

A phylogenetic method to assign ligand-binding relationships between 7TM receptors

Thomas M. Frimurer, Trond Ulven, Christian E. Elling, Lars-Ole Gerlach, Evi Kostenis and Thomas Högberg*

7TM Pharma AIS, Fremtidsvej 3, DK-2970 Horsholm, Denmark

Received 23 March 2005; revised 9 May 2005; accepted 26 May 2005

Available online 29 June 2005

Abstract—A computational protocol has been devised to relate 7TM receptor proteins (GPCRs) with respect to physicochemical features of the core ligand-binding site as defined from the crystal structure of bovine rhodopsin. The identification of such receptors that already are associated with ligand information (e.g., small molecule ligands with mutagenesis or SAR data) is used to support structure-guided drug design of novel ligands. A case targeting the newly identified prostaglandin D2 receptor CRTH2 serves as a primary example to illustrate the procedure.

© 2005 Elsevier Ltd. All rights reserved.

1. Introduction

The G protein-coupled receptors (GPCRs) allow signals from the exterior of cells to be communicated to second messenger systems within the cells. The receptors are integral membrane proteins characterized by seven transmembrane (7TM) helical segments traversing the membrane in an antiparallel way, with the N-terminal on the extracellular side of the membrane and the C-terminal on the intracellular side. This generic protein structure is extensively used for a variety of stimuli spanning from protons, fatty acids, monoamines, peptide hormones, glycoproteins to olfactory agents and light. The 7TM receptor superfamily is composed of many hundreds of receptors that can be further divided into smaller sub-families of receptors.¹ The largest of these sub-families is composed of the rhodopsin-like receptors, also termed family A receptors, named after the light-sensor in our eye.

Despite the fact that drugs have been successfully developed for 7TM receptors, efficient structure-based drug discovery is hampered by the lack of detailed structural information. Hitherto only one G protein-coupled receptor, bovine rhodopsin, has been subject to structural determination by X-ray crystallography at atomic resolution in its inactive conformation with 11-*cis*-retinal.²

Since the three-dimensional structure of only a single receptor has been solved to date, the helical lengths, and the beginning, centre and ends relative to the lipid bilayer membrane of each of the seven helices for target receptors are dissected by sequence analysis.³ Homology models of other rhodopsin-like 7TM receptors have been derived from the bovine rhodopsin structure.² These models have been used with varying success to drive the design and optimisation of novel ligands.⁴ Mutagenesis, metal-site engineering, biophysical and spectroscopic studies have supported the identification of a core-binding site located in the upper part of the transmembrane helical bundle.⁵

Much of the success in structure-based drug design in general for other target protein classes is driven by iterative processes utilising target protein as well as ligand information—especially ligand–protein complexes—in the design. To compensate for the lack of detailed structural information for the 7TM receptors it was reasoned that identification of receptors that already were associated with ligand information (e.g., small molecule ligands with mutagenesis or SAR data) and which were closely related to the target receptor could be used to improve the rational drug design process.

7TM receptor proteins have traditionally been classified based on their primary amino acid sequences, evolutionary phylogeny,⁶ or their pharmacological profile.¹ However, from a drug discovery perspective, it is more relevant to classify and characterize 7TM receptors according to their ligand recognition properties.

* Corresponding author. Tel.: +45 3925 7777; fax: +45 3925 7776; e-mail: th@7tm.com

Here, we describe a computational strategy which classifies 7TM receptors with respect to physicochemical features of selected amino acid residues located in a common structural framework constituting a core ligand-binding site as defined from the crystal structure of rhodopsin. The classification procedure involves the following steps: (i) primary sequence alignment (TM domain) of all 7TM receptors;^{3,7} (ii) identification and selection of core-binding site residues potentially involved in small molecule ligand binding and recognition;^{4,5,8} (iii) create binding site signatures and pseudo-sequence strings; (iv) assign physicochemical descriptors or indicator variables;^{9,10} (v) compare, rank and cluster 7TM receptors of interest.

2. 7TM sequence alignment (step i)

To allow for this type of comparison the protein sequences must be properly aligned using conventional alignment algorithms such as ClustalW.⁷ The resulting alignment is manually inspected and refined if necessary, so that conserved generic sequence signatures within the seven transmembrane helices are satisfied.

