JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

J. Agric. Food Chem. **2003,** *51,* 1977−1983 **1977**

Use of Catalyst in a 3D-QSAR Study of the Interactions between Flavor Compounds and β-Lactoglobulin

ANNE TROMELIN*,†,§ AND ELISABETH GUICHARD†

INRA-UMRA, 17 rue Sully, 21065 Dijon Cedex, France, and Université de Bourgogne-UFR Pharmacie, 7 boulevard Jeanne d'Arc, 21000 Dijon, France

This paper reports a 3D-QSAR study using Catalyst software to explain the nature of interactions between flavor compounds and β-lactoglobulin. A set of 35 compounds, for which dissociation constants were previously determined by affinity chromatography, was chosen. The set was divided into three subsets. An automated hypothesis generation, using HypoGen software, produced a model that made a valuable estimation of affinity and provided an explanation for the lack of correlation previously observed between the hydrophobicity of terpenes and the affinity for the protein. On the basis of these results, it appears that aroma binding to β-lactoglobulin is caused by both hydrophobic interactions and hydrogen bonding, which plays a critical role. Catalyst appears to be a reliable tool for the application of 3D-QSAR study in aroma research.

KEYWORDS: Flavor; aroma; β-lactoglobulin; 3D-QSAR; Catalyst

INTRODUCTION

Interaction between flavor compounds and proteins has been studied for many years (*1*, *2*). β-Lactoglobulin is one of the best-characterized milk proteins (*3*−*7*) and belongs, together with mammalian odorant binding proteins (OBP), to the lipocalin superfamily (*8*). Binding of a variety of ligands, particularly flavor compounds (*10*−*14*, *15*) has been demonstrated (*9*).

Competition studies were performed for different ligands (*16*−*18*), but interpretation of the results was difficult due to the lack of information concerning the location of binding sites. There have been conflicting results as to the binding site of retinol to β-lactoglobulin. Considering the close structural resemblance of β-lactoglobulin and retinol-binding protein, one might conclude that retinol binds inside the central cavity (*6*). However, other studies are in favor of an external binding site (*12*). Both sites seem to exist with a preference of fixation to the central cavity, the external binding site being occupied by retinol only in the presence of other ligands (*19*). The accepted interaction model was mainly based on hydrophobic interaction (*11*, *20*). However, recent studies (*10*, *21*, *22*) suggest that other factors are probably involved in aroma−protein interaction.

In a previous study, interactions between β-lactoglobulin and 35 flavor compounds were analyzed (*21*). Affinity chromatography was used to determine binding constants of flavor substances belonging to five different chemical classes (esters, pyrazines, phenolic compounds, terpenes, and furans). Within one chemical class, affinity for β-lactoglobulin increases with hydrophobic chain length and log *P* values except for the terpenic class, for which another explanation has to be found.

The aim of the present study is to explain these values of dissociation constants using Catalyst, a recent software product created for pharmacophore design. This focuses the modeling on the molecular behavior of a ligand interacting with a receptor from the point of view of the receptor, but using information from only the ligand (*23*). Because this procedure requires the considered ligands to have the same receptor site, we applied this approach to the interaction of flavor compounds with β-lactoglobulin.

MATERIALS AND METHODS

Compounds and Physicochemical Data. The binding between β-lactoglobulin and 35 compounds belonging to four chemical classes (terpenes, phenols, pyrazines, and furans) was previously investigated in our laboratory by affinity chromatography studies, which provided binding constant values (K_b) (*21*). In the present work, we used dissociation constants ($K_d = 1/K_b$). These 35 compounds, classified by decreasing affinity, and their corresponding chemical families and physicochemical data (log *P*, K_b, and K_d) are reported in **Table 1**.

Computational Methods. The 35 compounds were built with Catalyst (Catalyst version 4.6 software; Accelrys Inc., San Diego, CA, August 1999) running on a Silicon Graphics workstation (SGI-O_2). Catalyst considers molecular flexibility by considering each compound to be as a collection of conformers. For each compound, the conformers were generated using Catalyst/COMPARE. The "best conformer generation" procedure was applied to provide the best conformational coverage for a maximum number of conformers generated defaulted to 250 in a 0−20 kcal/mol range from the global minimum. The particularity of the minimization using the best conformation generation routine in Catalyst is the "poling" function, which is added to the molecular mechanics CHARMm-like all-atom force field implemented in the program (*24*) as an additional term. Minimization is performed on the entire system (*25*).

* Author to whom correspondence should be addressed [telephone (33)-380693512; fax (33)-380693227; e-mail tromelin@arome.dijon.inra.fr].
† INRA-UMRA.
§ Université de Bourgogne-UFR Pharmacie.

