Multiple Human Taste Receptor Sites: A Molecular Modeling Approach

Nicolas Froloff, Annick Faurion and Patrick Mac Leod

Laboratoire de Neurobiologie sensorielle, Ecole Pratique des Hautes Etudes, 1 Avenue des Olympiades, F-91305 Massy, France

Correspondence to be sent to: Nicolas Froloff, Laboratoire de Neurobiologie sensorielle, Ecole Pratique des Hautes Etudes, 1 Avenue des Olympiades, F-91305 Massy, France

Abstract

Numerous experimental data on the human peripheral taste system suggest the existence of multiple low-affinity and low-specificity receptor sites which are responsible for the detection and the complete discrimination of a very large number of organic molecules. According to this hypothesis, a given molecule interacts with numerous taste receptors and vice versa. Statistical analysis of taste intensities estimated by 58 human subjects for various molecules enables the calculation of taste intermolecular distances. For the present modeling study, we hypothesized that a short taste distance (i.e. taste similarity) between two distinct molecules indicates that they bind with similar distributions of affinities to the taste receptors, and hence display similar binding motifs. In order to find common molecular binding motifs among 14 selected organic tastants, hydrogen-bonding and hydrophobic interaction properties were mapped onto their molecular surfaces. The 14 surfaces were then cut in 240 fragments, most of which were made up of 2–4 potentially interacting zones. A correspondence index was defined to measure the analogy between two optimally superimposed fragments. The 75 most representative fragments were all matched pairwise. Twelve distinct clusters of fragments were isolated from the 2775 calculated comparisons. These 12 fragment types were used to calculate structural similarity distances. We then performed a combinatorial analysis to identify which fragment combination best reconciled structural and taste distances. We finally identified an optimal subset of seven fragment types out of the 12, which significantly and best accounted for the 91 pairwise taste distances between all 14 modeled tastants. These seven validated fragment types are therefore presented as good candidates to be recognized by the same number of distinct taste receptor sites. Potential applications of these identified binding motifs to tastant design are suggested. **Chem. Senses 21: 425–445, 1996.**

Introduction

Membrane receptor proteins enable human taste receptor cells to detect and discriminate a very large number of water-soluble organic molecules. However, fundamental features of these receptors, such as their number, their binding characteristics or their three-dimensional structures, have so far remained elusive.

Psychophysical studies were the earliest at addressing the receptor properties of the human taste peripheral system.

They provided the framework for studies aimed at understanding the principles of taste recognition, and early investigators started to look for common binding motifs that could be responsible for a given taste quality. Following three-quarters of a century of purely descriptive structure–activity relationship studies, Shallenberger and Acree (1967) postulated that two adjacent groups AH and B, located 2.5–4 Å apart, and able to form two antiparallel hydrogen bonds with a complementary receptor site, were required for a compound to taste sweet. Kier (1972) suggested a trifunctional binding motif made up of two adjacent hydrogen-bonding groups AH and B and a hydrophobic patch X, which could explain the similarly sweet tastes of cyclamate, saccharin, perillaldehyde oxime and some nitroanilines. The AH,B,X hypothesis elegantly rationalizes various experimental facts such as the low affinities which are involved in taste reception and the taste discrimination of chiral molecules. This motif can also be found on many unrelated molecular structures, due to its simplicity and its quasi-planar shape.

However, qualitative descriptors such as 'sweet' or 'bitter' have limitations for structure–activity relationship studies (Lee, 1987; Schiffman and Gatlin, 1993). Quantitative analyses of individual sensitivities emerged as a much more potent tool to address receptor-related events. These analyses take advantage of the large amount of information provided by the interindividual differences of taste sensitivities. It is indeed well established that taste detection thresholds to a given molecule usually vary within two orders of magnitudes in concentration from one individual to another (Blakeslee and Salmon, 1935). They vary up to a factor of 8000 in concentration for phenylthiocarbamide (PTC), which is almost tasteless for a significant percentage of humans. This 'taste-blindness' is inherited as a simple Mendelian recessive trait (Snyder, 1931). These observed interindividual differences cannot be explained solely in terms of differences in response amplification mechanisms, since ranking a set of subjects according to their thresholds to a given molecule cannot predict their thresholds to another one (Faurion *et al.*, 1980). Since Diamant *et al.* (1965) demonstrated that perceived intensities are directly proportional to taste nerve response amplitudes, and hence to the proportion of recruited taste receptors, interindividual threshold differences are likely to be also related to interindividual differences in genetic expression levels of taste receptors. In a few very stringent cases, like the autosomal recessive PTC taste-blindness, some individuals may simply not express the appropriate taste receptor.

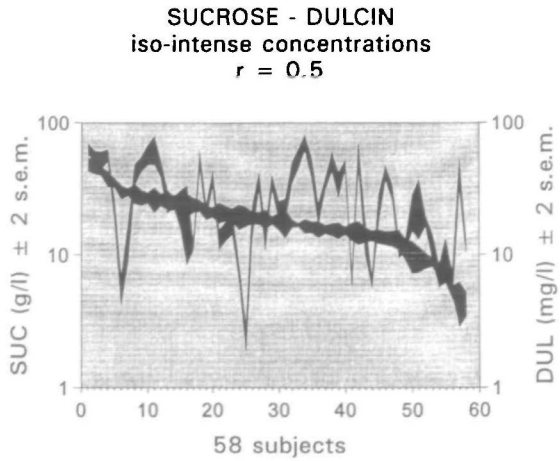

Figure 1 Sensitivity profiles of 58 human subjects to sucrose (gray) and dulcin (black) (concentrations individually matching the intensity of a 1.7 g/l NaCl solution). The width of each profile corresponds to the mean confidence interval on repeated estimates ($25 < n < 35$). Interindividual differences of sensitivity for the two sweet molecules are large ($R_p = 0.53$). The taste receptor sites binding sucrose molecules have to be, at least partially, different from those binding dulcin molecules.

In previous studies, we systematically exploited these interindividual differences by collecting intensity estimates made by many human subjects to various organic stimuli at suprathreshold concentrations (Faurion *et al.*, 1980; Faurion, 1993; Figure 1). Multidimensional statistical analyses of these responses led us to conclude that at least 10 distinct receptor mechanisms are involved in taste molecular recognition. Similar conclusions were drawn from the analysis of single unit responses to 18 compounds in hamster taste pores (Faurion and Vayssettes-Courchay, 1990). These results led to the proposal of multiple binding sites co-operating to elicit any gustatory sensation (Faurion *et al.*, 1980; Faurion and Mac Leod, 1982; Faurion, 1987). This hypothesis was further documented by other authors (Schiffman *et al.*, 1981; van der Heijden *et al.*, 1985a,b; Schiffman and Gatlin, 1993).

Biochemical studies showed that taste receptors are of low affinity for their ligands, with dissociation constants (K_ds) estimated to lie within 10^{-2}–10^{-5} M (for a review, see Froloff, 1994). Because of such low affinities, no mammalian taste receptor has been isolated using traditional biochemical binding techniques, and only impure fractions have been obtained (for detailed reviews, see Sato, 1987; Faurion, 1987). Promising results were obtained on non-mammalian animal models such as the

Table 1 Full chemical names, corresponding Chemical Abstract registry numbers, and abbreviations used throughout the paper for the 14 modeled tastants

Name	CAS no.	Abbr.	CSD ref. code	pK_{a1} (T)	pK_{a2} (T)	pK_{a3} (T)
3-Aminobenzoic acid	[99-05-8]	ABZ	AMBNZA	3.1 (25°C)[d]	4.7 (25°C)[d,g]	
Caffeine	[58-08-2]	CAF	CAFINE	1.2 (25°C)[e]		
Sodium cyclamate	[100-88-9]	CYC	none	1.4 (21°C)[f]	12.1 (25°C)[f]	
Dulcin	[150-69-6]	DUL	none	0.9 (25°C)[c]		
Glycine	[56-40-6]	GLY	GLYCIN15	2.3 (25°C)[b]	9.6 (25°C)[b]	
L-Threonine	[72-19-5]	LTH	LTHREO02	2.1 (25°C)[b,g]	9.1 (25°C)[b,g]	
3-Nitrobenzoic acid	[121-92-6]	MNB	MNBZAC01	3.5 (25°C)[b,d,g]		
3-Nitrobenzenesulfonic acid	[98-47-5]	NSA	DAGXOY	0.7 (25°C)[b,g,h]		
2-Nitrobenzoic acid	[552-16-9]	ONB	NBZOAO02	2.2 (25°C)[d,g]		
1-Propoxy-2-amino-4-nitrobenzene	[553-79-7]	PAN	none	2.5 (25°C)[g,i]		
Perillartine	[30950-27-7]	PER	VEJROR	none		
Picric acid	[88-89-1]	PIC	DICHUS	0.4 (25°C)[g]		
Saccharin	[81-07-2]	SAC	MGSACD10	1.6 (25°C)[a]		
Theophylline	[58-55-9]	TOF	DUXZAX	0.5 (25°C)[b]	2.5 (25°C)[b]	8.8 (25°C)[b]

Reference codes for geometries retrieved from the Cambridge Structural Database are given, as well as pK_a values taken from: [a]Albert and Serjeant (1984), [b]Budavari (1989), [c]Giffney and O'Connor (1975), [d]Kortüm *et al.* (1961), [e]Perrin (1965), [f]Spillane *et al.* (1977, 1982), [g]Weast (1974); pK_a values of [h]benzene sulfonic acid and [i]3-nitroaniline were used to estimate pK_a values of NSA and PAN respectively.

channel catfish (*Ictalurus punctatus*), where receptor-binding studies demonstrated the existence of two major independent classes of taste binding sites for L-alanine and for L-arginine (for a review, see Caprio *et al.*, 1993); measured apparent K_ds were in the micromolar range. The most significant result so far was obtained with the intensely sweet protein thaumatin: a 50 kDa thaumatin-binding protein was labeled by photoaffinity in monkey circumvallate papilla preparations (Shimazaki *et al.*, 1986). Unfortunately, its amino acid sequence is still unknown.

Genetic approaches recently appeared as a promising alternative to obtain amino acid sequences of taste receptors. Striem *et al.* (1989) suggested that taste receptors, like olfactory receptors (Pace *et al.*, 1985), might belong to the G protein-coupled receptor (GPCR) superfamily. Using this working hypothesis, Buck and Axel (1991) cloned and characterized 18 members of this superfamily which are specifically expressed in rat olfactory receptor cells. In addition, Raming *et al.* (1993) demonstrated that a member of this superfamily is indeed transcribed in olfactory receptor neurons, and that this GPCR can recognize odorants and couple to G proteins when expressed in non-neuronal surrogate cells (for recent reviews on olfactory receptors, see Lancet and Ben-Arie, 1993; Breer *et al.*, 1994). Similarly, Abe *et al.* (1993a,b) cloned 60 members of the GPCR superfamily expressed in rat tongue apical cells, and Matsuoka *et al.* (1993) reported the identification of other members expressed in bovine taste tissue.