3. Definition of core-binding site (step ii)

Based on the conserved key residues in each transmembrane segment, a generic numbering system has been suggested.^{3,8} For example, in TM-II the highly conserved aspartate (Asp) is given the generic number 10, i.e., AspII:10, and all other residues in the helix are hence numbered on this basis (cf. Fig. 1).⁸ A large body of research has been dedicated to identify which amino acid residues that are associated with binding of small molecules.^{4,5} A proposed core-binding site consisting of 22 amino acids facing the binding cavity is shown in Figure 2 for the CRTH2 receptor (vide infra).

4. Descriptors applied to binding site signatures (steps iii and iv)

The binding site signature, represented by a ‘pseudo-sequence string’ of the 22 amino acid positions, is encoded

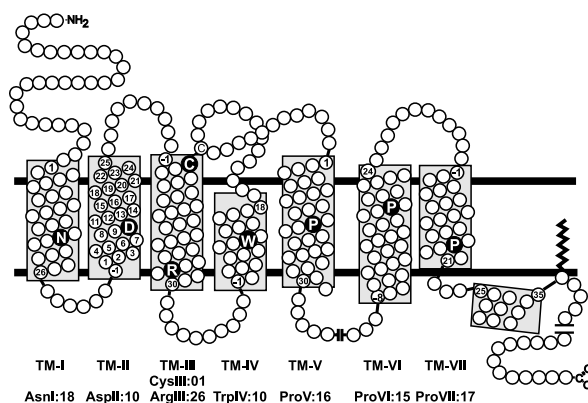


Figure 1. Schematic representation of the secondary structure elements of a rhodopsin-like 7TM receptor. The conserved key residues numbered according to the generic numbering system are highlighted in black.⁸

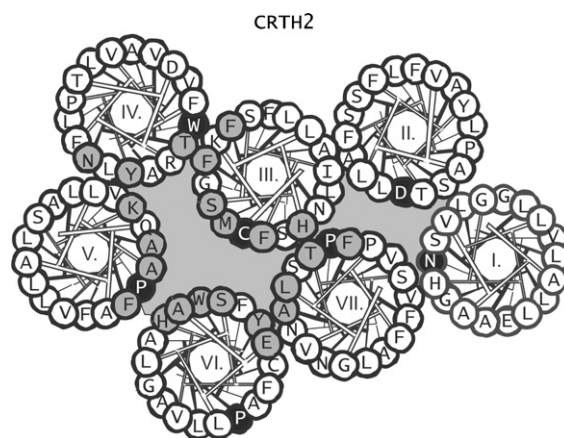


Figure 2. Helical wheel diagram of the amino acid positioning in the transmembrane domain of the CRTH2 receptor. Highly conserved sequence positions within the rhodopsin-like 7TM receptor family are shown in black and binding site residues in gray.⁸

with physicochemical descriptors relevant for ligand binding and recognition, such as ionic, ion–dipole, dipole–dipole, hydrogen bonding, hydrophobic, π -stacking, edge-on aromatic, and cation–interaction forces. The physicochemical descriptors can be experimentally derived and/or theoretically calculated.^{9,10} Such descriptors have successfully been employed in various types of quantitative structure–activity relationships (QSAR) analyses of drug/ligand responses.^{9,10}

A more simplistic analysis can be based on indicator variables, e.g., 1 and 0, to denote absence or presence of selected features as illustrated in Figure 3 for a tyrosine residue capable of hydrophobic, aromatic and polar interactions. The resulting 5-digit bit string 11001 will encode for its interaction features and the remaining residues are handled analogously to produce a combined bit string reflecting the interaction properties in the selected site. This binary representation allows for mathematically fairly simple Tanimoto-type comparisons between corresponding binding sites of large numbers of receptor proteins.¹¹

5. Computational classification procedure (step v)

Having defined the core ligand-binding site and associated physicochemical descriptors it is possible to com-

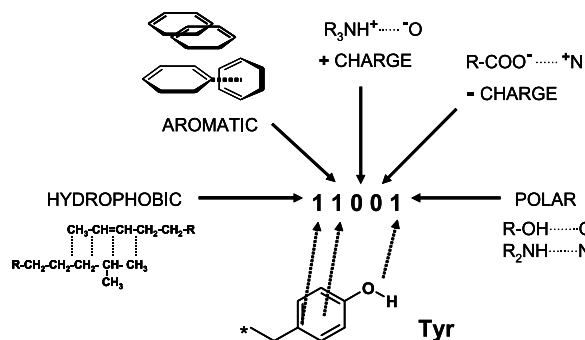


Figure 3. Bit map representation of potential ligand interactions with tyrosine.

pare, rank and cluster 7TM receptors with respect to their potential to interact with a given ligand or structurally similar ligands. By analogy to phylogenetic analysis, this process has been referred to as physicogenetic analysis.

