they are antiparallel. Whereas parallel coiled coils are typically oligomers of one helix, antiparallel coiled coils are generally composed of helices that are consecutive in the polypeptide chain and fold back onto each other. Because of this lack of sequence symmetry, antiparallel coiled coils have more diverse surfaces, can adapt more precisely to the interaction with asymmetric ligands and can more readily move around a hinge (in a homo-oligomer, any such motion would involve a symmetry-break). Not surprisingly, the coiled coils that are involved in the manipulation of polynucleotides are, therefore, mostly antiparallel (such as the 'thumb' in the Klenow fragment of DNA polymerase or the 'arms' of seryl- and valyl-tRNA synthases). With the recent crystal structures and biochemical studies of prefoldin and AAA+ proteins, coiled coils can now be seen to adopt similar functions in the manipulation of polypeptides.

Acknowledgements

We thank Lars-Oliver Essen for providing the thermosome coordinates and Ulrich Kahlow for assistance with the figure.

References

1 Lupas, A. (1996) Coiled coils: new structures and new functions. *Trends Biochem. Sci.* 21, 375–382
2 Hartl, F.U. and Hayer-Hartl, M. (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* 295, 1852–1858
3 Geissler, S. *et al*. (1998) A novel protein complex promoting formation of functional α- and γ-tubulin. *EMBO J.* 17, 952–966
4 Vainberg, I.E. *et al*. (1998) Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. *Cell* 93, 863–873
5 Siegers, K. *et al*. (1999) Compartmentation of protein folding *in vivo*: sequestration of non-native polypeptide by the chaperonin-GimC system. *EMBO J.* 18, 75–84
6 Leroux, M.R. *et al*. (1999) MtGimC, a novel archaeal chaperone related to the eukaryotic chaperonin cofactor GimC/prefoldin. *EMBO J.* 18, 6730–6743
7 Siegert, R. *et al*. (2000) Structure of the molecular chaperone prefoldin: unique interaction of multiple coiled coil tentacles with unfolded proteins. *Cell* 103, 621–632
8 Lundin, V.F. *et al*. (2004) Molecular clamp mechanism of substrate binding by hydrophobic coiled-coil residues of the archaeal chaperone prefoldin. *Proc. Natl. Acad. Sci. U. S. A.* 101, 4367–4372
9 Simons, C.T. *et al*. (2004) Selective contribution of eukaryotic prefoldin subunits to actin and tubulin binding. *J. Biol. Chem.* 279, 4196–4203
10 Llorca, O. *et al*. (2000) Eukaryotic chaperonin CCT stabilizes actin and tubulin folding intermediates in open quasi-native conformations. *EMBO J.* 19, 5971–5979
11 Llorca, O. *et al*. (2001) The 'sequential allosteric ring' mechanism in the eukaryotic chaperonin-assisted folding of actin and tubulin. *EMBO J.* 20, 4065–4075
12 Leroux, M.R. and Hartl, F.U. (2000) Protein folding: versatility of the cytosolic chaperonin TRiC/CCT. *Curr. Biol.* 10, R260–R264
13 Lee, S. *et al*. (2004) The ClpB/Hsp104 molecular chaperone – a protein disaggregating machine. *J. Struct. Biol.* 146, 99–105
14 Lee, S. *et al*. (2003) The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state. *Cell* 115, 229–240
15 Rohrwild, M. *et al*. (1996) HslV–HslU: a novel ATP-dependent protease complex in *Escherichia coli* related to the eukaryotic proteasome. *Proc. Natl. Acad. Sci. U. S. A.* 93, 5808–5813
16 Song, H.K. *et al*. (2000) Mutational studies on HslU and its docking mode with HslV. *Proc. Natl. Acad. Sci. U. S. A.* 97, 14103–14108
17 Nitsch, M. *et al*. (1998) Group II chaperonin in an open conformation examined by electron tomography. *Nat. Struct. Biol.* 5, 855–857
18 Sousa, M.C. *et al*. (2000) Crystal and solution structures of an HslUV protease–chaperone complex. *Cell* 103, 633–643