10.1021/jf0207981 CCC: $25.00
Published on Web 02/27/2003

Table 1. Flavor Compounds, Their Chemical Families, and Physicochemical Data

no.	flavor compound	family	log P[a]	K_b[a]	$K_d = 1/K_b$
1	*trans*-3-oxo-*p*-menthane-8-thiol	terpene	3.1[g]	1461	6.8×10^{-4}
2	2-methoxy-4-(2-propenyl)phenol (eugenol)	phenol	2.58[c]	1360	7.4×10^{-4}
3	*cis*-3-oxo-*p*-menthane-8-thiol	terpene	3.1[g]	1208	8.3×10^{-4}
4	4-ethenyl-2-methoxyphenol (4-vinylguaiacol)	phenol	2.08[e]	1165	8.6×10^{-4}
5	α-menthone	terpene	3.01[g]	1138	8.8×10^{-4}
6	nerol	terpene	3[g]	1134	8.8×10^{-4}
7	3-*sec*-butyl-2-methoxypyrazine	pyrazine	1.62[c]	912	1.1×10^{-3}
8	4-ethylphenol	phenol	2.26[d]	888	1.1×10^{-3}
9	pulegone	terpene	2.46[g]	857	1.2×10^{-3}
10	4-ethyl-2-methoxyphenol (4-ethylguaiacol)	phenol	2.38[e]	830	1.2×10^{-3}
11	3-isobutyl-2-methoxypyrazine	pyrazine	1.62[c]	795	1.3×10^{-3}
12	(–)-carvone	terpene	1.91[g]	748	1.3×10^{-3}
13	linalool	terpene	2.91[g]	565	1.8×10^{-3}
14	(–)-carveol	terpene	2.6[g]	542	1.8×10^{-3}
15	α-terpineol	terpene	3.15[g]	483	2.1×10^{-3}
16	3-ethoxy-4-hydroxybenzaldehyde (ethylvanillin)	phenol	1.76[f]	475	2.1×10^{-3}
17	3-isopropyl-2-methoxypyrazine	pyrazine	1.12[c]	452	2.2×10^{-3}
18	4-methylphenol (*p*-cresol)	phenol	1.95[d]	440	2.3×10^{-3}
19	ethyl pentanoate	ester	2.21[b]	366	2.7×10^{-3}
20	4-hydroxy-3-methoxybenzaldehyde (vanillin)	phenol	1.26[d]	319	3.1×10^{-3}
21	ethyl 2-methylbutyrate	ester	2.01[b]	288	3.5×10^{-3}
22	ethyl 3-methylbutyrate	ester	2.01[b]	284	3.5×10^{-3}
23	2-methoxyphenol (guaiacol)	phenol	1.33[d]	245	4.1×10^{-3}
24	3-ethyl-2-methoxypyrazine	pyrazine	0.82[c]	171	5.8×10^{-3}
25	ethyl butyrate	ester	1.71[b]	136	7.4×10^{-3}
26	ethyl isobutyrate	ester	1.51[b]	132	7.6×10^{-3}
27	2-phenylethanol	phenol	1.36[d]	132	7.6×10^{-3}
28	5-ethyl-3-hydroxy-4-methyl-2(5*H*)-furanone (abhexone)	furan	0.32[g]	82	1.2×10^{-2}
29	3-methyl-2-methoxypyrazine	pyrazine	0.32[c]	62	1.6×10^{-2}
30	2-methoxypyrazine	pyrazine	–0.24[c]	47	2.1×10^{-2}
31	2-ethyl-4-hydroxy-5-methyl-3(2*H*)-furanone (ethylfuraneol)	furan	1.32[g]	39	2.6×10^{-2}
32	3-hydroxy-4,5-dimethyl-2(5*H*)-furanone (sotolone)	furan	–0.22[g]	31	3.2×10^{-2}
33	4-methoxy-2,5-dimethyl-3(2*H*)-furanone (mesifuran)	furan	1.61[g]	19	5.3×10^{-2}
34	4-hydroxy-2,5-dimethyl-3(2*H*)-furanone (Furaneol)	furan	0.78[g]	16	6.3×10^{-2}
35	4-hydroxy-5-methyl-3(2*H*)-furanone (norfuraneol)	furan	0.24[g]	4	2.5×10^{-1}

[a] log P values are reported in a previous work (*21*) as follows: [b] calculated on the basis of the experimental value of ethyl propionate by applying the π-method; [c] calculated on the basis of the experimental value of pyrazine by applying the π-method; [d] means of the experimental values cited by Hansh and Leo; [e] calculated on the basis of the experimental value of guaiacol by applying the π-method; [f] calculated on the basis of the experimental value of vanillin by applying the π-method; [g] calculated by applying the fragment method.

In Catalyst, a hypothesis is a model, which describes a ligand as a set of chemical functions. These functions are defined within Catalyst in a dictionary using the CHM language based on atomic characteristics and include hydrophobic, hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), and positively and negatively ionizable sites (*26*). The hypotheses should be able to predict the activities of different compounds having the same receptor binding mechanism.

There are two ways to create hypotheses. In the first, molecular structures are used as templates to interactively build a hypothesis. In the second, HypoGen software attempts to automatically generate hypotheses from a set of molecules that explain variations in activity across the selected set of molecules (*27*). Starting with the most active molecules, HypoGen analyzes the set of "active" molecules first. The program performs a function mapping on each conformer using the selected function mapping and explores the hypothesis space that is accessible to the most active molecule. The most active compound set (usually five to eight compounds) is determined using the "uncertainty" parameter, noted Unc, so that

$$A_{max} \times \mathrm{Unc} - A/\mathrm{Unc} > 0.0$$

where A_{max} is the activity of the most active compound and A the activity of a compound of the most active set. HypoGen optimizes hypotheses, which are present in the highly active compounds in the training set.

HypoGen selects the best hypotheses from many possibilities by applying a cost analysis. The overall assumption is based on Occam's razor (between otherwise equivalent alternatives, the simplest model is preferred). Simplicity is defined using the minimum description length principle from information theory (*27*). The overall cost of a hypothesis is calculated by summing three cost factors: a weight cost, an error cost, and a configuration cost. HypoGen also calculates two theoretical costs, the null and fixed costs that can be used to determine the significance of the selected hypotheses. The fixed cost represents the simplest model that fits the data perfectly. The null cost represents the cost of a hypothesis with no features that estimates every activity to be the average activity.

The statistical relevance of the various hypotheses is moreover assessed on the basis of their cost relative to the null hypothesis and the fixed hypothesis. The goal of hypothesis generation is to generate a set of hypotheses with total costs as close as possible to the fixed cost.

In addition to the cost analysis, two parameters are involved: RMS represents the deviation of the log(estimated activities) from the log(measured activities) normalized by the log(uncertainties) and indicates the quality of "prediction" for the training set; correl is the linear regression derived from the geometric fit index.

Manual Construction and Automated Generation of Hypotheses. The two ways were used in our study. First, the manual construction of hypotheses allowed to us identify empirical structure families in order to sort out the compounds in several subsets. Then the automated generation of hypotheses was carried out on these compound subsets. Three chemical functions predefined in the Catalyst Feature Dictionary were used: hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), and hydrophobic.

