Protein Engineering vol.15 no.1 pp.7–12, 2001

Deriving structural and functional insights from a ligand-based hierarchical classification of G protein-coupled receptors

T.K.Attwood[1,2,3], M.D.R.Croning[1,2] and A.Gaulton[1]

[1]School of Biological Sciences, University of Manchester, 2.19 Stopford Building, Oxford Road, Manchester M13 9PT and [2]EMBL Outstation, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK

[3]To whom correspondence should be addressed in Manchester.
E-mail: attwood@bioinf.man.ac.uk

G protein-coupled receptors (GPCRs) constitute the largest known family of cell-surface receptors. With hundreds of members populating the rhodopsin-like GPCR superfamily and many more awaiting discovery in the human genome, they are of interest to the pharmaceutical industry because of the opportunities they afford for yielding potentially lucrative drug targets. Typical sequence analysis strategies for identifying novel GPCRs tend to involve similarity searches using standard primary database search tools. This will reveal the most similar sequence, generally without offering any insight into its family or superfamily relationships. Conversely, searches of most 'pattern' or family databases are likely to identify the superfamily, but not the closest matching subtype. Here we describe a diagnostic resource that allows identification of GPCRs in a hierarchical fashion, based principally upon their ligand preference. This resource forms part of the PRINTS database, which now houses ~250 GPCR-specific fingerprints (http://www.bioinf.man.ac.uk/dbbrowser/gpcrPRINTS/). This collection of fingerprints is able to provide more sensitive diagnostic opportunities than have been realized by related approaches and is currently the only diagnostic tool for assigning GPCR subtypes. Mapping such fingerprints on to three-dimensional GPCR models offers powerful insights into the structural and functional determinants of subtype specificity.

Keywords: drug targets/fingerprint database/motif analysis/multi-gene families/receptor subtypes

Introduction

G protein-coupled receptors (GPCRs) constitute a vast cell-surface receptor family that includes hormone, neurotransmitter and light receptors, all of which transduce extracellular signals through interaction with guanine nucleotide-binding (G) proteins (e.g. Stadel *et al.*, 1997; Bockaert and Pin, 1999; Marchese *et al.*, 1999). Several different types of GPCR are now known, which together constitute a 'clan' [by analogy with the complex family of peptidases (Rawlings and Barrett, 1993), we use the term clan to refer to a group of families for which there are indications of an evolutionary relationship, but between which there is no statistically significant sequence similarity]. Currently known clan members include the rhodopsin-like superfamily, the secretin-like receptor superfamily, the metabotropic glutamate receptor-like superfamily, the fungal pheromone mating factor families (STE2 and STE3), the cAMP-type receptors and the frizzled family of receptors (Bockaert and Pin, 1999). In the rhodopsin-like superfamily, ~60 families are populated by hundreds of receptor subtypes. These may invoke diverse cellular responses because individual receptor subtypes can couple to a variety of different effector systems and the same ligand may activate different family members (e.g. 13 receptor subtypes are known to be activated by serotonin).

GPCRs provide the targets for the majority of prescription drugs, whether β-blockers for high blood pressure, β-adrenergic agonists for asthma, anti-histamine (H_1 antagonist) for allergy, etc. Yet many therapies involving such drugs have some efficacy problems and limiting side effects, because the compounds do not differentiate between receptor subtypes. There is therefore considerable pharmaceutical interest in attaining therapeutic specificity by identifying the single receptor subtype that affects a particular physiology or pathophysiology and thereby defining an appropriate intervention point. Ultimately, the aim is to design drugs that eliminate or reduce unwanted effects, while still conferring the desired therapeutic benefit. For example, muscarinic agonists, especially those that activate the M_1 receptor subtype, are potentially useful in treating Alzheimer's disease because the cardiovascular and gastrointestinal side effects associated with non-specific muscarinic agents may be avoidable: the M_1 receptor, which is found in the brain, may be involved with cognition, while other subtypes regulate heart and gastrointestinal functions.

Another primary focus of many companies is the identification of novel GPCRs, and subsequent characterization of their cognate ligands and determination of their involvement in human physiology. With hundreds of receptors known and yet more to be discovered, there are clearly many opportunities ahead for pinpointing new drug targets (Herz *et al.*, 1997).

Typical computational strategies for identifying novel GPCR sequences tend to involve similarity searches using primary database search tools [e.g. BLAST (Altschul *et al.*, 1990)], sometimes coupled with searches of pattern databases [e.g. PROSITE (Hofmann *et al.*, 1999), BLOCKS (Henikoff *et al.*, 2000) and Pfam (Bateman *et al.*, 2000)]. However, while resources such as PROSITE provide patterns for some of the GPCR superfamilies (rhodopsin-like, secretin-like and metabotropic receptors), only one signature is offered at the family level, characterizing the opsins. Clearly, within large multi-gene families, a superfamily level diagnosis is of limited value should one's interest be, for example, in the aetiology of obesity and diabetes and one specifically wishes to identify melanocortin 4 receptors (e.g. Yeo *et al.*, 1998).

