed in animal tissues and body fluids [1]. The amino acid sequence and the identity of the active site residues of porcine PEP have recently been published [2]; however, the 3D structure of the enzyme is not yet known. This enzyme has a high hydrolysis specificity at the carboxylic side of prolyl residues in biologically active peptides with up to 30 amino acids and has been shown to play an important role in the metabolism of neurotransmitter peptides such as TRH [3]. Therefore much effort has been devoted to search for potent inhibitors of this enzyme. In fact, inhibitors of PEP

*Present address: Institut für Organische und Pharmazeutische Chemie der Universität Innsbruck, Innrain 52A, A-6020 Innsbruck, Austria.

**To whom correspondence should be addressed.

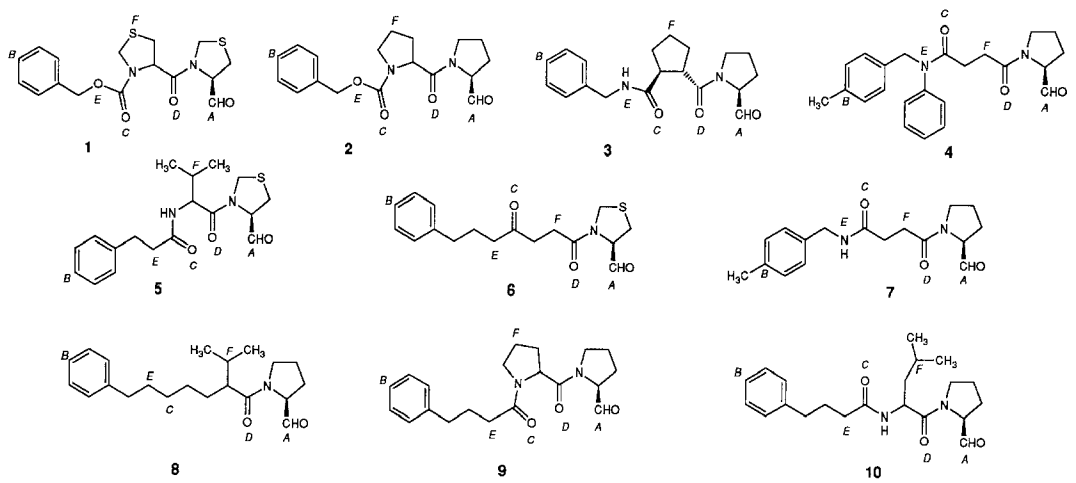

Fig. 1. Training set of molecules.

have been reported to exhibit antiamnesic effects [4], thus being of potential therapeutic interest for the treatment of senile dementia like Alzheimer's disease.

In parallel work [5] we reported the design and the synthesis of a new inhibitor for PEP based on a computer graphics analysis of several known potent and highly flexible compounds within a 0.1 nM to 100 μM inhibitory range. In this paper we describe the conformational analysis of several inhibitors of prolyl endopeptidase and the construction of a CoMFA model [6] permitting the prediction of the activity of compounds of a test set. For the training set we considered the molecules shown in Fig. 1. They are representative for the class of compounds under study and their IC_{50} values towards PEP (see Table 1) cover a broad range which enables development of a highly probable activity model.

As shown in Fig. 1, all prolyl endopeptidase inhibitors investigated have a common prolinal or thioprolinal C-terminal moiety, which seems to be crucial for enzyme recognition and is believed to occupy the P_1 site and to form the transition state analogue in the enzyme-substrate interaction [7]. Moreover, it has been suggested [7] that there is a hydrophobic pocket with limited size in the S_3 subsite of the enzyme which could interact hydrophobically with a certain N-blocked group of the inhibitor. In our investigation, the N-terminal phenyl rings are considered as the moieties responsible for hydrophobic interactions, and therefore distance parameters which are able to describe the relative position of these substructures in comparison to the prolinal or thioprolinal carbonyl group were chosen.

