

Research article

Open Access

## Exploring allosteric coupling in the $\alpha$ -subunit of Heterotrimeric G proteins using evolutionary and ensemble-based approaches

Kemal Sayar<sup>1,2</sup>, Özlem Uğur<sup>1</sup>, Tong Liu<sup>3</sup>, Vincent J Hilser<sup>3</sup> and Ongun Onaran\*<sup>1,2</sup>

Address: <sup>1</sup>Ankara University Faculty of Medicine, Department of Pharmacology and Clinical Pharmacology, Sıhhiye 06100, Ankara, Turkey, <sup>2</sup>Ankara University Faculty of Medicine, and Molecular Biology and Technology Research and Development Unit, Sıhhiye 06100, Ankara, Turkey and <sup>3</sup>Department of Biochemistry and Molecular Biology, and Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, Galveston, TX, 77555-1068 USA

Email: Kemal Sayar - kemalsayar@yahoo.com; Özlem Uğur - mugur@dialup.ankara.edu.tr; Tong Liu - toliu@ucsd.edu; Vincent J Hilser - vjhilser@utmb.edu; Ongun Onaran\* - onaran@dialup.ankara.edu.tr

\* Corresponding author

Published: 2 May 2008

Received: 4 December 2007

BMC Structural Biology 2008, 8:23 doi:10.1186/1472-6807-8-23

Accepted: 2 May 2008

This article is available from: <http://www.biomedcentral.com/1472-6807/8/23>

© 2008 Sayar et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Allosteric coupling, which can be defined as propagation of a perturbation at one region of the protein molecule (such as ligand binding) to distant sites in the same molecule, constitutes the most general mechanism of regulation of protein function. However, unlike molecular details of ligand binding, structural elements involved in allosteric effects are difficult to diagnose. Here, we identified allosteric linkages in the  $\alpha$ -subunits of heterotrimeric G proteins, which were evolved to transmit membrane receptor signals by allosteric mechanisms, by using two different approaches that utilize fundamentally different and independent information.

**Results:** We analyzed: 1) correlated mutations in the family of G protein  $\alpha$ -subunits, and 2) cooperativity of the native state ensemble of the  $G\alpha_i1$  or transducin. The combination of these approaches not only recovered already-known details such as the switch regions that change conformation upon nucleotide exchange, and those regions that are involved in receptor, effector or  $G\beta\gamma$  interactions (indicating that the predictions of the analyses can be viewed with a measure of confidence), but also predicted new sites that are potentially involved in allosteric communication in the  $G\alpha$  protein. A summary of the new sites found in the present analysis, which were not apparent in crystallographic data, is given along with known functional and structural information. Implications of the results are discussed.

**Conclusion:** A set of residues and/or structural elements that are potentially involved in allosteric communication in  $G\alpha$  is presented. This information can be used as a guide to structural, spectroscopic, mutational, and theoretical studies on the allosteric network in  $G\alpha$  proteins, which will provide a better understanding of G protein-mediated signal transduction.

### Background

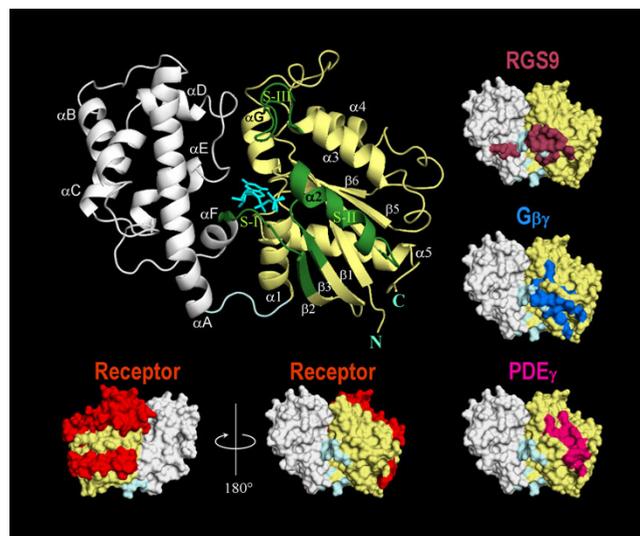
G proteins and G protein-coupled receptors (GPCR) constitute a large family of signaling proteins that transmit