0968-0004/$ - see front matter
doi:10.1016/j.tibs.2004.07.004

Protein Sequence Motif

NCD3G: a novel nine-cysteine domain in family 3 GPCRs

Long Yu[1,4], Songping Liang[2], Xianghua Liu[1], Quanyuan He[2], David J. Studholme[3] and Qi Wu[1]

[1]State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Handan Road 220, Shanghai 200433, P.R. China
[2]College of Life Science, Hunan Normal University, Changsha 410081, P.R. China
[3]The Wellcome Trust Sanger Institute, Wellcome Trust Genome campus, Hinxton, UK CB10 1SA
[4]Chinese National Human Genome Center at Shanghai (HGCS), Bi Bo Road 250, Shanghai 201203, P.R. China

The NCD3G [for nine-cysteine domain of family 3 G-protein-coupled receptors (GPCRs)] domain is a novel protein domain that is conserved in family 3 GPCRs, including metabotropic glutamate receptors, calcium-sensing receptors, pheromone receptors and taste receptors, with the exception of $GABA_B$ receptors. The NCD3G domain contains nine highly conserved cysteine residues. Structural predictions suggest that NCD3G might possess four β strands and three disulfide bridges. The structural model of NCD3G highlights the conserved residues co-segregated with certain familial diseases.

G-protein-coupled receptors (GPCRs) comprise a vast family of proteins that encompasses a wide range of functions. Family 3 of the GPCRs consists of metabotropic glutamate receptors (mGluRs), calcium-sensing receptors

Corresponding authors: Long Yu (longyu@fudan.edu.cn), Songping Liang (liangsp@public.cs.hn.cn).
Available online 3 August 2004

(CaSRs), pheromone receptors, taste receptors, γ-amino-butyric acid B ($GABA_B$) receptors and four orphan receptors [1,2]. All of these receptors possess a large N-terminal extracellular domain to which agonist binding takes place, a seven transmembrane domain (7tm_3; which forms the common core of all GPCRs), and a C-terminal intracellular domain. Here, we report a new domain that contains nine highly conserved cysteine residues and is present in many family 3 GPCRs.

Characterization of the NCD3G domain

We have found a pheromone receptor (amino acid residues 541–600 in BAA26126.1) containing a region of nine cysteine residues that did not match any entry in the Pfam 12.0 [3] and SMART 4.0 [4] databases. Using PSI–BLAST [5] with an inclusion threshold of 0.005, we searched the NCBI non-redundant protein database (http://www.ncbi.nih.gov/blast/) against this sequence. The first iteration retrieved many family 3 GPCRs with significant E-values, including CaSRs (e.g. NP_000379.2 with E-value 1e – 10), pheromone receptors (e.g. NP_064301.1 with E-value 1e – 08), sweet-taste receptors (e.g. AAK01937.1 with E-value 0.01), odorant receptors (e.g. AAQ64679.1 with E-value 3e – 10) and other uncharacterized GPCRs (e.g. NP_038831.1 with E-value 1e – 10). Many mGluRs with significant E-values (e.g. AAM47559.1 with E-value 9e – 12) and other GPCRs were retrieved by the second iteration. The search converged after five iterations. To summarize, 28 representative sequences were aligned using ClustalW [6] and manually edited (Figure 1). Owing to its composition and occurrence, we have named this novel domain NCD3G after nine-cysteine domain of family 3 GPCRs.

From searches against SMART and Pfam databases, it appears that the NCD3G domain is usually located between an extracellular receptor domain (ERD; e.g. ANF_receptor, Pfam: PF01094) and a transmembrane motif (e.g. 7tm_3, Pfam: PF00003) (Figure 2 and supplementary Figure 1). The NCD3G domain appears to be restricted to animals (from *Caenorhabditis briggsae* to *Homo sapiens*) because no significant matches to microbial, fungal or plant genomes were found using TBLAST (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi).

