

Proteochemometric Mapping of the Interaction of Organic Compounds with Melanocortin Receptor Subtypes

Maris Lapinsh, Santa Veiksina, Staffan Uhlén, Ramona Petrovska, Ilze Mutule, Felikss Mutulis, Sviatlana Yahorava, Peteris Prusis, and Jarl E. S. Wikberg

Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden

Received May 16, 2004; accepted September 27, 2004

ABSTRACT

Proteochemometrics was applied in the analysis of the binding of organic compounds to wild-type and chimeric melanocortin receptors. Thirteen chimeric melanocortin receptors were designed based on statistical molecular design; each chimera contained parts from three of the MC_{1,3–5} receptors. The binding affinities of 18 compounds were determined for these chimeric melanocortin receptors and the four wild-type melanocortin receptors. The data for 14 of these compounds were correlated to the physicochemical and structural descriptors of compounds, binary descriptors of receptor sequences, and cross-terms derived from ligand and receptor descriptors to obtain a proteochemometric model (correlation was performed using partial least-squares projections to latent structures; PLS). A well fitted mathematical model ($R^2 = 0.92$) with high predictive ability ($Q^2 = 0.79$) was obtained. In a further valida-

tion of the model, the predictive ability for ligands ($Q^2_{\text{lig}} = 0.68$) and receptors ($Q^2_{\text{rec}} = 0.76$) was estimated. The model was moreover validated by external prediction by using the data for the four additional compounds that had not at all been included in the proteochemometric model; the analysis yielded a $Q^2_{\text{ext}} = 0.73$. An interpretation of the results using PLS coefficients revealed the influence of particular properties of organic compounds on their affinity to melanocortin receptors. Three-dimensional models of melanocortin receptors were also created, and physicochemical properties of the amino acids inside the receptors' transmembrane cavity were correlated to the PLS modeling results. The importance of particular amino acids for selective binding of organic compounds was estimated and used to outline the ligand recognition site in the melanocortin receptors.

Melanocortin receptors (MCRs) are members of the seven transmembrane (TM)-spanning G protein-coupled receptor (GPCR) superfamily. To date, five MCR subtypes, MC_{1–5}, are recognized in mammals, and each of these subtypes stimulates cAMP signal transduction pathways. An endogenous group of peptides, the melanocyte-stimulating hormones (MSH) and corticotropin and agouti and agouti-related protein, bind to the MCRs with agonistic and antagonistic properties, respectively. However, an exception is the MC₂R, which binds only the corticotropin (Schioth et al., 1996a).

The MCRs have a wide range of physiological roles. The MC₁R regulates melanin pigment formation in the skin and has a regulating role in the immune system. The MC₂R regulates corticosteroid production of the adrenals. The MC₃ and MC₄R play important roles in controlling feeding and sexual behaviors, and the MC₅R is involved in the regulation of exocrine

glands (Wikberg et al., 2000, 2002; Wikberg, 2001). The potential of using the MCRs as targets for drugs to treat important medical conditions such as obesity/diabetes, inflammatory conditions, and sexual dysfunctions prompts the need for compounds that show high specificity for particular MCR subtypes. However, the design of selective drugs for subtly different receptor subtypes is a difficult task that would become simplified if one had in hand detailed knowledge of the determinants for ligand-receptor interaction.

We have recently invented the proteochemometric modeling technology for the analysis of protein-ligand interactions (Wikberg et al., 2003, 2004). This technology is particularly suited to gain knowledge in the differences in the ligand recognition of different targets. Contrary to traditional quantitative structure-activity relationship methods that aim to correlate a description of ligands with their affinity for one particular target, proteochemometrics considers many targets and ligands simultaneously. Proteochemometrics thus analyzes the experimentally determined interaction activity data of a series of ligands for a series of proteins by correlating the data to descriptors of ligands, proteins, and cross-

This work was support financially by Swedish Research Council grants 04X-045959 and 621-2002-4711.

Article, publication date, and citation information can be found at <http://molpharm.aspetjournals.org>.
doi:10.1124/mol.104.002857.

ABBREVIATIONS: MCR, melanocortin receptor; TM, transmembrane; GPCR, G protein-coupled receptor; MSH, melanocyte stimulating hormone(s); NDP, Nle⁴,D-Phe⁷; PCR, polymerase chain reaction; bp, base pair(s); PLS, partial least-squares projections to latent structures.

terms derived from ligand and protein descriptors. The differences in the properties of both interaction partners are accordingly exploited to explain ligand-receptor affinity.

In a previous study, we applied proteochemometrics to study the interactions of wild-type and chimeric and point-mutated α -adrenergic receptors with piperidyl oxazole derivatives (Lapinsh et al., 2001). In another study, we applied it onto 21 native amine GPCRs that interact with a set of diverse organic compounds (Lapinsh et al., 2002). We have also modeled the interactions of chimeric melanocortin receptors with natural and synthetic analogs of melanocortin peptides (α -MSH, NDP-MSH, and ^{125}I -NDP-MSH) (Prusis et al., 2001, 2002) and native MC_{1,3-5}Rs with a large series of organic compounds (Lapinsh et al., 2003). In all cases, statistically highly valid models that were interpretable in a chemical sense were obtained. For example, the most recent of our studies (Lapinsh et al., 2003) allowed us to reveal compound properties that influence the binding of organic compounds to different MCR subtypes. However, using only four receptors in the modeling did not allow us to map the receptors' ligand binding pockets. In the present study, we have undertaken a further analysis to model the recognition site for organic compounds in the MCRs by using 4 native and 13 multipart chimeric MCRs, which were designed using the principles of statistical molecular design.

Materials and Methods

Materials. Primers were from TAG Copenhagen A/S (Copenhagen, Denmark). Restriction endonucleases HindIII, XhoI, and XbaI were from Promega (Madison, WI). ^{125}I -NDP-MSH (^{125}I -[Nle⁴, D-Phe⁷] α -MSH) was custom-synthesized by Euro-Diagnostica AB (Malmö, Sweden). The genes of the MC₁ and MC₅Rs had been cloned earlier in our laboratory (Chhajlani et al., 1992, 1993). The MC₃ and MC₄R genes were gifts of Dr. Ira Gantz (University of Michigan Medical School, Ann Arbor, MI) (Gantz et al., 1993a,b).

Experimental Design of Chimeric Receptors. The chimeric receptors contained parts from the human MC₁, MC₃, MC₄, and MC₅Rs. The native receptors were divided into three parts (A, B, and C), which could then maximally be combined into $4^3 - 4 = 60$ three-part chimeras containing portions from three or two melanocortin receptor subtypes. However, we wanted to maximize the information content in the data gained from chimeric receptors with minimal experimental work. Therefore, we applied statistical molecular design using D-optimal design (Eriksson and Johansson, 1996) as generated by the MODDE 6.0 software (Umetrics AB, Umeå, Sweden). According to the design, 12 chimeras were selected that, together with the four native receptors, provided the best representation of all possible combinations of receptor portions in linear terms. The design thus contained all 4×4 possible pairs of different sequence combinations in the A/B, A/C, and B/C part of the receptors; the 12 selected chimeras were 1-3-4, 1-4-5, 1-5-3, 3-1-5, 3-4-1, 3-5-4, 4-1-3, 4-3-5, 4-5-1, 5-1-4, 5-3-1, and 5-4-3 (numbers represent respective MCR subtype).

To obtain chimeric receptors with unaltered folding and functionality, we elected to recombine receptors in the sequence stretches showing the highest conservation among all four melanocortin receptors. We identified four such highly conserved stretches, making it possible to divide the receptors into five segments. For technical reasons, the receptors were created as two sets of chimeras. For the first set (F-set), the combinations took place at the end of the third of the seven transmembrane regions (i.e., TM3) (residues 138–143 in the MC₁ receptor¹) and at the end of TM5 (residues 205–210). In the

second set (S-set), the combination sites were located at the end of TM2 (residues 73–81) and in the middle of TM6 (253–259).

