On Transversal Hydrophobicity of Some Proteins and Their Modules

Andrzej Galat*

Institute de Biologie et de Technologies de Saclay, IBiTec/DSV/CEA, CE-Saclay, F-91191 Gif-sur-Yvette Cedex, France

Received January 12, 2009

Hydrophobicity of proteins encoded in the genomes of diverse organisms was quantified using two novel concepts: (A) amino acid (AA) bulkiness-dependent hydrophobicity profiles and (B) spatial context of hydrophobicity distribution in AA triads. Both concepts were introduced into an algorithm that was used for extracting protein clusters from diverse genomic databases whose sequence attributes were similar to those in the multiple sequence alignment (MSA) of a given family of proteins. The sequences of the G protein-coupled receptors (GPCRs) encoded in different genomes were used as templates for testing the above concepts. The following sequence attributes were used for protein clustering: (A) sequence similarity scores (IDs); (B) amino acid composition (AAC); (C) hydrophobicity; (D) AA-bulkiness; and (E) α-helical propensity potentials. Diverse GPCRs display variable distributions of AA bulkiness-dependent buildups and declines in the hydrophobicity profiles that may be related to their function-dependent way of packing and allostery in the membrane. It is shown that intramolecular transversal nonbonded interactions between the TM segments in diverse GPCRs involve about 50% of hydrophobic atoms. Similar interaction networks exist between α-helices of tetratricopeptide (TPR) motifs-containing immunophilins and other proteins containing α-helical bundles.

INTRODUCTION

Nearly 30% of the expressed gene products in diverse eukaryotic cell phenotypes display a high hydrophobicity,[1] although not all of them belong to the membrane-anchored proteins, which usually contain from one to more than ten transmembrane segments (TMs)[2] with specific folding topology.[1,3−6] Some of the proteins embedded in the membrane of the cell's surface display selective affinity to diverse endocrine,[6] natural,[7,8] and synthetic pharmacologically relevant molecules[9−13] that control multiple functional aspects of the cell by signalization pathways transmitted to the cytoplasm/nuclear networks of proteins. For example, several hundreds of genes coding for G protein-coupled receptors (GPCRs) have been detected in the human genome. Average sequence similarity in the family of the human GPCRs is in the twilight zone (ID ≤ 25%).[14] The GPCRs bind functionally diversified arrays of molecules whose molecular mass may vary from hundreds of daltons, such as acetylcholine and other small neurotransmitters,[6,9,13] to higher mass molecules including diverse endocrine peptides[10,11] and large extracellular proteins.[6] Those diverse molecules display a wide spectrum of activities in the nervous system and other tissues. The majority of the metabotropic GPCRs have a similar membrane topology,[5,6,12] namely their seven TM α-helical segments are linked by extracellular and intracellular loops of different length. The TM segments contain on average about 60% of hydrophobic amino acid residues (AAs), but polar AAs and proline occur in the helical TM segments.[14] The extracellular loops and some amino acid (AA) patches localized in the TM segments constitute selective binding sites for different molecules that via allosteric effect[15,16] transmit some conformational alterations to the intracellular loops that are associated with various proteins involved in multiple networks of intracellular signalization pathways.[6,13] Some signalization networks have a more sophisticated assembly of membrane proteins, for example the hedgehog-signalization system[17,18] involving smoothened (7 TMs) and patched1 (12 TMs) assembly of proteins that transmit the signals from extracellular gradients of morphogens (hedgehog family of proteins) to intracellular pathways in phosphorylation-dependent manner.[19]

Here we describe several procedures that were used for the whole-genome analyses of hydrophobicity of proteins and their modules. The procedures utilize different hydrophobicity scales that were used for establishment of relative hydrophobicity levels of genomic proteins. We showed that some of the locally restricted sequence attributes such as sequence conservation level, hydrophobicity, bulkiness, pI, α-helical propensity potential, and AAC are useful in clustering functionally related hydrophobic proteins. In order to better elucidate the nature of intramolecular interactions involved in proteins containing diverse α-helical bundles, we computed two-dimensional (2D) distance maps from the X-ray structures of several GPCRs[20−29] the tetratricopeptide repeat (TPR) motifs-containing immunophilins[30−33] and several other proteins. Analyses of the 2D maps have revealed that these two hydrophobic sequence motifs (TMs and TPRs) display similar interaction patterns to those on 2D maps of helical bundles-containing structures.[34−36] The interhelical interaction clusters are stabilized via transversal networks of the hydrophobic AAs. Similar transversal nonbonded interactions between hydrophobic atoms constitute a major stabilizing force for diverse binary and ternary

* Corresponding author fax: 33-169089137; e-mail: galat@dsvidf.cea.fr. Corresponding author address: Bat. 152, SIMOPRO/IBiTec/DSV/CEA, CE-Saclay, F-91191 Gif-sur-Yvette Cedex, France.

10.1021/ci9001316 CCC: $40.75
Published on Web 07/01/2009

complexes between extracellular domain (ectodomain) of TGF-β-like receptors and their diverse ligands.[37] It is suggested that the TPR-containing hydrophobic immunophilins[38,39] may engage their FK506-like binding domains and TPRs in forming transversal intermolecular interaction networks that could link soluble multiproteic entities to some membrane proteins.

METHODS

Genomic Databases and Their Processing. Genomic databases produced in the National Center for Biotechnology Information (NCBI, National Institutes of Health, Bethesda, U.S.A.) were downloaded via the ftp server at the address (ftp.ncbi.nlm.nih.gov)[40] and were analyzed with an improved version of the Par_Seq modules,[14] called here the Pro_Hy program (Protein_Hydrophobicity). Two types of files were used, namely those that finish with an *.faa suffix (Fasta amino acid file) and *.gbk files (GenBank flat file format). The following genomic databases were used: the *Homo sapiens* genome (37742 sequences; 17172281 AAs), the *Maccaca mulata* genome (38001 sequences, 16867755 AAs), the *Bos taurus* genome (24853 sequences), the *Mus musculus* genome (26336 sequences, 12507256 AAs), the *Danio rerio* genome (34437 sequences, 15633279 AAs), the *Gallus gallus* genome (18529 sequences, 8729533 AAs), the *Drosophila melanogaster* genome (18218 sequences, 10059628 AAs), and the *Apis mellifera* genome (9257 sequences, 4729486 AAs). The current version of the program can use up to 70 peptides of length not exceeding 50 AAs for simultaneous comparisons of their sequence attributes (amino acid composition, overall hydrophobicity, AA-bulkiness, and α-helical propensity) to those embedded in the hydrophobic sequence segments derived from diverse genomic databases. Pro_Hy generates all the hydrophobic segments from each genomic sequence. Hydrophobic sequence patches that match the bait sequences are written to an external file in the Fasta format, and the file can be used as input to a multiple sequence alignment program. A detailed description of diverse modules of Pro_Hy is in the Supporting Information. All the MSAs were created with the Clustal W1.81 program[41] on a Macintosh computer.

Intramolecular Distance Maps Derived from Crystallographic Structures. Distance maps derived from several entries downloaded from the PDB server[42] were processed with the CORDAN_Pr program.[43] The following PDB entries were used: 2I37, photoactivated meta II-like bovine rhodopsin;[20] 3DQB, bovine rhodopsin-G_αCT complex;[21,22] 2Z73, squid rhodopsin,[23] 3D4S, 2RH1[24,25] and 2R4R[26] of the human β_2-adrenergic receptor; 2 VT4, β_1-adrenergic GPCR from the modified turkey *Meleagris gallopavo*;[27] 3EML, human A_{2A} adenosine receptor;[28] and the TPR motifs-containing PPIases:[29] 2IF4, a 41 kDa FKBP from *A. thaliana*;[30] 1IIP, bovine CyP40,[31] 1P5Q, FKBP52[32] and 1KT1, FKBP51.[33] Three high resolution structures containing four-helix bundle motifs were analyzed, namely 1M6T representing designed four-helix bundle in a soluble form of cytochrome[34] B562 (1.8 Å), the extracellular domain of human CD38 (1.60 Å) 3DZH,[35] and soluble human Flt3 ligand involved in hematopoiesis (1ETE).[36] The calculated triangular distance maps were decorated with the ClarisDraw program, whereas the crystallographic structures were drawn with the PyMol program.[44] Here we used the following three arbitrarily chosen ranges for sequence position-dependent intramolecular atomic interactions, namely short-range interactions are those between the atoms in AAs separated by not more than three other AAs (interactions i to i+4 where i corresponds to sequence position), medium range interactions for AAs separated from four to nine AAs (i with ($i \geq i+5$ to $i \leq i+10$)), and long-range interactions for sequence positions separated by at least AAs (i to $i > 10$). Interatomic interactions were sorted out into the following three types, namely 1° interactions between electron-donor and electron-acceptor atoms (diverse types of N and O atoms); 2° interactions between hydrophobic atoms (different types of C and S atoms); and 3° interactions between the atoms in groups 1° and 2° (mixed atom types). Atom types and their parameters were taken from the Amber force filed.[45]