3D structure model of NCD3G

Although the crystal structure of the ERD for mGluRs has been reported [7,8], the 3D structure of NCD3G is still unknown Using 3D-PSSM [9], FUGUE [10], mGenTHREADER [11], PDB-BLAST (http://bioinformatics.burnham-inst.org/pdb_blast/) and SAM-T99 (HMM) [12] through the meta-server @TOME (http://bioserv.cbs.cnrs.fr) [13], we searched for homologs within known 3D structures.

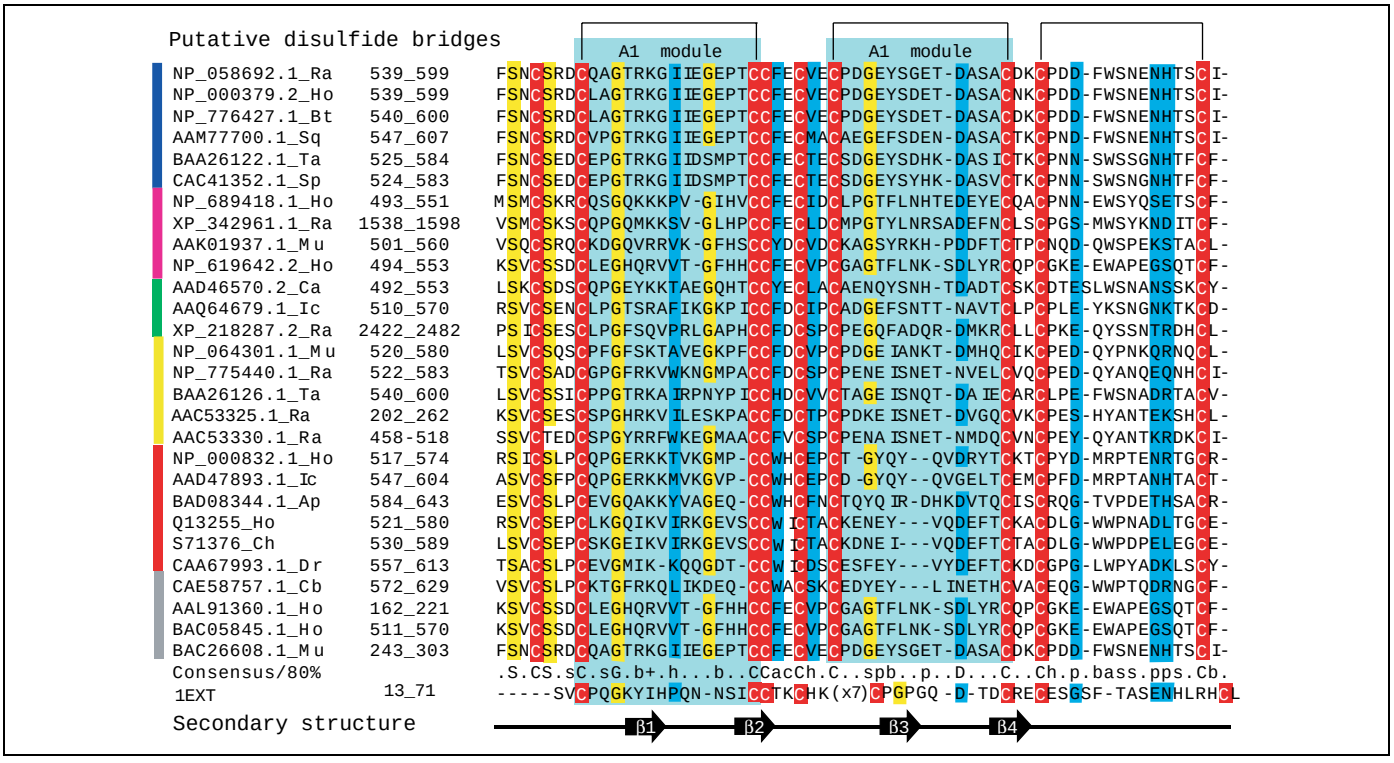

Figure 1. Representative alignment of NCD3G (for nine-cysteine domain of family 3 G-protein-coupled receptors) domain. The alignment was generated by ClustalW [6] plus manual editing from the selected protein sequences of PSI–BLAST [5] search. The 80% consensus sequence of the domain is calculated and colored using Chroma [30]. Capital letters represent amino acids. Lower-case letters: a, aromatic; b, big; h, hydrophobic; l, aliphatic; p, polar; s, small; t, tiny. The disulfide bridges are marked out above the alignment according to the 3D model. The secondary structure below the alignment is derived from the output of the prediction programmer Jpred [31] with black arrows indicating β strands. Structure codes for 1EXT are from the Protein Data Bank (http://www.rcsb.org/pdb/) and