Primer Design and Manufacture of DNA Constructs by PCR. Part A of the F-set was built by using vector-specific forward primer and receptor-specific reverse primers (GTACCTGTCCACT-GCGATGGC for MC₁, MC₃, and MC₅Rs and GTACCTGTCCACTGC-GATTGA for MC₄R) that recognize the sequence encoding the IAVDRY sequence stretch at the end of the third transmembrane helix, which is present in all the four melanocortin receptors. Part B of the F-set was built by using forward primers (ATCGCAGTGGACAGG-TACATCTCCA for MC₁R and ATCGCAGTGGACAGGTAC for MC₃₋₅Rs) that recognize IAVDRY and reverse primers (CATGTGGACG-TACAGCAC for MC₁R, CATGTGGACGTACAGGGT for MC₃R, and CATGTGGACGTACAGAGA for MC_{4,5}Rs) that recognize LYVHM at the end of the fifth transmembrane helix of the MC₁, MC₃, and MC₄Rs (for the MC₅R, the primer will induce a change in the sequence from LYIHM to LYVHM). Part C was built using forward primers (CTGTACGTCCACATGCTG for MC₁R and CTGTACGTCCACATGTTCTT for MC₃₋₅Rs) that recognize LYVHM and vector-specific reverse primers. The three sets of PCR amplifications resulted in DNA products for all the four melanocortin receptors with an 18-bp identical overlap between the end of the first segment and beginning of the second and a 15-bp identical overlap between the end of the second segment and beginning of the third. When segments were combined, their overlaps acted as primers for each other. For example, the combination of the MC₁R part A of the F-set and MC₃R part B and the T7 forward primer and LYVHM reverse primer resulted in the amplification of the MC₁/MC₃R chimeric part A/B segment. When this A/B segment was combined with part C of the MC₄R, together with the T7 primer (forward) and a reverse primer that recognizes the 3'-UT region (e.g., the T3 primer), the full-length MC₁/MC₃/MC₄ chimeric receptor gene was produced. The combination of parts A, B, and C from different melanocortin receptors could thus produce any one of the multiple chimeric receptor constructs of the F-set.

The S-set was built by taking advantage of the almost perfect overlap between the melanocortin receptors at the beginning of the second transmembrane helix (MYFFICSSLA; exact match in MC₄R) and at the sixth transmembrane helix (CWAPFFL; exact matches for the MC₃, MC₄, and MC₅Rs). For the annealing site between parts A and B, forward primer ATGTACTTCTTCATCTGCAGC-CTGGC and reverse primer GCCAGGCTGCAGATGAAGAAGTACAT were used. For the annealing site between parts B and C, forward primer TGCTGGGCCCCCTTCTTCTT and reverse primer AGGAAGAAGGGGGCCAGCA were used. The S-set of bands amplified by PCR resulted in a 26-bp overlap between parts A and B and a 20-bp overlap between parts B and C.

PCR was performed at 95°C for 2 min, followed by 26 cycles at 95°C for 1 min, 47°C for 40 sec, and 68°C for 1 min, using a Techne apparatus (Techne, Cambridge, UK). PCR products were isolated by agarose gel electrophoresis. Bands were recovered using a standard NaI/silica-based method (Vogelstein and Gillespie, 1979) and combined to produce chimeras by adding the outer primers and then running 31 PCR cycles.

Cloning, Sequencing, and Expression of Chimeric Receptors. The chimeras where the A segment had been taken from the MC₁ or MC₅R genes had a vector-specific HindIII site before the start codon, whereas those taken from the MC₃ or MC₄R genes had a XhoI site before the start codon. All the chimeras had a vector-specific XbaI site after the stop codon. The chimeras starting with MC₁ or MC₅R sequences were cloned with HindIII and XbaI into the pcDNA.3 expression vector, and chimeras starting with MC₃ or MC₄R sequences were cloned with XhoI and XbaI into the pCiNeo vector. To assure that the chimeras were correct, they were sequenced by using an ABI Prizm sequencer (Applied Biosystems, Foster City, CA) (full accounts on the manufacture of the DNA constructs will be given elsewhere).

For receptor expression, COS-7 cells were grown in Dulbecco's

¹ Amino acid numbering throughout refers to the position (or corresponding position) in the MC₁R.

modified Eagle's medium with 10% fetal calf serum. Confluent cultures (80%) were transfected on 100-mm dishes with the expression constructs of chimeric or wild-type melanocortin receptors [10 μ g of DNA per dish and mixed with liposomes as described previously (Schioth et al., 1996b)]. After 12 to 16 h of transfection, the serum-free medium was replaced with growth medium, and the cells were cultivated for about 48 h and then scraped off, centrifuged, and used for radioligand binding.

Data Set for Proteochemometric Modeling. Unfortunately, we failed to obtain full-length constructs for some chimeras, whereas others showed very low levels of expression, making their use unfeasible. To obtain a working set of receptors, we therefore combined the F- and S-sets so that the final set included 4 native and 13 chimeras (eight from the F-set and five from the S-set), as depicted schematically in Fig. 1. As seen, each of the receptors can be considered as consisting of five segments. Parts of all four native receptors are well represented in the chimeras, with the exception of the third segment of MC₄R and fifth segment of MC₅R, which are present only in the native receptors and one of the chimeric receptors. The number system given in Fig. 1 will be used to denote these multiple chimeric receptors.

Eighteen organic compounds that show binding activity for MCRs were synthesized in our laboratory (Fig. 2). Of these compounds, one (1) had been designed earlier by others (Sebhat et al., 2002), whereas the rest were our original designs. We reported compounds 2 and 4 earlier (Mutulis et al., 2002a, 2002b). Full accounts on the synthesis of the other compounds will be given elsewhere.

Interaction affinities (expressed as the negative logarithm of dissociation constants; pK_i values) were determined by using competition binding assays with the radioligand ¹²⁵I-NDP-MSH. Dissociation constants of the radioligand for each receptor were estimated by saturation assays, and the dissociation constants of competing compounds were then determined by competition assays. All calculations were based on nonlinear curve fitting assuming that ligands bind to a site according to the law of mass action, essentially using the approach described earlier (Schioth et al., 1995, 1996b). The radioligand binding was performed in Dulbecco's minimal essential medium containing 2 g/l albumin and 0.2 g/l phenanthroline without fetal calf serum. Cells were washed with binding buffer, scraped off, and distributed into 96-well non-culture-coated plates, which were centrifuged, and the binding buffer was removed. The cells were then immediately incubated with the radioligand and organic compound for 1 h at 37°C in 50 μ l of binding buffer/well. In the saturation experiments, the concentration of ¹²⁵I-NDP-MSH was varied by di-

lution in 2-fold intervals covering a range of about 6 pM to 12 nM. Nonspecific binding was defined in the presence of 3 μ M NDP-MSH. In the competition experiments, about 1 nM ¹²⁵I-NDP-MSH and various concentrations of competing ligand were added to the assays. After incubation, the cells were washed with 0.2 ml of ice-cold binding buffer and then detached from the plates with 0.2 ml of 0.1 N NaOH. The binding assays were performed in duplicates and repeated at least three times. Curve fitting for computing dissociation constants was performed using Prism 4 software (GraphPad Software, Inc., San Diego, CA). The dissociation constants (pK) for ¹²⁵I-NDP-MSH obtained from the saturation studies of the wild-type and chimeric melanocortin receptors were as follows: MC1 = 10.34, MC3 = 9.31, MC4 = 8.73, MC5 = 8.58, 11-5-33 = 10.38, 33-5-44 = 8.66, 44-1-33 = 8.64, 44-3-55 = 8.81, 44-5-11 = 8.54, 55-1-44 = 8.45, 55-3-11 = 10.85, 55-4-33 = 9.51, 1-333-4 = 9.98, 3-555-4 = 8.62, 4-555-1 = 9.35, 5-111-4 = 10.04, and 5-333-1 = 9.18. Results from the competition studies are given in Table 1.

Our data set obtained from 18 organic compounds and 17 receptors thus included $18 \times 17 = 306$ interaction affinity values (Table 1). In some cases, competition binding was not observed up to a concentration of 1 mM ($pK_i = <3$). In these cases, we arbitrarily assigned a pK_i value of 3. The large number of observations allowed us to divide the data into a work set comprising the receptor affinities of 14 compounds that were used for model creation and a test set comprising the receptor affinities of four compounds that were set aside and used after the creation of the proteochemometric model to assess the model's predictive ability.

Description of Organic Compounds. Compounds were characterized by 62 descriptors, which were calculated by the Dragon 2.1 software (Talet S.r.l., Milano, Italy). Descriptors represented different physicochemical properties (e.g., molecular weight, van der Waals volume, electronegativity, polarizability, molar refractivity, polar surface area, and log P, etc.) and the numbers of functional groups and structural fragments in the molecule. Before their use, the descriptors were checked for mutual correlation. For each pair of descriptors with a mutual correlation higher than 0.95, the one showing the highest correlation with any other descriptor was excluded. A fair number of the remaining descriptors showed invariant (constant) values for more than 4/5 of the compounds. For reasons suggested elsewhere (Q2 manual; Multivariate Infometric Analysis S.r.l., Perugia, Italy), these descriptors were also discarded. After these procedures, 31 descriptors remained in the data set (see list in Fig. 4 legend).