For hypothesis generation, the Unc value is usually defaulted to 3, but in our case, the value of 1.2 was preferred because of the close range of activities (K_d values of 6.8×10^{-4} to 0.25). With Unc = 1.2, compounds **1**−**6** constitute the most active set. Other modifications concern the MinPoints and MinSubsetPoints parameters. The MinPoints parameter controls the minimum number of location constraints required for any hypothesis. The MinSubsetPoints parameter defines the number

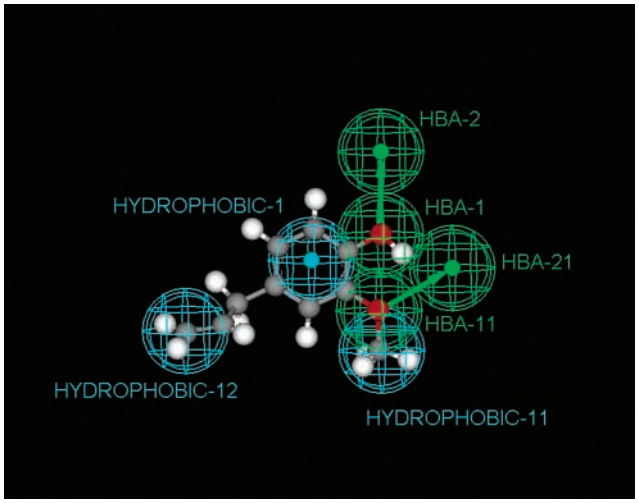

Figure 1. Mapping of eugenol on manually constructed hypothesis. HBA_1 and HBA_2 correspond to the two hydrogen bond acceptor features and HBA_2 and HBA_21 to their projection spheres.

of chemical features that a hypothesis must map in all of the compounds set. Several aroma molecules are small and rigid, and for this reason, these two parameters are set to 2 in addition to the default value of 4. In our case, the value of 3 for these parameters did not provide different results from those obtained with 2 or 4.

RESULTS

Manual Construction of Hypotheses. A hypothesis was constructed for each molecule, using the lower energy conformer as template and only two chemical functions, HBA and hydrophobic. Notice that any HBA function (e.g., alcohol and phenols) can also be described as HBD, but we retained only the most general case. Construction of the eugenol hypothesis is reported in **Figure 1**. Two green spheres (initial and project points) represent one HBA. The blue spheres represent the hydrophobic features and are located on hydrocarbon chains and the hydrophobic aromatic ring.

On the basis of the number of HBA and hydrophobic features mapped by each molecule, aromas are classified into five groups designated **H1**, **H2**, **H3**, **H4**, and **H5**:

H1 corresponds to hypotheses with three hydrophobic features and one HBA (compounds **5**, **6**, **9**, **12**, and **14**).

H2 corresponds to hypotheses with three hydrophobic features and two HBA (compounds **1**−**4**, **7**, **10**, and **21**).

H3 corresponds to hypotheses with two hydrophobic features and one HBA (compounds **8**, **13**, **15**, **18**, **24**, and **27**).

H4 corresponds to hypotheses with two hydrophobic features and two HBA (compounds **11**, **17**, **19**, **22**, **23**, **25**, **26**, and **29**).

H5 corresponds to hypotheses with three HBA (compounds **16**, **20**, **28**, and **30**−**35**).

Automated Generation of Hypotheses. The three features HBA, HBD, and hydrophobic were used for automated hypothesis generation.

First, among groups **H1**−**H5**, we chose two subsets: the first one comprised molecules corresponding to groups **H1** and **H2**, and the other one molecules from group **H4**. On these two groups, automated hypothesis generation was carried out (results not reported here). Then, on the basis of estimated affinity values, a semiempirical classifying procedure was performed to incorporate all compounds belonging to groups **H3** and **H5** in a subset. This led to the retention of three major groups for automatic hypothesis generation:

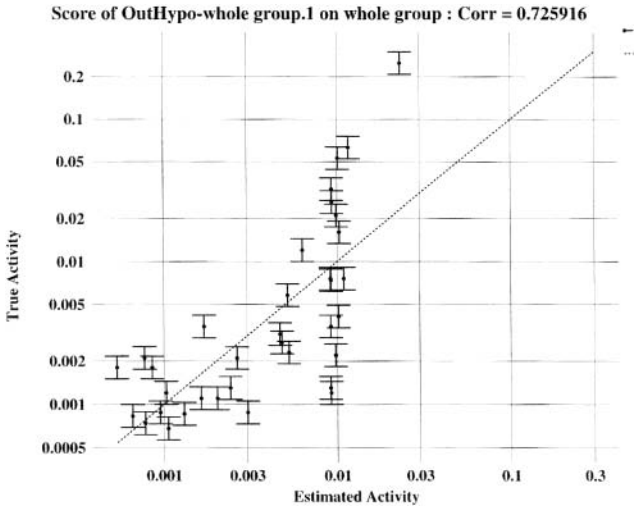

Figure 2. Graph regression obtained with best significant hypothesis (named OutHypo-whole group.1) generated using the whole group.

Group 1 is formed from 15 compounds (**1**−**12**, **15**, **17**, and **18**) and includes the most active compounds (**1**−**6**). The alternative group without 4-ethylphenol (**8**) is designated 1a.

Group 2 is formed from 14 compounds (**13**−**16** and **19**−**30**) in which the molecules have a relatively low affinity for β-lactoglobulin.

Group 3 is formed by the six furans. Several compounds belonging to group 1 (i.e., **1**, **3**, and **6**), or group 2 (i.e., **13**, **14**, and **16**) should be successfully added to group 3 to generate hypotheses, but we retained only the most active ligand **1** to form another subset. Group 3a is thus constituted by group 3 with the addition of *trans*-menthanethiol **1**.

Using the Whole Group (35 Compounds). We perform a hypothesis generation on the entire group setting the MinPoints and MinSubsetPoints parameters first to 2 and then to the default value 4. In these two cases, the same features (one HBA, two hydrophobic) in the same topology constitute the most significant hypothesis. With the MinPoints and MinSubsetPoints parameters equal to 4, it appears that some affinity values are under- and overestimated by the best significant hypothesis (cost = 588.869, RMS = 5.46684, correl = 0.725947; fixed cost = 62.9089; null cost = 1160.57). Moreover, several compounds of different experimental affinities are related to the same value of estimated K_d (**Figure 2**).