TABLE 1
ACTIVITY DATA OF THE TRAINING SET COMPOUNDS

Compound	IC_{50} (nM)	Reference	Compound	IC_{50} (nM)	Reference
1	0.035	7	**6**	16	21
2	0.74	19	**7**	23	20
3	1.4	20	**8**	80	22
4	1.9	20	**9**	80	23
5	5.7	21	**10**	120	23

The first step in a conformational analysis is to develop a method for generating the required conformations. One possibility would be the use of an exhaustive systematic conformational analysis for each molecule. This approach has been used in our previous work [5] for determining all possible conformers of the considered PEP inhibitors. Another approach would be the use of the Monte Carlo-type conformational search techniques [8–10]. However, both methods are unable to generate a statistically correct ensemble of possible conformers at given thermodynamic conditions. In some way molecular dynamics (MD) is more appropriate for this kind of problem since it generates a sample of the available conformational space at a certain temperature rather than all valid conformations as would be produced by a systematic search algorithm. MD has been applied to conformational searching mainly in the area of small peptides [11] but has also been used to explore the conformational space of small molecules [12–14] and appears to work well in both cases. Since the major issue of MD is its sampling efficiency, often it may require much computational effort to produce structures greatly different from the starting conformation unless the temperature is raised or the potential is softened.

Our approach to the conformational analysis of the inhibitors of PEP was divided into two parts; first, we investigated the conformational space available at room temperature by molecular dynamics and described the conformers obtained using classification algorithms based on distance parameters. Second, we generated conformers using a randomized search technique similar to the Monte Carlo method in order to compare the potential energies of the global minima obtained with both methods. For the construction of the CoMFA model we intended to start with the lowest energy structures of the molecules. Although these conformers are not necessarily the ones interacting with the enzyme, they are of course the most significant, since the 'active conformation' is probably among them.

METHODS

The entire molecular modelling study was done using the SYBYL [15] software package running on a VAX Station 4000 and the molecules were visualized on an E&S PS390 graphic terminal. All structures were built starting from the Tripos standard fragment library and before further analysis were fully minimized using the Tripos force field [16]. Charges were computed using the method of Berthod and Pullman [17] and electrostatic interaction terms were included both in minimization and in dynamics processes.

The MD studies were performed at constant temperature with an integration step of 1 fs. The structures were initialized at 300 K starting from a Boltzmann distribution set followed by 10 ps of equilibration and 90 ps of simulation, also at 300 K. The conformations were recorded every 100 fs. Data analysis was done using the Tripos QSAR module. The final classification of conformers obtained after MD was performed using the *FAMILY* algorithm implemented in SYBYL. The Monte Carlo-like conformational search was carried out with the *RANDOMSEARCH* option of SYBYL.

The CoMFA model was derived from global minima structures provided by the randomized conformational analysis and fitted using the field fit procedure on the global minimum conformation of the most potent compound **1**. First a crossvalidated partial least square (PLS) analysis using the CoMFA and log ($1/IC_{50}$) columns was performed to determine the optimum number of

TABLE 2
DISTANCE SET DEFINITION

Distance	Atoms[a]	Distance	Atoms[a]
D1	A – B	D6	A – E
D2	A – C	D7	A – D
D3	C – D	D8	A – F
D4	B – C	D9	B – F
D5	D – E		

[a] For corresponding atoms see Fig. 1.

components to use. Then a non-crossvalidated PLS run was used to deduce the model and to produce the CoMFA steric and electrostatic contribution contour plots.