Binary Description of Receptors. As discussed above, the receptors can be considered as consisting of five segments. We described each of these segments separately by using four binary descriptors. The first descriptor was equal to 1 when the segment was taken from MC₁R; otherwise, it was set to -1. The second descriptor was equal to 1 when the segment was taken from MC₃R; otherwise, it was set to -1, and so forth. In this way, each receptor was represented by $5 \times 4 = 20$ descriptors.

Ligand-Receptor Cross-Terms. Ligand-receptor recognition depends on the complementarity of properties of two interacting entities. Such complementarity cannot be explained by linear combinations of ligand and receptor descriptors because complex nonlinear processes govern it. In proteochemometrics, the nonlinearity may be accounted for by computing ligand-receptor cross-terms (Lapinsh et al., 2001, 2002; Prusis et al., 2001, 2002; Wikberg et al., 2003). Cross-terms herein were obtained by multiplying mean-centered descriptors of compounds and receptors. In this way, an additional descriptor block was obtained that comprised $31 \times 20 = 620$ descriptors. Cross-terms were also calculated between mean-centered receptor descriptors, representing different sequence segments. This block included $(20 \times 16)/2 = 160$ descriptors.

Preprocessing of Data. All descriptors were first mean-centered and scaled to unit variance. Because the data set comprised descriptors of different types (i.e., descriptors of ligands, receptors, and

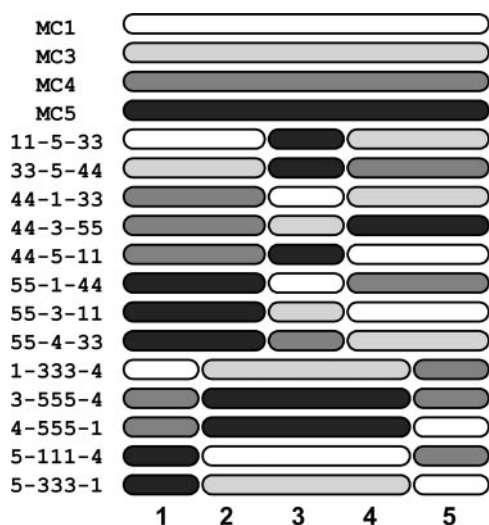


Fig. 1. Schematic representation of the 4 native and 13 chimeric melanocortin receptor variants used.

cross-terms), block scaling was applied. Although the variables of the same type kept equal variance, scaling weights between blocks were systematically varied until an optimal (i.e., the most predictive) model was obtained. The response variable (pK_i) was also mean-centered before use in the computations.

Partial Least-Squares Projections to Latent Structures. Descriptors were correlated to the affinity data by partial least-squares projection to latent structures (PLS). PLS is a multivariate analysis method that finds the relationship between predictor variables (X matrix) and response variables (Y matrix or vector; in our case, the Y corresponded to pK_i). The PLS analysis has the objective to approximate X and Y by simultaneous projecting to latent variables (components), with an additional constraint to maximize the covariance between projections of X and Y. For each response, PLS derives a regression equation, where regression coefficients reveal the direction and magnitude of the influence of X variables on the response (for a detailed description of PLS algorithms, see Geladi and Kowalski, 1986; Wold, 1995).

For a proteochemometric model comprising descriptors of receptors, ligands, ligand-receptor cross-terms, and intrareceptor cross-terms, the regression equation can be expressed as follows:

$$y = \bar{y} + \sum (\text{coeff}_r \times (x_r - \bar{x}_r)) + \sum (\text{coeff}_i \times (x_i - \bar{x}_i)) + \sum (\text{coeff}_{r,i} \times (x_r - \bar{x}_r) \times (x_i - \bar{x}_i)) + \sum (\text{coeff}_{r_1,r_2} \times (x_{r_1} - \bar{x}_{r_1}) \times (x_{r_2} - \bar{x}_{r_2})) \quad (1)$$

PLS analysis was carried out using the Q2 software (Multivariate Infometric Analysis S.r.l.).

The goodness of fit of the PLS models was characterized by the fraction of explained variation of Y (R^2Y). The predictive capability was characterized by the fraction of the predicted Y variation (Q^2) and assessed by cross-validation, as described previously (Baroni et al., 1993; Eriksson and Johansson, 1996). R^2Y may vary between 0

and 1; the value increases by each extracted PLS component. Q^2 values usually vary between 0 and 1; however, negative values can also be encountered, indicating nonpredictive models. A model of biological data is generally considered acceptable if $R^2Y > 0.7$ and $Q^2 > 0.4$ (Lundstedt et al., 1998). In the current study, cross-validation was performed using five randomly formed groups and repeated 100 times. The Q^2 estimates were used to adjust block-scaling weights and determine the optimal number of PLS components.

Three-Dimensional Modeling and Physicochemical Characterization of the Melanocortin Receptors' Ligand Binding Pocket. We constructed three-dimensional models of MCR transmembrane regions by using the crystal structure of bovine rhodopsin as a template (Palczewski et al., 2000). Sequence alignments of human MC₁₋₅Rs and bovine rhodopsin were taken from the GPCR database (Horn et al., 2003). An alignment of MCR transmembrane regions showed that over 40% (78 of 178) of the amino acids were conserved among all four MCR subtypes used herein. Using the three-dimensional models, we selected the residues that varied between the receptors and which faced the inside of the ligand binding cavity. Thirty-seven residues were chosen and subsequently coded by z-scale descriptors (Sandberg et al., 1998). These z-scales encapsulate 26 measured and computed physicochemical properties of amino acids, are obtained by principal component analysis from the original properties, and are accordingly 1) orthogonal to each other and 2) scaled in such a way that the same numerical difference in each z-scale corresponds to the same physicochemical difference between amino acids. Furthermore, z-scales are interpretable and represent essentially the hydrophobicity (z1), steric bulk properties and polarizability (z2), polarity (z3), and electronic effects (z4 and z5) of amino acids. Moreover, these five scales represent more than 95% of the original measured and computed properties of the amino acids. In

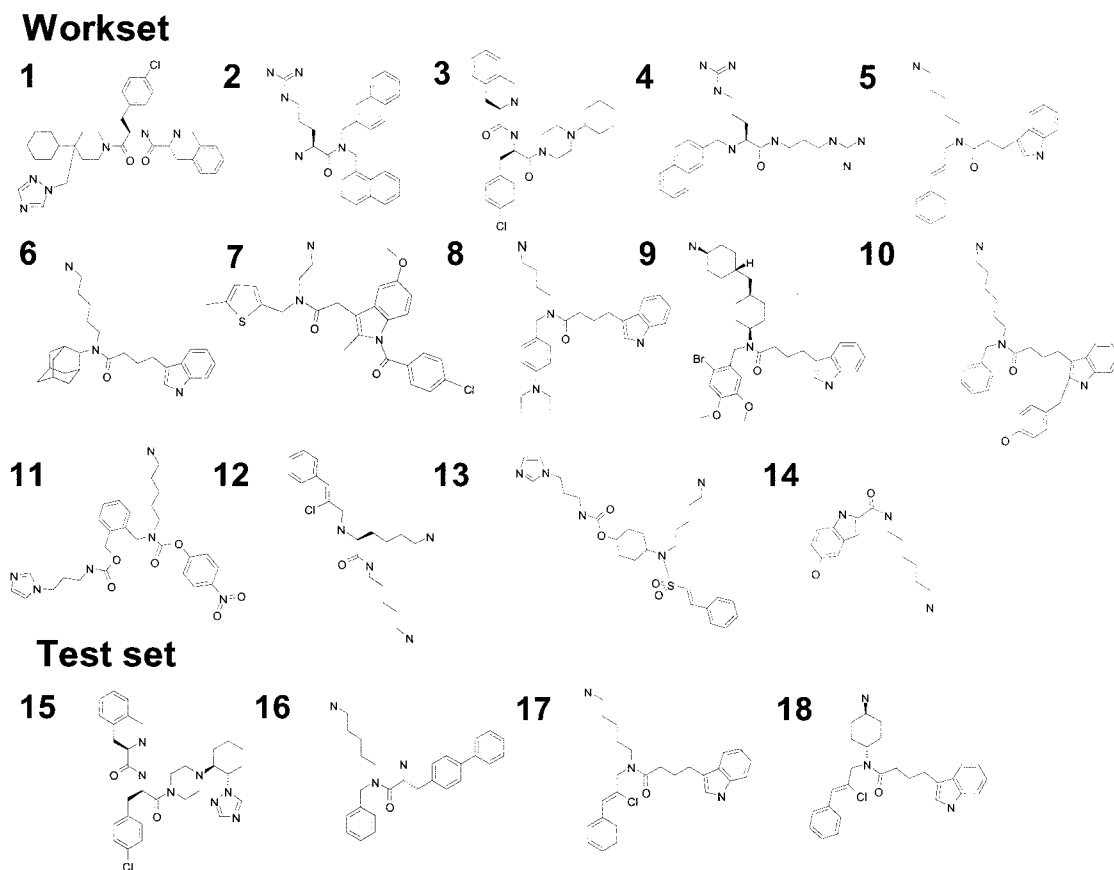


Fig. 2. Structures of organic compounds used herein.

this way, the differences in presumed MCR binding pockets were encoded by $37 \times 5 = 185$ descriptors.