Given their pharmaceutical relevance and the importance of being able to identify particular GPCR subtypes, part of the PRINTS fingerprint database (Attwood *et al.*, 2000) has been devoted to the development of a diagnostic resource for GPCRs (Attwood, 2001). To date, more than 250 fingerprints have been created that distinguish GPCRs at the levels of superfamily, family and specific receptor subtype. For a given query, it is therefore possible to determine to which GPCR superfamily

© Oxford University Press

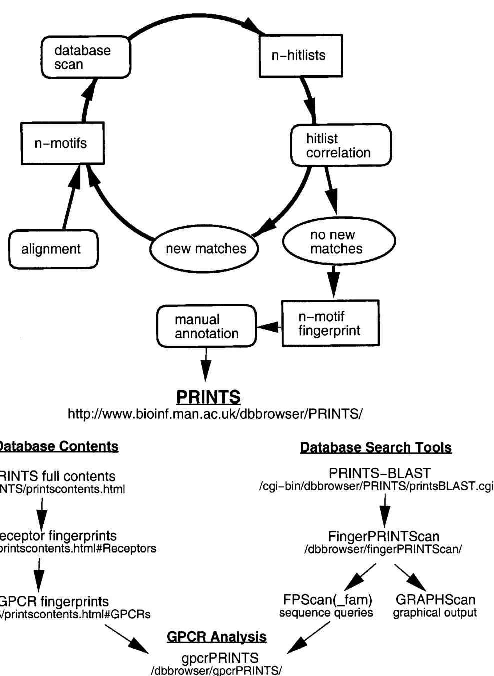

Fig. 1. Illustration of the iterative process by which fingerprints are generated from seed sequence alignments, before being annotated and deposited in the PRINTS database and made available via the PRINTS Web server. The Web resource includes tools both to access the full database contents or specific subsets of it and to search the database interactively. The combination of these tools has allowed us to build a GPCR-specific fingerprint facility, termed gpcrPRINTS.

Scan of sequence: O73667
MELANOCORTIN 4-RECEPTOR

Highest scoring fingerprints for O73667

Fingerprint	E-value	GRAPHScan
MCRFAMILY	6.659329e-52	Graphic
MELNOCORTN4R	6.444801e-46	Graphic
GPCRRHODOPSN	9.841077e-31	Graphic
MELNOCORTINR	3.499750e-29	Graphic

Ten top scoring fingerprints for O73667

Fingerprint	No.of Motifs	Pvalue	Evalue		GRAPHScan
MCRFAMILY	7 of 7	2.6e-57	6.7e-52	IIIIIII	Graphic
MELNOCORTN4R	5 of 5	2.5e-51	6.4e-46	IIIII	Graphic
GPCRRHODOPSN	6 of 7	3.8e-36	9.8e-31	iII.iIi	Graphic
MELNOCORTINR	6 of 6	1.4e-34	3.5e-29	IIIIII	Graphic
MELNOCORTN5R	2 of 5	3.2e-11	8.3e-06	...II	Graphic
CANNABINOIDR	3 of 9	1.1e-10	2.9e-05	.I...i..I	Graphic
MELNOCYTESHR	4 of 7	1.8e-09	0.00046	.i.I.II	Graphic
MELNOCORTN3R	2 of 5	1.4e-07	0.036	.I..I	Graphic
EDTRNSPORT	2 of 9	2.9e-06	0.76	.i.i.....	Graphic
TMPROTEINSRA	2 of 7	3.5e-06	0.9	...i.i.	Graphic

Fig. 2. Search output returned by FingerPRINTScan. For a given query sequence, the program makes an 'intelligent' best guess, based on the occurrence of the highest-scoring full or partial fingerprint matches; deeper levels of matches, pushing further into the Twilight Zone, are presented in additional HTML tables. In this example, the program's best guess confirms the query sequence O73667 (title panel) as being a rhodopsin-like GPCR and, additionally, diagnoses it as a member of the melanocortin receptor superfamily belonging to the melanocortin family, being specifically a type 4 receptor (middle panel). In the next level of output, the top 10 best-scoring matches are given (bottom panel). This table shows the number of motifs matched, the computed *P*- and *E*-values, and a thumb-nail sketch, which gives an instant visual diagnosis of the match (I represents a strong match to a motif, i a weak match and . indicates no match); hyperlinks to the graphical output option allow such sketches to be visualized in more detail (see Figure 3).

the sequence belongs (e.g. whether rhodopsin-like, secretin-like, etc.), of which family it is a member (e.g. whether muscarinic, adrenergic, etc.) and which subtype its sequence signature most resembles (e.g. whether M_1, M_2, M_3, etc.). We describe here a number of applications that illustrate the power of this hierarchical approach to receptor classification: http://www.bioinf.man.ac.uk/dbbrowser/PRINTS/printscontents.html#Receptors.