Hydrophobicity of Proteins. Several scales for computing hydrophobicity profiles of proteins were used[46−50] including the recently established unified hydrophobicity scale (UHS) and mammalian hydrophobicity scale (MHS).[51] An odd-number of AAs was used as a sliding frame, namely from 7 to 11 AAs were used in calculating the hydrophobicity profile of a protein that has a low overall HI (hydrophilic proteins), whereas longer AA-sliding frames were used for computation of hydrophobicity profiles of membrane proteins (15 to 21 AAs). Hydrophobicity values were averaged using rectangular weight function. Overall hydrophobicity of a protein (HI) was expressed as percentage of the AA residues (AAs) being in its hydrophobic segments. The HI factors are dependent on the hydrophobicity scale and the formula used for hydrophobicity profile computing. Amphipathic character of protein's segments was established with the procedure of Cornette and colleagues.[48] As an option, we also used the AA sliding triads within the principal sliding frame. All 8000 possible AA triads were calculated from the genomic databases using the cube.f program (Galat, unpublished). This procedure yielded fine distributions of buildups and declines of spatial hydrophobicity by taking into account the smallest step of 'overall' physicochemical character of given AA residue flanked by −1 and +1 AA residues. We established a simple formula describing buildups and declines of potential spatial hydrophobicity by taking into account either the bulkiness or surface of AAs in given sequence patch. The values of the bulkiness[52] of 20 natural AAs were used in establishing a series of scaling factors (Sf)

$$\mathrm{Sf_i}(\text{i AA-residue}) = (\text{bulkiness of i AA-residue})/(\text{bulkiness of Gly}) \quad (1)$$

Hydrophobicity profile (H_p) for a protein of length (L) as a function of the bulkiness of AAs constitutes a series of values that are derived from eq 2

$$\mathrm{H_{p(m)}} = \sum_{i=1}^{9} \mathrm{f_i}*\mathrm{Sf_i}/\left[\sum_{j=1}^{7} \mathrm{f_j}*\mathrm{Sf_j} + \sum_{k=1}^{4} \mathrm{f_k}*\mathrm{Sf_k}\right] \quad (2)$$

where $\mathrm{Sf_i}$, $\mathrm{Sf_j}$, and $\mathrm{Sf_k}$ are the bulkiness factors, and $\mathrm{f_i}$, $\mathrm{f_j}$, and $\mathrm{f_k}$ are the numbers of the following AA residues (AAs) being within the sliding frame of length m. If Σ(hydrophobic AAs) < Σ(hydrophilic AAs + charged AAs), then we take an inverse of the right side of eq 2. Twenty natural AAs were classified here in the following fashion: hydrophobic

AAs (i=A,C,F,I,L,M,W,Y,V); hydrophilic AAs and well soluble in aqueous solution (j=G,H,N,Q,P,S,T); and charged AAs (k=D,E,K,R). The solubility profiles of the common L-α-amino acids have been reported in numerous papers.[53,54] Although, proline should be accounted for as a hydrophobic residue, but probably due to heterogeneity of its ring puckering,[55] the crystalline packing forces are weakened that causes its high solubility in aqueous solution.[53,54] Moreover, a high propensity of Pro to occur in solvent-exposed β-turns prompted us to include Pro in the series of hydrophilic AAs. As a matter of fact, in the diverse hydrophobicity scales,[46−51] Pro was considered to be a hydrophilic residue.[55] Solubility potential of AAs was used for the establishment of the first hydrophobity scale by Nozaki and Tanford.[56] Since then diverse hydrophobicity scales have been proposed,[51] namely by applying different experimental methods,[47,50,57,58] knowledge-based hydrophobicity potentials of amino acid residues,[51,59] or using a combination of both above-mentioned approaches.[46] In this communication a novel hydrophobicity scale of 20 AAs was deduced from the nature of interaction patterns between diverse AAs and established from the X-ray structures of the following three groups of proteins: the FKBPs,[60] the cyclophilins,[61] and the TFPDs.[62] Each of these groups contains multiple paralogues whose overall hydrophobicity varies to a considerable level.[60−62] More than 100 X-ray structures of the above three groups of proteins (see Table S1, Supporting Information) were analyzed, and the nature of intramolecular interactions established. In Table S1 were summarized interaction types in the three sets of the crystallographic structures, namely the nature of the atoms (hydrophobic [C and S atoms] or hydrophilic [O and N atoms]) being at the range of distances 2.7−4.5 Å. The analyses showed that the hydrophobic atoms of the following residues Cys, Phe, Ile, Leu, Met, Trp, Tyr, and Val have the large numbers of nonbonded interactions with their counterparts in 20 natural AAs. For example, 43% interactions of a Tyr residue involve its hydrophobic atoms, whereas only about 18% interactions of a Ser residue are between hydrophobic atoms. About 11% of interactions are between the hydrophilic atoms of a Tyr residue with their counterparts in the other AAs, while the remaining interactions (46%) are between the mixed types of atoms. For each of the 20 natural AAs, we calculated the ratio between the numbers of its hydrophobic atoms involved in interaction networks (F1) to the numbers of the hydrophilic atoms being engaged in the networks (F2). The ordering of AAs according to the obtained series of values resembles that in the diverse hydrophobicity scales (Table S1); this series of data was tentatively called the intramolecular interaction-based hydrophobicity scale (IMI_HS). It was calculated that about 50% of intramolecular interactions between the hydrophobic atoms of charged Lys and Arg residues are with their counterparts in 20 natural AAs which shows that these two AAs also contribute to the hydrophobic interaction network. Moreover, it has been suggested that the sulfur atoms in Cys and Met residues may have a truly hydrophobic character.[63] As a matter of fact, in an *in silico* folding study,[64] it has been established that Cys is the most hydrophobic AAs. Our analyses of intramolecular interactions have revealed that the sulfur atoms in cystines and methionines often occur inside the hydrophobic cores of the proteins and have large numbers of interactions with the hydrophobic atoms.[37] We added the estimated surface of the S atom (in average 41 Å) to the hydrophobic surfaces of Cys and Met residues.[63] The ratios between the total volume and surface and their hydrophobic counterparts in 20 natural AAs were compared with the distributions of intramolecular interactions in the above three groups of proteins (Table S2). The comparison revealed that, in general, the ratios between the hydrophobic volume and its total for 20 natural AAs display a similar trend as that shown in Table S1.

Software Availability. All the programs presented here were written in standard Fortran 77. They were executed on a mainframe computer working under the Unix system, on a Power Macintosh G4 with the OX10.3 version of the operating system, and a PC_Windows station. The precompiled versions of the Pro_Hy program and the two database processing programs, himapi and data_proc, for the MacOS and Windows platforms accompanied by a manual and several input and output modules can be obtained from the author or can be extracted from a zip file attached in the Supporting Information. The CPU time for an a search of a genomic database containing about 30,000 protein using seven TMs does not exceed more than 2 min on a Powerbook G4. The programs were compiled using the Absoft F95 compiler (version 9.0) for the Macintosh and PC_Windows series of computers.