other sequences are from NCBI Entrez database (http://www.ncbi.nlm.nih.gov) for the NCD3G. The species abbreviations are: Ap, *Apis mellifera*; Bt, *Bos taurus*; Ca, *Carassius auratus*; Cb, *Caenorhabditis briggsae;* Ch, *Cherry salmon*; Dr, *Drosophila melanogaster*; Go, *Gorilla gorilla*; Ho, *Homo sapiens*; Ic, *Ictalurus punctatus*; Mu, *Mus musculus*; Pa, *Pan troglodytes*; Ra, *Rattus norvegicus*; Sp, *Sparus aurata*; Sq, *Squalus acanthias*; Ta, *Takifugu rubripes*. Different groups classified by function are shown by colored lines on the left of the alignment: blue, calcium-sensing receptors; magenta, sweet-taste receptors; green, odorant receptors; yellow, pheromone receptors; red, metabotropic glutamate receptors; gray, others. The consensus motif deduced for the NCD3G domain can be summarized $xSxCSx_2Cx_2Gx_{1-2}[R/x][K/R]x_{6-8}CCx[E/x]Cx_2Cx_{11-14}Cx_2Cx_2[D/x]x_{1-2}[W/x]x_3[N/X]x_3Cx$; the A1 module can be summarized as $Cx_2GxBxx_{4-9}C$, where 'x' is any residue with a range as indicated. This multiple sequence alignment has been deposited with the European Bioinformatics Institute (ftp://ftp.ebi.ac.uk/pub/databases/embl/align/) with the accession number ALIGN_000674.