Our previous results on humans and animals, together with biochemical and genetic data, all converge to a reasonable hypothesis of multiple taste receptor sites of low affinity and low specificity. We therefore designed the present study to look for several binding motifs on a selected set of organic tastants by taking advantage of available quantitative taste response data and atomic resolution crystallographic data. A few assumptions can be made *a priori* on these taste-binding motifs. First, taste receptors should be of low specificity, so as to be able to bind many different organic molecules. Conversely, a given molecule should bind to a specific subset of all available taste-binding sites at a given concentration. In addition, the low affinities involved in taste reception (K_ds within 10^{-2}–10^{-5} M, see above) may exclude the search for pharmacophores made of several interaction zones like those defined for drug molecules, which bind their receptors with K_ds within 10^{-6}–10^{-9} M (Verlinde and Hol, 1994). Indeed molecules as small as chloroform or ethylene glycol actually act as taste stimuli. It is therefore likely that taste receptors establish

Table 2 Ninety-one Pearson's correlation coefficient (R_p) values between the 14 paired tastants, as calculated from psychophysical data

	ABZ	CAF	CYC	DUL	GLY	LTH	MNB	NSA	ONB	PAN	PER	PIC	SAC
CAF	0.59												
CYC	0.07	0.00											
DUL	0.78	0.39	0.22										
GLY	0.23	0.09	0.55	0.40									
LTH	0.53	0.08	0.45	0.31	0.63								
MNB	0.54	0.52	0.15	0.38	0.14	0.36							
NSA	0.61	0.60	0.00	0.41	0.20	0.37	0.78						
ONB	0.79	0.64	0.20	0.63	0.42	0.34	0.60	0.63					
PAN	0.73	0.48	0.38	0.66	0.52	0.62	0.58	0.46	0.68				
PER	0.32	0.38	0.38	0.58	0.45	0.32	0.22	0.06	0.38	0.69			
PIC	0.27	0.51	0.00	0.22	0.25	0.08	0.65	0.65	0.42	0.29	0.01		
SAC	0.59	0.50	0.48	0.50	0.64	0.50	0.49	0.43	0.52	0.27	0.43	0.21	
TOF	0.51	0.50	0.46	0.27	0.45	0.44	0.53	0.45	0.47	0.36	0.00	0.32	0.58

The taste distances used as references in the present modeling study were defined and calculated as: $d = (1 - R_p^2)^{1/2}$.

interactions with limited portions of their ligands, mediated by two or three hydrogen bonds together with some hydrophobic contacts. Indeed, two hydrogen bonds are generally necessary to overcome the competition with water molecules and to lead to a productive ligand–receptor complex, whereas three hydrogen bonds seem a reasonable upper limit to achieve both low affinity and low specificity. We therefore tested the hypothesis of multiple low-affinity and low-specificity taste-binding sites. We developed a blind-searching methodology to find common binding motifs among 14 selected molecules, made up of a few hydrogen-bonding groups and hydrophobic patches. The isolated binding motifs were then validated against the taste similarity matrix between the 14 molecules, built from previously collected quantitative psychophysical data (Faurion, 1993).

Methods

Tastant selection and binding property modeling

Fourteen organic compounds were selected for molecular modeling (Table 1): five synthetic sweeteners [saccharin (SAC), cyclamate (CYC), perillartine (PER), dulcin (DUL) and 1-propoxy-2-amino-4-nitrobenzene (PAN)], two amino acids [glycine (GLY) and L-threonine (LTH)], five acidic substituted benzenes [picric acid (PIC), 3-aminobenzoic acid (ABZ), 3-nitrobenzenesulfonic acid (NSA), 2- and 3-nitrobenzoic acids (ONB and MNB respectively)] and two xanthines [caffeine (CAF) and theophylline (TOF)]. The 14 most rigid structures among 39 molecules previously tested during two supraliminal psychophysical studies in our laboratory (Faurion, 1993) were retained for the present modeling study. Flexible ligands were not used since they can undergo substantial geometrical distortions to achieve optimal binding (Jorgensen, 1991). Selecting flexible tastants would have led to intractable speculations due to their distinct binding conformations with a variety of taste receptor sites.

Table 2 gives a quantitative index of the taste similarity between the 14 selected tastants. The 91 listed pairwise correlation coefficients were calculated on paired results of perceived intensities, where each human subject was asked to determine, for all 14 monomolecular aqueous solutions, the concentrations best matching the memorized intensity of a reference 1.7 g/l NaCl solution.

Atomic Cartesian coordinates were retrieved from the Cambridge Structural Database (CSD) (Allen *et al.*, 1979, 1983). Crystal structures of small organics are usually very precise and are known to be close to a stable or a metastable state in aqueous solution (Gilli, 1992). Where no usable CSD entry existed for a given compound (CYC, DUL and PAN), a conformational search based on a Monte Carlo

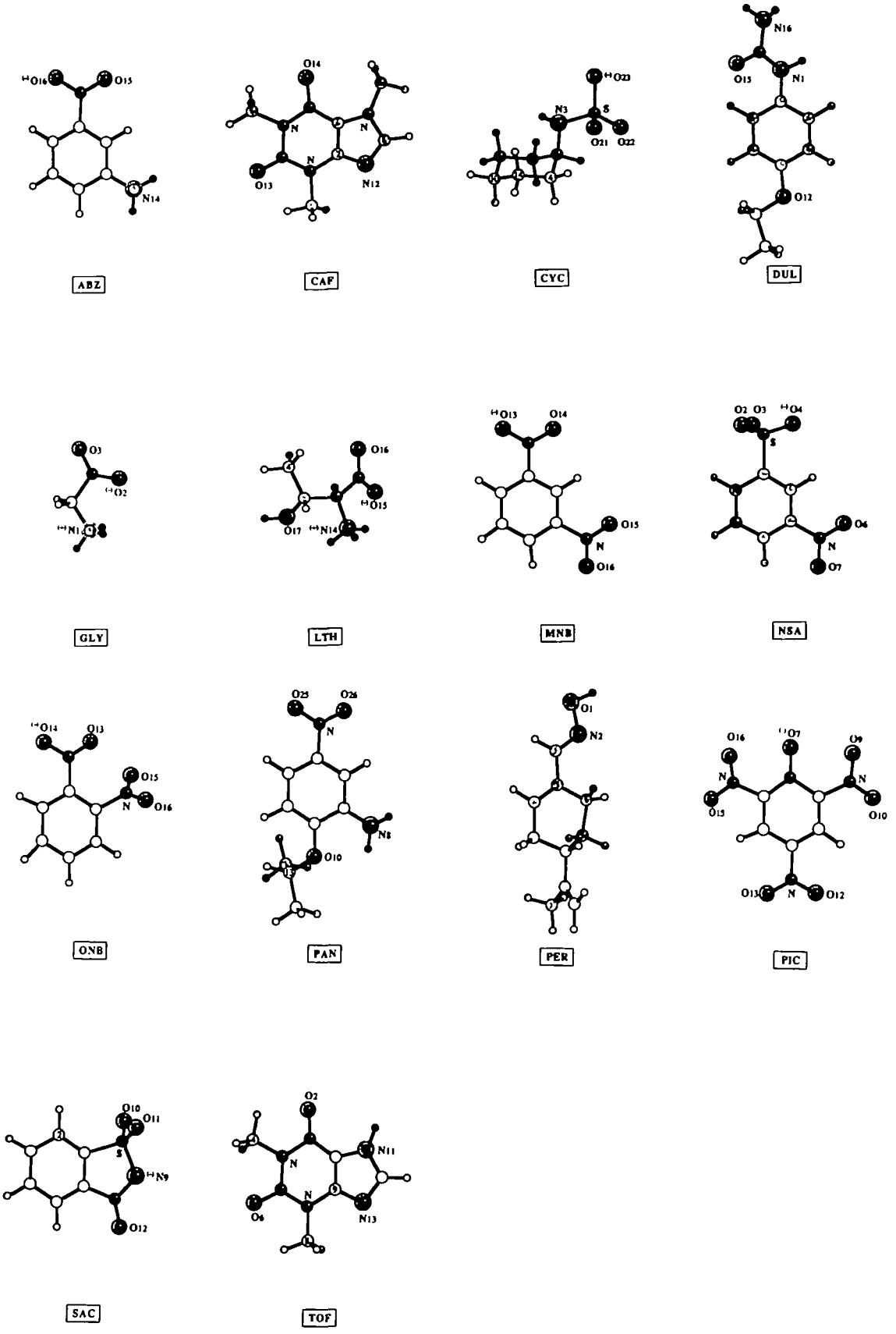

Figure 2 Depth-cued ball-and-stick molecular skeletons of the 14 modeled tastants. Potential hydrogen-bonding oxygen or nitrogen atoms are circled and black (acceptors), gray (donors) or hatched (acceptor/donor atoms). All polar atoms which are neither acceptor nor donor are black. All heteroatoms and all assumed ionization states are specified.

Metropolis optimization method was performed (Metropolis *et al.*, 1953), starting from standard geometric parameters and using a MM2-like force field (Allinger, 1977). This routine is implemented in the Molecular Advanced Design software (MAD) (Lahana, 1990). A further geometrical refinement was then performed in the Austin Model 1 semiempirical Hamiltonian (AM1) (Dewar *et al.*, 1985) available in MOPAC. Compatible MNDO parameters (Dewar and Thiel, 1977) were used for sulfur atoms.

Ionization states of acidic and basic chemical groups, which are crucial for the hydrogen-bonding properties of the corresponding oxygen and nitrogen atoms, were determined according to tables of acidity constants. Human whole saliva pH averages 7.3 ± 0.23 and is strongly buffered by bicarbonate ions (Norris *et al.*, 1984). The 14 selected tastants have pK_a values out of this range (see Table 1), so that the stimulating molecular structures considered for this study (Figure 2) were assumed to be the major ionic species in a buffered aqueous solution of neutral pH.

The molecular surface is defined as the locus of points available to the surface of a spherical water probe 'rolling' on the van der Waals surface of the solute (Richards, 1977). This surface is the most probable interaction interface of a ligand with a protein-binding site. It was constructed for each molecule using the analytical algorithm of Connolly (1981, 1983), included in MAD, with a 1.5 Å probe radius and a 6 dots/Å^2 density.

Potential hydrogen-bonding nitrogen and oxygen atoms (HBAs) were singled out and classified as H-acceptors, H-donors, or acceptor/donor atoms (see Figure 2). The central electron density of the phenyl moiety, which is known to be a weak hydrogen-bond acceptor in the gas phase (Suzuki *et al.*, 1992), was not included in the set of HBAs defined for the present study. Hydrophobic regions of the molecular surface were defined as the sets of molecular surface dots closest to non-polar atoms. These non-polar atoms have a near zero net charge as calculated by AM1 and mostly belong to alkyl chains and phenyl rings.

Common binding motif searching

Since most of the 14 selected molecules are of wide chemical diversity, but show substructural similarities indicating possible common binding motifs, we developed an algorithm to systematically split the molecular models into many quasi-planar overlapping fragments. Each triangular facet of the convex polyhedron wrapping the atomic nuclei of a molecule served as a sectioning plane of the molecular surface. This sectioning plane thus outlined a small patch with binding properties determined by the three atoms at the vertices of the facet and their nearest neighbors. HBAs situated 1 Å apart from the sectioning plane of a fragment were considered as parts of this fragment. Therefore, each fragment consisted of a patch of the molecular surface attached to a few atomic nuclei, some of which were HBAs.

Hydrogen-bonding and hydrophobic correspondences between two fragments were quantified using a correspondence index (C). Each fragment was initially projected onto a planar 1 Å square grid. The projected fragments were then superimposed pairwise. HBA geometrical correspondence was computed first: C was incremented by one unit each time two potential H-acceptors or donors were lying within 1 Å of each other. Distances between HBAs were minimized allowing planar translations and rotations of the second fragment using the downhill simplex algorithm (Nelder and Mead, 1965; Press *et al.*, 1986; NAG library routine E04CCF). Several initial positions were used to avoid trapping in a local optimum. C was then further optimized to include hydrophobic surface correspondence: C was incremented by 0.005 unit for each pair of grid-matched hydrophobic surface dots. In these conditions, given the 6 dots/Å^2 density of the molecular surfaces, a hydrogen-bond and a 30 Å^2 hydrophobic contact were made equivalent in the calculation of C (this choice is justified in the Discussion). Therefore, the final quantity to be optimized was:

$$C = N_a + N_d + 0.005N_x$$

with N_a, N_d and N_x representing the numbers of matched H-acceptors, H-donors and hydrophobic dots respectively.