RESULTS AND DISCUSSION

In order to test the reliability of results generated by MD, we performed several simulations for different time periods (50 ps and 100 ps) and with variable temperatures (700 K and 300 K) on representative compounds in the series. It was found to be sufficient to run the dynamics for a 100-ps time period at 300 K to get consistent results. This procedure permitted us to derive energetically preferred families of conformers since they are statistically more important due to their longer lifetime. For comparison and classification purposes we defined a set of 9 interatomic distances in each compound. The distances were evaluated for all conformers and the correlation matrix of their autoscaled values was submitted to principal component analysis (PCA) in order to detect highly correlated distance parameters. The atoms used for the definition of the distance set are listed in Table 2 and the cumulative percent variance of the first 3 principal components is reported in Table 3. The loading projections of these principal components suggested, that in some cases, less than 9 interatomic distances were necessary for the description of the total variability of the system without loosing useful information. On the other hand, the cumulative

TABLE 3
CUMULATIVE PERCENT VARIANCE OF THE FIRST THREE PRINCIPAL COMPONENTS OBTAINED BY PCA

Compound	PC1	PC2	PC3
1	29.7	49.9	64.9
2	24.6	48.7	65.4
3	36.7	56.9	70.1
4	40.3	59.8	75.1
5	25.1	45.4	63.7
6	34.2	56.5	72.0
7	30.1	48.9	64.3
8	39.5	61.2	74.5
9	28.0	51.7	66.3
10	28.7	46.1	62.0

TABLE 4
INTERATOMIC DISTANCE VALUES (Å) OF THE PREDOMINANT CLUSTER OF CONFORMERS GENERATED BY MOLECULAR DYNAMICS FOR COMPOUNDS **1–10**

Compound		D1	D2	D3	D4	D5	D6	D7	D8	D9
1	Mean	8.70	6.93	4.32	5.33	5.61	3.42	3.03	6.12	8.63
	High	11.79	8.01	6.07	7.32	7.06	5.30	4.03	7.35	10.29
	Low	5.02	5.26	3.13	3.53	3.30	2.22	2.28	4.10	5.49
2	Mean	7.67	6.95	4.15	6.26	3.17	5.58	3.05	6.08	8.71
	High	11.15	8.00	5.77	7.22	4.87	7.22	3.94	7.13	10.10
	Low	4.08	5.17	3.08	3.56	2.26	4.26	2.39	4.40	6.62
3	Mean	10.65	8.28	5.45	5.10	4.80	7.48	3.12	6.31	6.20
	High	12.41	9.03	5.99	6.28	5.83	8.96	3.75	7.05	7.81
	Low	8.78	6.70	4.35	3.60	3.55	6.01	2.58	5.53	3.57
4	Mean	10.71	6.61	4.51	5.52	5.12	7.37	2.96	5.30	6.67
	High	13.31	7.98	5.90	7.23	6.09	8.43	4.02	6.16	7.98
	Low	8.08	4.04	2.25	3.60	3.09	5.95	2.37	4.01	5.22
5	Mean	10.16	6.32	4.57	7.47	3.06	5.36	3.05	5.19	9.58
	High	13.13	7.60	5.48	9.13	3.81	6.16	3.88	6.17	11.34
	Low	7.01	5.00	2.55	4.52	2.30	4.10	2.35	4.03	7.07
6	Mean	10.84	7.16	4.95	7.36	5.29	7.67	3.13	4.49	8.06
	High	14.91	8.62	5.94	8.99	6.54	8.87	4.19	5.03	10.39
	Low	6.09	5.91	3.05	4.29	3.29	6.04	2.25	3.90	5.03
7	Mean	9.34	6.86	4.47	6.19	3.87	6.19	2.94	4.45	7.76
	High	13.66	7.88	5.87	7.46	6.00	8.41	4.05	4.96	9.27
	Low	3.88	4.41	2.62	3.69	2.28	4.03	2.37	3.88	5.97
8	Mean	10.85	6.79	4.58	6.45	5.54	7.57	2.92	4.95	10.20
	High	13.94	8.02	5.96	7.06	6.96	8.96	3.91	6.02	12.02
	Low	6.72	5.08	3.05	4.58	4.01	5.34	2.41	3.97	6.51
9	Mean	6.10	5.93	5.22	6.06	4.30	5.43	3.05	6.73	9.20
	High	10.12	6.79	5.65	7.74	5.29	6.83	3.68	7.10	11.29
	Low	2.89	4.67	4.77	4.00	3.42	4.20	2.59	6.25	6.22
10	Mean	8.10	6.95	4.86	6.46	3.34	5.68	2.98	5.61	9.22
	High	11.86	8.13	5.53	8.47	4.34	6.92	3.99	7.17	11.95
	Low	3.53	5.59	3.78	2.98	2.66	4.45	2.38	3.78	5.19