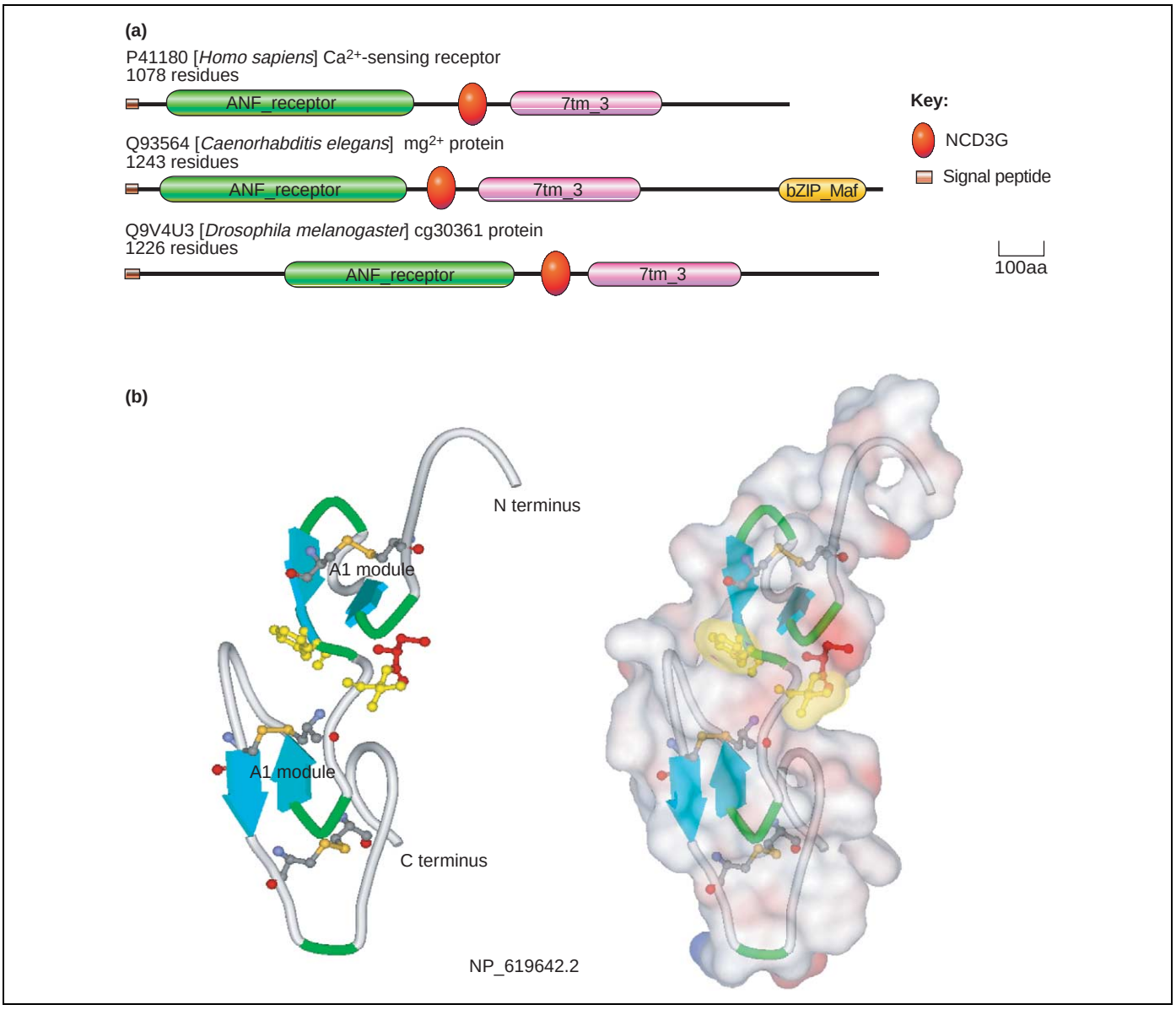

Figure 2. Architecture and 3D model of the NCD3G domain. **(a)** Examples of different motif arrangements in various proteins according to searches against SMART [4] and Pfam [3] databases. Name abbreviations of each domain are as follows: 7tm_3, metabotropic glutamate family 7 transmembrane receptor (Pfam: PF00003); ANF_receptor, receptor family ligand-binding region (Pfam: PF01094); bZIP_Maf, transcription factor (Pfam: PF03131) and signal peptide; NCD3G, nine-cysteine domain of family 3 G-protein-coupled receptors (which has been deposited in the Pfam database with accession number PF07562). **(b)** Representative structure of the modeled NCD3G domain (NP_619642). The three disulfide bridges in NCD3G are shown in dark yellow. The side chains and surface of the two conserved hydrophobic residues (Phe517 and Val 520) between two A1 modules are colored in yellow; the cysteine (Cys519) between these hydrophobic residues is red.

