**Protein Engineering** vol.15 no.11 pp.881–885, 2002

# Prediction of heterodimerization interfaces of G-protein coupled receptors with a new subtractive correlated mutation method

**Marta Filizola[1], Osvaldo Olmea and Harel Weinstein[1]**

Department of Physiology and Biophysics, Mount Sinai School of Medicine, One Gustave Levy Place, New York, NY 10029, USA

[1]To whom correspondence should be addressed.
E-mail: marta.filizola@physbio.mssm.edu; harel.weinstein@mssm.edu

**Recent studies employing differential epitope tagging, selective immunoprecipitation of receptor complexes and fluorescence or bioluminescence resonance energy transfer techniques provide direct evidence for heterodimerization between both closely and distantly related members of the G-protein coupled receptor (GPCR) family. Since heterodimerization appears to play a role in modulating agonist affinity, efficacy and/or trafficking properties, the molecular models of GPCRs required to understand receptor function must consider these oligomerization hypotheses. To advance knowledge in this field, we present here a computational approach based on correlated mutation analysis and the structural information contained in three-dimensional molecular models of the transmembrane regions of GPCRs built using the rhodopsin crystal structure as a template. The new subtractive correlated mutation method reveals likely heterodimerization interfaces amongst the different alternatives for the positioning of two tightly packed bundles of seven transmembrane domains next to each other in contact heterodimers of GPCRs. Predictions are applied to GPCRs in the class of opioid receptors. However, in the absence of a known structure of any GPCR dimer, the features of the method and predictions are also illustrated and analyzed for a dimeric complex of known structure.**

*Keywords*: correlated mutation analysis/dimerization/
G-protein coupled receptors/interface

## Introduction

Until fairly recently, the interaction between G-protein coupled receptors (GPCRs) and G-proteins had considered only 1:1 stoichiometric ratios between monomeric units of the receptors and their heterotrimeric targets. More recent experimental studies have suggested dimerization or even higher order oligomerization of GPCRs (Devi, 2001; Angers *et al*., 2002). In particular, such complexes can either involve identical proteins (homodimers) or be the result of the association of two non-identical proteins (heterodimers).

Recent reports on heterodimerization (Jordan and Devi, 1999; Devi, 2001) of closely and distantly related members of the GPCR family suggest potential roles for this phenomenon in modulating agonist affinity, efficacy and/or trafficking properties. Heterodimerization seems to be selective, so that GPCRs will interact with one type of receptors, but not another. Heterodimerization between closely related members of the GPCR family has been observed for $GABA_BR1$–$GABA_BR2$ (Jones *et al*., 1998; Kaupmann *et al*., 1998; White *et al*., 1998), $M_2$–$M_3$ muscarinic (Maggio *et al*., 1999; Sawyer and Ehlert, 1999), κ–δ opioid (Jordan and Devi, 1999), μ–δ opioid (George *et al*., 2000; Gomes *et al*., 2001), 5HT1B–5HT1D serotonin (Xie *et al*., 1999), SSTR1–SSTR5 somatostatin (Rocheville *et al*., 2000b), SSTR2A–SSTR3 somatostatin (Pfeiffer *et al*., 2001) and CCR2–CCR5 chemokine (Mellado *et al*., 2001) receptors. Recent examples of suggested heterodimerization between more distantly related members of the GPCR family are adenosine A1–D1 dopamine (Gines *et al*., 2000), angiotensin AT1–bradykinin B2 (AbdAlla *et al*., 2000), somatostatin SSTR5–D2 dopamine (Rocheville *et al*., 2000a), $\beta_2$-adrenergic–δ opioid (Jordan *et al*., 2001), $\beta_2$-adrenergic–κ opioid (Jordan *et al*., 2001) and metabotropic glutamate 1–α-adenosine A1 (Ciruela *et al*., 2001) receptors. Finally, examples of GPCR subtypes that have been shown not to produce heterodimers are μ opioid with κ opioid receptors (Jordan and Devi, 1999), somatostatin SSTR5 with SSTR4 (Rocheville *et al*., 2000b) and chemokine CCR2 with CXCR4 (Mellado *et al*., 2001) receptors.