Following the procedure above it is possible to quantify how similar a given receptor-binding site or subsite is to other receptor-binding sites and/or subsites. In general, binary descriptors are intuitively and computationally attractive due to their flexibility and the possibility to handle large data sets. In this work, we have used Cosine Coefficients to assess the similarity between binding site signatures providing a statistically easily normalized readout reflecting the percentage of overlap.^{11,12} More advanced types of descriptors and other similarity measures rely on pattern recognition methods and principal component analysis (PCA) (data not shown).¹¹

6. Results and discussion

The rhodopsin-like receptor, CRTH2, was recently identified as the second high-affinity prostaglandin D2 receptor.¹³ It is expressed on eosinophils, basophils and Th2 cells, and it has attracted attention as a novel target for treatment of allergic diseases like asthma and rhinitis. It was selected as an appropriate target for applying a structure-guided approach of identifying ligands since indomethacin was the only identified ligand for this receptor at the time we initiated the study described in this report.

It is important to emphasize that the relationship between proteins derived on these assumptions in most cases and for several reasons will be different from sequence alignment comparisons or a phylogenetic analysis based on substitution tables such as PARM, BLOSSUM, etc.⁶ Other means to compare monoamine related 7TM receptors based on ligand and chemogenomics input have also been devised.¹⁴ Examples of receptors where sequence alignment gives comparable relationships as physicogenetic approaches are typically illustrated by certain receptors with subclasses such as the neurokinin NK1 to NK4 receptors and muscarinic M1 to M5 receptors.

In this study, a model using binary codes describing absence or presence of aromatic, hydrophobic, negatively charged, positively charged and polar interactions was used to rank the human receptors with the Cosine Coefficient relative to the first noted receptor given in parenthesis.¹² Thus, the closest receptors to the muscarinic M3 was M2 (1.00), M4 (1.00), M1 (0.97) and M5 (0.97), which is in accordance with findings that muscarinic antagonists are in principle fairly subtype unselective.¹⁵ Likewise, the model ranks the receptors closest to the neurokinin NK1 as NK4 (0.88), NK3 (0.88), NK2 (0.87) and CXCR3 (0.63).

In contrast, the model applied to histamine H2 will rank the closest receptors as 5-HT₇ (0.76), α_{2C} (0.69), whereas the other histamine receptors are more remote, i.e., histamine H3 (0.61), H1 (0.59) and H4 (0.54).

The model applied to CRTH2 will rank the closest receptors as chemokine-like receptor 1 (0.79), angiotensin AT2 (0.74) and leukotriene 1 (0.64). Besides angiotensin AT1 (0.61), a few less investigated GPCRs were identified but since they are not associated with any ligand information they were omitted. In the ordinary phylogenetic analysis of the CRTH2 receptor, relationships with a number of chemotactic receptors (CysLTs, BLTs, FPRLs) and prostanoid receptors (TP, FP, IP, EPs, DP) are revealed.¹³ In the physicogenetic analysis, the chemotactic receptors are also identified, albeit with a more distant relationship. Notably, the AT1 and AT2 receptors are not identified as close neighbours according to the conventional evolutionary relationship model and hence provide novel opportunities.