Using Groups 1 and 1a. With the MinPoints and MinSubsetPoints parameter values equal to 2, two or three hydrophobic features, lacking HBA or HBD, make up some hypotheses obtained from both groups 1 and group 1a. There is the case for the best significant hypothesis obtained from group 1, composed of three hydrophobic features (cost = 47.3801, RMS = 1.17003, correl = 0.822228; fixed cost = 34.94; null cost = 55.1857).

Using the default value of 4 for the MinPoints and MinSubsetPoints parameters, the hypothesis generation on group 1 (15 compounds) provides hypotheses having HBA and/or HBD in addition to hydrophobic features. None are relevant hypothesis with satisfactory quality of predicted affinity (cost = 43.0945, RMS = 1.06827, correl = 0.85435; fixed cost = 31.9778; null cost = 55.1857 for the best significant hypothesis).

Hypothesis generation run on group 1a (14 compounds, without 4-ethylphenol **8**) provides the best result. Modifications of the MinPoints and MinSubsetPoints parameter values have no effect on features of the best significant hypothesis, constituted by one HBA and two hydrophobic (cost = 35.6008, RMS = 0.431673, correl = 0.979505; fixed cost = 30.242; null cost

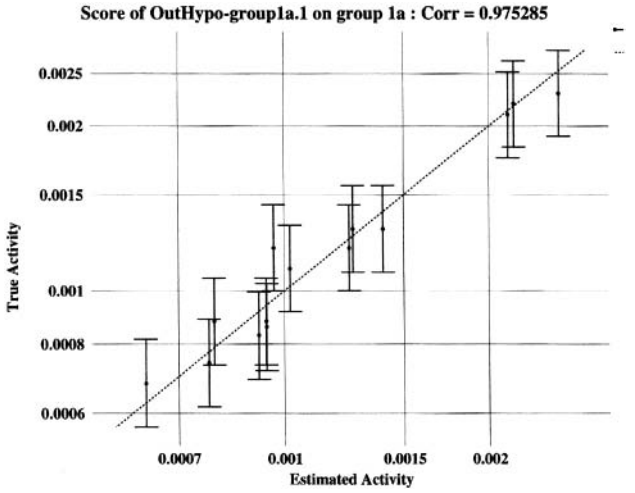

Figure 3. Graph regression obtained with best significant hypothesis for group 1a (OutHypo-group.1a).

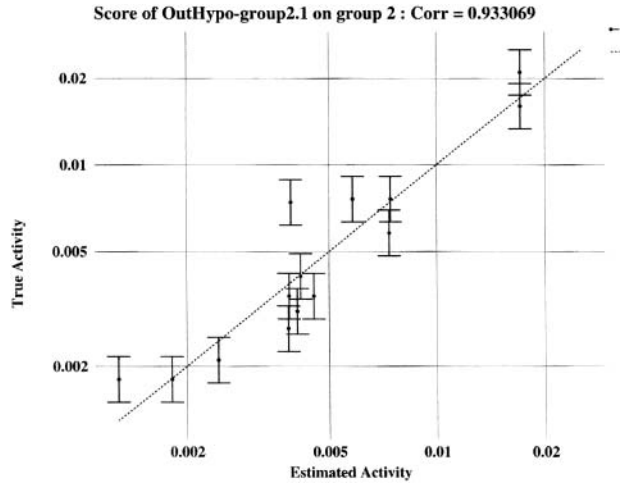

Figure 4. Graph regression obtained with best significant hypothesis for group 2 (OutHypo-group.2).

= 53.5795). There is very good estimation of affinities (**Figure 3**).

Using Group 2. Two hydrophobic, as for group 1a, but one HBD instead of HBA, constitute the best hypothesis, obtained with the MinPoints and MinSubsetPoints parameters set to 2. There are mean estimations of activities (**Figure 4**), despite the good values of cost and correlation parameters (cost = 49.0117, RMS = 1.45461, correl = 0.933142; fixed cost = 32.2987; null cost = 136.528).

Using Groups 3 and 3a. The most significant hypotheses produced by both group 3 (cost = 18.3025, RMS = 0.228563, correl = 0.998014; fixed cost = 18.0717; null cost = 48.7079) and group 3a (total cost = 25.1534, RMS = 0.181536, correl = 0.999605; fixed cost = 24.8042; null cost = 158.344) made an excellent estimation of affinities (**Figure 5**).

The comparison of the three best significant automatically generated hypotheses for groups 1a, 2, and 3 displayed in **Figure 6** shows that the distance constraints between the two hydrophobic spheres are very different for hypotheses generated from group 1a and 2 (5.43 and 7.22 Å, respectively), but relatively close for hypotheses from groups 1a and 3 (5.43 and 4.31 Å, respectively).

Statistical Validation. We performed a statistical cross-validation study to assess the significance of the best hypotheses

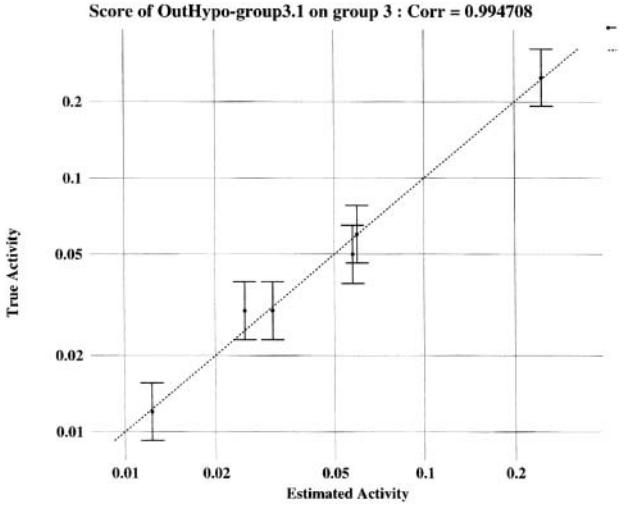

Figure 5. Graph regression of the best significant hypothesis from group 3 (OutHypo-group.3).