As the effect that GPCR heterodimerization has *in vivo* on the modulation of receptor function is not yet known, molecular models of interacting GPCRs can be used to advance knowledge in this field. A model of receptor interaction involving swapping of TM domains has been proposed for homodimers and symmetric chimeric heterodimers (Gouldson *et al*., 2001), but its validity has been put into question by experimental evidence (Lee *et al*., 2000; Schulz *et al*., 2000). In any case, even the original authors propose that receptor heterodimers are more likely to contain only 'contact dimers' (Gouldson *et al*., 2001). These dimers derive from the association of 1:1 stoichiometric molecular complexes of receptors and may involve their extracellular, transmembrane and/or C-terminal regions. In addition, this association may be due to a combination of both covalent (disulfide) and non-covalent interactions.

To obtain a structural model of a dimerization complex involving the transmembrane domains of GPCRs, the goal would be to pack the bundles of seven transmembrane segments against one another. There are at least 49 different configurations in which the bundles can be packed next to each other. The computational approach presented here offers to reduce the number of possible configurations to a limited number of the most likely interfaces for specific GPCR heterodimerization.

It has recently been demonstrated that oligomer interfaces are significantly conserved with respect to the protein surface (Valdar and Thornton, 2001). Moreover, correlated mutations have been shown to contain information about inter-domain contacts (Oliveira *et al*., 1993; Pazos *et al*., 1997). The correlation has been interpreted as a result of the tendency of positions in proteins to mutate in a coordinated manner if the interface has to be preserved for structural or functional reasons. Thus, sequence changes occurring during evolution at the interface of dimerization of a given monomer A would be compensated by changes in the interacting monomer B in

Downloaded from https://academic.oup.com/peds/article/15/11/881/1554007 by guest on 21 August 2022

© Oxford University Press 

order to preserve the interaction interface. Based on these observations and the computational methods for identifying correlated mutations (Olmea and Valencia, 1997), we have developed a new subtractive correlated mutation (SCM) method aimed at the identification of the most likely heterodimerization interfaces between interacting proteins that are structurally similar to each other, such as individual GPCRs in subfamilies of these receptors. The prediction of the interface is further refined by filtering the residues that are identified at the heterodimerization interface of interacting GPCRs by application of the SCM method, based on structural models of the individual GPCRs. Thus, a list of putative interface residues is pruned based on a criterion of solvent accessibility that identifies the residues on the outer (lipid-facing) surface of the transmembrane bundle. The resulting interfaces can be used in the consideration of alternatives for packing the GPCRs in dimers. The method and resulting predictions are illustrated for GPCRs in the class opioid receptors and evaluated for a protein dimer of known structure.

## Methods

Given two structurally similar interacting proteins A and B with different amino acid sequences, a list of pairs of correlated residues is predicted from sequence conservation criteria in their multiple sequence alignment. This list contains both intra- and intermolecular pairs and is obtained from the alignment of the sequences of the two proteins in various species, constructed as described below. The general algorithm for the identification of pairs of correlated mutations has been described in detail (Olmea and Valencia, 1997). Based on the principles used in this algorithm, we have developed a new SCM method. This method allows one to filter out the intramolecular pairs of correlated residues within A and within B from the complete list of intra- and intermolecular pairs of correlated residues predicted from the multiple alignment. Specifically, the method requires analysis of four different multiple sequence alignments: (1) the sequence alignment of A + B, obtained by appending the sequences of protein A from each of the species to the corresponding sequences of protein B, with each resulting A + B sequence from a particular species being treated as if it were from a single protein; (2) the multiple alignment of all known sequences of A from different organisms; (3) the multiple alignment of all known sequences of B from different organisms; and (4) the multiple alignment including all known sequences of A together with all known sequences of B. Use of the multiple sequence alignment of A + B as an input to calculate correlated mutations yields a list of all intra- and intermolecular pairs of correlated residues [CM(A + B)]. In contrast, correlated mutations based on the multiple sequence alignments of A or B will provide a list of likely intramolecular pairs of correlated residues [CM(A) and CM(B), respectively]. Since the two monomers A and B are structurally similar to each other, the correlated mutations [CM(A,B)] can be calculated using the multiple sequence alignment of all known sequences of A together with all known sequences of B. The resulting set of pairs provides an additional filter to eliminate pairs of correlated residues that are likely intramolecular. The pairs of intermolecular correlated residues (I) can then be obtained from the following equation that defines the SCM:

$$I = CM(A + B) - CM(A) - CM(B) - CM(A,B) \quad (1)$$

In order finally to identify the residues that are at the heterodimerization interface of A and B, the resulting set (I) obtained from the subtractive correlated mutation method is further pruned based on solvent accessibility values calculated for each residue of A and B from the atomic coordinates of their three-dimensional structures. Specifically, the intermolecular

Downloaded from https://academic.oup.com/peds/article/15/11/881/1554007 by guest on 21 August 2022

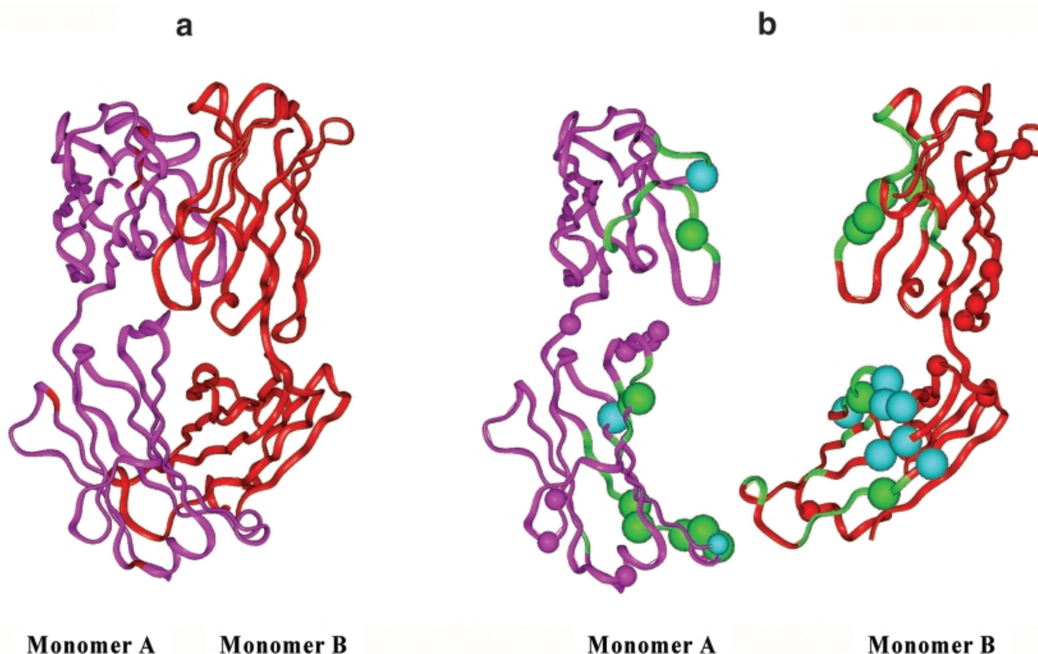


**Fig. 1.** Ribbon representations of the three-dimensional structures of monomers A and B in the 15C8 complex (**a**) and as separated entities (**b**). The interface of heterodimerization between monomers A and B as it appears in the crystal structure of 15C8, is shown in green. CPK elements indicate the residues predicted with the SCM method to be at the heterodimerization interface of the A and B monomers. Specifically, green CPK objects indicate the predictions that corresponded exactly to residues at the real heterodimerization interface between A and B, whereas light-blue CPKs indicate those that were close ($<i + 7$) to residues at the interface. False positives are reported as magenta CPK objects on monomer A and red CPKs on monomer B.

pairs where either one or both residues are completely or partially inaccessible to the solvent are eliminated from the list. The remaining residues of each monomer are then considered to be candidates for the interface of heterodimerization between the two proteins.