Having identified 7TM receptors which resemble each other with respect to the physicochemical environment in the binding site (physicogenetics), it is possible to utilise known ligands which interact with one receptor as chemical starting points for drug development on related receptors (target jumping). It is generally recognized that structurally related small organic molecules tend to bind the same biological target proteins.¹⁶ Conversely, related biological targets also tend to bind the same or similar small organic molecule ligands. The transfer of a specific chemical starting point from one target to another related target could be seen as an example of chemogenomics. Thus, an efficient physicogenetic method to compare 7TM receptors having known small molecule ligands (known or potential drug molecules) with novel receptors lacking identified ligands allows for possibilities to identify lead structures for drug development since no previous information regarding ligands binding to the new receptor under investigation is needed. Our recent identification of the related dopamine receptors D₂ and D₃ to provide structural input in the design of melanin-concentrating hormone (MCH) antagonists may serve as an example of this process.¹⁷ Alternatively, one can derive pharmacophore models based on specific ligands binding to the receptors related to the receptor under investigation. Furthermore, one can use information supported by mutagenesis and SAR on how known ligands are thought to be binding to a known receptor to derive a pharmacophore based on the new target receptor homology model and the ligand interaction features obtained from the related receptor–ligand information.

The latter approach was applied to CRTH2 which at the start of the project only had been associated with the ligand indomethacin besides prostaglandin D2.¹³ As described above, physicogenetic analysis identified the related receptors angiotensin AT2 and AT1, which are associated with a wealth of ligand and mutational information which were used in the construction of the pharmacophore model. Especially, AT1 ligands and mutational data^{18,19} were used to identify crucial interaction sites in the AT1 receptor, which were transformed into the corresponding sites in the CRTH2 homology model²⁰ of the receptor to construct the pharmacophore model. The assumed binding site is characterised by a hydrophobic pocket surrounded by several phenylalanines, e.g., PheVI:13, PheV:09, PheV:13, PheIII:08,

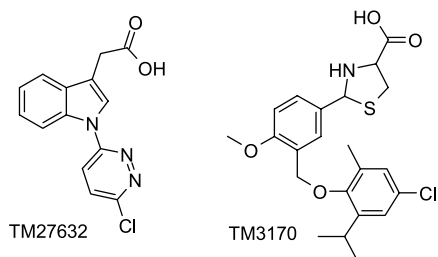


Figure 7. Some representative chemotypes identified after the *in silico* screening: TM27632 (IC₅₀, 2.4 μM) and TM3170 (IC₅₀, 1.9 μM).

of using a pharmacophore query that focuses on the interaction features of the ligands rather than on identified chemotypes or scaffolds from ligands to related receptors, e.g., candesartan. The graph also illustrates that an approach utilising AT ligands as structural starting points would have required substantial optimisation to suppress the AT affinities and turn them into useful lead structures. A diverse set of compounds with IC₅₀'s as low as 20–30 nM and selectivity over AT1 and AT2 were obtained. Several chemotypes have indeed served as starting points for internal drug discovery efforts.

In summary, a novel approach to identify 7TM receptors related with respect to ligand-binding features has been described. By incorporation of ligand information from related receptors identified by this protocol several approaches of structure-guided ligand design can be envisioned. A pharmacophore-based approach to retrieve a small library targeted for the novel receptor CRTH2 produced 10% hits having a large spread in chemotypes.

Conceptually, a smaller library requires less optimisation of the assay in comparison to the throughput and robustness needed for a conventional high-throughput screening (HTS) campaign. The smaller library size also permits more information-rich assays to be used and the post-screening work of validating the hits and sorting out non-wanted chemistries is considerably reduced. The limitation is obviously that the smaller number of compounds limits the possibility of covering enough chemotypes or providing SAR information upfront. The latter aspect is handled by using iterative loops to enrich the library with analogs of identified hits. An inherent risk with the small library size is that active compounds may be removed in the clustering and diversity analysis procedure, leaving only inactive representatives. However, we are convinced that the advantages in speed and saved work should in most cases outweigh such drawbacks, which is supported by this study on CRTH2 producing a large number of drugable hits.

Acknowledgment

The authors are grateful to Thue W. Schwartz for insightful support and stimulating discussions.

References and notes

1. (a) Fredriksson, R.; Lagerstrom, M. C.; Lundin, L. G.; Schioth, H. B. *Mol. Pharmacol.* **2003**, *63*, 1256; (b)

Fredriksson, R.; Schioth, H. B. *Mol. Pharmacol.* **2005**, *67*, 1414.

2. (a) Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M. *Science* **2000**, *289*, 739; (b) Filipek, S.; Teller, D. C.; Palczewski, K.; Stenkamp, R. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 375.

3. (a) Baldwin, J. M. *EMBO J.* **1993**, *12*, 1693; (b) Baldwin, J. M.; Schertler, G. F.; Unger, V. M. *J. Mol. Biol.* **1997**, *272*, 144.

