3D-QSAR/CoMFA Models as a Tool for Biocatalysis and Protein Engineering

J. D. Carballeira Rodríguez*, E. Alvarez Ruiz[#], and J. V. Sinisterra Gago
Biotransformations Group, Organic and Pharmaceutical Chemistry Department. Faculty of Pharmacy, Universidad Complutense de Madrid, Plaza de Ramón y Cajal s/n, 28040 Madrid, Spain. http://www.ucm.es/info/btg/
[#]Molecular Screening Department. Centro de Investigación Básica – GlaxoSmithKline, P.T.M., 28760 Tres Cantos, Madrid, Spain.

Original scientific paper
Received: August 26, 2005
Accepted: March 3, 2006

The x-ray structure of an enzyme is taken into account, when available, as the reference model to explain catalytic activity and selectivity. Unfortunately, in most of the cases the structure is available only as apostructure, i.e. without the substrate bound to the active site, and it is strange to find many different enzyme-substrate complexes of a specific enzyme as crystals. Moreover this structure is not the "real" structure of the protein during catalysis as the crystal is stationary.

In this paper we propose the use of CoMFA models to evaluate the differences between the crystal and the real structure of the enzyme under reaction conditions.

In addition to the stationary nature of a crystal, the experimental limitations of crystallographic techniques to obtain crystals in a fast and reliable manner, give a chance to the creation of CoMFA models by evaluating the easy to obtain catalytic properties of enzyme variants to provide information about the structural changes produced by the mutations. By means of the evaluation of different structures as substrates CoMFA models will not only provide information about the structure of the enzyme, but also about the flexibility and potential conformational changes of the substrate binding site.

Key words:
Biotransformations, activity models, enzyme structure

Introduction

Comparative Molecular Fields Analysis (CoMFA) is a method for 3D quantitative structure-activity relationships (3D-QSAR) developed at Tripos. Although the concept of the approach has been known as DYLOMMS (dynamic lattice-oriented molecular modelling system)[1] for over a decade, it was not until 90s that the method became widely used after it was named as CoMFA in 1988.[2-3] The methodology has been patented and the program is available as a QSAR package in SYBYL (Tripos Inc.; 1699 South Handley Rd., St. Louis, MO 63144).

Comparative Molecular Field Analysis (CoMFA), applied to enzyme-catalyzed biotransformations as a 3D-QSAR methodology, may picture a semi-quantitative structure of the active site. CoMFA has proven its ability to predict biological properties of systems not amenable to direct analysis. The fundaments of this methodology are well described in the literature[4–10] but CoMFA has been applied only a few times in biocatalysis.[11–12] The CoMFA model of an enzyme is a 3D representation of the steric and electrostatic zones of the active site, and it is directly related to the structure of the enzyme, when the alignment of the substrates is based on their real docking poses. The alignment of the ligands in the proper orientation is the critical step of 3D-QSAR/CoMFA analysis. In biocatalysis, the model is built according to the conversions obtained in the experimental reactions as a measurement of the enzyme-substrate affinity.

In this article the aim of the authors is to explain in a comprehensive way some potential utilities of the models to improve the understanding of enzyme structure and mechanism:

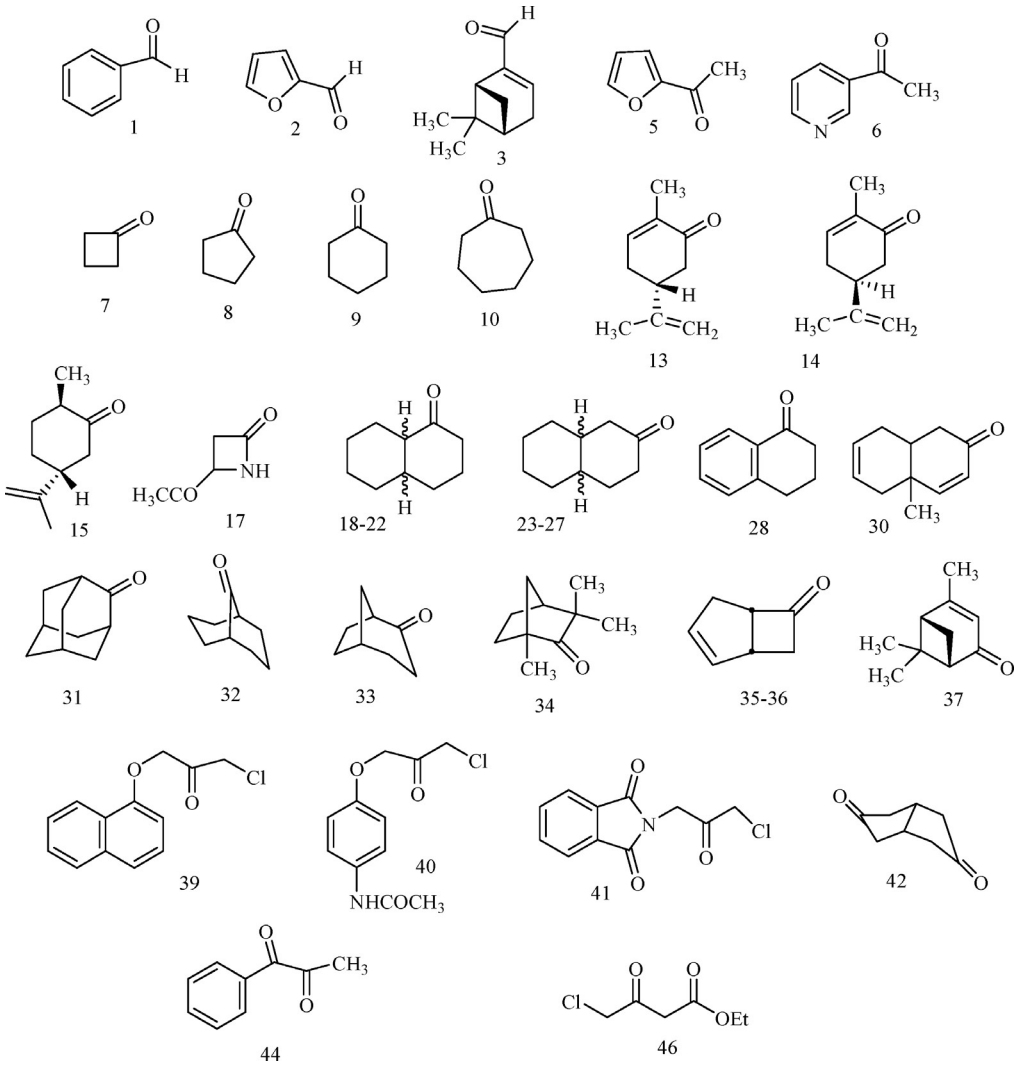
i) Application of the models, based upon empirical data, to properly dock substrates into the binding site of the enzyme.