Results

Results of Radioligand Binding

The affinities of the 18 compounds for the 4 wild-type MC_{1,3-5} and 13 multiple chimeric melanocortin receptors, determined by radioligand binding, are shown in Table 1. As seen, the affinities covered a range of more than four logarithmic units. Most of the compounds were MC₁R-selective, whereas three compounds showed their highest affinities for the MC₄R. As seen in the table, the data were divided into a work set comprising 14 compounds and a test set comprising four compounds. For the subsequent modeling, only the work set was used, whereas the test set was used to validate the model by using so-called external prediction (see below).

Creation of the Proteochemometric Model

PLS modeling of the work set (Table 1) using only descriptors of receptors and organic compounds resulted in a five-component model explaining $R^2Y = 0.77$ of the variance of compound affinities and having a predictive ability of $Q^2 = 0.70$. Ligand-receptor cross-terms were then included, allowing us to account for the nonlinearity of the ligand and receptor-affinity profiles. This resulted in a five-dimensional model explaining $R^2Y = 0.86$ at $Q^2 = 0.73$. Further improvement was obtained by including intrareceptor cross-terms. This was a reasonable measure because the creation of chimeras would be expected to result in receptors with altered folding having a negative influence on the ligand binding. Indeed, PLS modeling showed that some of the intrareceptor cross-terms attained large negative coefficients, indicating that some particular combinations of receptor segments diminish the ligands' affinities. However, most of the intrareceptor cross-terms were insignificantly small. The final model was therefore created by including only the 16 intrareceptor cross-terms showing the largest negative coefficients.

TABLE 1

Binding affinities (pK_i) of organic compounds for 4 native and 13 multiple chimeric melanocortin receptors. pK_i values represent the mean of at least three separate determinations. The number coding of chimeric receptors represents the presence of the corresponding sequence of the MC₁, MC₃, MC₄, or MC₅R in each of the five varying sequence stretches.

Comp. Nr	Receptor																
	MC1	MC3	MC4	MC5	11-5-33	33-5-44	44-1-33	44-3-55	44-5-11	55-1-44	55-3-11	55-4-33	1-333-4	3-555-4	4-555-1	5-111-4	5-333-1
Work set																	
1	6.1	5.3	7.1	5.3	5.3	5.3	5.8	6.9	5.9	4.9	5.4	5.5	7.0	4.5	4.6	5.8	5.2
2	5.8	5.5	5.4	5.5	5.0	5.1	5.6	5.3	5.5	5.7	4.9	5.7	5.6	5.1	5.5	5.7	5.0
3	4.9	5.3	5.8	5.0	4.8	4.6	5.2	7.4	5.3	4.8	4.8	4.8	5.6	4.4	4.5	5.1	4.8
4	5.3	5.0	5.0	4.1	5.0	4.9	5.1	4.6	5.6	6.0	4.4	5.2	5.4	5.2	5.3	4.7	4.5
5	5.4	4.6	5.1	5.1	4.2	4.9	4.9	4.7	4.9	4.9	4.7	5.1	4.8	4.5	5.2	4.7	4.6
6	5.2	4.6	4.9	4.3	4.2	4.5	4.9	4.7	5.0	4.6	4.9	4.9	4.5	4.4	5.0	4.9	4.6
7	5.2	4.7	4.7	4.5	4.0	4.4	4.8	4.6	4.7	4.7	4.3	4.8	4.5	4.3	4.7	5.3	4.6
8	5.2	4.5	4.6	4.5	4.1	4.6	4.6	4.5	4.9	4.6	4.7	4.8	4.5	4.3	4.9	4.5	4.4
9	5.4	4.8	4.8	4.8	4.0	4.1	4.3	4.1	4.1	4.3	4.8	5.2	4.2	4.1	4.8	4.6	4.8
10	5.1	4.2	4.4	4.4	4.0	4.3	4.7	4.4	4.8	4.5	4.4	4.7	4.5	4.2	4.6	4.8	4.2
11	4.6	3.8	4.2	4.3	4.2	3.9	4.4	4.4	4.6	4.5	4.0	4.0	4.1	4.0	4.2	4.5	3.9
12	4.4	4.1	4.0	3.5	3.7	3.9	4.1	3.6	4.6	5.0	3.6	3.9	4.1	3.9	3.8	4.1	3.6
13	4.7	4.1	4.0	3.9	3.6	3.4	4.2	4.0	4.4	4.2	3.7	3.7	3.7	3.5	3.6	5.1	3.4
14	3.9	3.5	3.3	3.4	3.2	2.7	3.6	3.2	3.7	3.4	3.0	3.7	3.0	3.0	3.5	3.0	3.2
Test set																	
15	6.5	5.1	6.5	4.9	4.8	4.6	5.4	7.4	5.3	5.0	4.4	4.9	6.8	4.0	4.2	5.4	5.1
16	5.4	4.7	4.9	4.7	4.3	4.7	5.0	4.7	5.5	5.2	5.0	5.3	5.0	4.7	5.2	5.1	4.6
17	5.1	4.7	4.8	4.5	4.1	4.3	4.9	4.7	4.9	4.9	4.5	5.1	4.6	4.8	5.1	4.6	4.8
18	5.5	4.3	4.7	4.8	4.0	4.7	4.6	4.4	4.7	4.7	5.0	5.0	4.8	4.5	4.9	4.6	4.5

The performances of the models after extracting different numbers of PLS components are summarized in Table 2. As seen, extracting five to seven PLS components led to models with the same predictive abilities ($Q^2 = 0.79$). Cross-validation was further performed so that all 17 observations of each compound were included in the same cross-validation group. In this way, the capacity of the model to predict the affinity of novel ligands (herein termed Q^2_{lig}) was assessed. Likewise, we assessed Q^2_{rec} by including all 14 observations of each receptor in the same cross-validation group. As seen in Table 2, very high Q^2_{rec} values of about the same magnitudes were obtained after extracting five to seven PLS components. However, the Q^2_{lig} value reached its largest value after extracting seven components. Closer inspection of the cross-validation results showed that one of the compounds (**4**; the only structure containing two guanidine groups) was systematically predicted with an affinity that was too low. Without including the predictions for this compound, Q^2_{lig} would have reached 0.73. Results for the final model (i.e., the model with seven extracted components) are illustrated graphically in Fig. 3.

External predictions for the four compounds, not included in the data set during centering of descriptors, calculations of cross-terms, and model elaboration, confirmed the high predictive capacities of the model ($Q^2_{\text{ext}} = 0.73$). The results are presented graphically in Fig. 3, where the observed versus predicted pK_i values are plotted. In the following sections, the seven-component model will be referred to as "the model" and is, unless otherwise stated, the one used in all subsequent analysis.

Interpretation of the Model

Analysis of Compound Properties of Importance for Melanocortin Receptor Binding. To analyze the influence of different properties of the compounds on their overall affinities to MCRs, we used the PLS regression equation of the model. The PLS coefficients for compound descriptors are shown in Fig. 4. As can be seen, the regression coefficients for

the numbers of nitrogen atoms, secondary aliphatic amines and amides, unsaturation index, number of unsubstituted aromatic sp² carbon atoms, molecular weight, and number of circuits attained the largest positive values. The presence of phenol in the molecule gave a large negative impact. The numbers of tertiary aliphatic amines, oxygen atoms, rotatable bond fraction, and mean electrotopological state also correlated negatively to the affinity, whereas the mean atomic van der Waals volume and log P correlated positively. A negative coefficient was also assigned to the number of halogen atoms in the molecule; however, halogen atoms attached to the aromatic ring were assessed positively. It is surprising that only minor positive correlation to the affinity was associated to the number of 6-, 9-, and 10-membered rings.