4. For reviews on structure based drug design targeting 7TM receptors see (a) Klabunde, T.; Hessler, G. *ChemBioChem* **2002**, *3*, 928; (b) Bikker, J. A.; Trumpp-Kallmeyer, S.; Humblet, C. *J. Med. Chem.* **1998**, *41*, 2911; (c) Flower, D. R. *Biochim. Biophys. Acta* **1999**, *1422*, 207.

5. Identification of residue positions of importance for ligand binding and recognition, e.g. (a) Kristiansen, K. *Pharmacol. Ther.* **2004**, *103*, 21; (b) tGRAP Mutant Database (<http://tinygrap.uit.no/>) containing 10,500 mutants from 1380 papers up to April 2001.

6. (a) Needleman, S. B.; Wunsch, C. D. *J. Mol. Biol.* **1970**, *48*, 443; (b) Dayhoff, M. O.; Schwartz, R. M.; Orcutt, B. C. In *Atlas of Protein Sequence and Structure*; Dayhoff, M. O. Ed.; National Biochemical Research Foundation: Washington, DC, 1978; Vol. 5, pp. 345–352; (c) Smith, T. F.; Waterman, M. S. *J. Mol. Biol.* **1981**, *147*, 195; (d) Pearson, W. R.; Lipman, D. J. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 2444; (e) Henikoff, S.; Henikoff, J. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 10915; (f) Henikoff, S.; Henikoff, J. G. *Proteins* **1993**, *17*, 49; (g) Altschul, S. F. *J. Mol. Biol.* **1991**, *219*, 555; (h) Altschul, S. F. *J. Mol. Evol.* **1993**, *36*, 290.

7. (a) Thompson, J. D.; Higgins, D. G.; Gibson, T. *J. Nucleic Acids Res.* **1994**, *22*, 4673; (b) Elling, C. E. Molecular dissection of 7TM receptor structure and function. Ph.D. thesis, University of Copenhagen, Denmark, 2000; (c) Sonnhammer, E. L.; von Heijne, G.; Krog, A. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **1998**, *6*, 175.

8. Schwartz, T. W. *Curr. Opin. Biotechnol.* **1994**, *5*, 434.

9. Representative papers: (a) Molnar, S. P.; King, J. W. *Int. J. Quant. Chem.* **2001**, *85*, 662; (b) Zaliani, A.; Gancia, E. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 525; (c) Sandberg, M.; Eriksson, L.; Jonsson, J.; Sjöström, M.; Wold, S. *J. Med. Chem.* **1998**, *41*, 2481; (d) Matter, H. *J. Pept. Res.* **1998**, *52*, 305; (e) Norinder, U.; Svensson, P. *J. Comput. Chem.* **1998**, *19*, 51; (f) Collantes, E. R.; Dunn, W. J., III *J. Med. Chem.* **1995**, *38*, 2705, and references given therein.

10. Norinder, U.; Högberg, T. In *Textbook of drug design and discovery*; Krosggaard-Larsen, P., Liljefors, T., Madsen, U., Eds.; Taylor & Francis: London, 2002, p 117.

11. Various similarity scores are described in Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T.. *Numerical Recipes: The Art of Scientific Computing*; Cambridge University Press: Cambridge, 1986.

12. Cosine Coefficient ($CC = BC/(B1*B2)^{1/2}$); $B1$ and $B2$ are the numbers of 1's in $B1$ and $B2$ bit maps, respectively, and BC is the number of 1's in common between $B1$ and $B2$. The resulting Cosine score CC is a value between 0 and 1. If $CC = 1$, $B1$ and $B2$ are identical. If $CC = 0$, $B1$ and $B2$ have no TRUE = 1 occasions in common.

13. (a) Hirai, H.; Tanaka, K.; Yoshie, O.; Ogawa, K.; Kenmotsu, K.; Takamori, Y.; Ichimasa, M.; Sugamura, K.; Nakamura, M.; Takano, S.; Nagata, K. *J. Exp. Med.* **2001**, *193*, 255; (b) Indomethacin effects: Hirai, H.; Tanaka, K.; Takano, S.; Ichimasa, M.; Nakamura, M.; Nagata, K. *J. Immunol.* **2002**, *168*, 981; (c) Review: Mitsumori, S. *Curr. Pharm. Design* **2004**, *10*, 3533.