A similarity index S, computed from C, was defined to remove redundant fragments within one molecule. S was computed as follows for two fragments i and j:

$$S_{ij} = C_{ij}/(C_{ii} + C_{jj} - C_{ij})$$

S varies as C, and is equal to unity for two identical fragments. The 14 intramolecular lists of similarity indices were computed, and only one representative was kept for pairs of fragments showing a value of $S > 0.7$. Finally, fragments with only one or no HBA were removed because of their poor informational content.

Correspondence indices were then computed pairwise

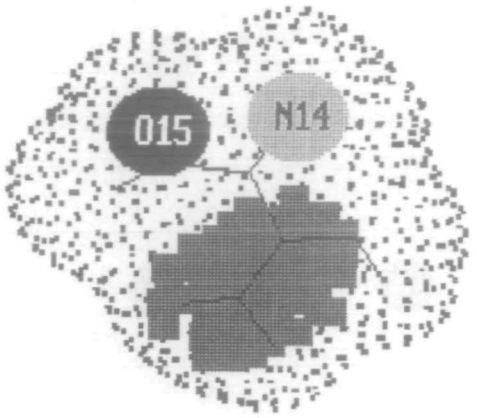

Figure 3 Planar projection of a fragment cut out from L-threonine. This fragment contains the H-acceptor oxygen no 15 (black circle), the H-donor nitrogen no 14 (gray circle) and a 10 Å^2 hydrophobic patch (dark gray area). The molecular skeleton (sticks) and the rest of the molecular surface (dots) are also shown.

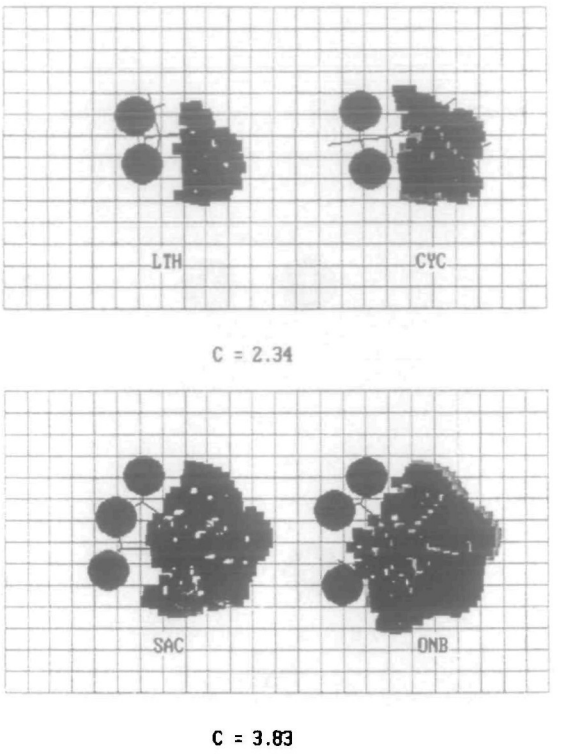

Figure 4 Two optimally superimposed fragment pairs. (Top) Comparison of two fragments belonging to L-threonine and to cyclamate respectively, each made of two H-acceptors (black circles) and a hydrophobic patch (gray areas). The correspondence index value (C = 2.34) quantifies the correct superposition of the two H-acceptors and of a common 11 Å^2 hydrophobic area on the 1 Å square grid. The light gray area surrounding the right fragment indicates the relative position of the left fragment hydrophobic patch. This type of fragment was found on 12 different molecules. (Bottom) Comparison of two other fragments respectively belonging to saccharin and to 2-nitrobenzoic acid. Three H-acceptors and a 28 Å^2 common hydrophobic area are correctly superimposed. This other type of fragment was found on five different molecules.

between all retained fragments. This list of values was used to compute groups of fragments showing the same geometrical arrangement of HBAs (within 1 Å resolution). Groups that contained fragments belonging to at least three distinct molecules were retained, and from now on are referred to as 'fragment types'.

Validating common binding motifs with experimental data

This step was aimed at assessing which fragment types would best account for the taste intermolecular similarities (Table 2). For this step, we hypothesized that the value of the Pearson's correlation coefficient between two molecules (see Figure 1 and Table 2), which measures their sensory similarity, accurately quantifies the degree of overlap between both subsets of taste receptor sites recognizing these two molecules (Faurion and MacLeod, 1982).

With this goal in mind, each molecule was represented by n digits corresponding to the n isolated fragment types. Each digit was set to 1 or 0 according to the presence or absence of the corresponding fragment on the molecule's surface. This binary representation enabled us to define and calculate a structural similarity distance between two molecules by computing a city-block distance between paired sets of molecular digits. All possible combinations of fragment types were automatically generated, and all resulting structural distance matrices were compared to the reference taste distance matrix. Since the probability distribution functions of both sets of distances were not known, each distance matrix was converted into a rank matrix. The Spearman rank-order correlation coefficient (R_s) was used to estimate the accuracy of the fit between the 91 reference taste ranks (r_i) and the 91 structural ranks (s_i). R_s is defined as their linear regression correlation coefficient:

$$R_s = \mathrm{Cov}(r_i, s_i)/\{\mathrm{Var}(r_i) * \mathrm{Var}(s_i)\}^{1/2}$$

A computer program examined all fragment combinations to determine the subset of fragment types maximizing R_s, and thus best reconciling structural distances with taste distances. The significance of a non-zero value of R_s was

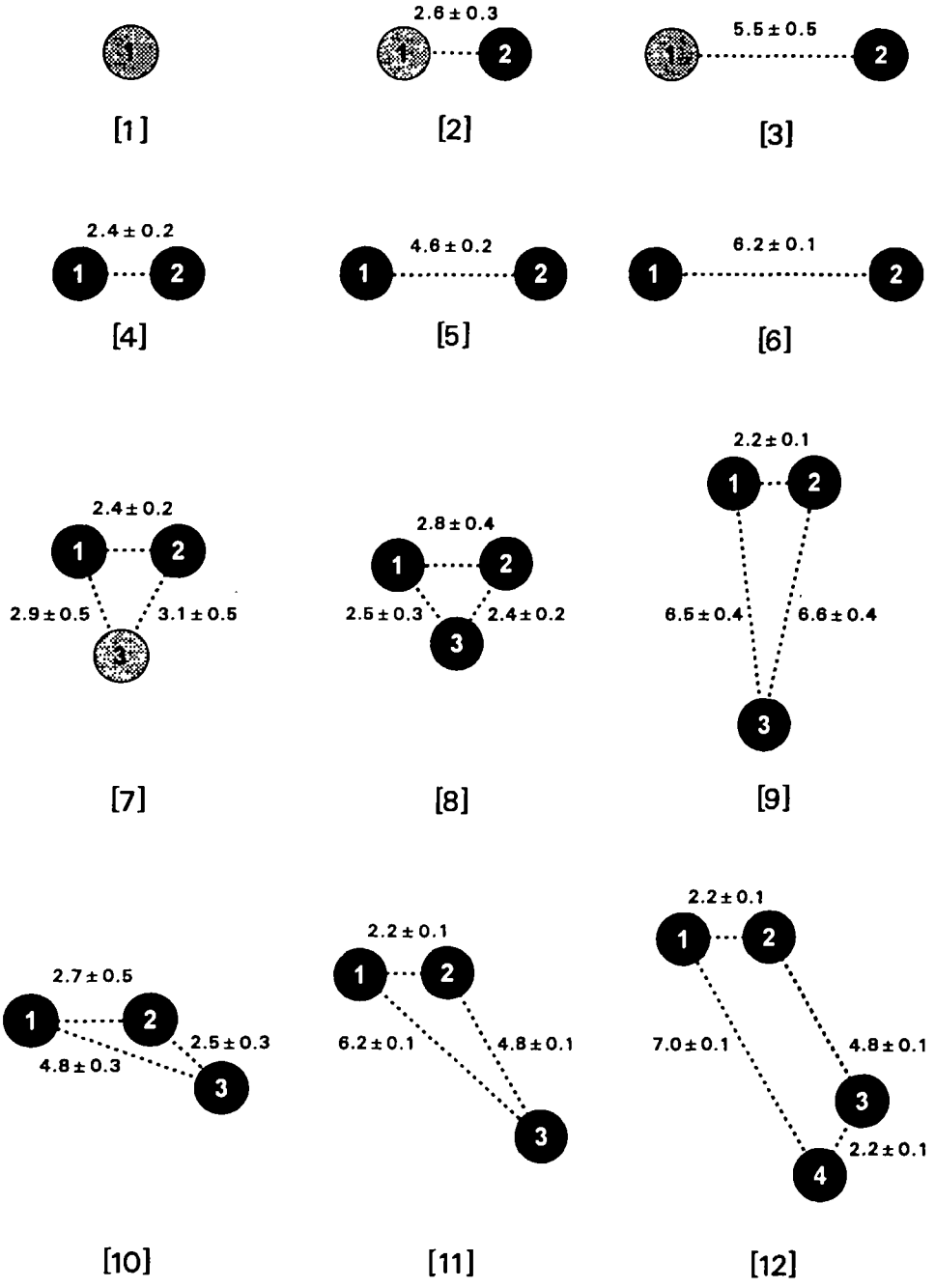

Figure 5 Acceptor and donor average positions for each of the 12 fragment types (numbered [1]–[12]) isolated from the 14 modeled tastants. Motifs formed by potential H-bonding atoms are all different. These atoms are either black (acceptors) or gray (donors), and are numbered for Table 3.

tested by computing the two-sided significance level of t defined as:

$$t = R_s \times \{(N-2)/(1-R_s^2)\}^{1/2}$$

which is distributed approximately as Student's distribution with $N - 2$ degrees of freedom (here $N = 91$) (Press *et al.*, 1986).

Results

Two hundred and forty fragments were cut out using all the faces of the 14 molecular models. The average area of each dissected molecular surface patch is 55 ± 18 Å^2. The total surface area of all fragments of a given molecule is about five times greater than the molecule's surface area. Thus, each fragment surface area corresponds to approximately

Table 3 List of molecular fragments forming the 12 isolated groups (columns numbered [1]–[12] as in Figure 5)