## Results

### *Testing the method*

*(a) Selection of structures for the test set*. To carry out a validation test of the new SCM method, crystallographic structures of dimeric complexes were retrieved from the Protein Quaternary Structure File Server (PQS; http://pqs.ebi.ac.uk) and considered for inclusion in the analysis if they fulfilled the following criteria: (1) sequences must contain >50 amino acids; this requirement excludes peptides from the test set; (2) the two proteins in the complex must have <80% amino acid sequence identity; this will ensure elimination of homodimers from the test set; (3) the mean loss of accessible surface area per chain upon assembly formation compared to the isolated chains must be >400 Å; (4) the two monomers must have similar 3D structures (r.m.s.d. <3.0 Å); this condition is required since meaningful 3D models of GPCRs (Ballesteros *et al*., 2001; Visiers *et al*., 2002) are currently built using the same rhodopsin crystal structure as a template; and (5) the sequences of the corresponding proteins in at least five species must be available for the sequence alignments.

*(b) Prediction of interfaces*. Among the initial 883 heterodimeric complexes retrieved from the PQS server on December 11, 2001, only four structures satisfied all the criteria listed above. Application of the SCM method to these four structures demonstrated the ability of the method to predict residues at the interface between structurally related proteins. As an example of the predictive ability of the SCM method, we report here the results obtained for one (PDB code: 15C8) of these four dimeric complexes.

The heterodimer corresponding to the 15C8 PDB code consists of two proteins (A and B) that share a 23% sequence identity and a 2.6 Å structural similarity. Twenty-seven corresponding species of A and B were appended to each other as described in Methods and treated as if they were one protein in order to identify the intra- and intermolecular pairs of correlated residues derived by their multiple sequence alignment. For the analysis, the dimerization interface was defined by residues in A and B that had Cα atoms within 8 Å distance.

Application of the SCM method identified likely intermolecular residues for both the A and B monomer. Figure 1 shows the three-dimensional structures of monomers A and B in the 15C8 complex (Figure 1a), as well as separated entities (Figure 1b). The interface of heterodimerization between the monomers A and B, as it appears in the crystal structure of 15C8, is represented in green in Figure 1b. On these ribbon representations, CPK elements were used to indicate the residues predicted with the SCM method to be at the heterodimerization interface of the monomers. Specifically, the method predicted correctly 36% of the residues of A (from a total of 36) appearing at the heterodimerization interface in the crystal structure and 44% of those of B (from a total of 34). Of the total number of residues predicted to be at the heterodimerization interface of the A monomer, 35% were exact and 47% were within ($i$ + 7) and hence considered correct. For the B monomer, the corresponding correct predictions totaled 65%. In Figure 1b, predictions that corresponded exactly to residues at the heterodimerization interface in the 15C8 crystal structure are shown as green CPK objects and those that were close (<$i$ + 7) are rendered in light-blue CPKs. False positives are reported in magenta CPK objects on monomer A and red CPKs on monomer B.

### *Application of the SCM method to opioid receptor heterodimers*

*(a) The δ–μ heterodimer*. The five available sequences of the δ opioid receptor from human, rat, mouse, pig and zebra fish were appended to the corresponding sequences of the μ opioid receptor and arranged in a multiple sequence alignment. Application of the SCM method to the δ–μ opioid receptor heterodimer identified more than one heterodimerization interface. As shown in Figure 2a, most of the residues of δ opioid receptor predicted to be at the interface of the heterodimer with the μ opioid receptor are within TM4, TM5 and TM6. In contrast, the interface residues predicted for the μ opioid receptor (Figure 2b) appear to involve mainly TM1 in the heterodimerization with δ opioid receptor. Based on these results, the number of alternatives for positioning the bundles of seven TM domains of δ and μ opioid receptors next to each other in a heterodimer is reduced to a small number of

Downloaded from https://academic.oup.com/peds/article/15/11/881/1554007 by guest on 21 August 2022