RESULTS

Hydrophobicity of Proteins Encoded in Diverse Genomic Databases. In Figure 1A,B are shown seven hydrophobicity profiles generated from the sequences of two GPCRs, namely the human rhodopsin (NP_000530) having 93% sequence similarity to the sequence of bovine rhodopsin[20] whose structure has been solved (2I37)[19] and the human adenosine A_{2A} receptor[29] (NP_000666, 3EMI). The profiles were calculated using the scales established by Hopp and Woods (HW),[47] Kyte and Doolittle (KD),[46] Wimley and White (WW),[49] the hydrophobicity scale developed by Hessa and collegues,[50] the sign-inverted knowledge-based hydrophobicity scales (UHS and MHS) established by Koehler and colleagues,[51] AA triad-based profile and AA-bulkiness buildups, and declines profile (black line). The profiles display seven major positive peaks whose sequence positions perfectly correspond to the experimentally established TMs in the X-ray structures[20,29] of 2I37 and 3EML. Each of the sequences has its specific distribution of the hydrophobicity peaks, namely in the human rhodopsin TM-I and TM-VI are highly hydrophobic, whereas TM-I and TM-V of the adenosine A_{2A} receptor have considerable loads of the hydrophobic AAs. The hydrophobicity profiles of the GPCRs generated with eq 2 lack any symmetry. Moreover, as higher is the value of the positive peak in the profile (eq 2) as higher is the extent of the spatial hydrophobicity embedded in given peptidic patch.

Equation 2 was also used in the Pro_Hy program for the whole-genome detection of the hydrophobic segments. Analyses of several genomic databases have revealed that, in general, diverse GPCRs belonging to class A have different profiles of buildups and declines of their spatial hydrophobicity (eq 2), while the distributions of the hydrophobic peaks lacked any symmetry. The Pro_Hy analyses also ordered the genomic sequences according to their overall

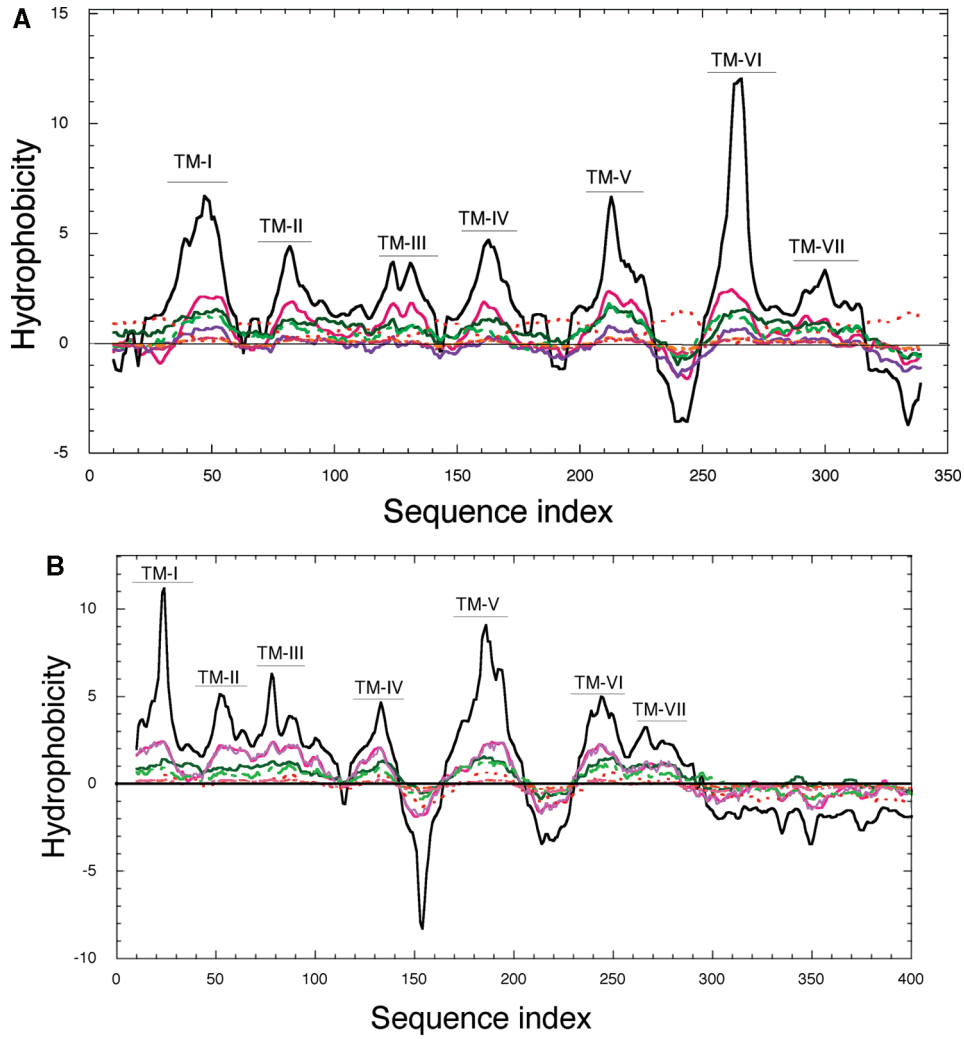

Figure 1. Hydrophobicity profiles generated using the scales: AA-triads (continuous scarlet lane); spatial hydrophobicity (continuous black line); HW (continuous green line); KD (continuous violet line); WW (dotted red line); IMI_HS (dashed green line); UHS (dashed orange line); MHS (dashed pink line); 19 AAs sliding frame was used.

hydrophobicity (HI). Distributions of the HIs were dependent on several factors such as the hydrophobicity scale used, the length of the AAs sliding frame, and the formula used for hydrophobicity profiles computing; a distribution of the HIs in the human genomic database is shown in Figure 2.

Using the KD scale[46] and a 17 AAs sliding frame, we established the HIs ranging from 35% to nearly 80% for the diverse human GPCRs. The hydrophilic edge (the HIs 30−40%) of the GPCRs is occupied by some of the human cholinergic, nicotinic, and NMDA receptors, namely NMDA receptor 1 (NP_067544), GPCR156 (NP_694547), or nicotinic α-cholinergic receptor (NP_000735). The hydrophobic edge of the GPCRs is populated with the olfactive and opioid receptors (the HIs from 70 to 80%, the KD scale). Thus, the HIs of some of the GPCRs are smaller than the HIs of some 'soluble' proteins such as the TPR-motifs containing FKBPs.[60] In Figure 3 is shown a distribution of the HIs in several sequence modules of the human membrane-bound immunophilin FKBP38[38,65] whose total HI=40% (the KD scale). FKBP38 has a capacity for modulation neural tube patterning by promoting bone morphogenetic protein (BMP, a morphogen) signalization via antagonism of sonic hedgehog (HH, another morphogen) pathway.[66]

The sequence of hFKBP38 may be divided for five small modules, namely hydrophobic FKBD, three tetratricopeptide

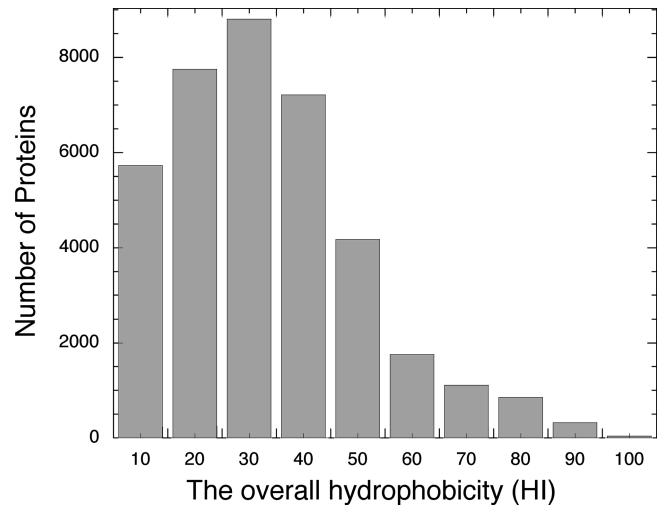

Figure 2. Overall hydrophobicity (HI) of proteins derived from the human genomic database; the KD scale was used with 17 AAs moving frame.

repeat units, and a highly hydrophobic C-terminal TM which is embedded in the membrane.[65] FKBP38 is associated with the mitochondria and interacts with the Bcl-2 protein.[67] Although it has been suggested that FKBP38 interacts with the Ras-like GTPase Rheb and activates the mammalian

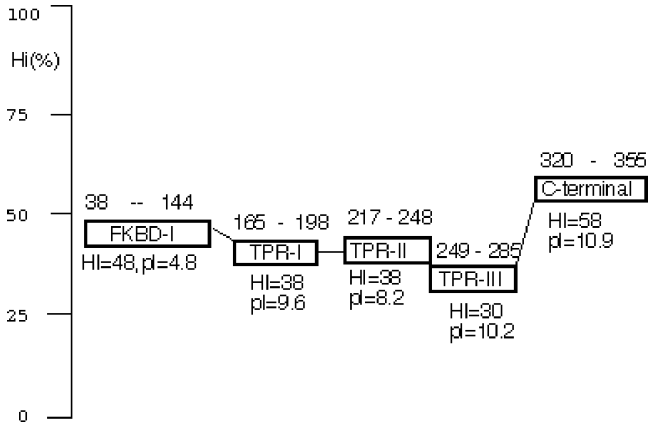

Figure 3. The calculated HIs of several sequence modules of the hFKBP38 (NP_036313); the KD-scale and 9-AAs sliding frame was used.