For most queries, the N-terminal region of chain A of 1EXT – an extracellular domain of the 55-kDa tumor necrosis factor receptor (TNFR) – was iterated first by 3D-PSSM and FUGUE 83 with distinctly better scores than others (E-values ranging from 1.89–9.60 and Z-scores of 2.84–3.98, respectively). The results from mGenTHREADER with relatively lower E-values (4.202–6.260) also suggested weak fold compatibility between 1EXT and NCD3G (supplementary Table 1). Based on 1EXT, some crude models were built by molecular structure tools including TITO [14], SCWRL [15] and MODELLER [16]. After energy minimizing by Deep View [17], these models were evaluated by amino acid environment evaluation tools such as Eval23D [18] and Verify3D [19]. Their Eval23D and Verify3D scores range from -0.045 to 0.041 and 0.064 to 0.124, respectively, which are comparable to the scores of chain A of 1EXT (Eval23D, 0.079 and Verify3D, 0.152). These models commonly consist of four β strands that are linked by turns and three intra-disulfide bridges. Further alignment of NCD3G and 1EXT (Figure 1) indicated that NCD3G might contain two A1 modules (Cys-Xaa$_2$-Gly-Xaa-b-Xaa-Xaa$_{4-9}$-Cys, where b represents any amino acid with a large side chain), which repeatedly presents in TNFR-related structures and epidermal growth factor-like domains [20]. And each module contains two antiparallel β strands and one disulfide bridge (Figure 2).

Postulated function

GPCRs mediate the sense of vision, smell, taste and pain in mammals. They are also involved in cell recognition and communication processes and, hence, are prominent drug

targets. Using mutagenesis and chimera approaches, Hu *et al.* [21] have demonstrated that NCD3G has a crucial role in the transmission of signals from ERD to 7tm_3. A naturally occurring inactivating mutation of the human CaSR (C582Y in NP_000379) associated with neonatal severe primary hyperparathyroidism was identified [22]. This suggests that the conserved cysteine residues, which have the potential to form disulfide bridges, might be essential for the function of NCD3G. Based on the structural prediction, the conserved glycine residues within the A1 modules of NCD3G might be important for correct folding of NCD3G domains because the backbone of NCD3G commonly turns greatly at these sites. Recently, it was found that the highly conserved glycine residues in the NCD3G domain are important for the bioactivity of human CaSR. Substitution of the glycine at codon 549 to arginine (G549R) and of the glycine at codon 557 to glutamic acid (G557E) in human CaSR was shown to co-segregate with familial hypocalciuric hypercalcemia [23,24]. These results suggest that, apart from the nine conserved cysteine residues, other conserved amino acids in NCD3G might also contribute to the function of family 3 GPCRs.

In contrast to the heterodimeric $GABA_B$ receptor, which lacks the NCD3G domain, it has been demonstrated that mGluR and CaSR form homodimers [25–29]. But whether the NCD3G domain plays a part in homodimer formation is still an unanswered question. Our 3D models highlight a region between two A1 models that contains two conserved hydrophobic amino acids and a conserved cysteine (e.g. Phe517, Val520 and Cys519 in NP_619642), all of which have the potential to form hydrophobic interactions or intermolecular disulfide links between two receptors in a homodimer (Figure 2).

Concluding remarks

The identification of NCD3G should provide new insight into the structure and function of the extra-membrane domain of GPCR family 3. Confirmation of its function and relationship to the other two domains should help our understanding of the mechanism of signaling via GPCRs, and benefit the development of novel therapeutics.

Acknowledgements

We thank Dianqing Wu, Qiang Yu and anonymous referees for critical reviewing of the article. We thank Youxing Qu for valuating 3D models. This work was supported by the National 973 Program, the National 863 Program and the National Natural Science Foundation (No. 30024001).

Supplementary data

Supplementary data associated with this article can be found at doi:10.1016/j.tibs.2004.07.009