ii) Creation of homology models taking into account catalytic data and not only structural similarities at the protein level.

iii) Comparison of different biocatalysts according to their substrate selectivity.

iv) Development of new approaches: A method, based on the statistical regression of the CoMFA models from different related enzymes or enzyme mutants is also proposed (Crosslink-CoMFA (Cr-CoMFA)).

* To whom correspondence should be addressed.

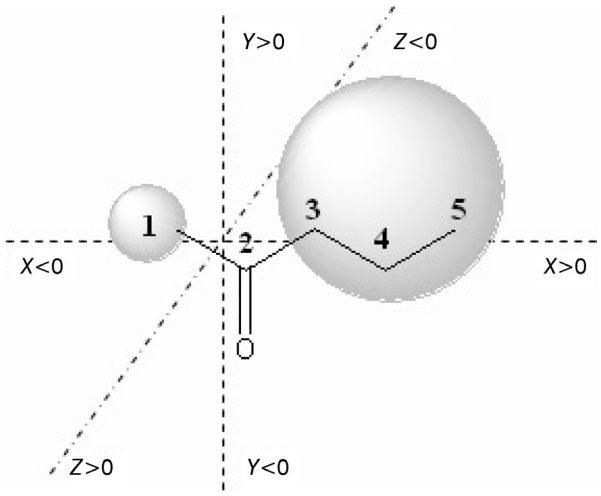


F i g . 1 – *Chemical structures of most of the carbonylic compounds used as substrates*

Development of the Crosslink CoMFA (Cr-CoMFA) approach

A collection of 416 micro-organisms from different taxonomic groups were tested versus a designed series of carbonylic compounds (Fig 1) in a hierarchical screening process looking for new active alcohol dehydrogenases.[13-14] From this work two filamentous fungi, *Diplogelasinospora grovesii* IMI 17018, *Gongronella butleri* CBS 157.25, and two yeasts, *Schizosaccharomyces octosporus* NCYC 427 and *Geotrichum candidum* NCYC 49, were selected due to their activity in the reduction of different compounds (Table 1).

From these data a 3D-QSAR/CoMFA model of each individual biocatalyst was generated selecting the alignment criteria explained in Fig 2. In every case q^2 (regression coefficient; over 0.5 the model is considered predictive) values over 0.5 were obtained. The statistical values from these analyses are shown in Table 2 and the CoMFA model ob-



F i g . 2 – *Criteria used for the alignment of the substrates. Using MULTIFIT within SYBYL® the molecules were aligned by different ways. The template substrate used for the study was 2-adamantanone. The best results were achieved superimposing the C = O to be reduced, then the bulkier part of the molecule is located in the X>0 (right) and the less voluminous in the X<0 (left) according to a coordinate system.*

Table 1 – *Results obtained using the selected microbial strains vs. the collection of substrates. [Substrate] = 2.5 m mol dm^{-3}, 250 min^{-1}, 28 °C, 48 h. The results are represented as conversion (%) of product (alcohol).*