variance data indicated that after the dimensionality reduction a considerable amount of information would be lost. Therefore, we decided to consider all distance parameters for each compound investigated and to use the gridded data space algorithm option *FAMILY* of SYBYL [15]. This procedure permitted a very fast definition of groups within the conformers. Interestingly, the most important cluster for each molecule was populated by 60–90% of the complete set of conformers. The mean distance values of these predominant conformational families exhibited high convergence and are listed in Table 4.

These results could be a hint for the existence of a resembling predominant conformation in all compounds. However, detailed conformational comparison of the main clusters between different molecules in this study revealed no significant overall similarity for all compounds investigated. Therefore we assumed that a self-consistent pharmacophore model for the inhibitors investigated was hardly deducable using this approach.

In order to gain more insight into the conformational behaviour of compounds **1–10**, a search using a Monte Carlo-like algorithm was performed. The global minima located by this procedure were found to be significantly lower than those obtained after full energy minimization of randomly selected conformers from MD (potential energies of the global minima obtained with MD and the randomized search method are listed in Table 5). Moreover, a high degree of similarity between the global minima structures of different compounds suggested that a reliable pharmacophore model could be obtained using these conformers as starting coordinates. Therefore we used these templates for the 3D-QSAR analysis.

Since the CoMFA approach is very sensitive to small differences between the compounds used for the construction of the model, before analysis, a field fit procedure was done in order to minimize the non-structural variance of the molecules. A stereoview of the superposed compounds **1–10** is given in Fig. 2.

The prolyl endopeptidase inhibitors dataset was examined in CoMFA with combined electrostatic and steric fields. Crossvalidated r^2 (r^2_{cv}) describes the relative predictive capability of a model that is strongly unsymmetric, i.e. has many more predictor variables than dependent observables. 'Press' is a measure for the difference between predicted and actual values from a crossvalidated analysis [18]. The r^2_{cv} value obtained of 0.429 (press = 0.631, 2 components) is within a range where satisfactory predictivity can be expected. In order to evaluate our model, a set of test compounds **11–13** (Fig. 3) was used to predict activities running a non-crossvalidated PLS analysis (r^2 = 0.825, s = 0.411). The parameters used for the CoMFA analysis are reported in Table 6; the predicted vs. the experimental affinity data of the training set compounds are listed in Table 7. As shown in the affinity correlation graph, acceptable predictivity of the CoMFA model can be assumed.

This 3D-QSAR study allowed us to establish a possible pharmacophore model of the inhibitors of prolyl endopeptidase investigated. As can be deduced from the contoured CoMFA coefficient, maps for the steric and electrostatic contribution to QSAR, regions with important steric interaction are located in the range of the residue of the central amino acid moiety. A 5-membered ring at this position as present in compounds **1, 2, 3** and **9** seems favourable for high activity. Molecules having a side chain with more than 4 carbon atoms at this position could interfere with a zone of unfavourable steric interactions; the relatively low IC_{50} of compound **10** could be in part explained by this fact. Moreover, our model seems to confirm the findings of Tsuru et al. [7] that the steric dimension of the N-blocking group is limited to a certain space. The existence of a 'forbidden' steric zone placed above the aromatic systems suggests, that in the case