Molecules	Fragment types											
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
ABZ	N_{14},Ph.		N_{14},O_{15},Ph. N_{14},O_{16},Ph.	O_{15},O_{16},Ph.								
CAF					O_{13},O_{14},C_2.							
CYC		N_3,O_{21},C_{10}.		O_{22},O_{23},C_5 O_{23},O_{21},C_{14}. O_{21},O_{22},C_4.			O_{23},O_{21},N_3.	O_{21},O_{23},O_{22}.				
DUL	N_{16},Ph.	N_1,O_{15},Ph.	N_1,O_{12},Ph.			O_{15},O_{12},Ph.						
GLY	N_1,C_5.	N_1,O_2,C_5.		O_3,O_2,C_5.			O_2,O_3,N_1. O_3,O_2,N_1.					
LTH	N_{14},C_4. O_{17},C_4.	N_{14},O_{15},C_4. N_{14},O_{17},C_4.		O_{16},O_{15},C_4.	O_{15},O_{17},C_4. O_{16},O_{17},C_4.		O_{15},O_{16},N_{14}.					
MNB				O_{13},O_{14},Ph.	O_{14},O_{15},Ph.	O_{13},O_{15},Ph. O_{14},O_{16},Ph.			O_{14},O_{13},O_{16}. O_{13},O_{14},O_{16} O_{15},O_{16},O_{13}.		O_{13},O_{14},O_{15}, Ph.O_{16},O_{15}, O_{14},Ph.	O_{13},O_{14},O_{15}, O_{16},Ph.
NSA				O_2,O_3,C_{13}. O_3,O_4. O_4,O_2.	O_6,O_4,Ph.	O_3,O_6,Ph. O_6,O_2,Ph.		O_2,O_4,O_3.	O_4,O_2,O_7. O_7,O_6,O_2. O_6,O_7,O_3.		O_3,O_4,O_6,Ph. O_7,O_6,O_4,Ph.	O_7,O_6,O_4,O_2,Ph.
ONB				O_{13},O_{14},Ph.	O_{14},O_{15},Ph.			O_{13},O_{16},O_{15}		O16,O_{13},O_{14},Ph		
PAN	N_8,C_{13}.	N_8,O_{10},Ph.	N_8,O_{26},Ph. N_8,O_{25},Ph.	O_{25},O_{26},Ph.		O_{26},O_{10},Ph. O_{25},O_{10},Ph.			O_{25},O_{26},O_{10}. O_{26},O_{25},O_{10}.			
PER	O_1,C_7.	O_1,N_2,C_3.										
PIC					O_{10},O_{12},Ph. O_{13},O_{15},Ph.	O_7,O_{12},Ph. O_7,O_{13},Ph.			O_{13},O_{12},O_7.	O_9,O_7,O_{16},Ph.	O_9,O_{10},O_{12}, Ph.O_{12},O_{13}, O_{15},Ph.	O_9,O_{10},O_{12}, O_{13},Ph.O_{12}, O_{13},O_{15},O_{16},Ph.
SAC				O_{11},O_{10},C_2	O_{10},O_{12},Ph. O_{12},O_{11},Ph.			N_9,O_{10},O_{11}.		O10,N_9,O_{12},Ph . O_{12},N_9,O_{11},Ph.		
TOF		N_{11},N_{13},C_4. N_{11},O_2,C_8.		N_{11},N_{13}.	O_6,O_2,C_9.					O_2,N_{11},N_{13},C_8. N_{13},N_{11},O_2,C_8.	N_{11},N_{13},O_6,C_4.	

Fragments are noted as lists of atoms separated by commas and terminated by full stops. Hydrogen-bonding atoms are ordered according to numbers in Figure 5. Atom numbers taken from Figure 2 are indicated. A numbered carbon or a phenyl ring (Ph) indicates a possible additional hydrophobic area. Fragments belonging to groups nos 4, 7 and 8 have a negatively charged H-acceptor. Some members of the same fragment group were cut out from different faces of the same molecule.

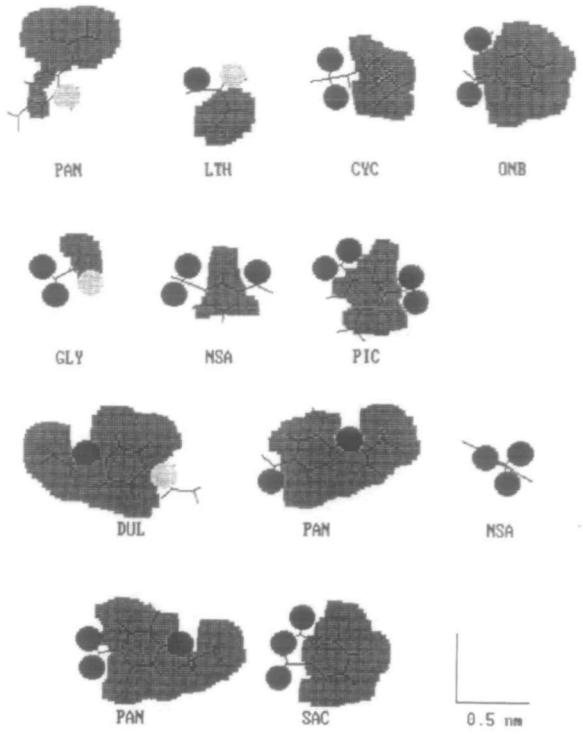

Figure 6 Twelve representatives of the 12 isolated fragment types. (Top) Seven representatives of the seven fragment types validated by the taste distances. (Bottom) Five representatives of the five rejected fragment types.

Table 4 Binary table tabulating the occurrence of fragments

	[1]	[2]	[4]	[5]	[7]	[11]	[12]
ABZ	1	0	1	0	0	0	0
CAF	0	0	0	1	0	0	0
CYC	0	1	1*	0	1	0	0
DUL	1	1	0	0	0	0	0
GLY	1	1	1	0	1*	0	0
LTH	1	1*	1	1	1	0	0
MNB	0	0	1	1	0	1	1
NSA	0	0	1	1	0	1*	1
ONB	0	0	1	1*	0	0	0
PAN	1*	1	1	0	0	0	0
PER	1	1	0	0	0	0	0
PIC	0	0	0	1	0	1	1*
SAC	0	0	1	1	0	0	0
TOF	0	1	1	1	0	1	0

Ninety-one intermolecular structural distances were defined and calculated as city-block differences between rows taken pairwise. These 91 distances were optimized according to the 91 taste distances. The seven fragment types numbered (1, 2, 4, 5, 7, 11, 12) significantly and best reconciled structural and taste distances. The asterisks (*) denote the seven representative fragments depicted in Figure 6 (top).

one-quarter of the whole surface area of a molecule. Most fragments contain 0–4 HBAs, and the area of their hydrophobic surface patch averages 28 ± 13 Å^2. A typical molecular fragment is shown in Figure 3.

One hundred and seven fragments with only one or no HBAs were removed. Among the 133 remaining fragments, 75 were selected after removal of redundant intramolecular fragments. The pairwise comparison of these 75 fragments resulted in a list of 2775 fragment pair correspondence indices. Fragments appeared to be in geometrical correspondence through one HBA (1062 pairs), two HBAs (1340 pairs), three HBAs (223 pairs) or four HBAs (26 pairs). Four additional fragment pairs were superimposed through more than four HBAs, but they concerned only picric acid fragments, whereas 120 pairs displayed no HBA correspondence. Two different pairs of superimposed fragments are shown as examples in Figure 4.

Twelve independent groups of fragments displaying similar HBA arrangements within 1 Å were isolated using the 2775 fragment pair list. The geometrical motifs formed by the HBAs for each fragment type are represented in Figure 5. A comprehensive list of the fragments forming the 12 groups is given in Table III. Twelve typical fragments are shown in Figure 6.

Though numerous fragments have rather complex structures (e.g. 10 of the 12 fragments cut out from MNB have four HBAs, and 11 of the 14 fragments cut out from PIC have 4–7 HBAs), HBA motifs common to several molecules that emerged from the extensive comparison of all fragments are usually simple (Figures 5 and 6): only the no. 12 fragment type has four HBAs. The other fragment types have one HBA (no. 1), two HBAs (nos 2–6) or three HBAs (nos 7–11), flanked by a hydrophobic area (except for no. 8 type). There are only three hydrogen donors among all 30 HBAs forming the 12 isolated motifs (Figure 5). This result stems from the proportion of each given type of HBAs that can be singled out from the 14 molecules: 45 atoms are H-acceptors, seven are H-donors and three are acceptor/donor atoms.

Fragment type no. 1 was found on molecules that have one H-donor flanked by a large hydrophobic area. This fragment type was retained though its informational content is poor. Indeed, the opposite type, consisting of one

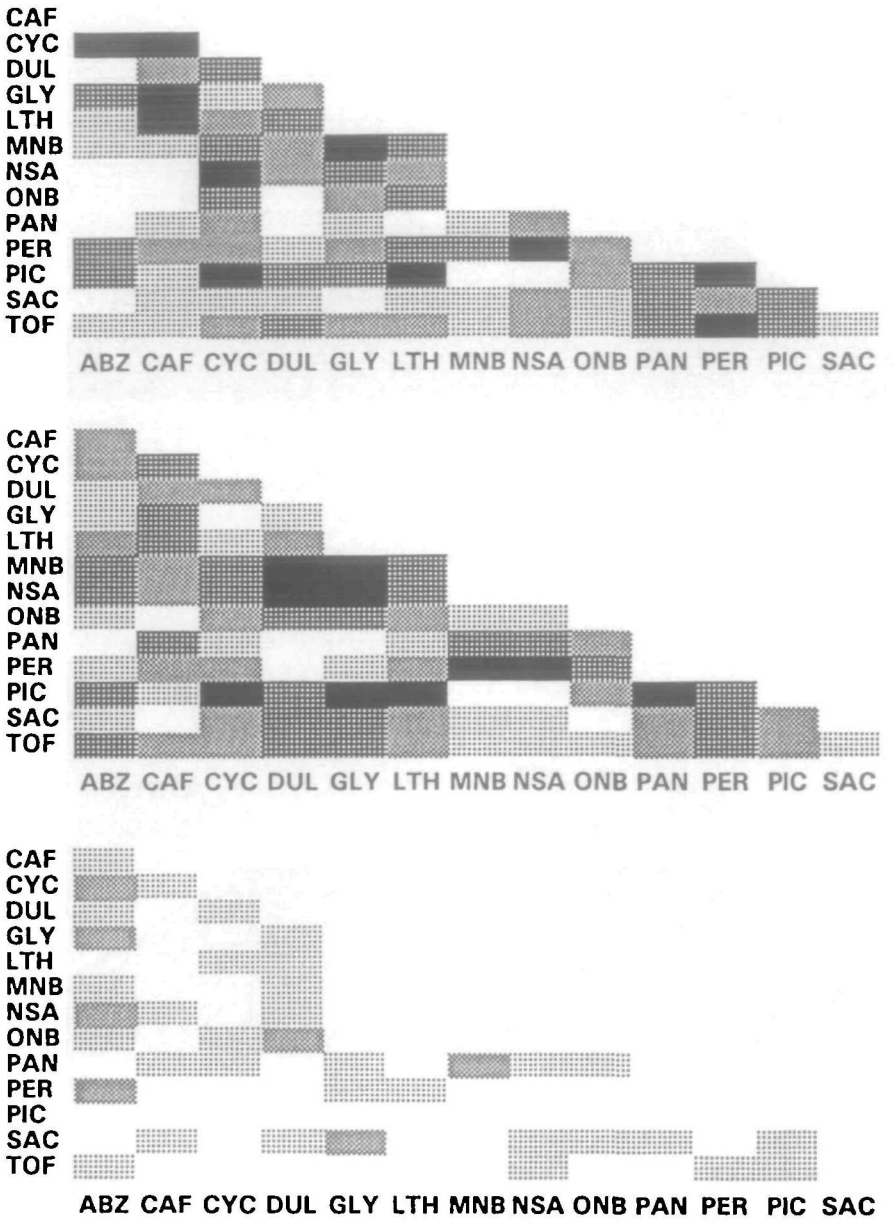

Figure 7 Comparison of taste and structural ranks. (Top) Ninety-one taste ranks, derived from the intermolecular taste distances (see Table 2). Ranks are color-coded in five levels from white to black. Accordingly, white blocks correspond to the 20 smallest distances whereas black ones correspond to the 10 highest. (Middle) Ninety-one structural ranks, derived from the optimal structural similarity distances. (Bottom) Ninety-one differences between taste and structural ranks. White, light gray and dark gray blocks thus correspond to rank differences within 0–20, 21–40 and 41–60 respectively. Rank differences involving ABZ are the highest and average 27.1, whereas those involving PIC are the lowest and average 9.9.

H-acceptor close to a large hydrophobic area, was found on all molecules, and thus was rejected since it is non-discriminating for this set of tastants. Fragment type no. 2, a representative of which is the L-threonine fragment shown in Figure 3, is similar to Kier's trifunctional binding motif; however, its symmetrical counterpart was also observed on the same molecules (i.e. CYC, DUL, LTH, PAN, TOF and possibly GLY and PER), due to another hydrophobic area located on the opposite side of the axis joining the two HBAs. A symmetrical counterpart of the no. 11 type can also be found on all molecules from which it was isolated.