N	Substrate[a]	*D. grovesii*	*G. butleri.*	*S. octosporus*	*G. candidum*
1	benzaldehyde	0	95	98	0
2	2-furaldehyde	98	98	0	0
3	*R*-myrtenal	73	90	42	0
4	tetradecanal	0	0	0	0
5	2-acetylfurane	0	11	39	80
6	2-acetylpyridine	14	13	73	0
7	cyclobutanone	77	67	10	0
8	cyclopentanone	14	33	2	0
9	**cyclohexanone**	**85**	**96**	**87**	**36**
10	cycloheptanone	9	26	8	0
11	cyclooctanone	0	18	0	0
12	cyclododecanone	0	25	0	0
13	4*R*-carvone*	90	87	65	77
14	4*S*-carvone*	93	51	60	82
15	**(1*R*,4*R*)-dihydrocarvone**	**79**	**88**	**75**	**76**
16	2-azetidinone	10	6	0	0
17	4-acetoxi-azetidin-2-one	12	0	0	0
18	1-decalone	90	65	34	76
19	(4a*R*, 8a*R*)-*cis*-1-decalone	43	36	3	86
20	**(4a*S*, 8a*S*)-*cis*-1-decalone**	**86**	**77**	**11**	**90**
21	(4a*R*, 8a*S*)-*trans*-1-decalone	96	71	8	41
22	**(4a*S*,8a*R*)-*trans*-1-decalone**	**96**	**86**	**14**	**92**
23	2-decalone	65	57	75	72
24	(4a*R*,8a*S*)-*cis*-2-decalone	81	0	86	20
25	**(4a*S*,8a*R*)-*cis*-2-decalone**	**24**	**65**	**90**	**26**
26	**(4a*R*,8a*R*)-*trans*-2-decalone**	**70**	**26**	**72**	**25**
27	(4a*S*,8a*S*)-*trans*-2-decalone	73	12	18	5
28	α-tetralone	0	6	9	5
29	**4-methyl-decal-2-en-1-one**	**3**	**1**	**0**	**0**
30	**4a*R*-methyl-decal-2-one**	**3**	**2**	**0**	**0**
31	**2-adamantanone**	**58**	**98**	**22**	**12**
32	**bicyclo-[3.3.1]-nonan-9-one**	**39**	**95**	**32**	**16**
33	bicyclo-[3.2.1]-octan-2-one	74	32	10	0
34	***R*-fenchone**	**0**	**0**	**0**	**0**
35	*cis*-bicyclo-[3.2.0]-hept-2-en-6-one	20	44	24	0
36	*trans*-bicyclo-[3.2.0]-hept-2-en-6-one	1	7	4	0
37	**Verbenone**	**0**	**0**	**0**	**0**
38	**2-aza-bicyclo[2.2.1]hept-5-en-3-one**	**0**	**2**	**0**	**0**
39	1-chloro-3-(1-naphtiloxy)-propan-2-one	99	0	93	9
40	N-[4-(3-chloro-2-oxo-propiloxi)-phenyl]-acetamida	99	98	0	0
41	1-chloro-3-(phtalimidyl)-propan-2-one	68	0	0	0
42	***cis*-bicyclo[3.3.0]octan-3,7-dione**	**0**	**0**	**0**	**0**
43	acetyl-ferrocene	0	0	0	0
44	1-phenyl-propano-1,2-dione	93	2	0	0
45	cholesten-3-one	0	0	0	0
46	4-chloro-ethyl-aceto-acetate	0	95	98	No

[a]The experimental and analytical conditions for these reactions have been previously described.[13-14]

R*-carvone and *S*-carvone were not considered for the Cr-CoMFA due to the reduction of the double bond that appears to certain extent in some of the catalyst, instead the dihydrocarvone **15 was considered.

Table 2 – *Statistical parameters of the CoMFA models of the biocatalysts tested*

Parameter	Analysis	*D. grovesii*	*G. butleri.*	*S. octosporus*	*G. candidum*
q^2	Leave one out	0.549	0.625	0.555	0.648
N° of components	Leave one out	7	11	14	8
R^2	No validation	0.991	1	0.995	1
F values	No validation	280.575	3405	472123	66008
Prob. $R^2 = 0$	No validation	0	0	0	0

tained from *G. candidum* alcohol dehydrogenase activity is shown as example in Fig 3.

During the evaluation of the results obtained in the reduction reactions and the CoMFA models we realized that the pattern of substrate affinity was similar between the different strains. Especially interesting was the high activity in the reduction of 1-decalone **18** and 2-decalone **23** while no one of the biocatalysts was able to reduce related structures as **28**, **29** and **30**. This fact was astonishing as far as alcohol dehydrogenases present in microbial strains, distant from the taxonomic point of view, were showing considerable similarities in substrate scope.

In the bibliography different publications point out the structural relationships between the different alcohol dehydrogenases and even the hypothesis about the existence of a common ancestor,[15-16] based in structural studies of different proteins of the same family.

The hypotheses suggested by these structural studies were similar to the conclusions reached evaluating our biocatalytic data, this was quite interesting for us and we attempt to go one step further as a test probe, in that sense we tried to developed a methodology that we call Crosslink CoMFA (Cr-CoMFA).