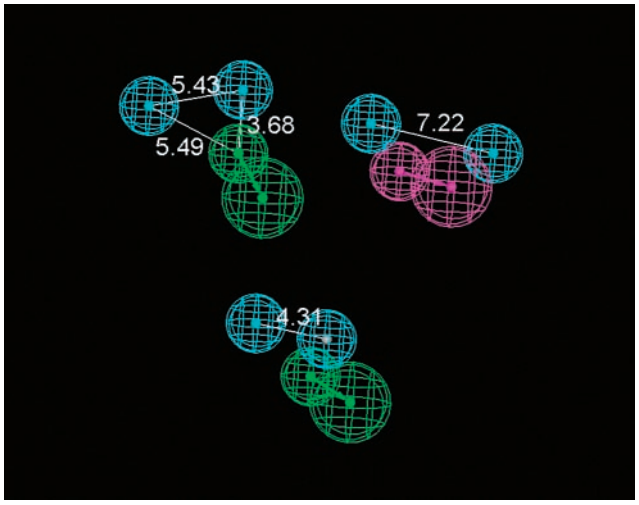

Figure 6. Comparison of the best hypothesis from groups 1a (upper left), 2 (upper right), and 3 (bottom). Distances between center of features are reported in angstroms.

using the catScramble program available in Catalyst. The statistical significance is given by the equation

$$\text{significance} = \left[1 - \frac{(1+x)}{y}\right] \times 100$$

where x = total number of hypotheses having a total cost lower than best significant hypothesis and y = number (HypoGen runs initial + random runs).

To obtain a 95% confidence level, 19 random spreadsheets are generated ($y = 20$) and every generated spreadsheet is submitted to HypoGen using the same experimental conditions (functions and parameters) as the initial run.

For groups 2 and 3, on the one hand, and groups 1 and 3a, on the other hand, the significance values are, respectively, 84 and 89%.

For the whole group and group 1a no generated hypothesis by random runs has a total cost lower than OutHypo1 of (x = 0)) and significance = 95%.

Moreover, we complete the internal validation of group 1a by 14 leave-one-out hypothesis generation runs. Except for the training set without α-terpineol, we obtain the same set of hypotheses (13 satisfactory cases of 14).

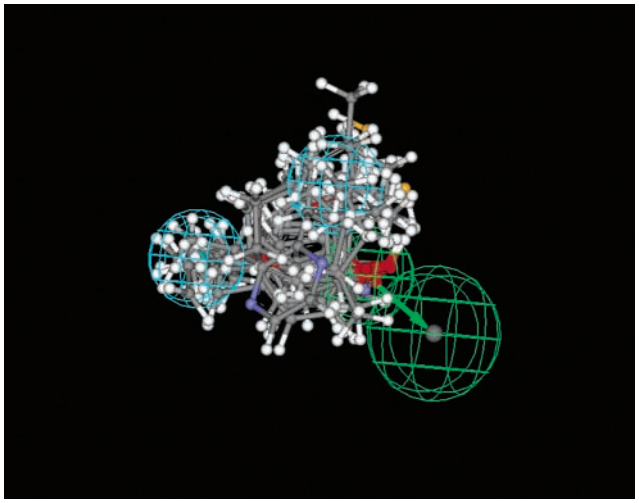

Figure 7. Alignment of compounds **1**–**3** on the best significant hypothesis of group 1a.

To perform a pseudo-external validation, we estimated affinities of compounds of group 1a with the best significant hypothesis of group 3. The result is satisfactory because errors of estimated activities were between −1 and 2.

Mapping of Compounds onto Hypothesis. An alignment of compounds from group 1a on the best significant hypothesis is displayed in **Figure 7** and shows that all of the compounds are disposed very closely in the same area. The same good alignment is also observed both for furans of group 3 (**Figure 8a**) and for compounds of group 3a (**Figure 8b**). On the other hand, compounds from group 2 are dispersed in a conformational space around the best significant hypothesis (**Figure 9**). Our attention was focused on some molecules of group 1.

*(a) α-Terpineol, trans-Menthanethiol, and 3-sec-Butyl-2-methoxypyrazinepyrazine (Compounds **1**, **7**, and **15**).* There are three aromas of particular interest: **1** (*trans*-menthanethiol) and **15** (α-terpineol) have very close log *P* values (3.1 and 3.15, respectively), but the affinity of α-terpineol **15** is 3 times lower than that of compound **1**. Inversely, **1** and **7** have close affinities, but pyrazine **7** presents a low hydrophobicity (log $P = 1.62$). **Figure 10** shows mapping of these three compounds on the best significant hypothesis of group 1a. The three molecules map the HBA sphere feature. Whereas **1** and **15** map the two hydrophobic features, α-terpineol **7** maps only one.

*(b) Trans and Cis Isomers of Menthanethiol and Phenols (Compounds **1**−**4** and **10**).* To explain the role of HBD in the binding of aromas to β-lactoglobulin, we examined the alignment of the most potent ligand **1**, its isomer **3**, and the phenols **2**, **4**, and **10** on another model as the best significant hypothesis (total cost = 48.1225, RMS = 1.33136, correl = 0.77992; fixed cost = 30.242; null cost = 53.5795). From the examination of the fit it appears that only the SH function of **1** maps the HBD sphere. Neither of the two hydrogen bond features (HBD or HBA) maps the OH functions of phenols (**Figure 11**).

*(c) 4-Ethylphenol (Compound **8**).* **Figure 12** shows the alignment of 4-ethylphenol **8** with isomers of menthanethiol (**1** and **3**) and α-terpineol (**15**) on the hypothesis which better estimates affinity of compound **8**. Mapping of 4-ethylphenol appears to be very different from those of the three other molecules.