Table 1. Ten Top Hits in Human Genomic Database Search Using Seven TM Segments from GPCR8 (NP_005277)

no.	code	SH	Seg	IDs/TMs	protein
1	NP_005276	2	7	0.86445	G protein-coupled receptor 7
2	NP_0010085	2	7	0.75638	μ1 opioid receptor, (MOP-R), P35372
3	NP_001044	3	7	0.75975	somatostatin receptor 5
4	NP_000903	1	6	0.79700	opioid receptor, κ1 (KOP-R), P41145
5	NP_000904	1	6	0.76262	opiate receptor-like 1, rat LC132 (KOR-3)
6	NP_000902	2	6	0.75583	δ1 opioid receptor, (DOP-R), P41143
7	NP_001041	4	4	0.78950	somatostatin receptor 2
8	NP_001548	2	4	0.75850	interleukin 8 receptor β
9	NP_000625	2	3	0.76625	interleukin 8 receptor α
10	NP_061822	2	3	0.76200	urotensin 2 receptor

target of rapamycin complex 1 (mTORC1),[68] some controversial points have been raised against any involvement of FKBP38 in the regulation of mTORC1.[69] As shown in Figure 3, FKBP38 is a hydrophobic protein that functions in membrane environment, and its potential *in vivo* targets may bind to several domains, such as PPIase cavity on the hydrophobic FKBD and via three TPR motifs, two of which are hydrophobic.

Clustering of the GPCRs According to the Sequence Attributes of Their TMs. In this work, the sequence attributes of the TMs extracted from the multiple sequence alignments (MSAs) consisting of 150 to 450 diverse GPCRs were compared with those embedded in the hydrophobic segments generated from several genomic databases by the Pro_Hy program. For example, using seven TM-segments taken from the human opioid-somatostatin-like receptor (NP_005277, GPCR8) that binds diverse neuropeptides,[70,71] the Pro_Hy algorithm clustered its TMs with the hydrophobic segments from 563 diverse proteins whose sequence attributes were similar to those of the seven TMs; ten top matches are shown in Table 1 (see the Supporting Information comprising a Pro_Hy output (GB.out) and Blast.out containing a full-sequence search of the human genomic database at the NCBI server using the Blast software[72]). The Pro_Hy established cluster of 10 top hits was ordered according to the decreasing numbers of the TM segments (Seg) that had the IDs, AACs, and α-helical propensity potential higher than the threshold values (17 AAs sliding frame, the ID cutoff 0.65; the F_j function 0.1). SH indicates the number of peaks in the hydrophobicity profile that were higher than 7.5 on the spatial-hydrophobicity scale (the maximal value is about 42.0 whereas the minimum reaches about -30.0). The GPCRs homologous to GPCR8 (listed below) contain somatostatin and opioid receptors.[8] At least three TM segments (Seg) from the Il8β receptor have similar sequence attributes to those of the opioid and somatostatin receptors. This could be due to the fact that these receptors are highly hydrophobic. For example, the peaks of the spatial hydrophobicity for TMs I and VI are higher in the hydrophobicity profiles generated from the sequences of GPCR7 and interleukin-8 receptor β-isoform (IL8Rβ) than the remaining TMs (Figure S1, Supporting Information). Moreover, all of these ten GPCRs have from one to four TMs loaded with bulky hydrophobic AAs increasing their spatial hydrophobicity that may be a crucial packing element for the 7-TMs assembly in the phospholipid bilayer. The sequence attributes of the TMs of the human GPCR8 were compared with those of the hydrophobic segments generated from the genomic databases of diverse organisms leading to a large pack of sequences, most of which were assigned as being the GPCRs or diverse orphan and hypothetical forms of the GPCRs (see the GBT.out file in the Supporting Information).

Do the TPR Motifs of the Immunophilins and the TMs in Some GPCRs Have Similar Structural Features. It has been suggested that some of the sequence attributes of the TPRs in diverse immunophilins and the TMs of some GPCRs have a moderate convergence level.[14] This prompted us to make some additional analyses of several recent crystallographic structures comprising the above sequence motifs, namely the structure of a 42 kDa TPR-containing the FK506-like binding protein from *A. thaliana* (2IF4, AtFKBP42),[30] the X-ray structure of photoactivated bovine rhodopsin (2I37),[20] and the other X-ray structures of diverse GPCRs.[21−29]

The 2D map of 2I37 (Figure 4A) comprises a classical distribution of the seven TM helical segments positioned along the diagonal (dotted yellow arrows) due to short-distance interactions within each of the α-helical TM segments. At the upper part of Figure 4A are shown extensive interactions between the N-terminal domain (green rectangle) and some of the loops of the receptor (shown as violet ellipses above the violet line). Six clusters that are perpendicular and close to the diagonal are due to interactions between α-helical TMs I−VII (dotted blue lines), while other interhelical interactions form either parallel or perpendicular clusters distant to the diagonal (blue dotted lines). Those latter types of clusters involve interactions between helix-II and helix-IV, helix-III and helix-V, helix-III and helix-VI, helix-II and helix-VI, and helix-I with helix-VII. Two other clusters comprise interactions between C-terminal and helix-I (yellow circle) and the loop linking helix-IV with helix-V and helix-VI (see Table S3A, Supporting Information). A quasi-similar distribution of interhelical interaction clusters are shown at the lower part (below the red line in Figure 4B) of the 2D map derived from 2IF4, the three TPR motifs-containing AtFKBP42.[30] The upper left part of the triangle comprises intramolecular interaction clusters in the FKBD that are mainly made of antiparallel β-strands (violet arrows perpendicular to the diagonal). In the lower part of the triangle occur five perpendicular clusters of interhelical interactions and two parallel clusters comprising interactions between helix-I and helix-III and helix-III with helix-V. Three small interaction