The Crosslink CoMFA approach

Cr-CoMFA is a method to obtain a model displaying the common characteristics of substrate affinity from a series of biocatalysts. The common pattern of substrate affinity may be obtained by the 3D statistical regression of the different single CoMFA models using techniques as partial least squares PLS.[2,7-9] As far as we know there is not any specific tool available for this purpose, due to this fact we decided to face our goal using an indirect

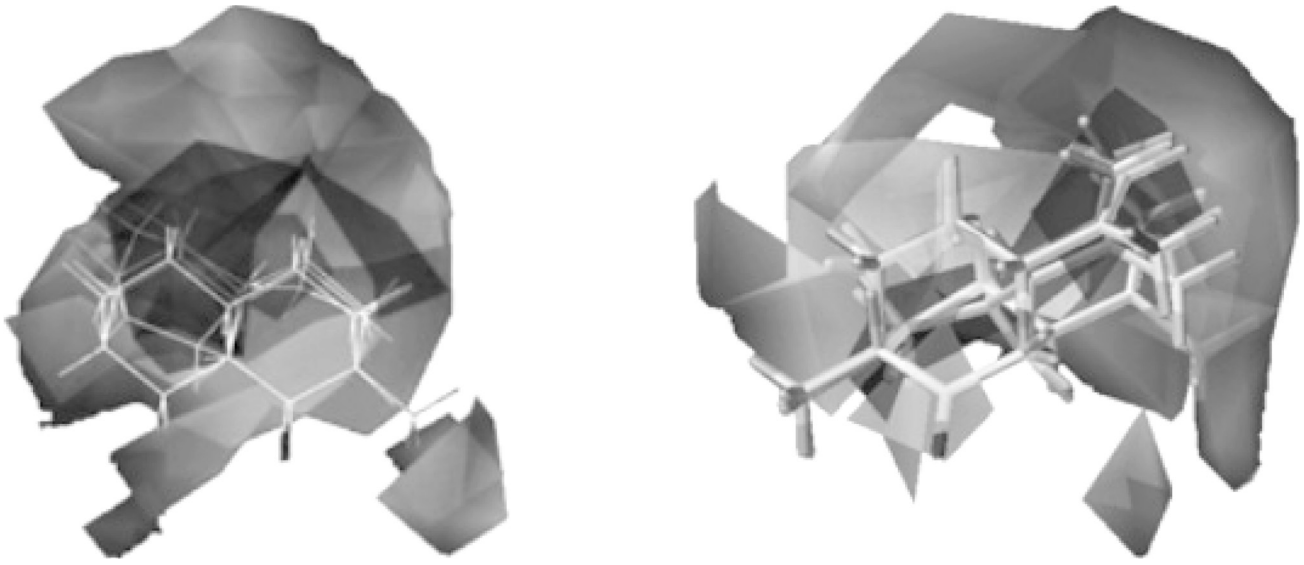


Fig. 3 – *3D-QSAR model of G.candidum (left) and S. octosporus (right) alcohol dehydrogenase activity. The colour code is as follows: Steric hindrance zones: green areas depict zones of space where occupancy by the substrates increases affinity, whereas yellow areas depict zones where occupancy decreases affinity. Electrostatic zones: areas where a high electron density provided by the ligand increases (red) or decreases (blue) the activity.*

method: A new data table was built containing only the data corresponding to those substrates where every biocatalyst presented the same catalytic behaviour (considering active >10 % conversion and not active substrates where the conversion is below 5 %). The substrates selected underlined in table 1. From this table the Cr-CoMFA model was generated with a q^2 of 0.8 (Tab 3), higher than the ones obtained for the individual models. The Cr-CoMFA model is displayed in Figure 4, where the important regions of steric hindrance (Lennard Jones) and electrostatic (coulombic) effects are well defined, displaying the overall characteristics common to all the enzymes tested.

In summary the Cr-CoMFA model shows the common pattern of substrate affinity of the considered biocatalysts. In specific cases, as when considering variants obtained from a wild-type enzyme by directed evolution or natural enzymes from the same family evolving from the same ancestor, the Cr-CoMFA models indicate the substrate scope characteristics conserved during evolution. The comparison of the CoMFA models of each enzyme and the Cr-CoMFA model generated from them could provide additional information to enhance our knowledge about evolution and directed evolution, as explained afterwards.

A diagram with a graphical overview of the Cr-CoMFA method is shown in Fig 5.

The analysis of the data from screening approaches by means of software tools like Sybyl/CoMFA could be interesting for protein engineers, especially in the case of evolutionary approaches, where is impossible to crystallize all the structural diversity generated, but where the catalytic profile, i.e. bioconversions against different structures, is easy to obtain.

Table 3 – *Statistical parameters of the Cr-CoMFA model*

Parameter	Analysis	*Cr-CoMFA*
q^2	Leave one out	0.798
N° of components	Leave one out	2
R^2	No validation	0.95
F values (n_1 = 2, n_2 = 11)	No validation	97.618
Prob. R^2=0 (n_1 = 2, n_2 = 11)	No validation	0

Enzyme mechanisms, CoMFA models and x-ray structure

The protein-ligand interaction were described at first by the key and lock theory of *Emil Fischer* in 1894,[17] where key (substrate) and lock (enzyme) were considered as rigid structures. In 1958 *Daniel E. Koshland* published the induced fit theory,[18] proposed in the following terms:[19] a) The precise orientation of the catalytic groups is required for enzyme action, b) the substrate causes an appreciable change in the three dimensional relationship of the amino acids at the active site, and c) the changes in the protein structure caused by the substrate will bring the catalytic groups into the proper alignment, whereas a nonsubstrate will not. Subsequently, the solvation co-solvation model[20] was developed considering as well the flexibility of the substrate binding site.

The "key and lock" model was evolved to the "hand in glove" model, which included Fischer idea of a fit but including the flexibility concept.