Materials and methods

Protein fingerprints

Fingerprints are groups of conserved ungapped motifs that are excised from multiple sequence alignments and used to derive potent signatures of family membership through iterative database scanning (Parry-Smith and Attwood, 1992; Attwood and Findlay, 1993, 1994). The source database for the fingerprinting process is a SWISS-PROT/TrEMBL (Bairoch and Apweiler, 2000) composite, minus fragments, which we term SPTr. The procedure commences with manual sequence alignment and excision of conserved motifs; these are used to trawl SPTr independently. For each motif, the scanning algorithm calculates a frequency matrix; in other words, no mutation or other similarity matrices are used to weight the searches. The scoring process uses a sliding-window approach, whereby each motif in a fingerprint is scanned across each database sequence in turn. For each position of the window (which, by definition, is the width of the motif), the algorithm simply sums the residue scores with reference to the motif frequency matrix. The best match is achieved when a position is found in the sequence where most of the residues within the sliding window match high-scoring terms in the frequency matrix. For each motif, results are stored in a hitlist that is rank-ordered by score; match probabilities are not calculated. Diagnostic performance is enhanced by iterative database scanning. The motifs therefore grow and mature with each database pass, as more sequences are matched and assimilated into the process. The procedure terminates when no more sequences that match all the motifs can be identified between successive database scans, i.e. when the scans have reached convergence. At this point, fingerprints are formatted and annotated prior to deposition in the PRINTS database.

Note that, during the iterative process, the population of the database determines how the motifs, and therefore fingerprints, evolve. As the scoring method is not biased by substitution matrices, pseudo-counts or more sophisticated schemes, it performs cleanly, with very little noise; the drawback is that its absolute scoring potential is low, depending on the depth of the motifs (which reflects the size of the family within the database). However, the approach derives potency principally from the use of multiple motifs, which can compensate for low-scoring elements; and the requirement to match all motifs in the correct order, with appropriate distances between them, reduces the chance of making random matches. Experience in building PRINTS (which currently contains 1550 fingerprints) demonstrates that the approach works well and provides a valuable complement to probabilistic techniques.

Deriving GPCR fingerprints

Sequence alignments were constructed manually, using the CINEMA colour alignment editor (Parry-Smith *et al.*, 1998). Alignments were created for each of the different clan members and for their constituent families and receptor subtypes. A number of different processes were used to determine which family members to use to seed the fingerprint process, but typically involved either simple text or BLAST searches of

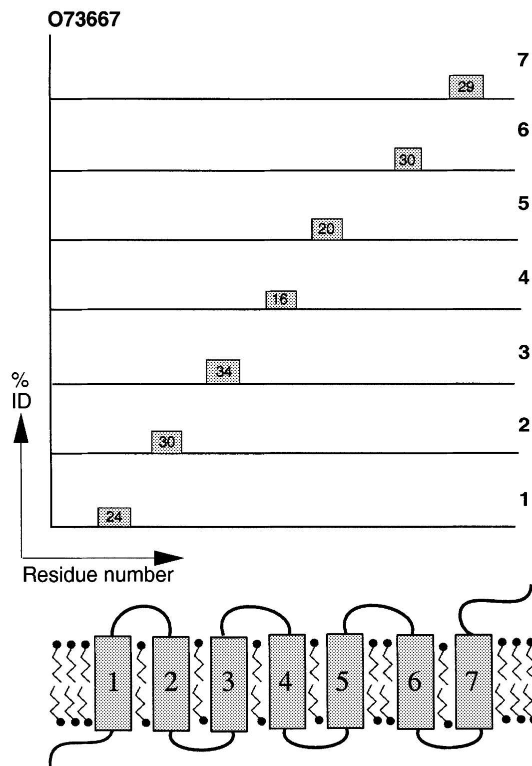

Fig. 3. Graphical output returned by FingerPRINTScan. Within each profile, the horizontal axis represents the sequence and the vertical axis the percentage score (identity) of each fingerprint element (0–100 per motif). Shaded blocks, whose widths denote motif widths, mark the positions of matches above a 15% threshold. The profile depicts the rhodopsin-like GPCR fingerprint of type 4 melanocortin receptor O73667. Blocks appearing in a systematic order along the length of the sequence and above the level of noise indicate matches with the constituent motifs. The sequence is clearly a true family member, matching all seven TM domains of the rhodopsin-like GPCR architecture (for the purpose of illustration, we have shown the TM domains schematically beneath the plot).