References

1 Bockaert, J. and Pin, J.P. (1999) Molecular tinkering of G protein-coupled receptors: an evolutionary success. *EMBO J.* 18, 1723–1729
2 Bräuner-Osborne, H. *et al.* (2000) Ligands for glutamate receptors: design and therapeutic prospects. *J. Med. Chem.* 43, 2609–2645
3 Bateman, A. *et al.* (2002) The Pfam protein families database. *Nucleic Acids Res.* 30, 276–280
4 Letunic, I. *et al.* (2002) Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res.* 30, 242–244
5 Altschul, S.F. *et al.* (1997) Gapped BLAST and PSI–BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402
6 Higgins, D. *et al.* (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680
7 Kunishima, N. *et al.* (2000) Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. *Nature* 407, 971–976
8 Tsuchiya, D. *et al.* (2002) Structural views of the ligand-binding cores of a metabotropic glutamate receptor complexed with an antagonist and both glutamate and Gd3+. *Proc. Natl. Acad. Sci. U. S. A.* 99, 2660–2665
9 Kelley, L.A. *et al.* (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.* 299, 499–520
10 Shi, J. *et al.* (2001) FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure dependent gap penalties. *J. Mol. Biol.* 310, 243–257
11 McGuffin, L.J. *et al.* (2000) The PSIPRED protein structure prediction server. *Bioinformatics* 16, 404–405
12 Karplus, K. *et al.* (1998) Hidden Markov models for detecting remote protein homologies. *Bioinformatics* 14, 846–856
13 Douguet, D. and Labesse, G. (2001) Easier threading through web based comparisons and cross-validations. *Bioinformatics* 17, 752–753
14 Labesse, G. and Mornon, J-P. (1998) A tool for incremental threading optimization (TITO) to help alignment and modelling of remote homologs. *Bioinformatics* 14, 206–211
15 Dunbrack, R.L., Jr. and Karplus, M. (1993) Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *J. Mol. Biol.* 230, 543–574
16 Sali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815
17 Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18, 2714–2723
18 Gracy, J. *et al.* (1993) Improved alignment of weakly homologous protein sequences using structural information. *Protein Eng.* 6, 821–829
19 Eisenberg, D. *et al.* (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol.* 277, 396–404
20 Jean-Luc, B. *et al.* (2002) The molecular architecture of the TNF superfamily. *Trends Biochem. Sci.* 27, 19–26
21 Hu, J. *et al.* (2000) Human Ca^{2+} receptor cysteine-rich domain. Analysis of function of mutant and chimeric receptors. *J. Biol. Chem.* 275, 16382–16389
22 Pearce, S.H.S. *et al.* (1995) Calcium-sensing receptor mutations in familial benign hypercalcemia and neonatal hyperparathyroidism. *J. Clin. Invest.* 96, 2683–2692
23 D'Souza-Li, L. *et al.* (1998) Four novel calcium-sensing receptor mutations in familial hypocalciuric hypercalcemia and autosomal dominant hypocalcemia. *Bone* 23S, W350
24 Nakayama, T. *et al.* (2001) A novel mutation in Ca^{2+}-sensing receptor gene in familial hypocalciuric hypercalcemia. *Endocrine* 15, 277–282
25 Möhler, H. and Fritschy, J-M. (1999) $GABA_B$ receptors make it to the top – as dimers. *Trends Pharmacol. Sci.* 20, 87–89
26 Romano, C. *et al.* (1996) Metabotropic glutamate receptor 5 is a disulfide-linked dimer. *J. Biol. Chem.* 271, 28612–28616
27 Tsuji, Y. *et al.* (2000) Cryptic dimer interface and domain organization of the extracellular region of metabotropic glutamate receptor subtype 1. *J. Biol. Chem.* 275, 28144–28151
28 Bai, M. *et al.* (1998) Dimerization of the extracellular calcium-sensing receptor (CaR) on the cell surface of CaR-transfected HEK293. *J. Biol. Chem.* 273, 23605–23610
29 Ray, K. *et al.* (1999) Identification of the cysteine residues in the amino-terminal extracellular domain of the human Ca^{2+} receptor critical for dimerization. *J. Biol. Chem.* 274, 27642–27650
30 Goodstadt, L. *et al.* (2001) CHROMA: consensus-based colouring of multiple alignments for publication. *Bioinformatics* 17, 845–846
31 Cuff, J.A. *et al.* (1998) Jpred: a consensus secondary structure prediction server. *Bioinformatics* 14, 892–893

0968-0004/$ - see front matter
doi:10.1016/j.tibs.2004.07.009