DISCUSSION

From these results, it appears that hypotheses automatically generated by the entire group fail to distinguish affinities of aromas for β-lactoglobulin, despite good statistical significance. A hypothesis generation run must be performed on subsets (1, 1a, 2, 3, and 3a) to provide relevant hypotheses. The best significance is obtained for group 1a.

Hypothesis generation runs provide some hypotheses constituted only by two or three hydrophobic features (without

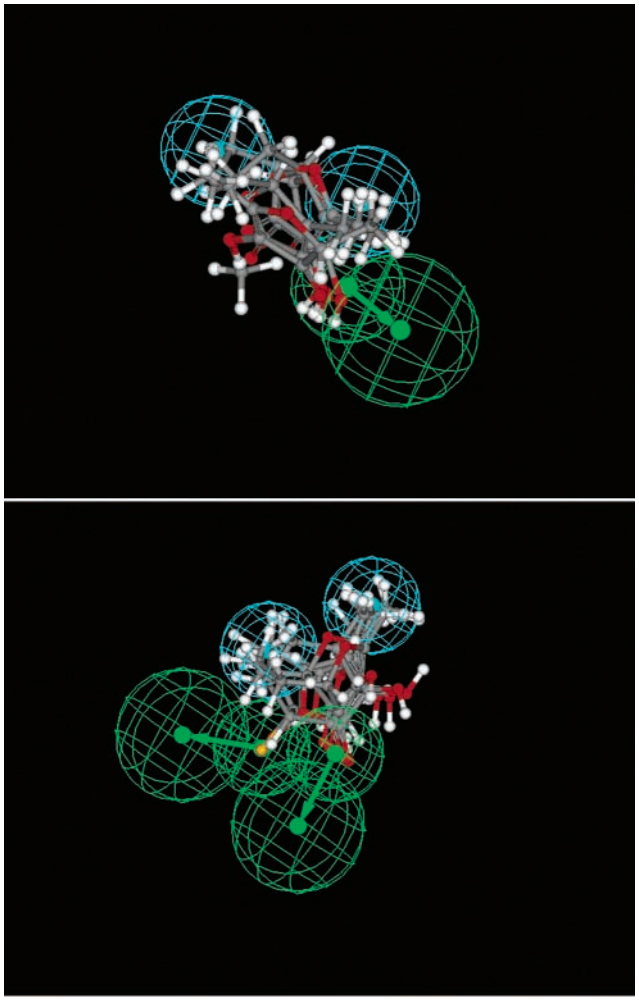

Figure 8. Alignment of the furans on the best significant hypothesis for group 3 (a, top) and of the furans and compound **1** on the best significant hypothesis for group 3a (b, bottom).

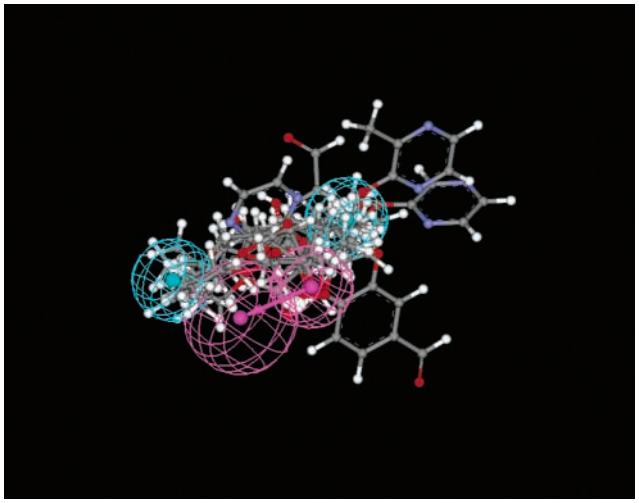

Figure 9. Alignment of compounds from group 2 on the best significant hypothesis for group 2.

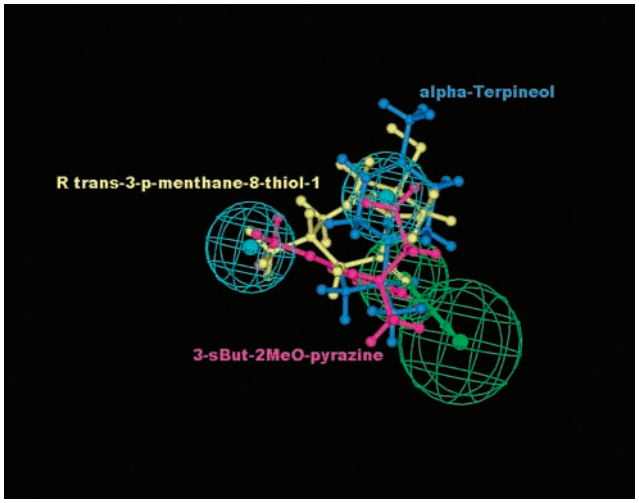

Figure 10. Alignment of compounds **1**, **7**, and **15** on the best significant hypothesis for group 1a.

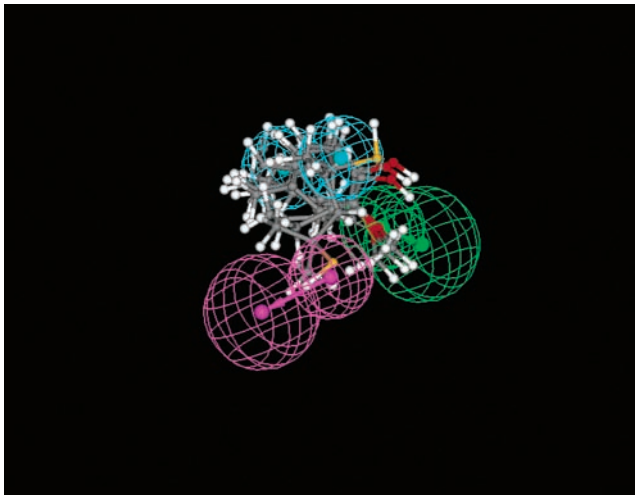

Figure 11. Alignment of menthanethiols **1** and **3** and phenols **2**, **4**, **10**, and **18** on the hypothesis that better estimates the affinity of compound **1**.

hydrogen bond feature). These models do not estimate the affinity of aromas for the β-lactoglobulin with a high reliability. The model constituted by a pair of hydrophobic and hydrogen bonding features is the most representative for the studied set of aroma compounds. It is obtained from either the entire group or groups 1, 1a, 2, or 3, although distances between hydrophobic features are different. The HBA feature plays an essential role. The HBD feature is rarely provided by automated hypothesis generation runs, and corresponding hypotheses fail to give a satisfactory estimation of affinity. However, the HBD feature seems to play a role by increasing the affinity of the trans isomer of menthanethiol for β-lactoglobulin (**Figure 11**).