SWISS-PROT to identify suitable candidates. Seed alignments did not need to be exhaustive (since the iterative process attracts further sequences), but included a representative selection of family members, including outliers, in order to highlight both conserved and gapped areas effectively. If motifs failed to perform well during the iterative process, the alignment was revisited to determine the cause and the motifs were re-seeded.

Individual alignments were compared visually to determine both the regions of similarity and, importantly, the regions of difference between them. Motifs were excised from these discriminatory regions and used to create a range of diagnostic fingerprints using the iterative technique outlined above. The process of alignment visualization and comparison was carried out by expert human inspection rather than algorithmically, as current pattern-recognition algorithms work on the principle of identifying areas of similarity shared between groups of sequences; for the purposes of this study, however, it was the regions of difference between family members that we particularly wished to identify.

Once fingerprints had been iteratively refined and annotated, they were deposited in PRINTS and made available through its quarterly releases. As an integral part of PRINTS, fingerprints may be searched with user-specified sequences, using the tools outlined below. The entire process of fingerprint generation and database searching is charted in Figure 1.

Searching PRINTS

Fingerprint diagnoses are made using the FingerPRINTScan suite (Scordis *et al.*, 1999). By contrast with the highly selective technique used to derive fingerprints, the algorithm employed by FingerPRINTScan to search PRINTS uses a sensitive ungapped profile approach and rank-orders hits according to combined motif expectation (*E*)-values. The suite provides facilities for individual and bulk sequence searches against PRINTS and for single sequence searches against individual fingerprints: http://www.bioinf.man.ac.uk/dbbrowser/fingerPRINTScan/. Two options are provided for individual sequence searches: FPScan and FPScan_fam. The latter details results in the context of the full PRINTS family hierarchy, so that familial and ancestral relationships between matched fingerprints can be understood more readily.

In an attempt to cater for both novice and expert users, results of individual searches against the database (whether using FPScan or FPScan_fam) are returned on different levels: first, an 'intelligent' best guess is provided, based on the occurrence of the highest-scoring fingerprint match above an *E*-value threshold (the default value, 0.0001, can be changed by the user); more detailed results are then provided in different layers via extended HTML tables, one of which gives the top 10 best-scoring matches, which necessarily include the 'best guess' match or matches, at the top of the table (Figure 2). The result of searching a single sequence with a named fingerprint takes the form of a graphical cartoon of the fingerprint profile, offering an instant diagnosis of the query (Figure 3).

PRINTS BLAST

Sequences matched in the current release of PRINTS are used to create a FASTA-format database and SRS indexing (Etzold *et al.*, 1996) is used to extract the relevant fingerprint- and sequence-specific information. An implementation of BLAST allows searches with either protein or DNA queries (Wright *et al.*, 1999) and results are again returned in formatted HTML tables: http://www.bioinf.man.ac.uk/dbbrowser/PRINTS/printsBLAST.cgi. For ease of interpretation, retrieved matches are linked directly to the sequence and fingerprint databases and to the graphical component of the FingerPRINTScan suite (Figure 4).

Results and discussion

Using the combination of manual alignment, visual inspection and iterative database searching described above, ~250 GPCR-specific fingerprints have been created, annotated and stored in the PRINTS database (http://www.bioinf.man.ac.uk/dbbrowser/ PRINTS/printscontents.html #Receptors). A typical result of querying PRINTS using the FingerPRINTScan suite is shown in Figure 2. For the TrEMBL query sequence O73667 (default parameters and *E*-value 10^{-6}), the output reveals a four-tiered diagnosis, indicating the sequence to be (i) a member of the rhodopsin-like GPCR superfamily, (ii) a member of the melanocortin receptor superfamily (which includes melanocortin-, adrenocorticotrophin- and melanocyte-stimulating hormone receptors), (iii) a member of the melanocortin family and specifically (iv) a type 4 melanocortin receptor (MC4R). For this same example, PROSITE and Pfam suggest only that the query is a rhodopsin-like GPCR.