The sign and magnitude of the PLS coefficient of the descriptor of the compounds reflects the impact of the underlying property of the compounds to the affinity to the receptor series. However, depending on the actual descriptor value for a particular compound, the contribution of the described property to the binding would for some compounds be positive, whereas for others it would be negative. Therefore, to reveal the contribution of the properties of particular compounds to their interaction activity, we multiplied each coefficient with the actual descriptor value for each given compound as follows:

TABLE 2

Results of the PLS modeling after extracting a different number ($A = 1-7$) of PLS components. Shown are the explained X and Y variance (R^2X and R^2Y) and predictive ability (Q^2) of the models. Q^2 was assessed by cross-validation with five randomly formed groups and repeated 100 times; Q^2_{lig} and Q^2_{rec} were assessed by cross-validation leaving out one ligand or receptor at a time, respectively; Q^2_{ext} was calculated from the predictions for the four compounds of the test set.

A	R^2X	R^2Y	Q^2	Q^2_{lig}	Q^2_{rec}	Q^2_{ext}
1	0.15	0.57	0.54	0.42	0.52	0.31
2	0.23	0.70	0.64	0.48	0.63	0.44
3	0.27	0.84	0.75	0.36	0.71	0.64
4	0.34	0.88	0.78	0.41	0.74	0.66
5	0.42	0.89	0.79	0.57	0.75	0.70
6	0.48	0.91	0.79	0.65	0.76	0.71
7	0.52	0.92	0.79	0.68	0.76	0.73

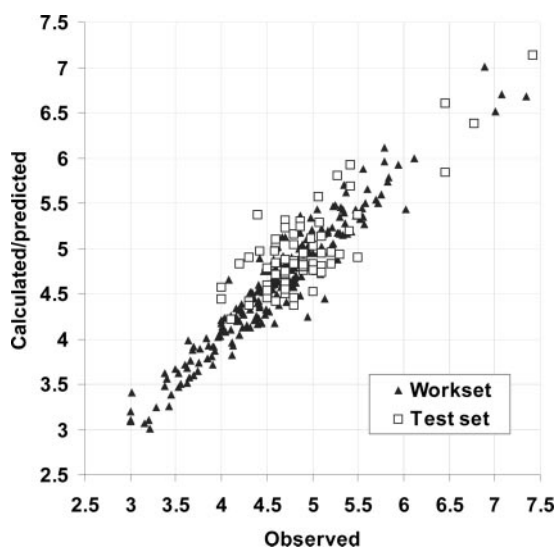


Fig. 3. Correlation of calculated and predicted pK_i versus measured pK_i values derived by the PLS modeling of receptor-ligand interactions.

$$\Delta pK_{i,1} = \text{coeff}_1 \times (x_1 - \bar{x}_1) \quad (2)$$

Using this approach, we found that the overall high affinity of compound **1** is associated with high numbers of nitrogen atoms and 6-membered rings, a low rotatable bond fraction (i.e., lack of long alkyl chains), and the presence of secondary and tertiary amides in the scaffold of the structure. Nevertheless, a negative influence is afforded by the presence of chlorine (and halogen, although the attachment of halogen to the aromatic ring is positively assessed). Thus, the model suggests that the high average affinity of the structure is not caused by the presence of chlorine and would not be lost by the replacement of the chlorobenzene group by, for example, naphthalene (such a modification, however, would essentially change the selectivity profile of the compound).

In compound **2**, the most positive influence is afforded by a high number of unsubstituted aromatic carbon atoms and a high unsaturation index value (i.e., properties that in the present case indicate the presence of two naphthalene moieties in the structure). Properties that indicate the presence of guanidine are also positively assessed; thus, we may conclude that a further increase in the affinity of **2** could be sought by modifying the scaffold of the structure. Similar analysis reveals that the affinity of compound **4** could be significantly improved by including more aromatic groups and increasing the hydrophobicity (log P) of the structure.

Contribution of Receptor Segments for Binding of Organic Compounds. We used the PLS regression equation of the model to calculate the change in pK_i for each compound for 20 hypothetical receptors, in which one of the sequence segments is taken from MC_{1,3-5}, whereas the descriptor values for the other four segments are replaced by

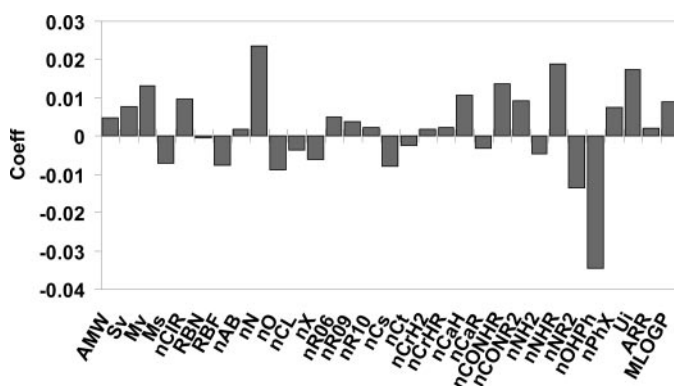


Fig. 4. PLS coefficients of the descriptors of organic compounds derived from the final proteochemometric model. Abbreviations correspond to descriptors as follows: AMW, average molecular weight; Sv, sum of atomic van der Waals volumes (scaled on carbon atom); Mv, mean atomic van der Waals volume (scaled on carbon atom); Ms, mean electrotopological state; nCIR, number of circuits; RBN, number of rotatable bonds; RBF, rotatable bond fraction; nAB, number of aromatic bonds; nN, number of nitrogen atoms; nO, number of oxygen atoms; nCL, number of chlorine atoms; nX, number of halogen atoms; nR06, number of 6-membered rings; nR09, number of 9-membered rings; nR10, number of 10-membered rings; nCs, number of total secondary C(sp³); nCt, number of total tertiary C(sp³); nCrH2, number of ring secondary C(sp³); nCrHR, number of ring tertiary C(sp³); nCaH, number of unsubstituted aromatic C(sp²); nCaR, number of substituted aromatic C(sp²); nCONHR, number of secondary amides (aliphatic); nCONR2, number of tertiary amides (aliphatic); nNH2, number of primary amines (aliphatic); nNHR, number of secondary amines (aliphatic); nNR2, number of tertiary amines (aliphatic); nOHP, number of phenols; nPhX, number of X-C on aromatic ring; Ui, unsaturation index; ARR, aromatic ratio; MLOGP, Moriguchi octanol-water partition coefficient (log P).

the mean value for the data set (i.e., zero after centering). Thus, 20 parameters were obtained for each compound, herein termed $\Delta pK_i(1,MC_1)$ to $\Delta pK_i(5,MC_5)$, allowing comparisons of contribution of different receptor sequence segments to the compounds' affinities.

To assess the importance of sequence segments for binding of the whole compound series, we calculated the averages of the $\Delta pK_i(1,MC_1)$ to $\Delta pK_i(5,MC_5)$ values for the 14 compounds [$\Delta pK_i(1,MC_1)$ to $\Delta pK_i(5,MC_5)$], as depicted in Fig. 5A). However, our data set included compounds with differing selectivity. Thus, two compounds (**1** and **3**) were MC₄R-selective, whereas most compounds showed their highest affinities for the MC₁R. Hence, we plotted the $\Delta pK_i(1,MC_1)$ to $\Delta pK_i(5,MC_5)$ for the 12 MC₁R-selective and two MC₄R-selective compounds separately (Fig. 4, B and C, respectively).

As seen in Fig. 5A, the first, second, and third segments strongly influence the affinity of the series of compounds. For example, the model predicts that the exchange of the first segment in the MC₃R with the corresponding segment in the MC₄R would result in an increase in the average affinity of the compounds by 0.15 pK_i units. The second and third segments of MC₄R also show positive influence. The higher affinity of the compounds for MC₁R versus MC₃R and MC₅R is explained mainly by differences in the third segment; e.g., the exchange of the third segment in the MC₁R with the corresponding segment in the MC₅R is estimated to reduce the affinity by about 0.2 pK_i units. Moreover, as seen in Fig. 5, B and C, the exchange of the third segment in the MC₁R with the corresponding segment in the MC₃R is estimated to reduce the affinity of MC₁R-selective compounds, whereas it would increase the affinity of the two MC₄R-preferring compounds. In fact, the pattern in Fig. 5C could suggest that the high-affinity binding of the latter compounds is accomplished by interactions with those residues in the third segment that are identical or similar in MC₃R and MC₄R.