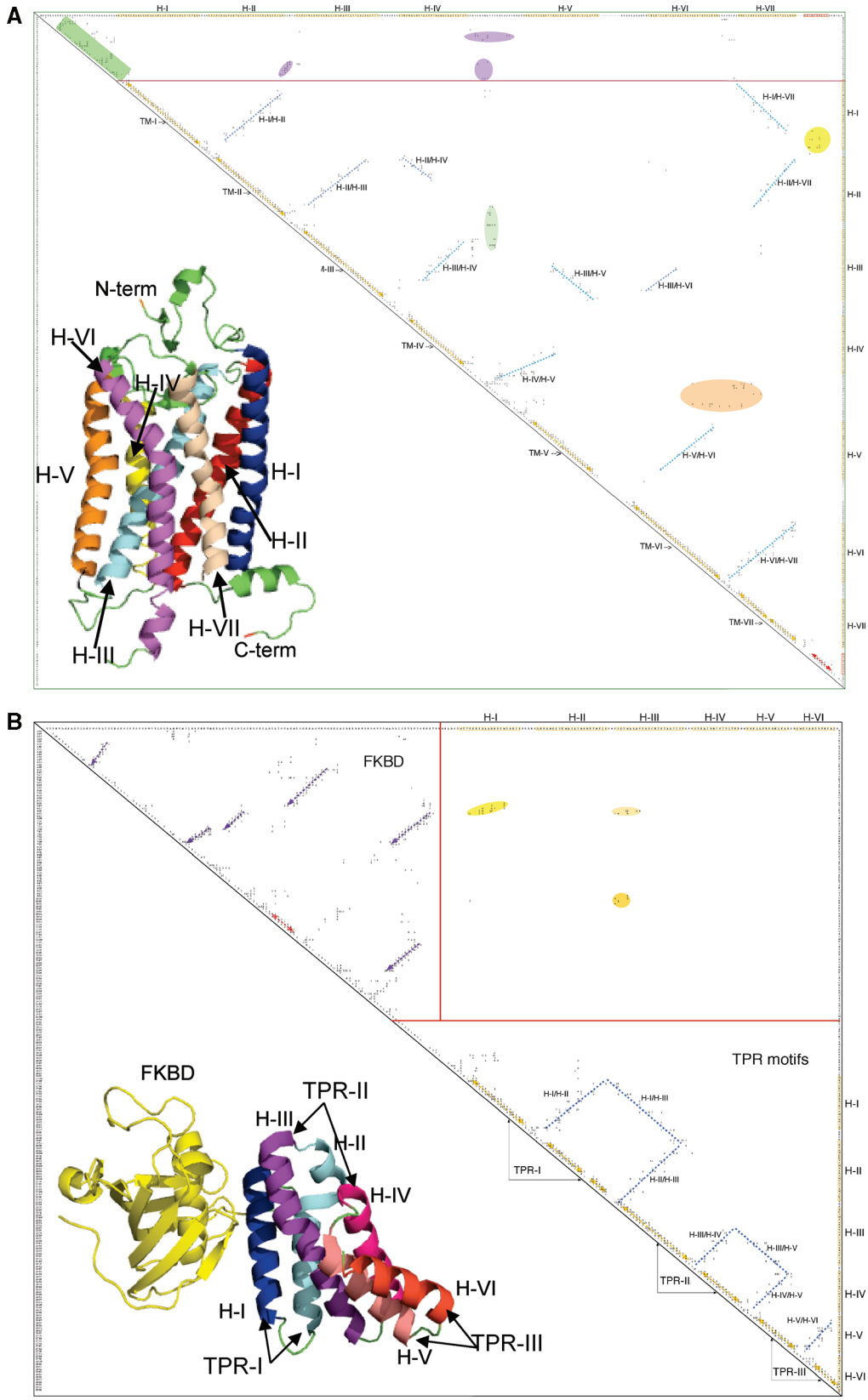

Figure 4. Intramolecular interaction patterns (upper triangles) derived from 2I37 (A)[20] and 2IF4 (B).[30] The upper and right coordinates bear the sequence with the indicated major secondary structures in color boxes. In the lower triangles are shown their respective structures drawn with the PyMol program[43] using the following color coding: **(A)** TM-I (blue); TM-II (red); TM-III (cyan); TM-IV (yellow); TM-V (orange); TM-VI (pink); TM-VII (wheat) and **(B)** FKBD (yellow); TPR-I comprises helix-I (blue) and helix-II (cyan); TPR-II comprises helix-III (violet) and helix-IV (pink); whereas TPR-III comprises helix-V (salmon) and helix-VI (red-brick).

clusters at the upper left corner are due to interactions between the FKBD and TPRs. Statistical distributions of the intramolecular interactions in 2I37 and 2IF4 are as follow: (A) 29.0 and 31.2% between hydrophobic atoms; (B) 19.0 and 16.6% between hydrophilic atoms; (C) 54.0 and 52.2% hydrophobic/hydrophilic atoms; (D) 45.0 and 41.3% (main-

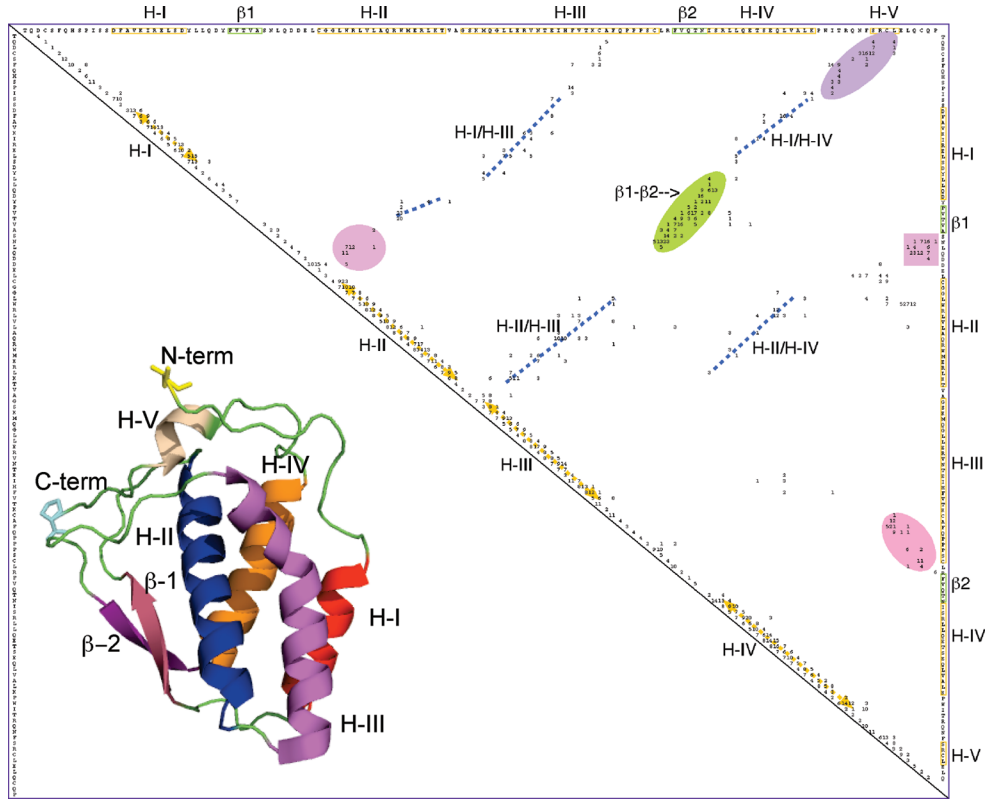

Figure 5. Intramolecular interaction patterns derived from the soluble part of Flt3 (1ETE).[36] H−I (red), H−II (blue) I, H−III (violet) and H−IV (orange), β1 (pink) and β2 (raspberry).

chain)/(main-chain) atoms; (E) 35.6 and 35.2% (main-chain)/(side-chain) atoms; and (F) 19.0 and 23.5% (side-chain/side-chain) atoms, respectively. The above statistics show that, in general, those two structures have similar distributions of the intramolecular interactions with a slight decrease of (side-chain)/(side-chain) interactions in the bovine rhodopsin (2I37). One could argue that in both cases, the clustered α-helical segments with their mutual spatial arrangements impose similar intramolecular interaction patterns.

Several other structures containing helical bundles were investigated, namely the 2D interaction maps were calculated from the X-ray structures of 1ETE, 1M6T, and 3DHZ. For example, on the upper triangle of Figure 5 are shown four α-helical segments (yellow arrows near the diagonal) of the soluble form of the Flt3 ligand.[36] Several transversal inter-helical networks are shown as dotted blue lines. Only helix-II is close to helix-III (perpendicular cluster of interaction close to the diagonal), whereas helix-I (H−I) interacts with helix-III and helix-IV. An antiparallel β-sheet is shown as a green ellipse whose interaction pattern is perpendicularly clustered to the diagonal. The interaction cluster between the N-terminal part and the C-terminus involving a short α-helix (H−V) is indicated as a violet ellipse (upper left corner), whereas scattered interactions between the C-terminus with helix-III (H−III) and a loop linking β1 to helix-II (H−II) are shown as pink clusters. Figures 5, S2A, and S2B displaying 1M6T and 3DZH, respectively (see the Supporting Information) show similar distributions of inter-helical transversal interaction clusters, but the 2D maps consist of other interactions patterns due to the presence of other secondary structures, e.g., in 1ETE and 3DZH.

Comparing the upper part of Figure 4B, that is rich in clusters comprising long-distance interactions, with its lower part containing mainly short- and medium-distance interaction clusters, reveals the stabilizing forces are differently distributed in these two domains. Sparse off-diagonal interaction clusters in the bovine rhodopsin (2I37) and the other GPCRs are mostly due to transversal interhelical nonbonded interaction network between the side chains of the hydrophobic AAs. These interactions change from one structure of the GPCR to another, which may suggest that the internal packing of diverse GPCRs may be subtly affected.[12] Inter-helical interactions between the TMs in different GPCRs always involve about 50% of hydrophobic AAs. Moreover, a high hydrophobicity of some of the TMs is due to the presence of more than 10 consecutive hydrophobic AAs that in part form a transversal hydrophobic interaction network between the TMs (see Table S3A). The remaining hydrophobic side chains should impose a large propensity of the TMs toward phospholipidic bilayers. In aqueous solution such helical assemblies would probably form randomly ordered macroaggregates. Our analyses revealed that the interhelical interactions between the α-helical segments of the TPRs in AtFKBP41 involve also about 50% of non-bonded interactions between hydrophobic atoms. It could signify that the TPRs in the immunophilins utilize a nonbonded interaction force for stabilizing their interhelical transversal interactions (see Table S3B); however, due to their amphipathic character and specific packing modes,[73,74] they remain soluble in aqueous phase and in that way they may be bridging molecules between soluble proteins and those associated with the membranes.