TABLE 5
POTENTIAL ENERGIES (KCAL/MOL) OF GLOBAL MINIMA CONFORMERS OF COMPOUNDS **1–10** FOUND BY MD AND BY RANDOM SEARCH

Compound	MD	Random search	Compound	MD	Random search
1	11.879	8.212	**6**	5.347	3.577
2	17.857	15.419	**7**	6.129	5.616
3	21.659	15.480	**8**	16.681	16.076
4	12.345	7.714	**9**	19.972	18.323
5	9.395	6.337	**10**	10.694	10.530

TABLE 6
PARAMETERS USED FOR THE CoMFA ANALYSIS

Dimension of the box:	15.25 Å × 17.69 Å × 19.49 Å
Gridspacing:	2 Å; Points: 722
Probe atom:	C.3; Charge: 1.0
Minimum_sigma value:	2.0 (crossvalidated), 0.0 (final analysis)
Crossvalidation groups:	9
Components used in the final analysis:	2
Fields:	steric, electrostatic
Type of dielectric function:	distance
Electrostatic term inside steric map:	not dropped
Lone pairs included:	in both electrostatic and steric field
Maximum steric energy value:	30.0 kcal
Maximum electrostatic energy value:	30.0 kcal
Repulsive VDW term:	12
No volume averaging PLS algorithm:	Nipals, 100 ltr., EPS = 0.0001

of compounds **8, 9,** and **10** conformers are existing in which a part of the phenyl rings may interfere with this space and therefore could be responsible for the low activity of these molecules. The electrostatic contribution coefficient contour plot clearly indicates a strong polarization necessary in the central part of the molecules for high activity as well as the importance of the amide function connecting the C-terminal amino acid to the rest of the molecule. The high activity of compounds **1, 2, 3,** and **4** may be attributed to the rest of the molecule. The high activity of compounds **1, 2, 3,** and **4** may be explained additionally in terms of electrostatic interaction by the aligned orientation of the carbonyl oxygen lone pairs forced in a negatively polarized region. On the other hand, a strongly positive polarized zone can be detected at the opposite side of this region. The lack of functional groups responsible for the formation of such an intramolecular polarization in compound **8** may account for the low activity, which is in total agreement with the model. The contoured CoMFA coefficient maps for the steric and electrostatic contributions to

TABLE 7
AFFINITIES OF COMPOUNDS **1–10** EXPERIMENTAL DATA VS. PREDICTED DATA (NON-CROSSVALIDATED PLS ANALYSIS)

Compound	Log ($1/IC_{50}$) experimental	Log ($1/IC_{50}$) predicted
1	1.46	1.36
2	0.13	0.10
3	−0.15	−0.17
4	−0.28	−0.30
5	−0.76	−0.77
6	−1.20	−1.16
7	−1.36	−1.35
8	−1.90	−1.88
9	−1.90	−1.85
10	−2.08	−2.03