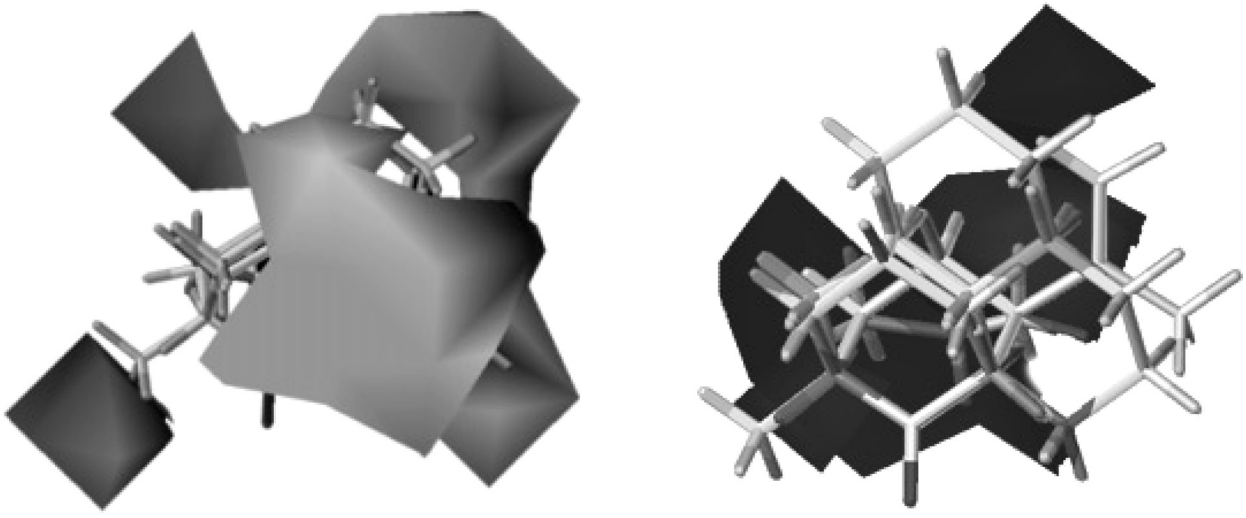


Fig. 4 – *Cr-CoMFA model generated. Steric hindrance zones: green areas depict zones of space where occupancy by the substrates increases affinity, whereas yellow areas depict zones where occupancy decreases affinity. Electrostatic zones: areas where a high electron density provided by the ligand increases (red) or decreases (blue) the activity.*