DISCUSSION

The presented method for clustering functionally related proteins is based on the usage of several global and local

sequence attributes and a novel concept of hydrophobicity profile computing. Analyses of several genomic databases have revealed that they contain a relatively large percentage (≥30%) of hydrophobic proteins (the HIs ≥ 40% calculated with the KD scale). Asymmetrical distributions of the hydrophobicity peaks in the sequences of the diverse GPCRs (Figures 1A, 1B, S1A, and S1B) may signify that their oligomerization potentials[75] and allostery[6,12,15,16,76] could be dependent on interhelical interaction modes and packing of the TMs.[12,20−29] Linear forms of the hydrophobicity profiles established from genomic sequences reveal nominal growth and decline of hydrophobic patches along the polypeptide chain. The profiles are dependent on the values assigned to each of the twenty natural amino acids in a given hydrophobicity scale and the method used for their averaging. The knowledge-based hydrophobicity scale presented in this paper was derived from spatial distribution of intramolecular interaction networks that involve the following types of atoms: 1) hydrophobic−hydrophobic, 2) hydrophilic−hydrophilic, and 3) hydrophilic−hydrophobic. It is shown that the hydrophobic atoms of the hydrophilic amino acid residues contribute to networks comprising interactions between the hydrophobic atoms. Likewise, the hydrophilic atoms of the hydrophobic AAs contribute to some extent to networks comprising the hydrophilic atoms. Those diverse interactions are crucial for formation of both 3D structures of proteins and their higher order molecular complexes.[77,78] For example, we showed that α-helical bundles are stabilized by transversal interhelical interactions between the side chains of the hydrophobic AAs in proteins as diverse as GPCRs, TPR-containing immunophilins, and diverse soluble proteins, for example Flt3.[36] Moreover, those transversal interactions are subtly affected by the local sequence contexts and mutual packing of the α-helices.

Analyses of diverse X-ray structures have revealed that even the charged AAs made some contributions to the nonbonded interaction networks involving the hydrophobic AAs. It further illustrates that the local sequence hydrophobicity and its global contexts such as spatial positions of hydrophobic and hydrophilic AAs on the surface layer[79] of a protein are crucial factors that impose its propensity toward aqueous solution or hydrophobic phase. For example, even if some of the cholinergic, adrenergic, and NMDA receptors have the comparable global HIs to those of the TPR-containing immunophilins, the former require the phospholipid bilayer for stabilizing their TMs due to their high hydrophobicity and the lack of amphipathic character. In contrast, the TPRs in the immunophilins or in the Bubr1 protein[80] have amphipathic character and different packing[80] than the TMs in GPCRs. These two attributes make that the TPR-containing immunophilins are soluble in aqueous solution. It has been shown, however, that some of the TPRs may have a sizable potential to interact with different hydrophobic molecules, as it is the case of the TPR motifs in PP5 interacting with the arachidonic acid[81] that controls its enzymatic activity.[82]

The TPRs and TMs display similar interhelical interaction networks that are mainly stabilized by nonbonded interactions between the hydrophobic side chains. The relatively high hydrophobicity levels of the TPRs and similar distributions of the intramolecular interaction clusters on 2D maps generated from the X-ray structures of the TPR motifs-containing immunophilins and the TMs of the GPCRs[12,20−29] may indicate for some structural relatedness of these two α-helical motifs. It could also imply that these two structural motifs might have had a common ancestor such as an α-helical bundles motifs that are present in diverse soluble proteins. Transformation of a TPR into two TMs would require an increase in its hydrophobicity concomitant with an elongation of the β-turn linking the two α-helical segments. Locally restricted sequence attributes such as hydrophobicity level, spatial distribution of hydrophobic side chains (bulkiness profiles), amphipathic character, and α-helical propensity potential may predispose the TPR-containing immunophilins to creating transversal hydrophobicity patches between the soluble entities and membrane proteins.

Abbreviations. AA, amino acid; AAC, amino acid composition; TPR, tetratricopeptide repeat; TM, transmembrane segment; ID, sequence similarity score; HI, overall hydrophobicity index; MSA, multiple sequence alignment; GPCR, G protein-coupled receptor; TFPD, three-fingered protein domain; FKBD, FK506-like domain.

ACKNOWLEDGMENT

I am indebted to SIMOPRO/IBiTec/DSV/CEA for financial support.

Supporting Information Available: (A) Figures S1A and S1B show the hydrophobicity profiles generated with the three scales: KD-scale (red line), triads, (blue line), and spatial hydrophobicity (violet line) for (S1A) human GPCR7 (NP_005276) and (S1B) human IL8βR (NP_001548); 17-AAs sliding frame was used. (B) Figures S2A and S2B display 2D interaction patterns calculated from 1M6T and 3DHZ, respectively. (C) GB.out is an output from comparing sequence attributes of proteins encoded in the human genome and the seven TMs of GPCR8 (NP_005277), whereas (D) Blast.out is the list of proteins found with the Blast software in the NCBI database with the threshold E ≤ 1.0. (E) GPT.out protein clusters extracted from several genomic databases using the same bait as above. (F) Table S1 containing analyses of crystallographic structures. (G) Table S2 contains the ratios of the total bulkiness of 20 natural AAs to the bulkiness of their hydrophobic counterparts. (H) Tables S3A and S3B contain interhelical interaction patterns derived from 2I37 and 2IF4, respectively, whereas Table S3C contains the interaction patterns derived from 3EML and 2Z73. (I) hymapi.zip file contains himapi, a precompiled version for a Macintosh platform of the Hi_Ma_Pi.f program for computing hydrophobicity profiles of proteins with different hydrophobicity scales and several files necessary for its usage, and datpro, a precompiled version of the program converting a cluster of files in Flat GenBank format into a database that can be used as input to the himapi program, whereas prohy.zip file contains a precompiled version of the Pro_Hy program accompanied with several input files and a short manual. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) von Heijne, G. Membrane-protein topology. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 909–918.