All possible combinations of fragment types among the 12 were automatically generated, and the resulting structural distance matrices were compared one after the other to the reference taste distance matrix. Our final result is a combination of seven fragment types among the 12 (nos

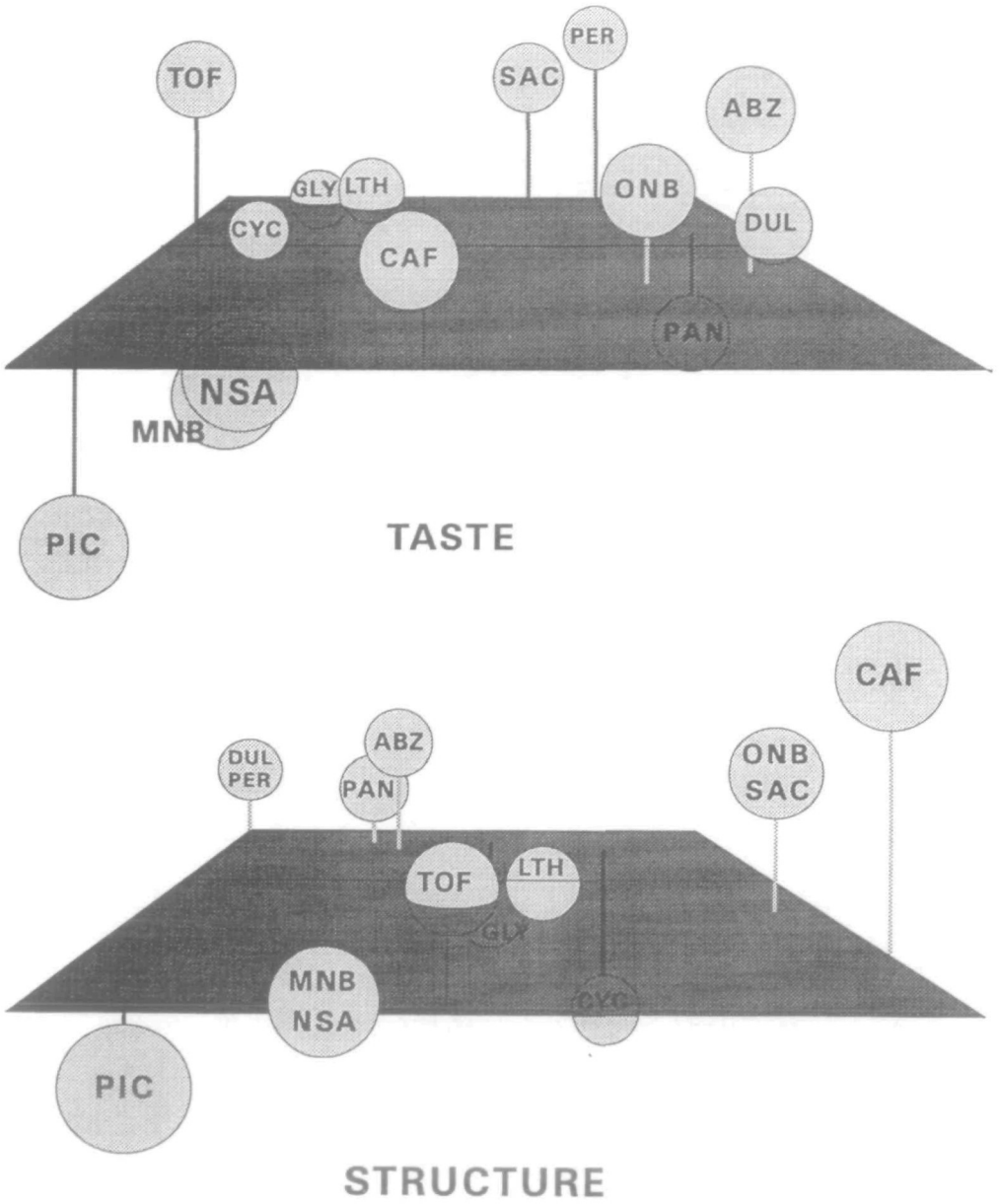

Figure 8 Comparison of taste and structural spaces. (Top) Taste space obtained by factorial analysis of the taste distance matrix (see Table 2). The three represented dimensions convey 45% of the distance information. (Bottom) Structural space obtained by factorial analysis of the contingency table (Table 4). The three represented dimensions convey 85% of the distance information.

1, 2, 4, 5, 7, 11 and 12) which best reconciled the 91 computed structural distances with the 91 experimental distances ($R_s = 0.62$, $P < 0.001$; Table 4).

Figure 6 shows that four of the five invalidated fragment types have a very large hydrophobic area. Also, fragments of type no. 8 exhibit three vicinal H-acceptors arranged in an equilateral triangle, which is a rather unusual HBA motif for organic molecules. On the other hand, validated fragment types are generally smaller, and made of simpler physicochemical motifs.

Three pairs of molecules out of 91 are not resolved by this optimal fragment type combination (DUL–PER, MNB–NSA and ONB–SAC), though their taste distances are not zero.

Figure 7 shows resulting rank differences, ranging from 0 to 55, between both sets of 91 ranks. Rank differences involving ABZ, PAN and DUL are the highest and average 27.1, 22.0 and 21.9 respectively, whereas those involving PIC, LTH and TOF are the lowest and average 9.9, 13.8 and 13.8 respectively.

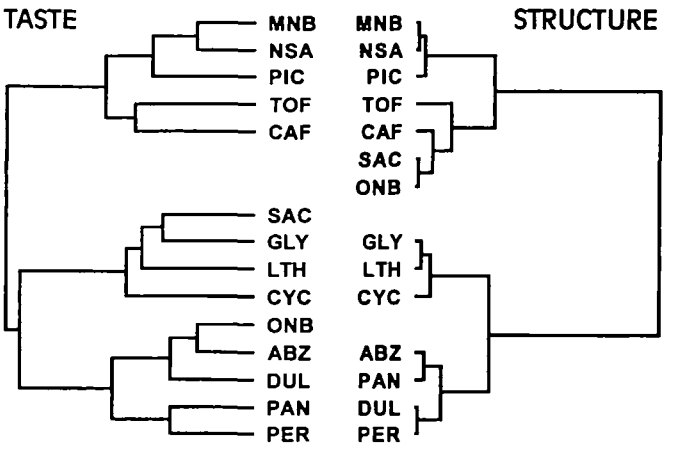

Figure 9 Comparison of taste and structural dendrograms. These hierarchical ascending dendrograms display the whole distance information contained in the corresponding multidimensional spaces.

In addition to the non-parametric statistical tests used to optimize the agreement between both matrices of 91 ranks, a correspondence factorial analysis (numerical routine ANCORR, ADDAD statistical library) was performed on the final binary table, which tabulates the presence or the absence of the validated fragments on the 14 molecules (Table 4). In addition, a factorial analysis (ANADIS, ADDAD) was performed on the taste distance matrix (Table 2). The two resulting spaces are drawn in Figure 8. This figure shows that molecules form similar groups in both spaces. However, the number of significant dimensions of the structural space is twice as low as that of the taste space: four dimensions, accounting for 94% of the distance information, as opposed to seven dimensions for 77% of the distance information.

The algorithm of nearest neighbors was then applied to the factors of the factorial analyses. The two resulting cluster analyses (CAHVOR, ADDAD) yielded dendrograms showing excellent agreement (Figure 9). Only two molecules are misclassified (SAC and ONB). In addition, branches of the structural dendrogram are shorter than those of the taste dendrogram, further indicating that intermolecular separation is lower in our model than in the experimental data.

Discussion

Modeling binding properties of tastants

The success of a search for common binding motifs on various tastants depends on the appropriate modeling of their non-covalent binding abilities with protein-binding sites.

The growing databank of three-dimensional coordinates of co-crystallized protein–ligand complexes deposited in the Brookhaven Protein Data Bank (PDB) (Bernstein *et al.*, 1977) clearly shows that shape complementarity, hydrogen bonds, salt bridges and hydrophobic contacts determine the specificity of binding. In computer-aided drug design, many authors explicitly try to satisfy these requirements in order to search for new drug molecules that fit in the binding site of a target protein receptor (Lewis and Dean, 1989a,b; Martin *et al.*, 1993; Böhm, 1994). The computer program GRID (Goodford, 1985), which produces energetically favorable regions in a binding site for various chemical probes, explicitly includes a hydrogen-bonding term in its molecular mechanics potential (Boobbyer *et al.*, 1989; Wade and Goodford, 1993; Wade *et al.*, 1993). This program is widely used in computer-aided drug design (Goodsell and Olson, 1990; Appelt *et al.*, 1991; Lawrence and Davis, 1992; Varney *et al.*, 1992; von Itzstein *et al.*, 1993). Therefore, numerous authors describe molecules with their hydrogen-bonding and hydrophobic interaction properties to find common binding motifs among different ligands of an unknown target protein structure (Danziger and Dean, 1985; Papadopoulos and Dean, 1991; Jain *et al.*, 1994).

Another way to describe the non-covalent binding properties of a molecule is to represent the electrostatic potential produced by the electrons and the atomic nuclei at the surface of the molecule. The search for common binding motifs then consists in finding similar patterns of electrostatic potentials between ligands (Dean *et al.*, 1988), rationalized by the comparative molecular field analysis method (CoMFA) (Cramer *et al.*, 1988; for a review, see Cramer *et al.*, 1993). However, we observed that describing ligands by hydrogen bonds and hydrophobic regions or by their surface electrostatic potential are somewhat equivalent (N. Froloff *et al.*, unpublished results), since peak values of the electrostatic potential generally map around hydrogen donor and acceptor atoms, whereas medium values delineate regions around aliphatic groups. Furthermore, hydrogen-bonding atoms provide well-defined points in space to superimpose efficiently two dissimilar molecules (Danziger and Dean, 1985; Jain *et al.*, 1994). We therefore decided to describe tastants by their hydrogen bond and hydrophobic interaction properties, since the 14 tastants that we chose for modeling display between two (PER) and seven

hydrogen-bonding groups (PIC) (Figure 2). Possible salt bridges are also taken into account, since charged atoms which could establish such interactions with some taste-binding sites are all included in the selected hydrogen-bonding groups (Figure 2).

While there is not much controversy on the determinants of *specificity* (related to the geometrical disposition of elementary interactions at the binding interface), the representation of properties that determine *affinity* (related to the binding free energy) on ligands is much more challenging. Indeed, methods to calculate the binding free energy between a protein of known structure and some of its ligands are still inaccurate, since the various enthalpic and entropic terms compensating each other to yield binding free energies mostly within 0–10 kcal/mol are of the order of several tens of kilocalories per mol or more (Novotny *et al.*, 1989; Williams *et al.*, 1991, 1993; Krystek *et al.*, 1993; Vajda *et al.*, 1994; Janin, 1995; Shen and Wendoloski, 1995; for a review, see Ajay and Murcko, 1995). Moreover, some of these terms are not only dependent on the structure of the ligands, but also on the atomic details of the receptor-binding site. Therefore, it is probably not possible to represent hot spots of binding energy on ligands without prior knowledge of the structure of their receptors. Furthermore, the identification of a common binding motif on some ligands (characterized by their shape, hydrogen-bonding atoms and hydrophobic patches) is probably not sufficient to estimate their relative affinities with a protein receptor whose three-dimensional structure is unknown.