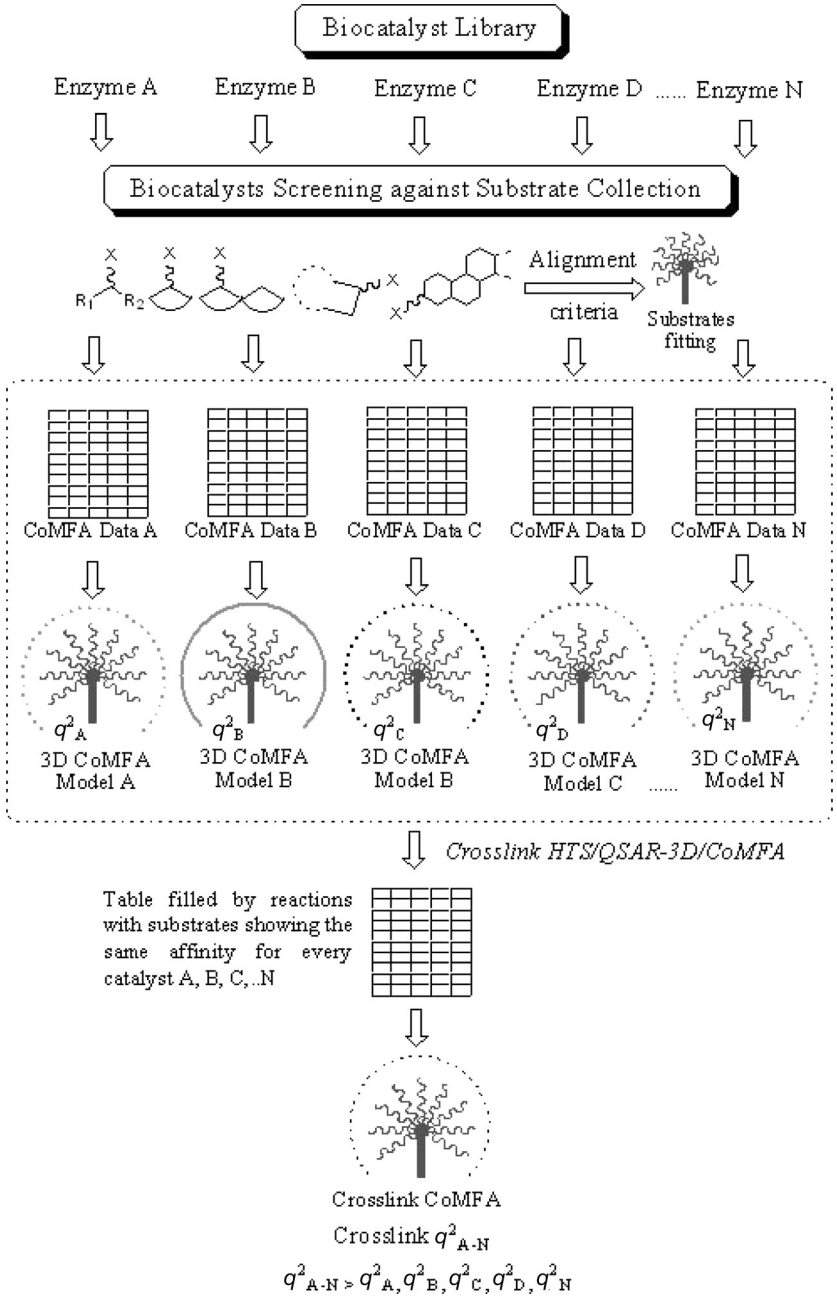


Fig. 5 – *Crosslink CoMFA. Overview.*

Today almost every enzyme has been shown to undergo significant ligand induced changes.[21] This potential flexibility of the substrate binding site of the enzymes is also a limitation when using the x-ray structure as reference even when a specific substrate is included in the pdb file. Different substrate structures shall lead to different binding site conformations, this fact supports the idea of using CoMFA models and fitting models (aligned substrates) of the accepted structures to enhance the understanding about the binding site behaviour, and add new information to the pdb files (x-ray structures), according to a global evaluation of the enzyme capabilities.

3D-QSAR/CoMFA models and biocatalysis: Potential applications

The building of 3D-QSAR/CoMFA models has several potential utilities that justify the use of this theoretical method for biotransformations and directed evolution.

When an enzyme or library of enzymes is screened against a collection of substrates it is not easy to have a 3D overview of the capabilities of the enzyme. CoMFA is a good option to display a comprehensive model from the large amount of data generated, that could serve as reference for the rational interpretation of the results and, further-

more, for the selection a priori of new potential substrates. The results from further experiments would provide as well new information to refine the model.

3D-QSAR/CoMFA models and Directed Evolution: Potential applications

The use of enzymes in organic reactions is restricted due to their substrate specificity, so it could not be assumed as a general synthetic methodology. Nowadays, new biocatalysts could be created by directed evolution enhancing the activity and selectivity of a known enzyme structure[22-30] and overcoming some of the catalytical restrictions of the natural protein. There are two general ways to face the improvement of an enzyme characteristic by directed evolution:

i) site directed mutagenesis[31] (as a traditional protein engineering method). This method is strongly recommended when the crystal structure of an enzyme is available and it is the ideal option when that structure has been obtained with a reference substrate bound at the active site. Practically, molecular biology techniques permit the substitution of a specific amino acid by another one at a defined position. According to this, in site directed mutagenesis the researcher needs detailed information concerning the 3D structure of the active-site of the enzyme to predict the proper amino acid substitution and information about the more convenient amino acid to be inserted.[28] 3D-QSAR/CoMFA models could provide interesting data to complete the information required for a successful directed evolution process based on rational design, especially, when using a homology model as reference or when the crystal structure of the enzyme is only available as apostructure.

ii) random mutagenesis methods[32-35] as well known error-prone PCR, are easier to perform and strongly recommended as first step, especially when the crystal structure or the enzyme is not available and not reliable homology models could be built, but with a huge problem, the library size. This drawback can only be overcome when a selection system or a efficient high throughput screening system are available to detect the hits.

The evaluation of the structural changes of the enzyme mutants generated either by site-directed or random mutagenesis, is a major point to understand the enzyme behaviour. When it is possible to obtain the crystal structure of the wild-type enzyme and its mutants, critical information can be achieved.[36] However, it is not a trivial thing to obtain the crystal structure of the enzymes, as some enzymes are unstable in their purified form or really difficult to crystallize, e.g. *Pseudomonas aeruginosa* lipase.[37] In those cases where the obtaining of the crystal is not straight forward, the comparison of different CoMFA models generated from the catalytic profile of wild type and mutants may provide additional information.

a) *CoMFA and site directed mutagenesis*

CoMFA models are especially interesting when the crystal structure of the enzyme is not available with the substrate docked at the active site. In this case, it is necessary to calculate the proper position where the substrate should be located. This operation is the key step of rational design approaches in site directed mutagenesis, as the docked substrate is the main reference for selecting the amino acids that shall be replaced to modify the activity or selectivity of the enzyme.

Usually, a well accepted substrate is selected to locate it into the active site as reference supported by computational tools. Anyway, the limitation is that, several different orientations of the substrate at the active site could be justified. In this case CoMFA models generated from different docking poses may help in the proper location of the substrate at the substrate binding site.

b) *CoMFA and homology models: 3D-models as reference for the creation of homology models.*

The main limitation for the rational design of the mutations appears when the enzyme of interest does not have the x-ray structure available. In these cases, the first recommended step is to use an error-prone PCR[29,35] as a random mutagenesis method, that create large libraries of mutants that have to be screened afterwards. However, the different activities or selectivities of the enzyme variants created by random mutagenesis methods could not be explained in the absence of a reference structure

Due to this fact, when the x-ray structure of the enzyme is not available, a related protein of known structure is used as template to build up a homology model. This homology is sometimes established taking into account the primary structure of the protein, and assuming from the beginning that the tertiary structure of the enzyme could be strongly different. As well, it has been widely described in literature that topology and homology sometimes are not related, so same overall structural topology could correspond with statistically insignificant sequence homology.[38] CoMFA models represent the enzyme affinity to different substrates according to i) electrostatic and steric properties and ii) organic reactions yields, giving as a result a 3D map of the

topology, as steric and electrostatic fields, of the active site involved in the biotransformation process.