From the output shown in Figure 2, the actual PRINTS hierarchy can only be inferred, based on the *E*-values of the four 'best guess' matches because, for convenience, FPScan

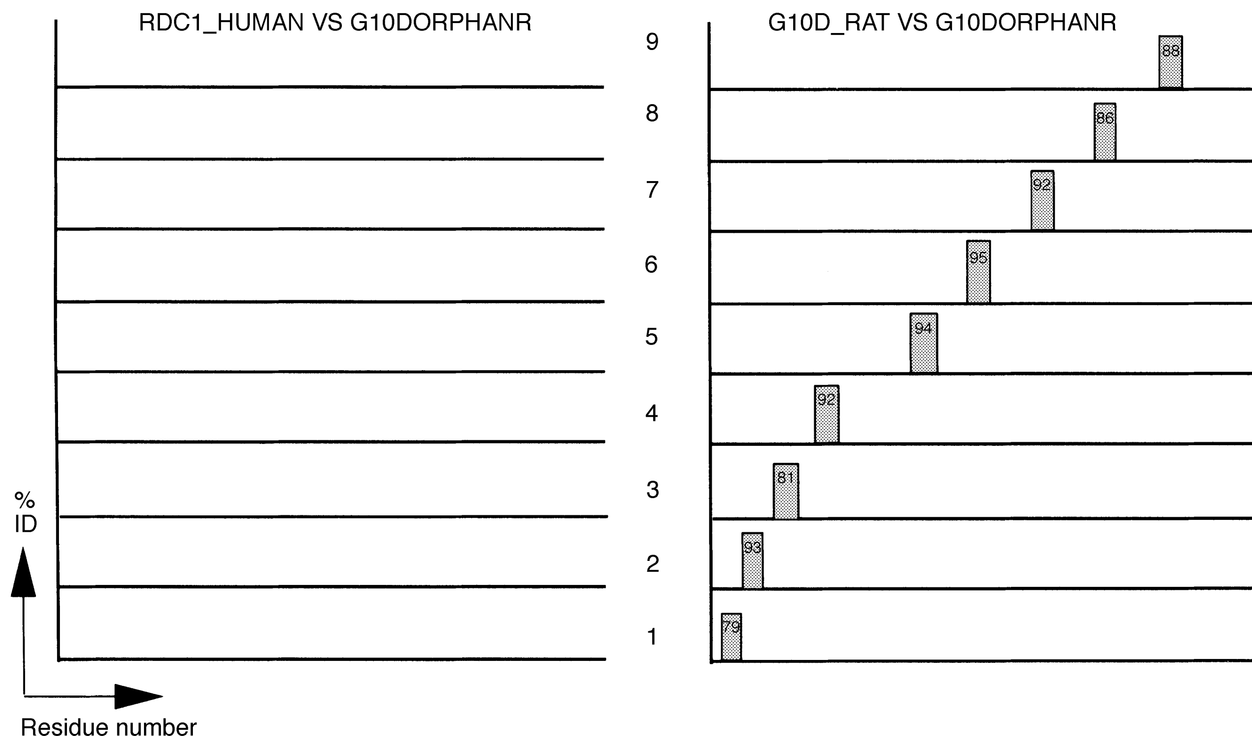

Fig. 4. Comparative fingerprint profiles invoked from PRINTS BLAST, following a search with query sequence RDC1_HUMAN (P25106). The top non-RDC1 match in the hitlist is G10D_RAT (Q64166), which matches all nine motifs of its own family fingerprint, as illustrated in the right-hand profile; by contrast, the query sequence, in the left-hand profile, matches none of the motifs. This result illustrates the difference between the diagnostic potential of BLAST, which identifies the shared 7TM receptor scaffold, and fingerprints, which are able to make more precise, family diagnoses (see Figure 4).

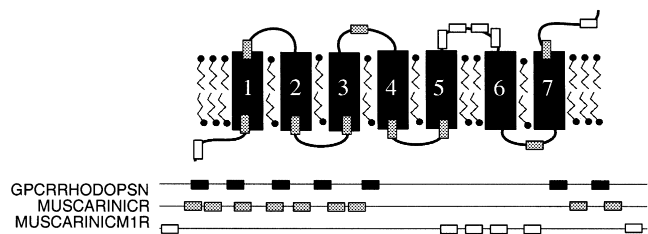

Fig. 5. GPCR fingerprints mapped schematically on to the 7TM architecture (the shaded bars denote the relative locations of the constituent motifs of each fingerprint). The depicted fingerprints are those for (i) the rhodopsin-like superfamily (black), (ii) the muscarinic family (grey) and (iii) the muscarinic M_1 receptor subtype (white). The different regions that characterize the receptors at each level are clearly evident and provide an insight into the different perspectives of BLAST and family-specific fingerprint diagnoses: whereas the former 'sees' the shared 7TM scaffold, the latter sees particular parts of TM and loop regions, which are likely to encode many of their functional determinants. Here, for example, motifs 3, 4, 5, 8 and 9 include residues known to be involved in either agonist or antagonist binding (Kuipers *et al.*, 1997; Beukers *et al.*, 1999).

was used to generate these results. The complete, extended hierarchy can be visualized by means of FPScan_fam (however, this output is not shown because the resulting table is very large).