(2) Liu, Y.; Engelman, D. M.; Gerstein, M. Genomic analysis of membrane protein families: abundance and conserved motifs. *Genome Biol.* **2002**, *3*, 1–12.
(3) White, S. H.; Wimley, W. C. Membrane protein folding and stability: physical principles. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 319–365.
(4) Wolfenden, R. Experimental measures of amino acid hydrophobicity and the topology of transmembrane and globular proteins. *J. Gen. Physiol.* **2007**, *129*, 357–362.
(5) Müller, D. J.; Wu, N.; Palczewski, K. Vertebrate membrane proteins: structure, function, and insights from biophysical approaches. *Pharmacol. Rev.* **2008**, *60*, 43–78.
(6) Westfall, T. C.; Westfall, D. P. Neurotransmission: the autonomic and somatic motor nervous systems. In *Goodman & Gilman's: the Pharmacology Basis of Therapeutics*, 11th ed.; Brunton, L., Lazo, J., Parker, K., Eds.; McGraw-Hill, Medical Publishing Division: New York, 2005; pp 137−181.
(7) Nichols, D. E. Hallucinogens. *Pharmacol. Ther.* **2004**, *101*, 131–181.
(8) Waldhoer, M.; Bartlett, S. E.; Whistler, J. L. Opioid receptors. *Annu. Rev. Biochem.* **2004**, *73*, 953–990.
(9) Charney, D. S.; Mihic, S. J.; Harris, R. A. Hypnotics and sedatives . *In Goodman & Gilman's: the Pharmacology Basis of Therapeutics*, 11th ed.; Brunton, L., Lazo, J., Parker, K., Eds.; McGraw-Hill Medical Publishing Division: New York, 2005; pp 401−427.
(10) Jensen, R. T.; Battey, J. F.; Spindel, E. R.; Benya, R. V. Mammalian bombesin receptors: nomenclature, distribution, pharmacology, signaling, and functions in normal and disease states. *Pharmacol. Rev.* **2008**, *60*, 1–42.
(11) Pennefather, J. C.; Lecci, A.; Candenas, M. L.; Patak, E.; Pinto, F. M.; Maggi, C. A. Tachynins and tachynin receptors: a growing family. *Life Sci.* **2004**, *74*, 1445–1463.
(12) Mustafi, D.; Palczewski, K. Topology of class A G protein-coupled receptors: insights gained from crystal structures of rhodopsins, adrenergic and adenosine receptors. *Mol. Pharmacol.* **2009**, *75*, 1–12.
(13) Mertens, I.; Vandingene, A.; Meeusen, T.; De Loof, A.; Schoofs, L. Postgenomic characterization of G-protein-coupled recepters. *Pharmacogenomics* **2004**, *5*, 1–16.
(14) Galat, A. Involvment of some large immunophilins and their ligands in the protection and regeneration of neurons: a hypothetical mode of action. *Comp. Biol. Chem.* **2006**, *30*, 348–359.
(15) Kuriyan, J.; Eisenberg, D. The origin of protein interactions and allostery in colocalization. *Nature* **2007**, *450*, 983–990.
(16) Goodey, N. M.; Benkovic, S. J. Allosteric regulation and catalysis emerge via a common route. *Nat. Chem. Biol.* **2008**, *4*, 474–482.
(17) Karhadkar, S. S.; Bova, G. S.; Abdallah, N.; Dhara, S.; Gardner, D.; Maitra, A.; Isaacs, J. T.; Berman, D. M.; Beachy, P. A. Hedgehog signalling in prostate regeneration, neoplasia and metastasis. *Nature* **2004**, *431*, 707–712.
(18) Ogden, S. K.; Fei, D. L.; Schilling, N. S.; Ahmed, Y. F.; Hwa, J.; Robbins, D. J. G protein $G_{\alpha i}$ functions immediately downstream of smoothened in hedgehog signalling. *Nature* **2008**, *456*, 967–970.
(19) Zhao, Y.; Tong, C.; Jiang, J. Hedgehog regulates smoothened activity by inducing a conformational switch. *Nature* **2007**, *450*, 252–258.
(20) Palczewski, K.; Kumasaka, T.; Hiori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M. Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* **2000**, *289*, 739–745.
(21) Scheerer, P.; Park, J. H.; Hildebrand, P. W.; Kim, Y. J.; Krauss, N.; Choe, H. W.; Hofmann, K. P.; Ernst, O. P. Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **2008**, *455*, 497–502.
(22) Park, J. H.; Scheerer, P.; Hofmann, K. P.; Choe, H. W.; Ernst, O. P. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* **2008**, *454*, 183–187.
(23) Murakami, M.; Kouyama, T. Crystal structure of squid rhodopsin. *Nature* **2008**, *453*, 363–368.
(24) Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. G.; Thian, F. S.; Kobilka, T. S.; Choi, H. J.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; Stevens, R. C. High-resolution crystal structure of an engineered human β_2-adrenergic G protein-coupled receptor. *Science* **2007**, *318*, 1258–1265.
(25) Rosenbaum, D. M.; Cherezov, V.; Hanson, M. A.; Rasmussen, S. G.; Thian, F. S.; Kobilka, T. S.; Choi, H. J.; Yao, X. J.; Weis, W. I.; Stevens, R. C.; Kobilka, B. K. GPCR engineering yields high-resolution structural insights into β_2-adrenergic receptor function. *Science* **2007**, *318*, 1266–1273.
(26) Rasmussen, S. G.; Choi, H. J.; Rosenbaum, D. M.; Kobilka, T. S.; Thian, F. S.; Edwards, P. C.; Burghammer, M.; Ratnala, V. R.; Sanishvili, R.; Fischetti, R. F.; Schertler, G. F.; Weis, W. I.; Kobilka, B. K. Crystal structure of the human β_2 adrenergic G-protein-coupled receptor. *Nature* **2007**, *450*, 383–387.
(27) Warne, T.; Serrano-Vega, M. J.; Baker, J. G.; Moukhametzianov, R.; Edwards, P. C.; Henderson, R.; Leslie, A. G.; Tate, C. G.; Schertler, G. F. Structure of a β_1-adrenergic G-protein-coupled receptor. *Nature* **2008**, *454*, 486–491.
(28) Hanson, M. A.; Cherezov, V.; Griffith, M. T.; Roth, C. B.; Jaakola, V. P.; Chien, E. Y.; Velasquez, J.; Kuhn, P.; Stevens, R. C. A specific cholesterol binding site is established by the 2.8 Å structure of the human β_2-adrenergic receptor. *Structure* **2008**, *16*, 897–905.
(29) Jaakola, V. P.; Griffith, M. T.; Hanson, M. A.; Cherezov, V.; Chien, E. Y.; Lane, J. R.; Ijzerman, A. P.; Stevens, R. C. The 2.6 Å crystal structure of a human A_{2A} adenosine receptor bound to an antagonist. *Science* **2008**, *322*, 1211–1217.
(30) Granzin, J.; Eckhoff, A.; Weiergraber, O. H. Crystal structure of a multidomain immunophilin from *Arabidopsis thaliana*: a paradigm for regulation of plant ABC transporters. *J. Mol. Biol.* **2006**, *364*, 799–809.
(31) Taylor, P.; Dornan, J.; Carrello, A.; Minchin, R. F.; Ratajczak, T.; Walkinshaw, M. D. Two structures of cyclophilin 40: folding and fidelity of the TPR domains. *Structure* **2001**, *9*, 431–438.
(32) Wu, B.; Li, R.; Liu, Y.; Lou, Z.; Ding, Y.; Shu, C.; Ye, S.; Bartlam, M.; Shen, B.; Rao, Z. 3D structure of human FK506-binding protein 52: implications for the assembly of the glucocorticoid receptor/Hsp90/immunophilin heterocomplex. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 8348–8353.
(33) Sinars, C.; Cheung-Flynn, J.; Rimerman, R. A.; Scammell, J. G.; Smith, D. F.; Clardy, J. Structure of the large FK506-binding protein FKBP51, an Hsp90-binding protein and a component of steroid receptor complex. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 868–873.
(34) Chu, R.; Takei, J.; Knowlton, J. R.; Andrykovitch, M.; Pei, W.; Kajava, A. V.; Steinbach, P. J.; Ji, X.; Bai, Y. Redesign of a four-helix bundle protein by phage display coupled with proteolysis and structural characterization by NMR and X-ray crystallography. *J. Mol. Biol.* **2002**, *323*, 253–262.
(35) Liu, Q.; Kriksunov, I. A.; Jiang, H.; Graeff, R.; Lin, H.; Lee, H. C.; Hao, Q. Covalent and noncovalent intermediates of an NAD utilizing enzyme, human CD38. *Chem. Biol.* **2008**, *15*, 1068–1078.
(36) Savvides, S. N.; Boone, T.; Karplus, P. A. Flt3 ligand structure and unexpected commonalities of helical bundles and cystine knots. *Nat. Struct. Biol.* **2000**, *7*, 486–491.
(37) Galat, A.; Gross, G.; Drevet, P.; Sato, A.; Ménez, A. Conserved structural determinants in three-fingered protein domains. *FEBS J.* **2008**, *275*, 3207–3225.
(38) Lam, E.; Martin, M.; Wiederrecht, G. Isolation of a cDNA encoding a novel human FK506-binding protein homolog containing leucine zipper and tetratricopeptide repeat motifs. *Gene* **1995**, *160*, 297–302.
(39) Ma, Q.; Whitlock, J. P. A novel cytoplasmic protein that interacts with the Ah receptor contains tetratricopeptide repeat motifs, and augments the transcriptional response to 2,3,7,8-tetrachlorodibenzo-p-dioxin. *J. Biol. Chem.* **1997**, *272*, 8878–8884.
(40) Wheeler, D. L.; Church, D. M.; Federhen, S.; Lash, A. E.; Madden, T. L.; Pontius, J. U.; Schuler, G. D.; Schriml, L. M.; Sequeira, E.; Tatusova, T. A.; Wagner, L. Database resources of National Center for Biotechnology. *Nucleic Acids Res.* **2003**, *31*, 28–33.
(41) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680.
(42) Berman, H. M.; Henrick, K.; Nakamura, H.; Markley, J. L. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.* **2007**, *35*, D301−D303.
(43) Galat, A. Functional drift of sequence attributes in the FK506-binding proteins (FKBPs). *J. Chem. Inf. Model.* **2008**, *48*, 1118–1130.
(44) DeLano, W. L. *The PyMOL Molecular Graphics System*; DeLano Scientific: San Carlos, CA, U.S.A., 2002. www://pymol.sourceforge.net/ (accessed Jan 12, 2009).
(45) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
(46) Kyte, J.; Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–132.
(47) Hopp, T. P.; Woods, K. R. Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 3824–3828.
(48) Cornette, J. L.; Cease, K. B.; Margalit, H.; Spouge, J. L.; Berzofsky, J. A.; DeLisi, C. Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. *J. Mol. Biol.* **1987**, *195*, 659–685.
(49) Wimley, W. C.; White, S. H. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Biol.* **1996**, *3*, 842–848.
(50) Hessa, T.; Kim, H.; Bihlmaier, K.; Lundin, C.; Boekel, J.; Andersson, H.; Nilsson, I.; White, S. H.; von Heijne, G. Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **2005**, *433*, 377–381.