The comparison of the CoMFA models of the enzyme under analysis and the homolog enzyme used as template should provide further data about the real homology degree of both enzymes from the catalytic point of view. Consequently the homology model may be refined at the active site level. This is the fundamental region where the evolutionary approaches are successful as can be seen in the recent reviews about directed evolution.[39-43] The summary of the work published point out that the most critical mutations in an enzyme for modifying substrate scope, for enantioselectivity enhancement and even for biocatalytic promiscuity,[40-41] lie in a radius of 1 nm (10 Å) around the binding pocket.[42-43] Nowadays the creation of focused libraries by rational design is considered the most reliable approach in directed evolution.[44] From our point of view, in the evolution of an enzyme by rational design, amino acid positions close to the active site should be considered first, and then remote positions. So the way to focus over the protein structure should follow a microscope-like approach, where the coarse focus knob (closer mutations) has to be used first to obtain a preliminary image (desired effect) that could be improved using the fine focus knob (remote mutations). The proper combination of focuses libraries (for close mutations) and random ep-PCR approaches (looking for not expected effects far away) will lead in general to successful results.

c) *CoMFA and evolution. Enzyme ancestors*

According to the divergent evolution theory the enzymes belonging to a specific class may present a related structure, more similar as closer they are in evolution. Some examples can be found in the literature about enzyme families that are supposed to evolve from a common ancestor, as alcohol dehydrogenases[17-18,45] kinases[46] peptidases, lipases, estereases, epoxide hydrolases and other serine hydrolases (PLEES).[47]

In this paper we suggest an indirect method, based in the use of the information provided from the CoMFA models of different evolutionary-related enzymes, to create an interpolation model (Cr-CoMFA) that will show the basic electrostatic and steric regions of the active site that are common, and which ones are not; namely which ones are conservative (critical for the activity) and non-conservative during evolution. This could support the creation of a 3D model of a simplified enzyme-ancestor structure, preserved in the actual enzymes.

Conclusions

The combined use of x-ray structures of the enzymes and the CoMFA models obtained from biocatalytic data could be of great interest to understand the real shape and the potential conformational changes of the substrate binding of the enzymes under reaction conditions. Due to this fact we propose the development of a software package able to handle the data from CoMFA, Cr-CoMFA and derived methods. This could represent a major improvement for protein databases as more information about the substrate profile, as well as analogies and differences between enzymes and enzyme variants, may be available.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the Centro de Investigación Básica of GlaxoSmithKline Pharmaceuticals in Tres Cantos (Madrid) and the cooperation of Dr. Leonardo Pardo and Dr. Mercedes Campillo from the Computational Medicine Laboratory of the Faculty of Medicine – Universidad Autonoma de Barcelona.

References

1. *Wise, M., Cramer, M. R. D., Smith, D., Exman, I.,* Progress in three-dimensional drug design: The use of real-time colour graphics and computer postulation of bioactive molecules in DYLOMMS. in J. C. Deardon, ed. Quantitative Approaches to Drug Design. Elsevier, Amsterdam. **1983**, pp. 145-146.
2. *Cramer, R. D., Bunce, J. D,. Patterson, D. E.,* Quant. Struct.-Act. Relat. **7** (1988) 18.
3. *Cramer, R. D., DePriest, S. A., Patterson, D. E., Hecht, P.,* The Developing Practice of Comparative Molecular Field Analysis, in "3D QSAR in Drug Design," Kubinyi, H., Ed.; ESCOM; Leiden, **1993**, pp. 443-485.
4. *Klebe, G., Abraham, U.,* J. Med. Chem. **36** (1993) 70.
5. *Cornell, W. D., Cieplak, P., Bayly, C., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W,. Kollman, P. A.,* J. Am. Chem. Soc. **117** (1995) 5179.
6. *Miertus, S., Scrocco, E., Tomasi, J.,* Chem. Phys. **55** (1981) 117.
7. *Dunn, W. J., Wold, S., Edlund, U., Hellberg, S., Gasteiger, J.,* Quant. Struct.-Act. Relat. **3** (1984) 131.
8. *Wold, S., Ruhe, A., Wold, H., Dunn, W. J.,* SIAM J. Sci. Stat. Comput. **5** (1984) 735.
9. *Wold, S., Albano, C., Dunn, W. J., Edlund, U., Esbensen, W., Geladi, P., Hellberg, S., Johansson, E., Lindberg, W., Sjöström, M.,* Multivariate data analysis in chemistry. In Chemometrics: mathematics and statistics in chemistry; B. R. Kowalsky, Ed.; Reidel: Dorrdrecht, The Netherlands, 1984, pp. 17-95.
10. *Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Zakrzewski, V. G., Montgomery, J. A. Jr., Stratmann, R. E., Burant, J. C., Dapprich, S., Millam, J. M., Daniels, A. D., Kudin, K. N., Strain, M. C., Farkas, O., Tomasi, J., Barone, V., Cossi,*

M., Cammi, R., Mennucci, B., Pomelli, C., Adamo, C., Clifford, S., Ochterski, J., Petersson, G. A., Ayala, P. Y., Cui, Q., Morokuma, K., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Cioslowski, J., Ortiz, J. V., Baboul, A. G., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Gomperts, R., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A,. Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Andres, J. L., Gonzalez, C., Head-Gordon, M., Replogle, E. S., Pople, J. A., Gaussian, Inc., Pittsburgh PA, 1998.