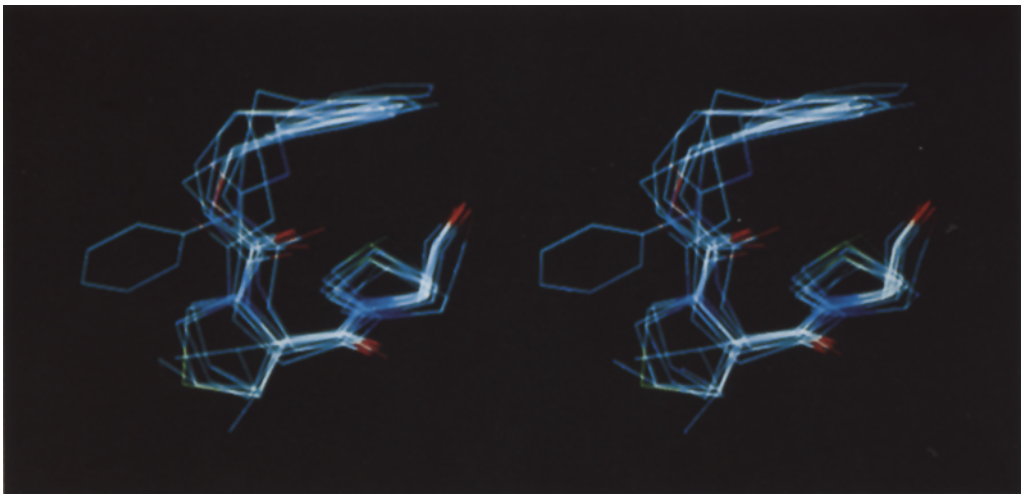

Fig. 2. Stereoview of compounds **1–10** as used for construction of the CoMFA model.

QSAR are shown in Fig. 4 and in Fig. 5. In the steric field coefficient plot, green regions represent zones where increased interaction leads to higher biological activity. Red regions represent zones, where increased interaction would lead to lower activity. In the electrostatic field coefficient plot, negatively polarized zones are contoured in blue and positively polarized regions in yellow. Enhanced negative or positive charge in the negatively or positively polarized region, respectively, would lead to higher biological activity and vice versa.

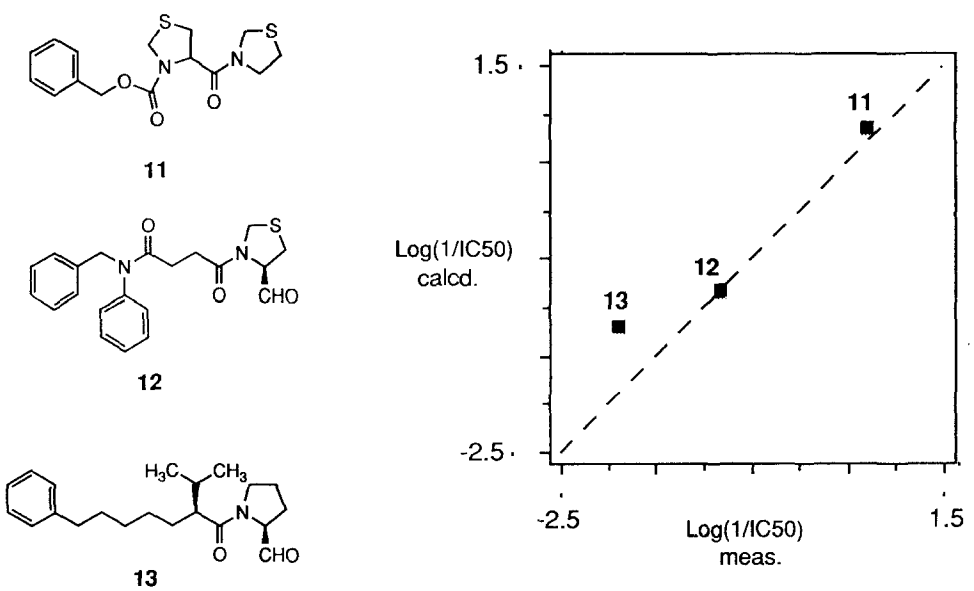

Fig. 3. Test set compounds and affinity correlation graph.