By contrast, the fourth and fifth segments only slightly influence the average affinity of the compound series. However, analyzing the MC₁R- and MC₄R-selective compounds separately (Fig. 5, B and C) shows that the third and fifth

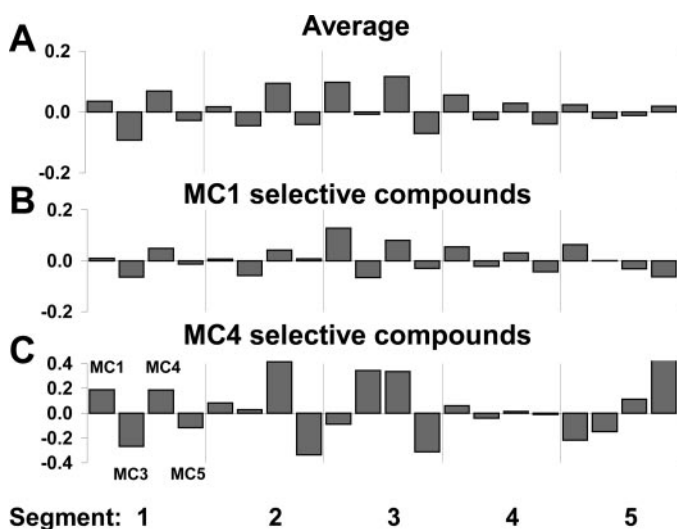


Fig. 5. Contribution of receptor segments for binding of organic compounds derived from the proteochemometric model. A shows $\Delta pK_i(1-5,MC_{1-5})$ values for all 14 compounds, whereas in B and C the $\Delta pK_i(1-5,MC_{1-5})$ is plotted separately for the 12 MC₁ and two MC₄R-selective compounds (see text for further details).

segments are responsible for differentiating the compounds into MC₁R- and MC₄R-selective ones. As seen in Fig. 5C, the affinity of the MC₄R-selective compounds for some chimeras is actually expected to be even higher than for the native receptor. This agrees with the observed high affinity of compounds **1** and **15** to, e.g., chimera 44-3-55 (see Table 1).

Contribution of Single-Sequence Residues for Ligand Affinity. As described above, the four ΔpK_i values for each sequence segment characterize the changes in ligand affinity once the segment is exchanged between MC_{1,3-5}Rs. Needless to say, these affinity changes originate from the differences in physicochemical properties of nonconserved residues of the given segment in the MC_{1,3-5}Rs. In a further analysis of the model, we wanted to find out which particular amino acids and amino acid properties are involved in creating selectivity for the melanocortin receptor ligands. This was done by correlating z-scale descriptors of the nonconserved amino acids located inside the TM cavity of the MCRs to the $\Delta pK_i(1,MC_1)$ to $\Delta pK_i(5,MC_5)$ values calculated from the proteochemometric model.

The analysis was performed separately for each of the five receptor segments by applying PLS. Thus, in each PLS model, the Y vector (a single column with four rows) comprised the four $\Delta pK_i(\text{segment Nr}, MC_{1,3-5})$ values, whereas the X matrix ($5 \times n$ columns, four rows) comprised the z-scales of n amino acids in the respective segment of the four (MC_{1,3-5}) receptors. Because the numerical differences in z-scales reflect physicochemical differences between amino acids, only centering of descriptors, but not rescaling to unit variance, was performed before the PLS modeling.

The analysis revealed that the $\Delta pK_i(1,MC_{1,3-5})$ values for the first segment essentially correlate to the physicochemical properties of amino acid position 41 (Ser, Lys, Ser, and Ala in MC_{1,3-5}, respectively), the most important being z-scales 3 and 4 for this amino acid (the results are presented graphically in Fig. 6A; as seen, the third z-scale shows the largest positive coefficient, whereas z4 shows the largest negative coefficient). An inspection of z-scale values for particular amino acids suggests that it is the differences in polarity between nucleophilic Ser (z3 = 1.15; z4 = -1.39) and electrophilic Lys (z3 = -2.49; z4 = 1.49) that are responsible for the compounds' preference for MC₄R and MC₁R versus MC₃R and, to a lesser extent, versus MC₅R (z-scale values for Ala being z3 = 0.60 and z4 = -0.14). Position 41, however, is occupied by the same amino acid in MC₁ and MC₄R and can thus not explain selectivity between these two receptors. An explanation is instead found by the large negative coefficient for z1 of residue 38 (Val, Val, Leu, Met), which positively assesses the more hydrophobic Leu (z1 = -4.28) in this position of the MC₄R.

For the second segment (Fig. 6B), several sequence residues seem to jointly explain the differences in selectivity of the compounds for MC₄R versus MC₅R and MC₃R. The largest negative coefficient is assigned to z1 of sequence position 114 (Gln, Gln, Val, Arg). This z-scale differentiates hydrophobic Val (z1 = -2.59) in the MC₄R from the hydrophilic Arg (z1 = 3.52) and, to a somewhat lesser extent, Gln (z1 = 1.75). High negative coefficients are also assigned to z2 and z3 at sequence position 120 (Ile, Phe, Ile, Phe) (i.e., a binary variation), preferring Ile in MC₁ and MC₄ over Phe in MC₃ and MC₅Rs. It is also noticeable that the other residues located at the extracellular end of TM2 (residue 99) and TM3

(116) are assigned higher absolute values of coefficients than the residues located deeper in the TM cavity (positions 83, 91, 124, 128, 129, 132, and 136).

For the third segment, the affinities are strongly increased if this segment is taken from the MC₁R or MC₄R, whereas they are decreased if it is taken from the MC₅R (see above; Fig. 5A). Moreover, the third segment (together with the fifth segment; see below) is of major importance in the creation of the selectivities of the two MC₄R-selective compounds. Therefore, the data corresponding to Fig. 5, B and C, were interpreted rather than those corresponding to Fig. 5A. Therefore, two separate PLS models were created. The results from this analysis are depicted in Fig. 6C. As can be seen, several residues, namely 171 (Ala, Cys, Ala, Phe), 175 (Phe, Cys, Ser, Cys), and 187 (Ala, Met, Ala, Tyr) were assessed to be important by both models, although the coefficients for particular z-scales are different or even opposite.

Thus, it can be inferred from the model assessing the

binding of MC₄R-selective compounds that the Phe in position 171 affords a negative influence because of the hydrophobicity ($z_1 = -4.22$) and high polarizability ($z_2 = 1.94$) of this residue. The third and fifth z-scales of this position, separating Cys from Ala, show less importance. By contrast, it can be inferred from the model for MC₁R-selective compounds that the role of the Cys ($z_3 = 3.71$; $z_5 = -2.65$) of the MC₃R at this position is to diminish the ability of the MC₁R-selective compounds to bind. Together, Ala is a preferable amino acid at this position for the binding of both MC₁R- and MC₄R-selective compounds. A corresponding analysis for position 175 shows that MC₄R-selective compounds prefer Ser instead of Phe, whereas MC₁R-selective compounds prefer Ser instead of Cys. In addition to the three abovementioned residues, a few additional sequence positions seem to be somewhat important in the third segment, namely for MC₁R-selective compound residues 200 and 184 and MC₄R-selective compound residue 203.

The fourth segment has only a marginal influence on selectivity. However, the fifth segment gives an opposite effect on the MC₁R- and MC₄R-selective compounds. The analysis shows that this effect is caused primarily by the amino acid in position 264 (Ile, Ile, Tyr, Met) (Fig. 6D). Thus, according to the analysis, the Met ($z_4 = 1.94$; $z_3 = 0.47$) in the MC₅R has a larger positive impact to the binding of MC₄R-selective compounds than the Tyr ($z_4 = 0.04$; $z_3 = 0.43$) in the MC₄R. The presence of an Ile ($z_4 = -0.84$; $z_3 = -1.71$) at this position is estimated to diminish the affinity of the MC₄R-selective compounds, whereas it increases the affinity of all the other compounds. Coefficients with opposite signs are also assigned to z_1 and z_2 of residue 285 (Ala, Val, Ile, Ile) in the two models. For residues 265 (Val, Ile, Ile, Leu) and 292 (Ile, Val, Ile, Val) (i.e., positions occupied by physicochemically similar aliphatic amino acids), the signs are also opposite, but the coefficients of the z-scales are rather small, making it unlikely that the latter amino acids have any larger impact on the compounds' selectivity.

Three-Dimensional Model of the Binding Pocket in Melanocortin Receptors. A three-dimensional model of the MC₁R is shown in Fig. 7 (see *Materials and Methods* for details on three-dimensional modeling), with the residues described above being important for receptor subtype selectivity marked in color. As seen, most of the marked amino acids are located close to the extracellular border of the receptor transmembrane cavity, but residues 171, 175, 200, and 203 form a cluster located deeper inside the transmembrane bundle. Moreover, two other clusters of significant residues can be recognized, suggesting binding regions of functional groups of organic compounds. One of these regions is outlined by residues 38, 41, 114, 120, and 285; the other region is outlined by residues 184, 187, and 264. Thus, although the proteochemometric model cannot assess the involvement of nonvarying residues to the ligand binding, it reveals amino acids with a dominant role in creating ligand selectivity and thereby tentatively mapping out the recognition site for the organic compounds in the MCRs.