Alignment of furans and *trans*-menthanethiol **1** on the best significant hypothesis provided by group 3a shows that one HBA sphere is mapped by all of the molecules (**Figure 8a**). Only SH and carbonyl functions of mesifurane are present in the second HBA sphere. An example of a hypothesis comprising HBA and HBD features shows that the HBD sphere is mapped only by the SH function of *trans*-menthanethiol. The oxygens of hydroxyl or carbonyl functions of all furans fit the HBA feature of the best significant hypothesis provided by this training set. It is interesting to note that some oxygens of furans never fit a hydrogen bond acceptor or donor feature but are close to the hydrophobic spheres, and their negative charge density is probably a disadvantage for binding to protein.

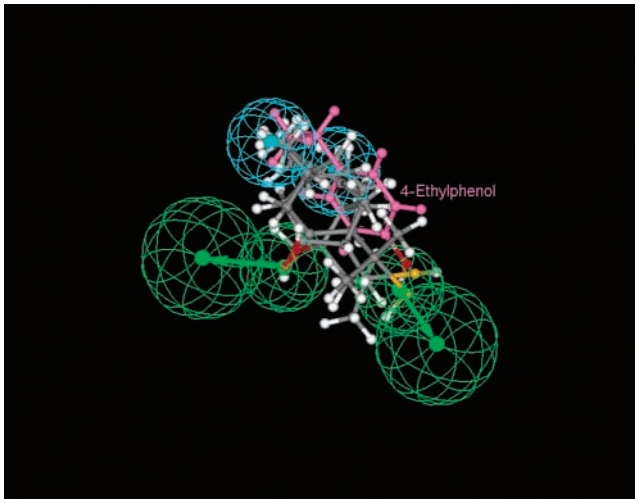

Figure 12. 4-Ethylphenol (**8**): alignment of compounds **1**, **3**, **15**, and **8** (pink) on the best hypothesis for affinity estimation of 4-ethylphenol.

When compounds of groups 1a, 3, and 3a are closely aligned to the best significant hypothesis related to each group, compounds of group 2 are dispersed in the space described by its hypothesis. That could be explained by the potential binding for molecules of group 2 on different sites and/or nonspecific surface sites of protein. Note that the significance for hypotheses from group 2 is <87%. Good affinity estimation in addition to statistical significance of generated hypotheses is obtained for only group 1a. In this way, it is possible that only compounds of group 1a could to be bound to the same inner site in the central cavity (*6*). Despite the similar mapping of hydrophobic and HBA hypotheses features, furans do not bind because of their high electronic density.

Alignments of *trans*-menthanethiol, α-terpineol, and 3-*sec*-butyl-2methoxypyrazine on the most significant hypothesis of group 1a provide an explanation of observed affinity for these compounds. Despite good hydrophobicity, α-terpineol appears to be incorrectly oriented to map the two hydrophobic sites, but only one of them. In contrast, the low hydrophobe pyrazine **7** maps the pair of hydrophobic spheres of the hypothesis. Hydrophobic chain length seems to have a minor importance in comparison to the orientation of the chain in space, whereas hydrogen bonding plays a crucial role.

4-Ethylphenol (**8**) seems to be a particular case, and its alignment is very different from those of the other aromas of the studied set (**Figure 12**). Introduction of this molecule in a training set dramatically disrupts automated hypothesis generation, and thus the affinity is rarely well estimated. This compound should belong to another family of aromas, which probably recognizes another receptor site on the protein.

Our results provide a model constituted by three features triangularly disposed: two hydrophobic and one HBA. Alignment of aroma molecules on this model gives a good explanation for the absence of correlation between log *P* values and affinity for β-lactoglobulin noted for terpenes. Moreover, it justifies the good affinity observed for some pyrazines despite their low hydrophobicities. This emphasizes the power of hydrophobic chain topology and hydrogen bonding.

CONCLUSION

To our knowledge, this work is the first application of Catalyst in aroma study. In the case of aromas, the main difficulty is the

choice of a subset used for hypothesis generation, and this can be related to the existence of many different binding sites on β-lactoglobulin. The multiple potential binding sites on β-lactoglobulin are not in good agreement with the required conditions for the use of Catalyst. It was necessary to first divide the initial training set into several subsets in order to obtain satisfactory automatically generated hypotheses.

Despite these limits, we succeeded in providing significant hypotheses by automated generation that very well estimated the affinity of the studied aromas for β-lactoglobulin. A pair of hydrophobic features and a hydrogen bond acceptor constitute the minimum components of the best hypothesis model. The hydrogen bond donor plays only a minor role in binding. This provides an explanation for the observed binding constants, which are not in relation to the molecules' hydrophobicities. It is important to note that the hydrophobicity is not the only important feature; the topology of the hydrocarbon chain and hydrogen bonding should be also essential.

This study can open the way for the use of Catalyst in aroma research for modeling interaction and improving the understanding of interactions between aromas and protein receptors.

ACKNOWLEDGMENT

We gratefully thank Dr. Vihn Tran (Unité de Physico-Chimie des Macromolécules, INRA, Nantes, France) for scientific aid and helpful discussion.

LITERATURE CITED

(1) Lubbers, S.; Landy, P.; Voilley, A. Retention and release of aroma compounds in foods containing proteins. *Food Technol.* **1998**, *52*, 68−74, 208−214.