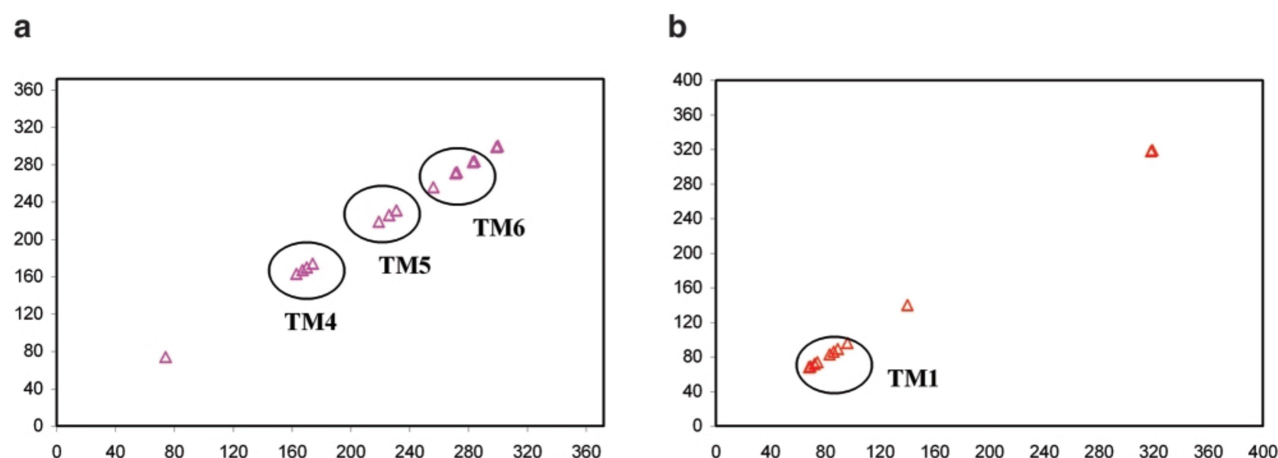


**Fig. 2.** Residues of δ (magenta) and μ (red) opioid receptors predicted to be at the most likely heterodimerization interfaces of the δ–μ complex by the SCM method. The specific helices to which the most of these residues belong are indicated in the black circles.

seemingly equally possible configurations. Specifically, the most likely heterodimerization interfaces of the δ–μ pair involve TM4, TM5 and TM6 of the δ opioid receptor with TM1 of the μ opioid receptor. Given the common template of the GPCRs defined by the bundle of transmembrane helices and their orientation in the membrane, any geometrically feasible combination of interfaces involving these TMs is identified as a possible configuration of the heterodimer. The number of such possible configurations is small.

*(b) The μ–κ heterodimer.* The same procedure using the SCM to identify the interface was applied to the μ–κ opioid receptor pair, which is known for failing to exhibit heterodimerization (Jordan and Devi, 1999). Interestingly, the SCM-based method correctly predicts that no residues are likely to be at the heterodimerization interface.

## Discussion

Although the functional implications of GPCR heterodimerization *in vivo* are not known yet [however, see elsewhere (He *et al*., 2002) for proposed functions of μ opioid receptor homodimers *in vivo*], a growing body of evidence from studies *in vitro* suggests that GPCR heterodimerization may be required for either efficient agonist binding and signaling or to generate novel binding sites. Since a novel pharmacology can be generated from this phenomenon, molecular models of interacting GPCRs can guide experiments to understand the basis of receptor function.