Discussion

We recently applied proteochemometrics to analyze the binding of natural melanocortin receptor ligands, namely the α -MSH peptide and some of its synthetic analogs to chimeric

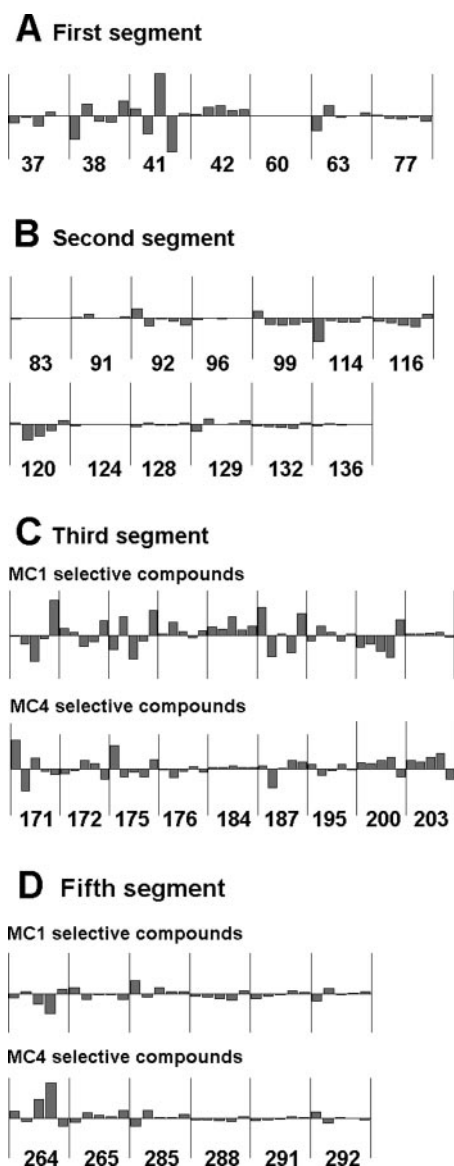


Fig. 6. Regression coefficients of PLS models correlating the physicochemical properties of varying sequence residues to the ΔpK_i (segment Nr, MC_{1,3-5}) parameters computed from the proteochemometric model (see text for further details).

MC₁/MC₃ receptors (Prusis et al., 2001, 2002). However, these studies could exploit only a binary description of the receptors because they were chimeras of two receptors, MC₁/MC₃; therefore, these studies allowed only a rough mapping of the ligand binding site for the melanocortin peptides. In yet another study, we applied proteochemometrics to study the interactions of a series of 54 organic compounds to native melanocortin receptors (Lapinsh et al., 2003). In this study, we could separately reveal chemical properties of the organic compounds that are important for their affinity and receptor subtype selectivity. However, because this study did not include any systematic variations of the receptor sequences, beyond that which is present in the wild-type receptors, it was not possible to reveal any information about the receptor properties that are responsible for creating affinity and of those that are involved in discriminating selective from non-selective compounds.

Therefore, in the present study, we evaluated a larger number of organic compounds on 4 native and 13 multiple chimeric melanocortin receptors, where the latter had been created by using statistical design methods, the purpose being to represent all possible combinations of three parts in four receptors in the best possible way in a minimal set. The binding data were then analyzed with the major aim of mapping out the receptors' binding pocket for the organic compounds. Applying proteochemometrics onto the data produced a model with high predictive ability. The standard deviation of errors of prediction corresponding to the obtained $Q^2 = 0.79$ of the model is 0.32 pK_i units. In view of the

intrinsic statistical error of the biological measurements, the modeling accuracy is thus very high.

Alongside with the conventional Q^2 parameter, we introduced two additional estimates of model predictive ability, the Q^2_{rec} and Q^2_{lig} . This seemed rational because one purpose of proteochemometric models is to make predictions for novel ligands and/or receptors rather than to merely predict affinity of untested combinations of moieties already present in the data set (i.e., filling gaps in a data table). The high values of Q^2_{rec} (0.76) and Q^2_{lig} (0.68), as well as the ability of proteochemometrics to perform accurate predictions for the four compounds not included during data preprocessing (i.e., centering of variables before calculating cross-terms) and model creation (i.e., adjusting block-scaling weights), affirms a high reliability of the proteochemometric modeling approach. Thus, these results indicate the usefulness of proteochemometrics for a priori drug design.

Although the present model was based on binary descriptors of receptors, we succeeded in a further analysis to reveal particular sequence residues that are the most likely to contribute to ligand selectivity. For this purpose, three-dimensional models of the receptors were first created, and properties of nonconserved residues that could form ligand binding pocket(s) were characterized by physicochemical descriptors. The values of these descriptors in the MC_{1,3-5}-Rs were correlated to the contributions of binary descriptors [$\Delta pK_i(\text{segment}, \text{MC}_{1,3-5})$] for each of the five receptor segments by PLS analysis. As shown in Fig. 7, 12 residues were found that potentially influence ligand selectivity. An interesting finding of these investigations was that, for the MC₁R-selective compounds, only residues from the third sequence segment showed up to be largely important. Three-dimensional modeling further indicates that only the two clusters shown on the left side in Fig. 7 are important for these compounds. By contrast, the region between residues 38, 41, 114, 120, and 285 also seems to be important for the two MC₄R-selective compounds, suggesting a more complex binding mechanism for these compounds.

Several studies were previously performed using site-directed mutagenesis in attempts to identify determinants for the MC₄R selectivity of melanocortin peptides. Thus, Nickolls et al. (2003) elucidated the effects of point mutations in MC₄R on the binding of 11 natural and synthetic peptides. It was found that an Ile125Phe mutation [corresponding to residue 120 (Ile, Phe, Ile, Phe) in MC_{1,3-5}; in our model, Ile is preferred to Phe] results in a 2- to 5-fold decrease in the affinity of MC₄-selective ligands, whereas the affinities of α -MSH and NDP-MSH are not significantly affected. The naturally occurring mutation of Ile137 [herein corresponding to 132 (Leu, Ile, Ile, Met)] to Thr significantly decreased the binding of most ligands. Because Thr shows higher values of z1 to z4 and lower z5 scale values compared with Ile, a drop of affinity by such a mutation agrees with our results. In a study by Haskell-Luevano et al. (2001), several mutations in the MC₄R were evaluated. It was found that the mutation of a Ser to Phe [herein position 175 (Phe, Cys, Ser, Cys); according to our model, MC₄R-selective compounds prefer Ser instead of Phe] resulted in a 4- to 6-fold decrease in the affinity of two cyclic peptides, whereas the affinity of NDP-MSH remained unchanged. The mutation of Met to Phe [herein position 195 (Phe, Met, Met, Met); unimportant according to our model] was reported to have no effect to agonist binding

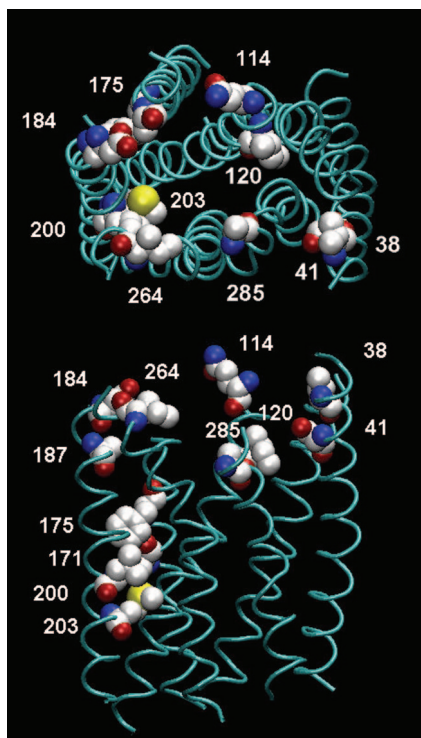


Fig. 7. Three-dimensional model of the MC₁ receptor with amino acids computed from the proteochemometric model found to be important for the creation of selective binding of organic amines. Top, transmembrane domains viewed from the extracellular side. Indicated are the side chains of the sequence residues found to be important for selective binding of organic compounds (also discussed in detail in the text). Bottom, side view of transmembrane domains, the extracellular portion facing up and the intracellular down, with the same residues indicated.

or potency; however, the mutation resulted in a constitutively active receptor. The mutation of the same residue to Ala in the study by Yang et al. (2000) was reported to decrease the affinity of α -MSH but not that of NDP-MSH.