(2) Guichard, E. Interactions between flavor compounds and food ingredients and their influence on flavor perception. *Food Rev. Int.* **2002**, *18*, 49−70.

(3) Batt, C. A.; Brady, J.; Sawyer, L. Design improvements of β-lactoglobulin. *Trends Food Sci. Technol.* **1994**, *5*, 261−265.

(4) Brownlow, S.; Morais Cabral, J. H.; Cooper, R.; Flower, D. R.; Yewdall, S. J.; Polikarpov, I.; North, A. C.; Sawyer, L. Bovine β-lactoglobulin at 1.8 Å resolution−still an enigmatic lipocalin. *Structure* **1997**, *5*, 481−495.

(5) Kuwata, K.; Hoshino, M.; Forge, V.; Era, S.; Batt, C. A.; Goto, Y. Solution structure and dynamics of bovine β-lactoglobulin A. *Protein Sci.* **1999**, *8*, 2541−2545.

(6) Papiz, M. Z.; Sawyer, L.; Eliopoulos, E. E.; North, A. C. T.; Findlay, J. B. C.; Sivaprasadarao, R.; Jones, T. A.; Newcomer, M. F.; Kraulis, P. J. The structure of β-lactoglobulin and its similarity to plasma retinol-binding protein. *Nature* **1986**, *324*, 383−385.

(7) Sawyer, L.; Brownlow, S.; Polikarpov, I.; Wu, S. Y. β-lactoglobulin: structural studies, biological clues. *Int. Dairy J.* **1998**, *8*, 65−72.

(8) Flower, D. Beyond the superfamily: the lipocalin receptors. *Biochim. Biophys. Acta* **2000**, *1482*, 327−336.

(9) Sawyer, L.; Kontopidis, G. The core lipocalin, bovine β-lactoglobulin. *Biochim. Biophys. Acta* **2000**, *1482*, 136−148.

(10) Sostmann, K.; Guichard, E. Immobilized β-lactoglobulin on a HPLC column: a rapid way to determine protein−flavour interactions. *Food Chem.* **1998**, *62*, 509−513.

(11) Pelletier, E.; Sostmann, K.; Guichard, E. Measurement of interactions between β-lactoglobulin and flavour compounds (esters, acids, and pyrazines) by affinity and exclusion size chromatography. *J. Agric. Food Chem.* **1998**, *46*, 1506−1509.

(12) Monaco, H. L.; Zanotti, G.; Spadon, P.; Bolognesi, M.; Sawyer, L.; Eliopoulos, E. E. Crystal structure of the trigonal form of bovine β-lactoglobulin and of its complex with retinol at 2.5 Å resolution. *J. Mol. Biol.* **1987**, *197*, 695−706.

(13) Guichard, E.; Langourieux, S. Interactions between β-lactoglobulin and flavour compounds. *Food Chem.* **2000**, *71*, 301−308.

(14) Charles, M.; Bernal, B.; Guichard, E. *8th Weurman Flavour Research Symposium, Reading (U.K.)*; Royal Society of Chemistry: London, U.K., 19xx; pp 433−436.

(15) Wu, S. Y.; Pérez, M. D.; Puyol, P.; Sawyer, L. β-lactoglobulin binds palmitate within its central cavity. *J. Biol. Chem.* **1999**, *274*, 170−174.

(16) Dufour, E.; Haertlé, T. Binding affinities of β-ionone and related flavor compounds to β-lactoglobulin: effects of chemical modifications. *J. Agric. Food Chem.* **1990**, *38*, 1691−1695.

(17) Muresan, S.; van der Bent, A.; de Wolf, F. A. Interaction of β-lactoglobulin with small hydrophobic ligands as monitored by fluorimetry and equilibrium dialysis: Nonlinear quenching effects related to protein−protein association. *J. Agric. Food Chem.* **2001**, *49*, 2609−2618.

(18) Narayan, M.; Berliner, L. J. Fatty acids and retinoids bind independently and simultaneously to β-lactoglobulin. *Biochemistry* **1997**, *36*, 1906−11.

(19) Cho, Y.; Batt, C. A.; Sawyer, L. Probing the retinol-binding site of bovine β-lactoglobulin. *J. Biol. Chem.* **1994**, *269*, 1102−1107.

(20) O'Neill, T. E.; Kinsella, J. E. Binding of alkanone flavors to β-lactoglobulin: effects of conformational and chemical modification. *J. Agric. Food Chem.* **1987**, *35*, 770−774.

(21) Reiners, J.; Nicklaus, S.; Guichard, E. Interactions between β-lactoglobulin and flavour compounds of different chemical classes. Impact of the protein on the odour perception of vanillin and eugenol. *Lait* **2000**, *80*, 347−360.

(22) Guth, H.; Buhr, K.; Fritzler, R. Descriptors for structure−property correlation studies of odorants. In *Aroma Active Compounds in Foods. Chemistry and Sensory Properties*; Takeoka, G. R., Güntert, M., Engel, K., H., Eds.; American Chemical Society: Washington, DC, 2001; pp 93−108.

(23) Sprague, P. W. Automated chemical hypothesis generation and database searching with Catalyst^R. In *Perspective in Drug Discovery and Design*; B. V. K., Ed.; ESCOM Science Publishers: Leiden, The Netherlands, 1995; pp 1−20.

(24) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaninathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.

(25) Smellie, A.; Teig, S. L.; Tobwin, P. Poling: Promoting conformational variation. *J. Comput. Chem.* **1995**, *16*, 171−187.

(26) Greene, J.; Kahn, S. D.; Savoj, H.; Sprague, P. W. Chemical function queries for 3D database search. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1297−1308.

(27) Kurogi, Y.; Guner, O. F. Pharmacophore modeling and three-dimensional database searching for drug design using catalyst. *Curr. Med. Chem.* **2001**, *8*, 1035−1055.

Received for review July 18, 2002. Revised manuscript received November 13, 2002. Accepted November 17, 2002. We thank the Conseil Regional de Bourgogne for financial support.

JF0207981