Fragment definition and comparison

Molecular fragments were defined in order to generalize Kier's proposal (1972) to an ensemble of common binding motifs all made up of a limited number of interaction zones. The algorithmic idea of using the facets of the convex polyhedron wrapping a molecule's atomic nuclei to cut the molecular surface reduced the difficulty of comparing molecules of different shapes and sizes, by describing them with a redundant set of easily comparable near-planar local maps. At the same time, it led to the definition of fragments made up of a few neighboring zones and thus satisfying the minimal requirements formulated in the Introduction. This blind-searching strategy for common binding motifs was originally designed to avoid considering intermolecular taste similarities at the beginning of the modeling process. It was also designed to circumvent biases due to human intervention in the early stages of fragment definition and comparison. Chances of missing relevant binding motifs were thus minimized as much as possible.

However, this fragment dissection technique suffers from a few drawbacks. First, it does not consider fragments made up of interaction zones located on opposite faces of a molecule that could fit into a pocket-shaped binding site. It thus restricts the diversity of common binding motifs that can be extracted from the 14 modeled tastants. It also discards the molecular volume information, and more generally the information on molecular features that surround a given fragment. Finally, it is not well suited to near-planar molecular skeletons (e.g. ABZ, MNB, ONB and PIC), where most dissected fragments are too large and nearly identical, since they basically correspond to both faces of such molecules. An additional technique may be needed to divide these large fragments into smaller, non-identical pieces.

Fragment comparison rules were designed to favor hydrogen-bonding over hydrophobic correspondences. It is known that leaving an unpaired hydrogen-bonding group upon binding weakens the binding free energy by 0.5 up to 4.5 kcal/mol (Fersht *et al.*, 1985), and may consequently increase the dissociation constant by up to three orders of magnitude. Since taste receptor–ligand binding energies are rarely more favorable than –5 kcal/mol (see the Introduction), it is likely that only receptor–ligand complexes where hydrogen bonds are satisfied contribute to taste responses. We therefore decided to penalize heavily any hydrogen-bonding mismatch between two fragments.

It is well established that the hydrophobic effect is a major driving force of biomolecular associations (Kauzmann, 1959), and that its magnitude is related to the amount of buried surface at the binding interface. However, its precise contribution to the binding free energy is still much debated and is estimated to lie within 25–50 cal/mol/Å^2 of buried solvent-accessible surface (SAS) area (Chothia, 1974; Sharp *et al.*, 1991). This roughly corresponds to 40–80 cal/mol/Å^2 of buried molecular surface (MS) area (Jackson and Sternberg, 1995), given the ratio of ~2 between the SAS area as defined by Lee and Richards (1971) and the MS area as defined by Richards (1977). Thus, 1 Å^2 of MS on a ligand can bury another 1 Å^2 on the receptor and thereby contribute by up to 80–160 cal/mol to the binding free energy. In our fragment comparison scheme, a 30 Å^2 overlap of hydrophobic MS patches between two fragments was made equivalent to the correct match of a pair of hydrogen-bonding atoms. A 30 Å^2 overlap corresponds to a good

match of both hydrophobic patches of two typical fragments, since the hydrophobic patch area of the fragments defined in our study averages 28 ± 13 Å^2. Furthermore, a 30 Å^2 MS area on a ligand may contribute up to 2.5–4.5 kcal/mol to the binding free energy, which is thus comparable to the 0.5–4.5 kcal/mol energy penalty associated with a single hydrogen-bond mismatch. A scoring function where hydrogen bonds and hydrophobic interactions were similarly weighted was previously proposed by Böhm (1992a,b) in order to rate the binding potencies of various organic ligands of the same enzyme, and led to the design of improved inhibitors of HIV-protease and dihydrofolate reductase.

Fragment type selection—validation by experimental data

We formulated the idea of a blind search for common binding motifs among organic tastants in the early 1980s (Faurion and Mac Leod, 1982). This idea independently emerged in pharmacology, where the chemical structures of ligands are usually known before the determination of the three-dimensional structures of their common receptors. For example, Dean and co-workers blind searched for physicochemical correspondences between two chemically unrelated neurotoxins (tetrodotoxin and saxitoxin) which compete for the same binding site of unknown shape. These authors used either hydrogen-bonding properties, or molecular electrostatic potentials mapped onto the molecules' surfaces (Danziger and Dean, 1985; Dean *et al.*, 1988). However, taste reception is a more complex situation since each tastant is likely to be recognized by several distinct taste-binding sites. To our knowledge the present work is the first attempt to blind search for multiple binding motifs shared by several molecules.

In this blind search for multiple common patterns, we had to deal with the inductive ambiguity underlying every classification process (Watanabe, 1985). Indeed, a continuum of fragment correspondences, rather than independent groups of tightly related fragments, appeared after pairwise comparison of all 75 molecular fragments. Thus, we had to choose between several possible partitions to propose a reasonable set of fragment groups. Furthermore, there was no obvious reason to suppose that all selected fragment types were pertinent to taste reception. It was thus necessary to validate these fragment types using quantitative experimental data on taste reception for the 14 modeled tastants.

The experimental taste distances between all 14 tastants were used as the criterion to validate the various isolated fragment groups. Indeed, we can assume a direct relationship between the taste distance between two molecules and the proportion of common taste receptor sites binding them. To account faithfully for this hypothesis, the city-block distance used to calculate the structural similarity distance between two molecules based on their common fragments seemed appropriate. Since both sets of 91 intermolecular distances were of distinct mathematical structures, we preferred a semi-quantitative non-parametric assessment of their similarity by converting them into ranks. Hence, despite some loss of information in replacing the original distances by ranks (e.g. the loss of the biological significance level contained in the value of the Pearson's correlation coefficient between two molecules), we could evaluate and optimize the degree of similarity between both sets of 91 distances using the robust Spearman rank-order correlation coefficient (R_s), and interpret the significance level of R_s without ambiguity (Press *et al.*, 1986).

Comparison of the reference space of taste responses and the calculated structural space

The taste peripheral system performs the amazing task of detecting and discriminating almost any organic molecule (Faurion and Vayssettes-Courchay, 1990). Remarkably, we were unable to find any tasteless molecules as controls: provided that a molecule was soluble enough, it always elicited a significant taste nerve response at high concentration. Therefore, taste receptor proteins probably focus on simple structural features to recognize organic tastants. Indeed, comparing molecular fragments of the 14 modeled tastants enabled us to find structural correspondences between chemically unrelated molecules that elicit similar taste responses. It is interesting to note that the rejected fragment types (Figure 6) generally have a wide hydrophobic patch, or a very unusual disposition of hydrogen-bonding atoms like the three vicinal oxygens in NSA. On the other hand, the validated fragment types have a hydrophobic patch made up of <10 atoms, indicating that taste receptors might instead exploit the diversity of hydrogen-bonding patterns to perform their recognition task. Although we were able to identify precisely common hydrogen-bonding patterns (Figure 5) characterized by tight distance constraints and small standard deviations, we were unable to do so for the location and shape of the common hydrophobic patches. Comparing more molecules will lead

to further conclusions, such as whether fragment type no. 2 actually comprises several distinct subtypes (recognized by distinct taste receptor sites), which would depend on the locations and shapes of additional hydrophobic patches, as suggested by van der Heijden *et al.* (1985a,b).

Distances between hydrogen-bonding atoms in the proposed hydrogen-bonding patterns basically fall into three ranges: 2.5, 4.5 and 6.5 Å. These values might be coincidental, since we modeled essentially rigid structures. But we can also reasonably suggest that taste receptors might take advantage of the fact that atoms of organic molecules basically fall into hexagonal or diamond lattices, so that a taste receptor recognizing such key molecular features would then be able to bind many chemically unrelated organic molecules. The fact that the distances are of three ranges also questions the independence of the isolated fragment types, since, for example, hydrogen-bonding pattern no. 5 is a subset of hydrogen-bonding pattern no. 11 (Figure 5). Our model does not actually exclude the possibility that some of the extracted taste-binding motifs could be recognized by distinct or partially overlapping subsites of the same taste receptor protein. However, all hydrogen-bonding groups that are buried upon binding should be involved in a hydrogen bond, since the presence of a buried, unpaired hydrogen-bonding group is unfavorable (Honig and Yang, 1995) and weakens the binding free energy by 0.5 up to 4.5 kcal/mol (Fersht *et al.*, 1985). Therefore, the interaction between a tastant displaying pattern no. 5 and a binding site being complementary to pattern no. 11, or vice versa, might often lead to an unproductive receptor–ligand complex with a dissociation constant larger by several orders of magnitude.

It is interesting to note that the validated fragment type no. 2 (of which a representative is the LTH fragment depicted in Figure 3) is very similar to the trifunctional binding motif suggested by Kier (1972). Kier suggested the same average distance (2.6 Å) between the hydrogen-bond donors and acceptors. He also explicitly indicated atoms forming this motif on various sweet molecules (CYC, PAN, PER and SAC), based on the previous study by Shallenberger and Acree (1967). However, SAC does not display any fragment of this type, according to our model, since it does not have any hydrogen donor group at physiological pH (its pK_a is 1.6 at 25°C). Furthermore, the atoms indicated by Kier on PAN and PER are not the ones that are suggested by our model (see Table 3).

Table 4 also illustrates the absence of relationship between molecular structures and the semantic description of their taste. If we consider molecules described as 'sweet' (ABZ, CYC, DUL, GLY, LTH, ONB, PAN, PER and SAC), we notice that our model does not include a fragment type common to all of them: ABZ and CYC, for example, show similar fragments of type no. 4, but DUL does not display any fragment of this type. Furthermore, any 'sweet' molecule is characterized by at least two fragments of different types. Therefore, our results show that the hypothesis of a unique binding motif responsible for the sweet taste (Shallenberger and Acree, 1967; Shallenberger *et al.*, 1969; Kier, 1972; for reviews, see Birch, 1976; Lee, 1987) is challenged by quantitative taste response data. Moreover, the fact that sweet molecules can share up to three common fragment types with bitter molecules (as is the case for LTH and TOF) contradicts the hypothesis that several types of taste receptor sites might exclusively co-operate to produce the sweet taste (cf., for example, the studies by van der Heijden *et al.*, 1985a,b). Instead, it illustrates the idea of a continuum of taste sensations. As for modeled molecules described as 'bitter' (CAF, MNB, NSA, PIC and TOF), they all share fragment type no. 5; however, this type of fragment is also found on LTH, ONB and SAC—three sweet molecules.

Our results rather reflect similarities in taste responses elicited by the 14 modeled tastants (Table 2). Within human taste responses (Faurion, 1987, 1993), sweet molecules tend to cluster loosely into two distinct groups, the first one comprising CYC, GLY and LTH, and the second one ABZ, DUL, ONB and PAN (Figure 8). Our model suggests that CYC, GLY and LTH share fragment types nos 2, 4 and 7, where no. 7 appears to be specific to these three molecules (Table 4). However, no fragment type specific to the second group of sweet molecules was found in the present study. Such a fragment type would have shortened the structural similarity distances between these molecules, and further separated this group from the first one in the structural space (Figure 8). This may be the reason why structural distances between these molecules do not well agree with their experimental taste distances (Figure 7). On the other hand, fragment types nos 11 and 12 appear to be characteristic of MNB, NSA and PIC, which form a tight group in the taste space. Fragment type no. 11 is simpler than no. 12 in terms of potential elementary interactions (three hydrogen acceptors instead of four), and is more likely to correspond to a taste-binding site, since it was also found on a chemically unrelated molecule (TOF).