We can gain a better appreciation of what such diagnoses mean by plotting graphical profiles of query sequences against their matched fingerprints (the 'graphic' or 'GRAPHScan' options in Figure 2). Figure 3 illustrates the profile produced when the rhodopsin-like GPCR fingerprint, which encodes the seven transmembrane (TM) domains of the GPCR scaffold, is scanned against the type 4 melanocortin receptor. Where a motif matches above a given threshold, a shaded block is used to mark its location. In the example shown, seven blocks are clearly observed, from N- to C-terminus, indicating matches with each of the seven TM domains.

The MC4R example, although perhaps highlighting the limitations of the more generic pattern recognition approaches, is not particularly outstanding when we know that BLAST can easily diagnose the query as a type 4 melanocortin receptor (albeit without shedding light on its family and superfamily relationships). However, BLAST does not always provide clear answers. Consider, for example, the human RDC1 orphan receptor. BLAST reveals the top non-RDC1 match to be a G10D receptor, with a confident *P*-value (6×10^{-53}). However, when we scan the RDC1 sequence against the G10D receptor fingerprint, we find no match at all, as shown in Figure 4.

At first sight, this may seem curious. Yet it reveals something extremely important about the way in which BLAST and fingerprints 'see' sequence similarity: BLAST can identify generic similarities between sequences (based on high-scoring sequence pairs), but cannot reveal differences between them. Here, the greatest generic similarity to RDC1 is seen in the

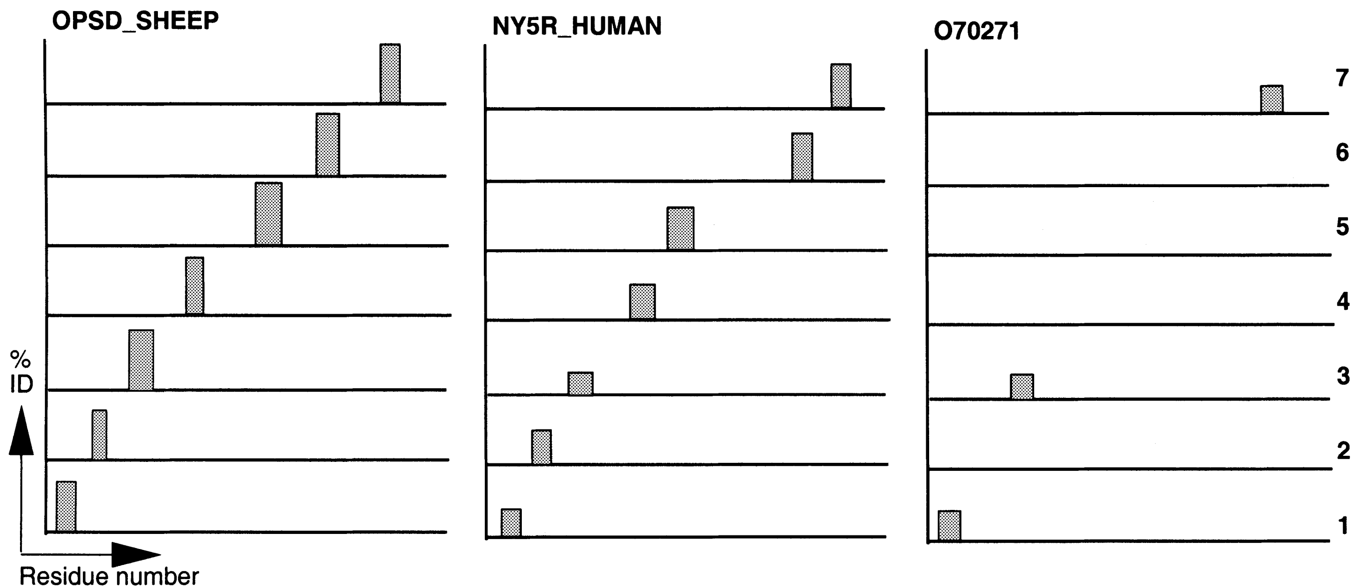

Fig. 6. Comparative fingerprint profiles depicting rhodopsin-like GPCR fingerprints of OPSD_SHEEP (P02700), NY5R_HUMAN (Q15761) and O70271. OPSD_SHEEP and NY5R_HUMAN are known true-positive family members, matching all seven TM domains, but differences in TM domain 3 mean that NY5R_HUMAN is not diagnosed by PROSITE; and O70271 is an outlying partial match that demonstrates a poor, but nevertheless tantalizing, set of matches with motifs 1, 3 and 7.