(51) Koehler, J.; Woetzel, N.; Staritzbichler, R.; Sanders, C. R.; Meiler, J. A unified hydrophobicity scale for multispan membrane proteins. *Proteins* **2009**, *76*, 13–29.
(52) Tsai, J.; Taylor, R.; Chothia, C.; Gerstein, M. The packing density in proteins: standard radii and volumes. *J. Mol. Biol.* **1999**, *290*, 253–266.
(53) Gekko, K.; Idota, Y. Amino acid solubility and protein stability in aqueous maltitol solutions. *Agric. Biol. Chem.* **1989**, *53*, 89–95.
(54) Amend, J. P.; Helgeson, H. C. Solubilities of the common L-α-amino acids as a function of temperature and solution pH. *Pure Appl. Chem.* **1997**, *69*, 935–942.
(55) Gibbs, P. R.; Radzicka, A.; Wolfenden, R. The anomalous hydrophilic character of proline. *J. Am. Chem. Soc.* **1991**, *113*, 4714–4715.
(56) Nozaki, Y.; Tanford, C. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. Establishment of a hydrophobicity scale. *J. Biol. Chem.* **1971**, *246*, 2211–2217.
(57) Wolfenden, R.; Andersson, L.; Cullis, P. M.; Southgate, C. C. Affinities of amino acid side chains for solvent water. *Biochemistry* **1981**, *20*, 849–855.
(58) Jayasinghe, S.; Hristova, K.; White, S. H. Energetics, stability, and prediction of transmembrane helices. *J. Mol. Biol.* **2001**, *312*, 927–934.
(59) Eisenberg, D.; Weiss, R. M.; Terwilliger, T. C. The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 140–144.
(60) Galat, A. A note on clustering the functionally-related paralogues and orthologues of proteins: a case of the FK506-binding proteins (FKBPs). *Comp. Biol. Chem.* **2004**, *28*, 129–140.
(61) Galat, A. Function-dependent clustering of orthologues and paralogues of cyclophilins. *Proteins* **2004**, *56*, 808–820.
(62) Galat, A. The three-fingered protein domain of the human genome. *Cell. Mol. Life Sci.* **2008**, *65*, 3481–3493.
(63) Karplus, P. A. Hydrophobicity regained. *Protein Sci.* **1997**, *6*, 1302–1307.
(64) Brylinski, M.; Konieczny, L.; Roterman, I. Hydrophobic collapse in (in silico) protein folding. *Comput. Biol. Chem.* **2006**, *30*, 255–267.
(65) Nakagawa, T.; Shirane, M.; Iemura, S.; Natsume, T.; Nakayama, K. I. Anchoring of the 26S proteasome to the organellar membrane by FKBP38. *Genes Cells* **2007**, *12*, 709–719.
(66) Cho, A.; Ko, H. W.; Eggenschwiler, J. T. FKBP8 cell-autonomously controls neural tube patterning through a Gli2- and Kif3a-dependent mechanism. *Dev. Biol.* **2008**, *321*, 27–39.
(67) Shirane, M.; Nakayama, K. I. Inherent calcineurin inhibitor FKBP38 targets Bcl-2 to mitochondria and inhibits apoptosis. *Nat. Cell Biol.* **2003**, *5*, 28–37.
(68) Ma, D.; Bai, X.; Guo, S.; Jiang, Y. The switch I region of Rheb is critical for its interaction with FKBP38. *J. Biol. Chem.* **2008**, *283*, 25963–25970.
(69) Wang, X.; Fonseca, B. D.; Tang, H.; Liu, R.; Elia, A.; Clemens, M. J.; Bommer, U. A.; Proud, C. G. Re-evaluating the roles of proposed modulators of mammalian target of rapamycin complex 1 (mTORC1) signaling. *J. Biol. Chem.* **2008**, *283*, 30482–30492.
(70) Mazzocchi, G.; Rebuffat, P.; Ziolkowska, A.; Rossi, G. P.; Malendowicz, L. K.; Nussdorfer, G. G. G protein receptors 7 and 8 are expressed in human adrenocortical cells, and their endogenous ligands neuropeptides B and W enhance cortisol secretion by activating adenylate cyclase- and phospholipase C-dependent signaling cascades. *J. Clin. Endocrinol. Metab.* **2005**, *90*, 3466–3471.
(71) Tanaka, H.; Yoshida, T.; Miyamoto, N.; Motoike, T.; Kurosu, H.; Shibata, K.; Yamanaka, A.; Williams, S. C.; Richardson, J. A.; Tsujino, N.; Garry, M. G.; Lerner, M. R.; King, D. S.; O'Dowd, B. F.; Sakurai, T.; Yanagisawa, M. Characterization of a family of endogenous neuropeptide ligands for the G protein-coupled receptors GPR7 and GPR8. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 6251–6256.
(72) Altschul, S. F.; Wootton, J. C.; Gertz, E. M.; Agarwala, R.; Morgulis, A.; Schäffer, A. A.; Yu, Y. K. Protein database searches using compositionally adjusted substitution matrices. *FEBS J.* **2005**, *272*, 5101–5109.
(73) Wilson, C. G.; Kajander, T.; Regan, L. The crystal structure of NlpI. A prokaryotic tetratricopeptide repeat protein with a globular fold. *FEBS J.* **2005**, *272*, 166–179.
(74) Kajander, T.; Cortajarena, A. L.; Mochrie, S.; Regan, L. Structure and stability of designed TPR protein superhelices: unusual crystal packing and implications for natural TPR proteins. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2007**, *63* (Pt 7), 800–811.
(75) Szidonya, L.; Cserzo, M.; Hunyady, L. Dimerization and oligomerization of G-protein-coupled receptors: debated structures with established and emerging functions. *J. Endocrinol.* **2008**, *196*, 435–453.
(76) May, L. T.; Leach, K.; Sexton, P. M.; Christopoulos, A. Allosteric modulation of G protein-coupled receptors. *Annu. Rev. Pharmacol. Toxicol.* **2007**, *47*, 1–51.
(77) Bolen, D. W.; Rose, G. D. Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annu. Rev. Biochem.* **2008**, *77*, 339–362.
(78) Voet, D.; Voet, J. G. *Biochemistry*, 2nd edition; John Wiley & Sons: New York, 1995; pp 141−214.
(79) Shaytan, A. K.; Shaitan, K. V.; Khokhlov, A. R. Solvent accessible surface area of amino acid residues in globular proteins: correlation of apparebt transfer free energies with the experimental hydrophobicity scales. *Biomacromolecules* **2009**, *10*, 1224–1237.
(80) Beaufils, S.; Grossmann, J. G.; Renault, A.; Bolanos-Garcia, V. M. Characterization of the tetratricopeptide-containing domain of BUB1, BUBR1, and PP5 proves that domain amphiphilicity over amino acid sequence specificity governs protein adsorption and interfacial activity. *J. Phys. Chem. B* **2008**, *112*, 7984–7991.
(81) Yang, J.; Roe, S. M.; Cliff, M. J.; Williams, M. A.; Ladbury, J. E.; Cohen, P. T.; Barford, D. Molecular basis for TPR domain-mediated regulation of protein phosphatase 5. *EMBO J.* **2005**, *24*, 1–10.
(82) Sinclair, C.; Borchers, C.; Parker, C.; Tomer, K.; Charbonneau, H.; Rossie, S. The tetratricopeptide repeat domain and a C-terminal region control the activity of Ser/Thr protein phosphatase 5. *J. Biol. Chem.* **1999**, *274*, 23666–23672.

CI9001316