11. *Faber, K., Griengel, H., Höning, H., Zuegg, A*., Biocatalysis **9** (1994) 227.
12. *French, J. B., Holland, G., Holland, H. L., Gordon, H. L*., J. Mol Cat: B, **31**(4-6) (2004) 87.
13. *Carballeira, J. D.*, Thesis Ph. D. Universidad Complutense de Madrid. 2003.
14. a) *Carballeira, J. D., Valmaseda, M., Alvarez, E., Sinisterra, J. V.,* Enz. Microb. Technol, **34** (2004) 611. b) *Carballeira, J. D., Alvarez, E., Campillo, M., Pardo, L., Sinisterra, J. V.,* Tetrahedron.: Asymmetry, **15**(6) (2004) 951.
15. *Jörnvall, H.,* Eur. J. Biochem. **72** (1977) 443.
16. *Eklund, H., Bränden, C. I., Jörnvall, H.,* J. Mol. Biol. **102** (1976) 61.
17. *Fischer, E.,* Ber. Deutsch. Chem. Gesellsch. **27** (1894) 2985, 3479.
18. *Koshland, D. E.,* Jr. Proc. Natl. Acad. Sci. USA **44** (1958) 98.
19. *Koshland, D. E.,* Angew. Chem. Int. Ed. Engl. **33** (1994) 2375.
20. *Dewar, M. J. S.,* Enzyme **36** (1986) 8.
21. *Gerstein, M., Lesk, A. M., Clothia, C.,* Biochemistry **33** (1994) 6739.
22. *Reetz, M. T.,* Prot. Eng, **388** (2004) 238.
23. *Reetz, M T., Jaeger, K. E.,* Superior Biocatalysts by Directed Evolution, Springer-Verlag Berlin Heidelberg. *W. D. Fessner* (ed.). Biocatalysis, from discovery to application. Springer Verlag. **2** (2000) 31.
24. *Schmidt, M., Baumann, M., Henke, E., Konarzycka-Bessler, M., Bornscheuer, U. T.,* Prot. Eng. **388** (2004) 199.
25. *Franke, D., Hsu, C. C., Wong, C. H.,* Prot. Eng. **388** (2004) 224.
26. *Jaeger, K. E., Eggert, T.,* Curr. Opin. Biotechnol. **15**(4) (2004) 305.
27. *Farinas, E. T., Bulter, T., Arnold, F. H.,* Curr. Opin. Biotechnol. **12**(6) (2001) 545.
28. *Reetz, M. T,.* Tetrahedron **58** (2002) 6595.
29. *Arnold, F. H.,* Nature **409** (2001) 253.
30. *Petrounia, I. P., Arnold, F. H.,* Curr. Opin. Biotechnol. **11**(4) (2000) 325.
31. *Ferhst, A. R*., Biochemistry **26** (1987) 8031.
32. *Leung, D. W., Chen, E., Goeddel, D. V.,* Technique **1** (1989) 11.
33. *Eckert, K. A., Kunkel, T. A*., PCR Methods Appl. **1** (1991) 17.
34. *Cadwell, R. C., Joyce, G. F.,* PCR Methods Appl. **3** (1994) 136.
35. *Stemmer, W. P. C.,* Nature **370** (1994) 389.
36. *Rajakumara, E., Acharya, P., Ahmad, S., Shanmugam, V. M., Rao, N. M., Sankaranarayanan, R.,* Acta Crystallographica (2004) **D60** 160.
37. *Nardini, M., Lang, D. A., Liebeton, K., Jaeger, K. E., Dijkstra, B. W.,* J. Biol. Chem. **275** (2000) 31219.
38. *Warren, M. S., Benkovic, S. J.,* Prot. Eng. **10** (1997) 63.
39. *Miller, O. J., Dalby, P. A.,* Trends Biotechnol. **22** (2004) 203.
40. *Bornscheuer, U. T., Kazlauskas, R. J.,* Angew. Chem. Int. Ed. **43** (2004) 6032.
41. *Kazlauskas, R. J.,* Current Opinion in Chemical Biology **9** (2005) 1.
42. *Morley, K. L., Kazlauskas, R. J.,* Trends Biotechnol, **23** (2005). In Press.
43. *Park, S., Morley, K. L., Horsman, G. P., Holmquist, M., Hult, K., Kazlauskas, R. J.,* Chemistry & Biology **12** (2005) 45.
44. *Kazlauskas, R. J.,* Nature **436** (2005) 1096.
45. *Ohlsson, I., Nordström, B., Bränden, C. I.,* J. Mol. Biol. **89** (1974) 339.
46. *Eventoff, W., Rossman, M. G.,* Crit. Rev. Biochem. **3** (1975) 111.
47. *Puente, X., López-Otín, C.,* Biochem. J., **322** (1997) 947.