Significance and incompleteness of our model

Several lines of evidence indicate that the results yielded by our model are significant yet incomplete. First, the Spearman rank-order correlation coefficient between the 91 optimal structural distances and the 91 experimental ones (0.62) is highly significant according to the computed Student's test ($P < 0.001$), but is still far from unity. This is substantiated by the residual rank differences between the two sets of distances in Figure 7, where ABZ appears to be the worst represented by the model. Secondly, an independent statistical evaluation of the information content of the two sets of distances shows a lower dimensionality of the structural space. Indeed, the structural space built from Table 4 has four significant dimensions accounting for 94% of the distance information, whereas the experimental space built from Table 2 has seven significant dimensions corresponding to 77% of the distance information. This points to a weaker discrimination of the molecules in our model, where three pairs of molecules out of 91 are not resolved (DUL–PER, MNB–NSA and ONB–SAC). This weaker discrimination translates into shorter structural distances within groups in the structural dendrogram (Figure 9), though both structural and experimental dendrograms show excellent agreement (except for SAC and ONB). The most probable cause of this weaker discrimination is to be found in the number of fragment types used to describe structural similarities between molecules. Modeling more molecules will help to identify other relevant types of fragments, so as to account fully for their taste responses. In this respect, results tend to show that an efficient extraction of relevant fragment types is conditioned not only by the selection of rather rigid structures for modeling, but also by the selection of molecules eliciting similar taste responses. Indeed, only one fragment was found on CAF, since this molecule is poorly correlated experimentally to all others (Table 2).

Conclusion and future prospects

Kier's original idea in 1972 was to suggest a trifunctional binding motif for all molecules perceived as sweet. However, quantitative experimental data on taste responses instead suggest that organic molecules are recognized by multiple distinct low-affinity and low-specificity taste receptor sites. Based on a firmly established set of such quantitative taste data for 14 molecules, we generalized Kier's proposal to an open system of seven distinct binding motifs made of approximately three elementary interaction zones. As discussed above, this system is obviously not complete and may be enriched by modeling more molecules and enlarging the fragment definition criteria. However, the experimental and structural dendrograms of Figure 9 show excellent agreement, indicating that the identified motifs are good candidates for binding to actual taste receptor sites, and may also be largely responsible for the taste recognition of the 14 selected molecules. We are now confirming the relevance of these binding motifs via direct competitive inhibition experiments (cross-adaptation) (Lloret *et al.*, 1995).

Efforts should now be focused on the definition and the calculation of 'average' binding motifs to develop and optimize the predictivity of the model. Calculated average distances between potential hydrogen-bonding atoms for each fragment type (Figure 5) are a first step towards the calculation of fragment representatives. Average binding motifs could then be used to probe new molecules, including flexible ones like sucrose or aspartame, and detect relevant taste fragments on their surfaces. Furthermore, average binding motifs could be used as additional information for the direct modeling of taste receptor proteins, once their three-dimensional structures and their binding properties are characterized. For example, Nordvall and Hacksell (1993) constructed the muscarinic m1 acetylcholine receptor on the basis of its postulated three-dimensional structural homology to bacteriorhodopsin. They refined the structure of the binding site using results from an indirect model that describes a proposed active agonist conformation of acetylcholine and a number of related compounds (Schulman *et al.*, 1983). The m1 receptor binding site thus modeled was accurate enough to predict experimentally determined stereoselectivities. Finally, average binding motifs may prove useful in tastant molecular design. A tastant designed to elicit human taste responses similar to those elicited by a known molecule (e.g. sucrose) would have to display binding motifs similar to those of the reference molecule, so as to show the same distribution of binding constants for the taste receptor repertoire.

ACKNOWLEDGEMENTS

We are grateful to Professor P. Laszlo and Professor J.-M. Lefour (DCFI and DCMR, Ecole Polytechnique, France) for many helpful discussions, and for kindly providing access to experimental and computing facilities. Part of this work was presented at the XIth International Symposium on Olfaction and Taste (Sapporo, Japan, July 12–16, 1993) (Froloff *et al.*, 1994), and at the XIth Congress of the European Chemoreception Research Organization (Blois, France, July 25–30, 1994) (Froloff *et al.*, 1995). N.F. was supported by a research fellowship from the Ecole Polytechnique (Paris).

REFERENCES

Abe, K., Kusakabe, Y., Tanemura, K., Emori, Y. and Arai, S. (1993a) Multiple genes for G protein-coupled receptors and their expression in lingual epithelia. *FEBS Lett.*, **316**, 253–256.

Abe, K., Kusakabe, Y., Tanemura, K., Emori, Y. and Arai, S. (1993b) Primary structure and cell-type specific expression of a gustatory G protein-coupled receptor related to olfactory receptors. *J. Biol. Chem.*, **268**, 12033–12039.

ADDAD (Association pour le Développement et la Diffusion de l'Analyse des Données), 4 Place Jussieu, 75005 Paris, France.

Ajay and Murcko, M.A. (1995) Computational methods to predict binding free energy in ligand-receptor complexes. *J. Med. Chem.*, **38**, 4953–4967.

Albert, A. and Serjeant, E.P. (eds) (1984) *The Determination of Ionization Constants*, 3rd edn. Chapman & Hall, London.

Allen, F.H., Bellard, S., Brice, M.D., Cartwright, B.A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B.G., Kennard, O., Motherwell, W.D.S., Rodgers, J.R. and Watson, D.G. (1979) The Cambridge Crystallographic Data Centre: computer-based search, retrieval, analysis and display of information. *Acta Crystallogr.*, **B35**, 2331–2339.

Allen, F.H., Kennard, O. and Taylor, R. (1983) Systematic analysis of structural data as a research technique in organic chemistry. *Accts. Chem. Res.*, **16**, 146–153.

Allinger, N.L. (1977) Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V_1 and V_2 torsional terms. *J. Am. Chem. Soc.*, **99**, 8127–8134.

Appelt, K., Bacquet, R.J., Bartlett, C.A., Booth, C.L.J., Freer, S.T., Fuhry, M.A.M., Gehring, M.R., Herrmann, S.M., Howland, E.F., Janson, C.A., Jones, T.R., Kan, C.-C., Kathardekar, V., Lewis, K.K., Marzoni, G.P., Matthews, D.A., Mohr, C., Moomaw, E.W., Morse, C.A., Oatley, S.J., Ogden, R.C., Reddy, M.R., Reich, S.H., Schoettlin, W.S., Smith, W.W., Varney, M.D., Villafranca, J.E., Ward, R.W., Webber, S., Webber, S.E., Welsh, K.M. and White, J. (1991) Design of enzyme inhibitors using iterative protein crystallographic analysis. *J. Med. Chem.*, **34**, 1925–1934.

Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.E., Jr, Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*, **112**, 535–542.

Birch, G.G. (1976) Structural relationships of sugars to taste. *CRC Crit. Rev. Food Sci. Nutr.*, **8**, 57–95.

Blakeslee, A.F. and Salmon, T.N. (1935) Genetics of sensory thresholds: individual taste reactions for different substances. *Proc. Natl. Acad. Sci. USA*, **21**, 84–90.

Böhm, H.-J. (1992a) The computer program LUDI: a new method for the *de novo* design of enzyme inhibitors. *J. Comput.-Aided Mol. Des.*, **6**, 61–78.

Böhm, H.-J. (1992b) LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *J. Comput.-Aided Mol. Des.*, **6**, 593–606.

Böhm, H.-J. (1994) The development of a simple empirical scoring function to estimate the binding constant for a protein–ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Des.*, **8**, 243–256.

Boobbyer, D.N.A., Goodford, P.J., McWhinnie, P.M. and Wade, R.C. (1989) New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure. *J. Med. Chem.*, **32**, 1083–1094.

Breer, H., Raming, K. and Krieger, J. (1994) Signal recognition and transduction in olfactory neurons. *Biochim. Biophys. Acta*, **1224**, 277–287.

Buck, L. and Axel, R. (1991) A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell*, **65**, 175–187.

Budavari, S. (ed.) (1989) *The Merck Index*, 11th edn. Merck, Rahway.

Cambridge Structural Database (CSD) Cambridge Crystallographic Data Centre (CCDC), University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, UK.

Caprio, J., Brand, J.G., Teeter, J.H., Valentincic, T., Kalinoski, D.L., Kohbara, J., Kumazawa, T. and Wegert, S. (1993) The taste system of the channel catfish: from biophysics to behavior. *Trends Neurosci.*, **16**, 192–197.

Chothia, C. (1974) Hydrophobic bonding and accessible surface area in proteins. *Nature*, **248**, 338–339.

Connolly, M.L. (1981) *Molecular Surface Program (QCPE program no. 429)*. The Quantum Chemistry Program Exchange, Indiana University, Bloomington, IN 47405, USA.

Connolly, M.L. (1983) Analytical molecular surface calculation. *J. Appl. Crystallogr.*, **16**, 548–558.

Cramer, R.D., III, Patterson, D.E. and Bunce, J.D. (1988) Comparative Molecular Field Analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.*, **110**, 5959–5967.

Cramer, R.D., III, DePreist, S.A., Patterson, D.E. and Hecht, P. (1993) The developing practice of comparative molecular field analysis. In Kubinyi, H. (ed.), *3D QSAR in Drug Design: Theory, Methods and Applications*. ESCOM, Leiden, pp. 443–485.

Danziger, D.J. and Dean, P.M. (1985) The search for functional correspondences in molecular structure between two dissimilar molecules. *J. Theor. Biol.*, **116**, 215–224.

Dean, P.M., Callow, P. and Chau, P.-L. (1988) Molecular recognition: blind-searching for regions of strong structural match on the surfaces of two dissimilar molecules. *J. Mol. Graphics*, **6**, 28–34.

Dewar, M.J.S. and Thiel, W. (1977) Ground states of molecules. 38. The MNDO method. Approximations and parameters. *J. Am. Chem. Soc.*, **99**, 4899–4907.

Dewar, M.J.S., Zoebisch, E.G., Healy, E.F. and Stewart, J.J.P. (1985) AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.*, **107**, 3902–3909.

Diamant, H., Oakley, B., Ström, L., Wells, C. and Zotterman, Y. (1965) A comparison of neural and psychophysical responses to taste stimuli in man. *Acta Physiol. Scand.*, **64**, 67–74.

Faurion, A. (1987) Physiology of the sweet taste. In Ottoson, D. (ed.), *Progress in Sensory Physiology 8*. Springer-Verlag, Berlin, pp. 129–201.

Faurion, A. (1993) The physiology of sweet taste and molecular receptors. In Mathlouthi, M. and Birch, G.G. (eds), *Sweet Taste Chemoreception*. Elsevier, Amsterdam, pp. 291–316.

Faurion, A. and Mac Leod, P. (1982) Sweet taste receptor mechanisms. In Birch, G.G. and Parker, K.J. (eds), *Nutritive Sweeteners*. Applied Science, London, pp. 247–273.

Faurion A. and Vayssettes-Courchay C. (1990) Taste as a highly discriminative system: a hamster intrapapillar single unit study with 18 compounds. *Brain Res.*, **512**, 317–332.

Fersht, A.R., Shi, J.-P., Knill-Jones, J., Lowe, D.M., Wilkinson, A.J., Blow, D.M., Brick, P., Carter, P., Waye, M.M.Y. and Winter, G. (1985) Hydrogen bonding and biological specificity analysed by protein engineering. *Nature*, **314**, 235–238.

Froloff, N. (1994) Modélisation moléculaire des sites récepteurs du goût. Thèse de Doctorat, Ecole Polytechnique, Paris.

Froloff, N., Faurion, A. and Mac Leod, P. (1994) Molecular modelling approach to the structure of taste receptor sites. In Kurihara, K., Suzuki, N. and Ogawa, H. (eds), *Olfaction and Taste XI*. Springer-Verlag, Berlin, p. 88.

Froloff, N., Faurion, A. and Mac Leod, P. (1995) Molecular modelling approach to the structure of taste receptor sites. *Chem. Senses*, **20**, 82 (abstract ECRO XI).

Giffney, C.J. and O'Connor, C.J. (1975) Spectrophotometric determination of basicity constants. Part III. Phenylureas. *J. Chem. Soc. Perkin Trans. II*, 1206–1209.

Gilli, G. (1992) Molecules and molecular crystals. In Giacovazzo, C. (ed.), *Fundamentals of Crystallography*. Oxford University Press, Oxford, pp. 465–534.