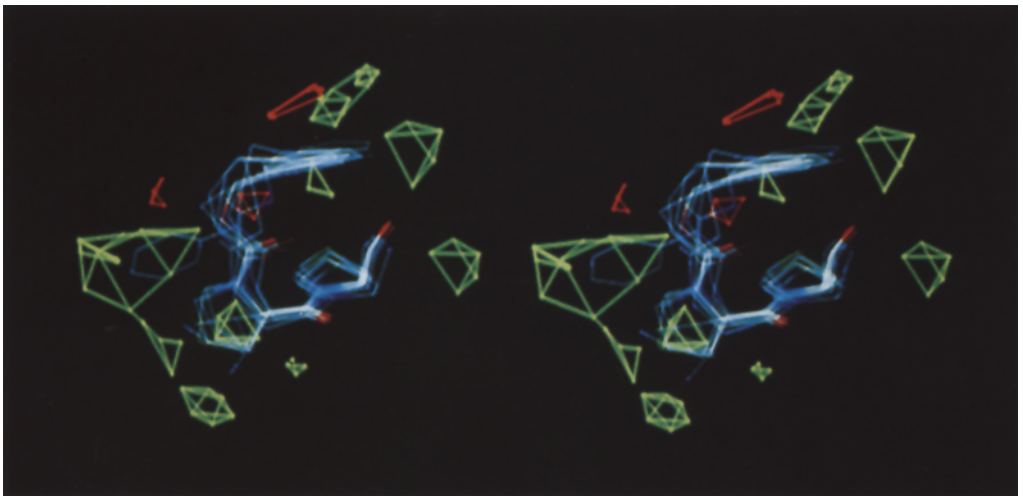

Fig. 4. Contoured CoMFA coefficient maps for the steric contribution to QSAR. The map is calculated from the three-field CoMFA experiment at 2 Å grid spacing. Green regions represent molecular modifications where increased steric interaction leads to an increased target value (i.e. biological activity); red regions represent zones where steric interaction leads to a decreased target value.

CONCLUSION

In this study, molecular dynamics has been shown to be a valuable method to produce a representative sample of conformations permitting the comparison and the classification of different compounds with common biological properties in simple geometric terms. The randomized conformational search procedure was able to detect global minimum structures with lower poten-

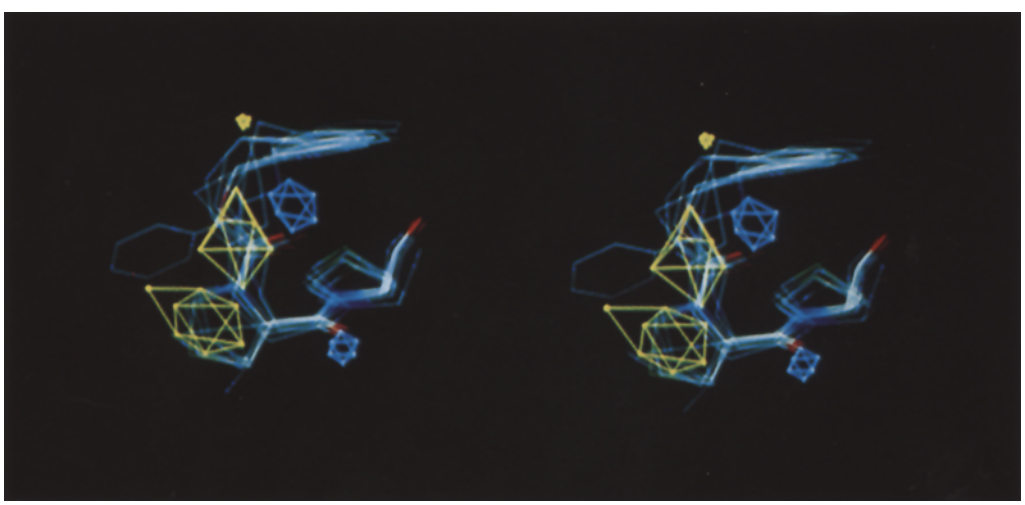

Fig. 5. Contoured CoMFA coefficient map for the electrostatic contribution to QSAR. Negatively polarized zones are contoured in blue and positively polarized ones in yellow. Enhanced negative charge in the negative polarized region, or positive charge in the positive polarized one, respectively, leads to a higher target value (i.e. increased biological activity).