Interpretations of mutagenesis data are straightforward when the change is one-dimensional and the effect it causes is a simple "additive" one (e.g., involving only one alteration and/or causing only a direct effect). When changes cause many simultaneous effects (e.g., by multiple interactions with the ligand and/or inside the receptor), the relations of the changes in activity to the changes in structure may become difficult or even impossible to reveal from just a few scattered observations. Using instead a set of mutated proteins that are designed to cover as much as possible a selected region of structural variation in conjunction with mathematical multivariate analysis, as applied herein, thus constitutes a solution. The data of the present study indicate indeed that the proteochemometrics modeling applied on data derived from the interactions of organic molecules with statistically designed multiple chimeric proteins is useful to map ligand recognition. Moreover, the models are quantitative and reveal the underlying chemical properties that determine ligand recognition, which is information that is highly desired in ligand design. Because the proteochemometrics approach is general, it could be applied to analyze the molecular recognition processes of any set of proteins.

References

- Baroni M, Costantino G, Cruciani G, Riganelli D, Valigi R, and Clementi S (1993) Generating optimal linear PLS estimations (GOLPE): an advanced chemometric tool for handling 3D-QSAR problems. *Quant Struct-Act Relat* **12**:9–20.
- Chhajlani V, Muceniec R, and Wikberg JES (1993) Molecular cloning of a novel human melanocortin receptor. *Biochem Biophys Res Commun* **195**:866–873.
- Chhajlani V and Wikberg JES (1992) Molecular cloning and expression of the human melanocyte stimulating hormone receptor cDNA. *FEBS Lett* **309**:417–420.
- Eriksson L and Johansson E (1996) Multivariate design and modeling in QSAR. *Chemom Intell Lab Syst* **34**:1–19.
- Gantz I, Konda Y, Tashiro T, Shimoto Y, Miwa H, Munzert G, Watson SJ, DelValle J, and Yamada T (1993a) Molecular cloning of a novel melanocortin receptor. *J Biol Chem* **268**:8246–8250.
- Gantz I, Miwa H, Konda Y, Shimoto Y, Tashiro T, Watson SJ, DelValle J, and Yamada T (1993b) Molecular cloning, expression and gene localization of a fourth melanocortin receptor. *J Biol Chem* **268**:15174–15179.
- Geladi P and Kowalski BR (1986) Partial least-squares regression: a tutorial. *Anal Chim Acta* **185**:1–17.
- Haskell-Luevano C, Cone RD, Monck EK, and Wan YP (2001) Structure activity studies of the melanocortin-4 receptor by in vitro mutagenesis: identification of agouti-related protein (AGRP), melanocortin agonist and synthetic peptide antagonist interaction determinants. *Biochemistry* **40**:6164–6179.
- Horn F, Bettler E, Oliveira L, Campagne F, Cohen FE, and Vriend G (2003) GPCRDB information system for G protein-coupled receptors. *Nucleic Acids Res* **31**:294–297.
- Lapins M, Prusis P, Gutcaits A, Lundstedt T, and Wikberg JES (2001) Development of proteochemometrics: a novel technology for the analysis of drug-receptor interactions. *Biochim Biophys Acta* **1525**:180–190.
- Lapins M, Prusis P, Lundstedt T, and Wikberg JES (2002) Proteochemometrics modeling of the interaction of amine G-protein coupled receptors with a diverse set of ligands. *Mol Pharm* **61**:1465–1475.
- Lapins M, Prusis P, Mutule I, Mutulis F, and Wikberg JES (2003) QSAR and proteochemometric analysis of the interaction of a series of organic compounds with melanocortin receptor subtypes. *J Med Chem* **46**:2572–2579.
- Lundstedt T, Seifert E, Abramo L, Thelin B, Nyström A, Pettersen J, and Bergman R (1998) Experimental design and optimization. *Chemom Intell Lab Syst* **42**:3–40.
- Mutulis F, Mutule I, Lapins M, and Wikberg JES (2002a) Reductive amination products containing naphthalene and indole moieties bind to melanocortin receptors. *Bioorg Med Chem Lett* **12**:1035–1038.
- Mutulis F, Mutule I, and Wikberg JES (2002b) N-Alkylaminoacids and their derivatives interact with melanocortin receptors. *Bioorg Med Chem Lett* **12**:1039–1042.
- Nickolls SA, Cismowski MI, Wang X, Wolff M, Conlon PJ, and Maki RA (2003) Molecular determinants of melanocortin 4 receptor ligand binding and MC4/MC3 receptor selectivity. *J Pharmacol Exp Ther* **304**:1217–1227.
- Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, et al. (2000) Crystal structure of rhodopsin: a G protein-coupled receptor. *Science (Wash DC)* **289**:739–745.
- Prusis P, Lundstedt T, and Wikberg JES (2002) Proteochemometrics analysis of MSH peptide binding to melanocortin receptors. *Protein Eng* **15**:305–311.
- Prusis P, Muceniec R, Andersson P, Post C, Lundstedt T, and Wikberg JES (2001) PLS modeling of chimeric MS04/MSH-peptide and MC₁/MC₃-receptor interactions reveals a novel method for the analysis of ligand-receptor interactions. *Biochim Biophys Acta* **1544**:350–357.
- Sandberg M, Eriksson L, Jonsson J, Sjöström M, and Wold S (1998) New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. *J Med Chem* **41**:2481–2491.
- Schiøth HB, Chhajlani V, Muceniec R, Klusa V, and Wikberg JES (1996a) Major pharmacological distinction of the ACTH receptor from other melanocortin receptors. *Life Sci* **59**:797–801.
- Schiøth HB, Muceniec R, and Wikberg JES (1996b) Characterisation of the melanocortin 4 receptor by radioligand binding. *Pharmacol Toxicol* **79**:161–165.
- Schiøth HB, Muceniec R, Wikberg JES, and Chhajlani V (1995) Characterisation of melanocortin receptor subtypes by radioligand binding analysis. *Eur J Pharmacol* **288**:311–317.
- Sebbat IK, Martin WJ, Ye Z, Barakat K, Mosley RT, Johnston DB, Bakshi R, Palucki B, Weinberg DH, MacNeil T, et al. (2002) Design and pharmacology of N-[(3R)-1,2,3,4-tetrahydroisquinolinium-3-ylcarbonyl]-(1R)-1-(4-chlorobenzyl)-2-[4-cyclohexyl-4-(1H-1,2,4-triazol-1-ylmethyl)piperidin-1-yl]-2-oxoethylamine (1), a potent, selective, melanocortin subtype-4 receptor agonist. *J Med Chem* **45**:4589–4593.
- Vogelstein B and Gillespie D (1979) Preparative and analytical purification of DNA from agarose. *Proc Natl Acad Sci USA* **76**:615–619.
- Wikberg JES (2001) Melanocortin receptors: new opportunities in drug discovery. *Exp Opin Ther Patents* **11**:61–76.
- Wikberg JES, Lapins M, and Prusis P (2004) Proteochemometrics—a tool for modelling the molecular interaction space, in *Methods and Principles in Medicinal Chemistry* (Müller G and Kubinyi H eds) vol 22, pp 289–309, Wiley-VCH, Weinheim, Germany.
- Wikberg JES, Muceniec R, Mandrika I, Prusis P, Lindblom J, Post C, and Skottner A (2002) New aspects on the melanocortins and their receptors. *Pharmacol Res* **42**:393–420.
- Wikberg JES, Mutulis F, Mutule I, Veiksina S, Lapins M, Petrovska R, and Prusis P (2003) Melanocortin receptors: ligands and proteochemometrics modeling. *Ann NY Acad Sci* **994**:21–26.
- Wold S (1995) PLS for multivariate linear modeling, in *Chemometric Methods in Molecular Design* (van de Waterbeemd H ed) vol 2, pp 195–218, VCH Verlagsgesellschaft, Weinheim, Germany.
- Yang YK, Fong TM, Dickinson CJ, Mao C, Li Y, Tota MR, Mosley R, Van Der Ploeg LH, and Gantz I (2000) Molecular determinants of ligand binding to the human melanocortin-4 receptor. *Biochemistry* **39**:14900–14911.

Address correspondence to: Dr. Jarl Wikberg, Pharmaceutical Pharmacology, Box 591, BMC, SE751 24 Uppsala, Sweden. E-mail: jarl.wikberg@farmbio.uu.se