An association of 1:1 stoichiometric molecular complexes of receptors is required to achieve contact dimers in GPCR heterodimerization. This association may also involve extracellular, transmembrane and/or C-terminal regions, but our analysis has been limited to the seven TM regions of GPCRs because at the present stage of the research, meaningful 3D models of interacting GPCRs can only contain the TM helices that have sufficiently high homology with the corresponding regions of the only GPCR crystallographic structure known to date, rhodopsin. In fact, detailed analyses of the rhodopsin structure together with the results of both sequence analysis and molecular modeling (Ballesteros *et al*., 2001; Visiers *et al*., 2002) support the use of the crystal structure of rhodopsin (Palczewski *et al*., 2000) as a template to model only the transmembrane domain of other rhodopsin-like GPCRs. Using such models, the SCM method was shown here to predict dimerization interfaces that significantly limit the choice of possible configurations from the $\geqslant$49 different configurations in which two interacting bundles of seven TM domains of GPCRs can be positioned next to each other. That correlated mutation analysis of lipid-facing residues may be used in an attempt to identify dimerization interfaces has recently been demonstrated (Gouldson *et al*., 2001). The new SCM method identifies the most likely heterodimerization interfaces of GPCRs amongst the different alternatives. Importantly, the method recognizes subtype specificity in GPCR heterodimerization, as demonstrated by the control case of μ–κ opioid receptors that had be shown experimentally not to dimerize (Jordan and Devi, 1999) and were correctly predicted with SCM to have no residues likely to be at the heterodimerization interface.

In the present application of the method to δ–μ opioid receptors (Figure 2), most of the correlated residues of δ opioid receptor that have been identified on the outer (lipid-facing) surface of the receptor bundle are in TM4, TM5 and TM6, whereas in the μ opioid receptor TM1 is the helix that is likely to be involved in the heterodimerization with δ. The structural interpretation is feasible even with low-resolution models of the TM region of the receptors. Here it was based on models derived from the rhodopsin structure (Palczewski *et al*., 2000) as a template. The results indicate that there are at least two mutually exclusive configurations of the δ–μ heterodimeric complex that can be formed on this basis. Specifically, TM1 of the μ opioid receptor cannot interact simultaneously with both TM6 and TM4 of the δ opioid receptor. These predictions are testable experimentally and have additional implications for the study of GPCR interactions. Thus, if experimental evidence were to implicate both TM4 and TM6 in the heterodimerization of δ opioid receptor with μ, then our results would indicate that oligomerization, rather than dimerization, is occurring between these opioid receptors.

Finally, it is important to emphasize that the predictive ability of the method can be influenced by many factors. First, the analysis requires multiple sequence alignments of the same GPCR cloned from different organisms. Based on the small but growing body of evidence on subtype specificity in GPCR heterodimerization, the sequence alignment has to be limited strictly to the specific receptor for which dimerization is considered. Second, only a few sequences from different organisms are known for each GPCR. As a result, the number of sequences in the multiple sequence alignments is often inadequate for a statistical analysis of the data. Third, predictions are limited to the TM regions of the GPCRs under study, owing to the low sequence identity of extracellular and intracellular loops among GPCRs. Fourth, the validation efforts for the SCM approach could achieve statistical significance only upon availability of more structures of heterodimeric complexes of structurally similar proteins that are eligible by the criteria we have defined.

## Acknowledgements

We are grateful to Dr Fabien Campagne for providing Bioinformatics tools to prune the list of the crystallographic structures used to test the method. The work was supported by NIH grants DA-12923 and K05 DA-00060. M.F. is supported by NRSA T32 DA07135 from NIH–NIDA.