14. (a) Schuffenhauer, A.; Floersheim, P.; Acklin, P.; Jacoby, E. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 391; (b) Lapinsh, M.; Prusis, P.; Lundstedt, T.; Wikberg, J. E. S. *Mol. Pharmacol.* **2002**, *61*, 1465; (c) Attwood, T. K.; Croning, M. D.; Gaulton, A. *Protein Eng.* **2002**, *15*, 7.
15. van Zwieten, P. A.; Doods, H. N. *Cardiovasc. Drugs Ther.* **1995**, *9*, 159.
16. Cf. Martin, Y. C.; Kofron, J. L.; Traphagen, L. M. *J. Med. Chem.* **2002**, *45*, 4350.
17. Receveur, J.-M.; Bjurling, E.; Ulven, T.; Little, P. B.; Nørregaard, P. K.; Högberg, T. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5075.
18. Excellent review over AT1 antagonists: Wexler, R. R.; Greenlee, W. J.; Irvin, J. D.; Goldberg, M. R.; Prendergast, K.; Smith, R. D.; Timmermans, P. B. M. W. M. I. *J. Med. Chem.* **1996**, *39*, 625.
19. Identification of AT ligand-binding sites, e.g. (a) Yamano, Y.; Ohyama, K.; Chaki, S.; Guo, D. F.; Inagami, T. *Biochem. Biophys. Res. Commun.* **1992**, *187*, 1426; (b) Noda, K.; Saad, Y.; Kinoshita, A.; Boyle, T. P.; Graham, R. M.; Husain, A.; Karnik, S. S. *J. Biol. Chem.* **1995**, *270*, 2284; (c) Schambye, H. T.; Hjorth, S. A.; Weinstock, J.; Schwartz, T. W. *Mol. Pharmacol.* **1995**, *47*, 425; (d) Nirula, V.; Zheng, W.; Sothinathan, R.; Sandberg, K. *Br. J. Pharmacol.* **1996**, *119*, 1505; (e) Vianello, B.; Clauser, E.; Corvol, P.; Monnot, C. *Eur. J. Pharmacol.* **1998**, *347*, 113.
20. Homology models were constructed by use of the comparative modeling program MODELLER [(a) Sali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, *234*, 779; (b) Marti-Renom, M. A.; Stuart, A.; Fiser, A.; Sánchez, R.; Melo, F.; Sali, A. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291.] and the structural framework of rhodopsin solved to 2.8 Å resolution (PDB Protein Database code 1F88). The sequence alignment program ClustalW [Higgins, D. G.; Thompson, J. D.; Gibson, T. J. *Methods Enzymol.* **1996**, *266*, 383] was used to obtain an alignment which was in agreement with generic 7TM receptor fingerprints [Baldwin, J. M. *EMBO J.* **1993**, *12*, 1693]. Docking of ligands was done in accordance with strategies outlined in: Rarey, M.; Kramer, B.; Lengauer, T. *Bioinformatics* **1999**, *15*, 243.
21. Grid interaction using different hydrophobic and polar probes was obtained using the SBF module of Cerius 2, Molecular Simulations Inc.: San Diego, CA.
22. Catalyst version 4.6; Molecular Simulations Inc.: San Diego, CA, 1997.
23. The 3D multi-conformational library was constructed from commercial compound collections including InterBioScreen Maybridge ChemBridge and ChemDiv to contain a maximum of 255 conformers per molecule with an energy threshold not exceeding 10 kcal/mol to the lowest energy conformation using CATALYST's conformational search module.
24. Radioligand-binding assay was conducted with stably transfected CHO cells expressing human CRTH2 receptor by competition binding using [³H]-PGD2. Nonspecific binding was determined as the binding in the presence of 10 M PGD2. Data were analyzed and IC₅₀ values determined by non-linear regression. AT receptor binding was done according to Perlman, S.; Costa-Neto C. M.; Miyakawa, A. A.; Schambye, H.; Hjorth, S. A.; Paiva, A. C.; Rivero, R. A.; Greenlee, W. J.; Schwartz, T. W.; *Mol. Pharmacol.* **1997**, *51*, 301.
25. SYBYL[®] 7.0, Tripos Inc., 1699 South Hanley Rd., St. Louis, Missouri 63144, USA.