Goodford, P.J. (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.*, **28**, 849–857.

Goodsell, D.S. and Olson, A.J. (1990) Automated docking of substrates to proteins by simulated annealing. *Proteins*, **8**, 195–202.

Honig, B. and Yang, A.-S. (1995) Free energy balance in protein folding. *Adv. Protein Chem.*, **46**, 27–58.

Jackson, R.M. and Sternberg, M.J.E. (1995) A continuum model for protein–protein interactions: application to the docking problem. *J. Mol. Biol.*, **250**, 258–275.

Jain, A.N., Koile, K. and Chapman, D. (1994) Compass: predicting biological activities from molecular surface properties. Performance comparisons on a steroid benchmark. *J. Med. Chem.*, **37**, 2315–2327.

Janin, J. (1995) Elusive affinities. *Proteins*, **21**, 30–39.

Jorgensen, W.L. (1991) Rusting of the lock and key model for protein–ligand binding. *Science*, **254**, 954–955.

Kauzmann, W. (1959) Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.*, **14**, 1–63.

Kier, L.B. (1972) A molecular theory of sweet taste. *J. Pharmacol. Sci.*, **61**, 1394–1397.

Kortüm, D., Vogel, W. and Andrussow, K. (eds) (1961) *Dissociation Constants of Organic Acids in Aqueous Solution*. Butterworths, London.

Krystek, S., Stouch, T. and Novotny, J. (1993) Affinity and specificity of serine endopeptidase–protein inhibitor interactions. Empirical

free energy calculations based on X-ray crystallographic structures. *J. Mol. Biol.*, **234**, 661–679.

Lahana, R. (1990) MAD, un programme français de modélisation moléculaire ouvert. *Le Technoscope de Biofutur*, no. 34, 14–15.

Lancet, D. and Ben-Arie, N. (1993) Olfactory receptors. *Curr. Biol.*, **3**, 668–674.

Lawrence, M.C. and Davis, P.C. (1992) CLIX: a search algorithm for finding novel ligands capable of binding proteins of known three-dimensional structure. *Proteins*, **12**, 31–41.

Lee, B. and Richards, F.M. (1971) The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.*, **55**, 379–400.

Lee, C.-K. (1987) The chemistry and biochemistry of the sweetness of sugars. In Tipson, R.S. and Horton, D. (eds), *Advances in Carbohydrate Chemistry and Biochemistry*. Academic Press, San Diego, CA, Vol. 45, pp. 199–351.

Lewis, R.A. and Dean, P.M. (1989a) Automated site-directed drug design: the concept of spacer skeletons for primary structure generation. *Proc. R. Soc. Lond. B Biol. Sci.*, **236**, 125–140.

Lewis, R.A. and Dean, P.M. (1989b) Automated site-directed drug design: the formation of molecular templates in primary structure generation. *Proc. R. Soc. Lond. B Biol. Sci.*, **236**, 141–162.

Lloret E., Martinez, J.-M. and Faurion, A. (1995) Cross-adaptation taste studies in the human. *Chem. Senses*, **20**, 144 (Abstract ECRO XI).

MAD (Molecular Advanced Design). Oxford Molecular Ltd, The Magdalen Centre, Oxford Science Park, Oxford OX4 4GA, UK.

Martin, Y.C., Bures, M.G., Danaher, E.A., DeLazzer, J., Lico, I. and Pavlik, P.A. (1993) A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists. *J. Comput.-Aided Mol. Des.*, **7**, 83–102.

Matsuoka, I., Mori, T., Aoki, J., Sato, T. and Kurihara, K. (1993) Identification of novel members of G-protein coupled receptor superfamily expressed in bovine taste tissue. *Biochem. Biophys. Res. Commun.*, **194**, 504–511.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953) Equation of state calculations by fast computing machines. *J. Chem. Phys.*, **21**, 1087–1092.

MOPAC. The Quantum Chemistry Program Exchange, Indiana University, Bloomington, IN 47405, USA.

NAG (Numerical Algorithms Group Ltd), Mayfield House, 256 Banbury Road, Oxford OX2 7DE, UK.

Nelder, J.A. and Mead, R. (1965) A simplex method for function minimisation. *Comput. J.*, **7**, 308–313.

Nordvall, G. and Hacksell, U. (1993) Binding-site modeling of the muscarinic m1 receptor: a combination of homology-based and indirect approaches. *J. Med. Chem.*, **36**, 967–976.

Norris, M.B., Noble, A.C. and Pangborn, R.M. (1984) Human saliva and taste responses to acids varying in anions, titratable acidity, and pH. *Physiol. Behav.*, **32**, 237–244.

Novotny, J., Bruccoleri, R.E. and Saul, F.A. (1989) On the attribution of binding energy in antigen-antibody complexes McPC 603, D1.3, and HyHEL-5. *Biochemistry*, **28**, 4735–4749.

Pace, U., Hanski, E., Salomon, Y. and Lancet, D. (1985) Odorant-sensitive adenylate cyclase may mediate olfactory reception. *Nature*, **316**, 255–258.

Papadopoulos, M.C. and Dean, P.M. (1991) Molecular structure matching by simulated annealing. IV. Classification of atom correspondences in sets of dissimilar molecules. *J. Comput.-Aided Mol. Des.*, **5**, 119–133.

Perrin, D.D. (ed.) (1965) *Dissociation Constants of Organic Bases in Aqueous Solution (Supplement, 1972)*. Butterworths, London.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (eds) (1986) *Numerical Recipes. The Art of Scientific Computing*. Cambridge University Press, Cambridge.

Raming, K., Krieger, J., Strotmann, J., Boekhoff, I., Kubick, S., Baumstark, C. and Breer, H. (1993) Cloning and expression of odorant receptors. *Nature*, **361**, 353–356.

Richards, F. M. (1977) Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioengng*, **6**, 151–176.

Sato, M. (1987) Taste receptor proteins. *Chem. Senses*, **12**, 277–283.

Schiffman, S.S. and Gatlin, C.A. (1993) Sweeteners: state of knowledge review. *Neurosci. Biobehav. Rev.*, **17**, 313–345.

Schiffman, S.S., Cahn, H. and Lindley, M.G. (1981) Multiple receptor sites mediate sweetness: evidence from cross adaptation. *Pharmacol. Biochem. Behav.*, **15**, 377–388.

Schulman, J.M., Sabio, M.L. and Disch, R.L. (1983) Recognition of cholinergic agonists by the muscarinic receptor. 1. Acetylcholine and other agonists with the NCCOCC backbone. *J. Med. Chem.*, **26**, 817–823.

Shallenberger, R.S. and Acree, T.E. (1967) Molecular theory of sweet taste. *Nature*, **216**, 480–482.

Shallenberger, R.S., Acree, T.E. and Lee, C.Y. (1969) Sweet taste of D and L-sugars and amino-acids and the steric nature of their chemo-receptor site. *Nature*, **221**, 555–556.

Sharp, K.A., Nicholls, A., Fine, R.F. and Honig, B. (1991) Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. *Science*, **252**, 106–109.

Shen, J. and Wendoloski, J. (1995) Binding of phosphorus-containing inhibitors to thermolysin studied by the Poisson–Boltzmann method. *Protein Sci.*, **4**, 373–381.

Shimazaki, K., Sato, M. and Nakao, M. (1986) Photoaffinity labeling of thaumatin-binding protein in monkey circumvallate papillae. *Biochim. Biophys. Acta*, **884**, 291–298.

Snyder, L.H. (1931) Inherited taste deficiency. *Science*, **74**, 151–152.

Spillane, W.J. and Thomson, J.B. (1977) Studies of the protonation equilibria of sulphamates using ^{13}C and ^{1}H nuclear magnetic resonance spectroscopic, potentiometric, and conductimetric methods. *J. Chem. Soc. Perkin Trans. II*, 580–584.

Spillane, W.J., Hannigan, T.J. and Shelly, K.P. (1982) Basicity of nitrogen–sulphur(VI) compounds. Part 3. Protonation equilibria of sulphamates using potentiometric and ^{13}C nuclear magnetic resonance methods. *J. Chem. Soc. Perkin Trans. II*, 19–24.

Striem, B.J., Pace, U., Zehavi, U., Naim, M. and Lancet, D. (1989) Sweet tastants stimulate adenylate cyclase coupled to GTP-binding protein in rat tongue membranes. *Biochem. J.*, **260**, 121–126.

Suzuki, S., Green, P.G., Bumgarner, R.E., Dasgupta, S., Goddard, W.A., III and Blake, G.A. (1992) Benzene forms hydrogen bonds with water. *Science*, **257**, 942–945.

Vajda, S., Weng, Z., Rosenfeld, R. and DeLisi, C. (1994) Effect of conformational flexibility and solvation on receptor-ligand binding free energies. *Biochemistry*, **33**, 13977–13988.

van der Heijden, A., van der Wel, H. and Peer, H.G. (1985a) Structure–activity relationships in sweeteners. I. Nitroanilines, sulphamates, oximes, isocoumarins and dipeptides. *Chem. Senses*, **10**, 57–72.

van der Heijden, A., van der Wel, H. and Peer, H.G. (1985b) Structure–activity relationships in sweeteners. II. Saccharins, acesulfames, chlorosugars, tryptophans and ureas. *Chem. Senses*, **10**, 73–88.

Varney, M.D., Marzoni, G.P., Palmer, C.L., Deal, J.G., Webber, S., Welsh, K.M., Bacquet, R.J., Bartlett, C.A., Morse, C.A., Booth, C.L.J., Herrmann, S.M., Howland, E.F., Ward, R.W. and White, J. (1992) Crystal-structure-based design and synthesis of benz[*cd*]indole-containing inhibitors of thymidylate synthase. *J. Med. Chem.*, **35**, 663–676.

Verlinde, C.L.M.J. and Hol, W.G.J. (1994) Structure-based drug design: progress, results and challenges. *Structure*, **2**, 577–587.

von Itzstein, M., Wu, W.-Y., Kok, G.B., Pegg, M.S., Dyason, J.C., Jin, B., Van Phan, T., Smythe, M.L., White, H.F., Oliver, S.W., Colman, P.M., Varghese, J.N., Ryan, D.M., Woods, J.M., Bethell, R.C., Hotham, V.J., Cameron, J.M. and Penn, C.R. (1993) Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature*, **363**, 418–423.

Wade, R.C., Clark, K.J. and Goodford, P.J. (1993) Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 1. Ligand probe groups with the ability to form two hydrogen bonds. *J. Med. Chem.*, **36**, 140–147.

Wade, R.C. and Goodford, P.J. (1993) Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 2. Ligand probe groups with the ability to form more than two hydrogen bonds. *J. Med. Chem.*, **36**, 148–156.

Watanabe, S. (ed.) (1985) *Pattern Recognition: Human and Mechanical*. Wiley, New York.

Weast, R.C. (ed.) (1974) *Handbook of Chemistry and Physics*, 55th edn. CRC Press, Cleveland, OH, pp. D126–D130.

Williams, D.H., Cox, J.P.L., Doig, A.J., Gardner, M., Gerhard, U., Kaye, P.T., Lal, A.R., Nicholls, I.A., Salter, C.J. and Mitchell, R.C. (1991). Toward the semiquantitative estimation of binding constants. Guides for peptide–peptide binding in aqueous solution. *J. Am. Chem. Soc.*, **113**, 7020–7030.

Williams, D.H., Searle, M.S., Mackay, J.P., Gerhard, U., Maplestone, R.A. (1993) Toward an estimation of binding constants in aqueous solution: studies of associations of vancomycin group antibiotics. *Proc. Natl. Acad. Sci. USA*, **90**, 1172–1178.

Received on December 19, 1995; accepted on May 7, 1996