## References

AbdAlla,S., Lother,H. and Quitterer,U. (2000) *Nature*, **407**, 94–98.
Angers,S., Salahpour,A. and Bouvier,M. (2002) *Annu. Rev. Pharmacol. Toxicol.*, **42**, 409–435.
Ballesteros,J.A., Shi,L. and Javitch,J.A. (2001) *Mol. Pharmacol.*, **60**, 1–19.
Ciruela,F. *et al.* (2001) *J. Biol. Chem.*, **276**, 18345–18351.
Devi,L.A. (2001) *Trends Pharmacol. Sci.*, **22**, 532–537.
George,S.R., Fan,T., Xie,Z., Tse,R., Tam,V., Varghese,G. and O'Dowd,B.F. (2000) *J. Biol. Chem.*, **275**, 26128–26135.
Gines,S. *et al.* (2000) *Proc. Natl Acad. Sci. USA*, **97**, 8606–8611.
Gomes,I., Jordan,B.A., Gupta,A., Trapaidze,N., Nagy,V. and Devi,L.A. (2001) *J. Neurosci.*, **20**, RC110 (111–115).
Gouldson,P.R., Dean,M.K., Snell,C.R., Bywater,R.P., Gkoutos,G. and Reynolds,C.A. (2001) *Protein Eng.*, **14**, 759–767.
He,L., Fong,J., von Zastrow,M. and Whistler,J.L. (2002) *Cell*, **108**, 271–282.
Jones,K.A. *et al.* (1998) *Nature*, **396**, 674–679.
Jordan,B.A. and Devi,L.A. (1999) *Nature*, **399**, 697–700.
Jordan,B.A., Trapaidze,N., Gomes,I., Nivarthi,R. and Devi,L.A. (2001) *Proc. Natl Acad. Sci. USA*, **98**, 343–348.
Kaupmann,K. *et al.* (1998) *Nature*, **396**, 683–687.
Lee,S.P., O'Dowd,B.F., Ng,G., Varghese,G., Akil,H., Mansour,A., Nguyen,T. and George,S.R. (2000) *Mol. Pharmacol.*, **58**, 120–128.
Maggio,R., Barbier,P., Colelli,A., Salvadori,F., Demontis,G. and Corsini,G.U. (1999) *J. Pharmacol. Exp. Ther.*, **291**, 251–257.



Downloaded from https://academic.oup.com/peds/article/15/11/881/1554007 by guest on 21 August 2022

Mellado,M., Rodriguez-Frade,J.M., Vila-Coro,A.J., Fernandez,S., Martin de Ana,A., Jones,D.R., Toran,J.L. and Martinez-A,C. (2001) *EMBO J.*, **20**, 2497–2507.

Oliveira,L., Paiva,A.C.M. and Vriend,G. (1993) *J. Comput.-Aided Mol. Des.*, **7**, 649–658.

Olmea,O. and Valencia,A. (1997) *Fold. Des.*, **2** (Suppl. 1), S25–S32.

Palczewski,K. *et al.* (2000) *Science*, **289**, 739–745.

Pazos,F., Helmer-Citterich,M., Ausiello,G. and Valencia,A. (1997) *J. Mol. Biol.*, **271**, 511–523.

Pfeiffer,M., Koch,T., Schroder,H., Klutzny,M., Kirscht,S., Kreienkamp,H.J., Hollt,V. and Schulz,S. (2001) *J. Biol. Chem.*, **276**, 14027–14036.

Rocheville,M., Lange,D.C., Kumar,U., Patel,S.C., Patel,R.C. and Patel,Y.C. (2000a) *Science*, **288**, 154–157.

Rocheville,M., Lange,D.C., Kumar,U., Sasi,R., Patel,R.C. and Patel,Y.C. (2000b) *J. Biol. Chem.*, **275**, 7862–7869.

Sawyer,G.W. and Ehlert,F.J. (1999) *J. Pharmacol. Exp. Ther.*, **289**, 464–476.

Schulz,A., Grosse,R., Schultz,G., Gudermann,T. and Schoneberg,T. (2000) *J. Biol. Chem.*, **275**, 2381–2389.

Valdar,W.S. and Thornton,J.M. (2001) *Proteins: Struct. Funct. Genet.*, **42**, 108–124.

Visiers,I., Ballesteros,J.A. and Weinstein,H. (2002) *Methods Enzymol.*, **343**, 329–371.

White,J.H., Wise,A., Main,M.J., Green,A., Fraser,N.J., Disney,G.H., Barnes,A.A., Emson,P., Foord,S.M. and Marshall,F.H. (1998) *Nature*, **396**, 679–682.

Xie,Z., Lee,S.P., O'Dowd,B.F. and George,S.R. (1999) *FEBS Lett.*, **456**, 63–67.

*Received March 9, 2002; revised June 15, 2002; accepted August 3, 2002*

Downloaded from https://academic.oup.com/peds/article/15/11/881/1554007 by guest on 21 August 2022