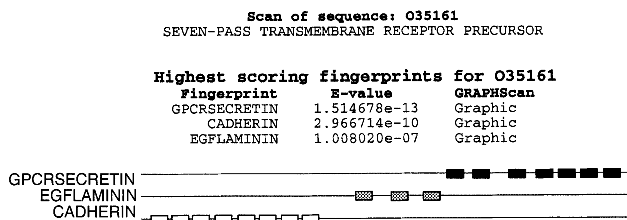

Fig. 7. Diagnosis of a mosaic GPCR. The example illustrates the murine 7TM O35161, which contains multiple N-terminal cadherin repeats (white), a central EGF-containing domain (grey), followed by a C-terminal 7TM secretin-like signature (black).

TM signature of the G10D receptor. However, the sequences are different in their loop and N- and C-terminal regions, which is where we might expect to discover many of their functional determinants: it is these features that family-specific fingerprints encode; and clearly, these tell-tale traits are not shared by RDC1 and G10D. This result has important ramifications for off-the-shelf automatic genome analysis packages, highlighting the danger of reliance on top-scoring BLAST hits to provide functional diagnoses.

We can perhaps better understand these different perspectives by mapping superfamily, family and receptor subtype fingerprints on to the 7TM architecture. Figure 5 compares fingerprints for the rhodopsin-like superfamily, for the muscarinic family and for its M_1 receptor subtype. The different regions that characterize the receptors at each level are clearly evident: the superfamily fingerprint focuses on the shared 7TM scaffold; the family fingerprint encodes specific parts of TM and loop regions; and the subtype fingerprint is drawn from the third cytoplasmic loop and the N- and C-terminal domains. This is consistent with our expectations that the highly conserved TM segments are likely to constitute the ligand-binding domain, whereas the large intracellular region, unique to each subtype, is likely to constitute part of the receptor–effector coupling domain (Peralta *et al.*, 1987).

Fingerprints gain diagnostic potency by virtue of the mutual context provided by motif neighbours. A sequence matching all the motifs in a fingerprint can therefore be confidently diagnosed as a true match. Perhaps more importantly, a sequence matching only some of the motifs can still be diagnosed as a distant relative, provided that the motifs are matched in the correct order, with appropriate distances between them. This offers a significant advantage over single-motif pattern-matching methods, which can only report that a match has been made and provide no biological context within which to assess its significance, leaving the user to verify this manually.

Regular expression methods, such as those embodied in PROSITE, suffer the further limitation that patterns do not tolerate similarity: a sequence either matches or not, because the patterns are encoded explicitly. Thus, for example, a query that shows only a single residue difference from a pattern will be treated as a mis-match. This problem is addressed in PROSITE by annotating such sequences as false negatives where it is known that matches have been missed. The difficulty arises with hypothetical sequences, where it is not realized that the pattern has missed them. Consider, for example, sequences OPSD_SHEEP (P02700), NY5R_HUMAN (Q15761) and O70271, whose fingerprint profiles are illustrated in Figure 6. OPSD_SHEEP is a known true member of the rhodopsin-like superfamily, matching all seven TM domains; NY5R_HUMAN is again a clear family member, but is not diagnosed by PROSITE because it contains changes in the third TM domain, which alone provides the basis for the PROSITE pattern; and O70271 makes a partial fingerprint match, lacking significant matches with TM domains 2, 4, 5 and 6 (this sequence fails to match the PROSITE pattern, is not annotated as a false negative, but falsely matches the class-II aminoacyl-transfer RNA synthetase pattern). For Twilight matches such as that for O70271, sequence analysis techniques cannot provide unequivocal functional diagnoses: such tentative matches must always be followed up by appropriate laboratory experiments. Nevertheless, for this sequence, FingerPRINTScan indicates strong similarity to the olfactory

receptors, thereby revealing a relationship that is missed by PROSITE and Pfam.

A final example of the utility of the GPCR resource is in the detection and depiction of GPCR mosaics. There are now many known examples of such molecules, including members of the rhodopsin-like superfamily that contain N-terminal leucine-rich repeats (e.g. gonadotrophin receptors such as bovine thyrotropin receptor, Q27987), N-terminal leucine-rich and LDL-receptor repeats (e.g. the GRL101 GPCR from *Lymnaea stagnalis*, P46023) and the more recently characterized secretin-like receptors that contain cadherin and laminin EGF-like repeats (e.g. the human myeloblast KIAA0279 protein, Q92566 and rat MEGF2 protein, O88278) (Stacey *et al.*, 2000). By virtue of the GPCR resource being embedded within the PRINTS database, the detection of additional domains is possible with a single database query. The example shown in Figure 7 illustrates the result of searching PRINTS with the murine 7TM O35161, which contains N-terminal cadherin repeats, a central EGF-containing domain, followed by the C-terminal 7TM secretin-like signature.