tial energy than those found by MD. This fact seems to be a consequence of the algorithm used in such an analysis, since the main goal represents the finding of the global minimum structure rather than the statistically correct sample of the conformational hyperspace. On the one hand, the results obtained from the dynamics simulation runs suggested predominant clusters with highly similar geometric features existing from all compounds investigated. On the other hand, the conformational similarity between different compounds obtained by this iterative procedure of generating and classifying conformers seemed not to be high enough to deduce a reliable 3D-QSAR, based on these results. The structures in the lowest potential energy range afforded by the random search method were shown to be better suited for the construction of the CoMFA model. We are aware that the biologically active conformation may not necessarily be among the structures found in our study. However, the proposed CoMFA model may constitute a useful tool for the design of novel inhibitors of prolyl endopeptidase. Work on the synthesis of new compounds with high inhibitory activity predicted is in progress; the results will be published elsewhere.

ACKNOWLEDGEMENTS

T.L. gratefully acknowledges financial support by the Fonds zur Förderung der wissenschaftlichen Forschung (Erwin Schrödinger Stipendium J0629-MOB).

REFERENCES

1 Yoshimoto, T., Ogita, K., Walter, R., Koida, M. and Tsuru, D., Biochim. Biophys. Acta, 569 (1979) 184.
2 Rennex, D., Hemmings, B.A., Hofsteenge, J. and Stone, S.R., Biochemistry, 30 (1991) 2195.
3 a. Jackson, I.M.D., N. Engl. J. Med., 306 (1982) 145.
b. Griffiths, E.C. and Bennett, G.W., Thyrotropin Releasing Hormone, Raven Press, New York, 1983.
4 Tsuru, D. and Yoshimoto, T., Bio. Ind., 4 (1987) 788.
5 Langer, T., Zhang, D., Rival, Y. and Wermuth, C.G., manuscript in preparation.
6 Cramer III, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
7 Tsuru, D., Yoshimoto, T., Koriyama, N. and Furukawa, S., J. Biochem., 104 (1988) 580.
8 Saunders, M.J., J. Am. Chem. Soc., 109 (1987) 3150.
9 Chang, G., Guida, W.C. and Still, W.C., J. Am. Chem. Soc., 111 (1989) 4379.
10 Ferguson, D.M. and Raber, D.J., J. Am. Chem. Soc., 111 (1989) 4371.
11 Osguthorpe, D.S., Dauber-Osguthorpe, P., Sessions, R.B., Paul, P.K.C. and Burney, P.A., In Leeming, P.R. (Ed.), Topics in Medicinal Chemistry (RCS Special Publ. No. 65), Royal Society of Chemistry, London, 1988.
12 Lybrand, T.P., McCammon, A. and Wipff, G., Proc. Natl. Acad. Sci. U.S.A., 83 (1986) 833.
13 Auffinger, P. and Wipff, G., J. Comput. Chem., 11 (1990) 190.
14 Hudson, B.D., George, A.R., Ford, M.G. and Livingstone, D.J., J. Comput.-Aided Mol. Design, 6 (1992) 191.
15 Version 5.5, Tripos Assoc., St. Louis, MO.
16 Clark, M., Cramer III, R.D. and Van Opdenbosh, N., J. Comput. Chem., 10 (1989) 982.
17 Berthod, H. and Pullmann, A., J. Chem. Phys., 62 (1965) 942.
18 a. Wold, S., Ruhe, A., Wold, H. and Dunn III, W.J., SIAM J. Sci. Stat. Comput., 5 (1984) 735.
b. Wold, S., Technometrics, 20 (1987) 397.
19 Wilk, S. and Orlowski, M., J. Neurochem., 41 (1983) 69.
20 Toda, M., Ohuchida, S. and Ohno, S., Eur. Pat., 0 268 190 A1 (1987).
21 Toda, M., Ohuchida, S. and Ohno, S., Eur. Pat., 0 280 956 A1 (1988).
22 Toda, M., Ohuchida, S. and Ohno, S., Eur. Pat., 0 275 482 A2 (1987).
23 Tanaka, T., Saitoh, M., Higushi, N. and Hashimoto, M., Eur. Pat., 0 201 743 A2 (1986).