Conclusion

The diagnostic resource described here has two main strengths. First, the use of multiple motifs to build characteristic signatures offers a biological context within which to assess the significance of a given match. Thus, a distantly related sequence that lacks matches with some components of a fingerprint may still be identified, by virtue of the diagnostic framework provided by neighbouring motifs; such a framework is not afforded by single-motif approaches. Second, by exploiting differences as well as similarities between related sequences, we have been able to create a range of potent GPCR fingerprints, encoding individual subtypes through to family and superfamily levels; no other diagnostic resource currently available offers such a powerful hierarchical discriminatory system for this fundamentally important class of cell-surface receptors. Moreover, by focusing on conserved loop and N- and C-terminal traits, such fingerprints offer the potential to make highly specific functional diagnoses. Fingerprint selectivity thus offers new opportunities to explore in more detail correlations between specific motifs and ligand binding and G protein coupling and consequently may provide insights in the ongoing quest to characterize orphan receptors. The resource is therefore valuable in cases where primary and other secondary database searches either produce ambiguous results or fail completely to return a match. Used wisely, as part of an integrated analysis strategy, GPCR fingerprints provide sensitive diagnostic opportunities that have not been realized by other computational approaches.

Acknowledgements

PRINTS is built and maintained at the University of Manchester with support from the Royal Society (T.K.A. is a Royal Society University Research Fellow). We are grateful for an EU grant to support M.D.R.C. and to the MRC for a studentship (A.G.).

References

Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) *J. Mol. Biol.*, **215**, 403–410.
Attwood,T.K. (2001) *Trends Pharmacol. Sci.*, **22**, 162–165.
Attwood,T.K. and Findlay,J.B.C. (1993) *Protein Eng.*, **6**, 167–176.
Attwood,T.K. and Findlay,J.B.C. (1994) *Protein Eng.*, **7**, 195–203.
Attwood,T.K., Croning,M.D.R., Flower,D.R., Lewis,A.P., Mabey,J.E., Scordis,P., Selley,J.N. and Wright,W. (2000) *Nucleic Acids Res.*, **28**, 225–227.
Bairoch,A. and Apweiler,R. (2000) *Nucleic Acids Res.*, **28**, 45–48.
Bateman,A., Birney,E., Durbin,R, Eddy,S.R., Howe,K.L. and Sonnhammer,E.L.L. (2000) *Nucleic Acids Res.*, **28**, 263–266.
Beukers, M.W., Kristiansen, I., Ijzerman, A.P. and Edvardsen, I. (1999) *Trends Pharmacol. Sci.*, **20**, 475–477.
Bockaert,J. and Pin,J.P. (1999) *EMBO J.*, **18**, 1723–1729.
Etzold,T., Ulyanov,A. and Argos,P. (1996) *Methods Enzymol.*, **266**, 114–128.
Henikoff,J.G., Greene,E.A., Pietrokovski,S. and Henikoff,S. (2000) *Nucleic Acids Res.*, **28**, 228–230.
Herz,J.M., Thomsen,W.J. and Yarbrough,G.G. (1997) *J. Recept. Signal Transduct. Res.*, **17**, 671–776.
Hofmann,K., Bucher,P., Falquet,L. and Bairoch,A. (1999) *Nucleic Acids Res.*, **27**, 215–219.
Kuipers, W., Oliveira, L., Vriend, G. and Ijzerman, A.P. (1997) *Receptors Channels*, **5**, 159–174.
Marchese,A., George,S.R., Kolakowski,L.F., Lynch,K.R. and O'Dowd,B.F. (1999) *Trends Pharmacol. Sci.*, **20**, 370–375.
Parry-Smith,D.J. and Attwood,T.K. (1992) *CABIOS*, **8**, 451–459.
Parry-Smith,D.J., Payne,A.W.R, Michie,A.D. and Attwood,T.K. (1998) *Gene*, **211**, GC45–GC56.
Peralta,E.G., Ashkenazi,A., Winslow,J.W., Smith,D.H., Ramachandran,J. and Capon,D.J. (1987) *EMBO J.*, **6**, 923–929.
Rawlings,N.D. and Barrett,A.J. (1993) *Biochem. J.*, **290**, 205–218.
Scordis,P., Flower,D.R. and Attwood,T.K. (1999) *Bioinformatics*, **15**, 799–806.
Stacey,M., Lin,H-H., Gordon,S. and McKnight,A.J. (2000) *Trends Biochem. Sci.*, **25**, 284–289.
Stadel,J.M., Wilson,S. and Bergsma,D.J., (1997) *Trends Pharmacol. Sci.*, **18**, 430–437.
Wright,W., Scordis,P. and Attwood,T.K. (1999) *Bioinformatics*, **15**, 523–524.
Yeo, M., Farooqi, I.S., Aminian, S., Halsall, D.J., Stanhope, R.G. and O'Rahilly, S. (1998) *Nature Genet.*, **20**, 111–112.

Received April 17, 2001; revised September 3, 2001; accepted September 27, 2001