extracellular signals to the intracellular milieu where the signals are integrated and transformed to a variety of biological responses. The receptor is activated by the binding

of agonists, which are neurotransmitters, hormones, autoids, odorants, taste or drug molecules present in the extracellular environment. The receptor then activates its cognate heterotrimeric G protein, which in turn transmits the signals to intracellular effectors, such as second-messenger generating enzymes or ion channels. The GDP-bound  $G\alpha$  subunit complexed with tightly bound  $G\beta\gamma$  (i.e. the heterotrimer) is generally considered as the inactive state of the G protein. The agonist-activated receptor catalyses the release of bound-GDP from  $G\alpha$  and leads to the binding of GTP to the nucleotide binding site of  $G\alpha$  (nucleotide exchange). The conformational change in  $G\alpha$  that occurs upon nucleotide exchange results in dissociation of GTP- $G\alpha$  from the  $G\beta\gamma$  subunit. Dissociated subunits interact with downstream effectors and modulate their activity. Hydrolysis of bound-GTP to GDP by the intrinsic GTPase of  $G\alpha$  and subsequent association of GDP- $G\alpha$  with  $G\beta\gamma$  restore the inactive heterotrimer (see [1] for review). Accordingly, the strength of the signal transmitted to the cellular effectors is determined by the detailed kinetics and energetics of this activation-inactivation cycle. Besides the receptor, signaling proteins such as RGS (regulator of G protein signaling) [2] and AGS (activator of G protein signaling) [3] proteins are also capable of regulating or modulating G protein activation in a non-trivial manner by interfering with the GTPase activity and/or nucleotide exchange rates of  $G\alpha$ , or by modifying binding of the partner proteins, namely  $G\alpha$ ,  $G\beta\gamma$  and receptor, to each other.

High resolution crystal structures of different  $G\alpha$  subunits [4-8] have revealed that  $G\alpha$  consists of two main folding domains: A Ras-like GTPase domain, which is conserved in the superfamily of GTPases, with extended amino and carboxyl terminal helices, and a helical domain folded into an orthogonal bundle of six  $\alpha$  helices which is unique to heterotrimeric G proteins. A deep guanine nucleotide binding pocket resides between the two domains that are connected with two linkers. Three flexible regions in the GTPase domain that change conformation upon nucleotide exchange have been identified and designated as switch I, II, and III (see Figure 1 for a summary). Most of the binding sites for signaling partners of  $G\alpha$  have been found or anticipated to be on the GTPase domain, although some of them, such as the one for RGS14 [9] or RGS9 [10], extend to the helical domain as well (see Figure 1). The interdomain interface is thought to be involved in guanine nucleotide exchange and heterotrimer dissociation (i.e. G protein activation) [11].

From a molecular point of view, regulation of G protein-mediated signal transduction relies on a complex interplay among different functional and structural domains of  $G\alpha$ ; a perturbation at one functional domain modifies the structural, energetic and eventually functional proper-



**Figure 1**  
**Main structural elements of  $G\alpha$  subunit and its interaction sites with its partners.** Ribbon representation of transducin ( $G\alpha t$ ) is shown in the centre as a prototypical example of  $G\alpha$  subunits. Ras-like GTPase and helical domains are shown as pale yellow and white, respectively. Three switch regions, indicated as S-I to S-III are colored green. Bound-nucleotide (GTP $\gamma$ S) is shown as cyan sticks in the centre of the molecule. The nomenclature for secondary structure is also indicated on the picture. The long N-terminal helix (26 residues) is missing in the structure. Pictures on the sides are surface representation of the molecule with the same orientation as the ribbon representation. Contact sites with  $G\beta\gamma$  [15], RGS9 and phosphodiesterase (PDE $\gamma$ ) [10], and sites involved in receptor recognition and interaction [1, 41, 42] are shown with different colors as indicated in the picture. All the pictures were rendered by using PyMol (DeLano Scientific LLC, South San Francisco, California, USA) with the deposited coordinates in Protein Data Bank (PDB code [1TAG-chainA](#)).

ties of the others. Although some of these interactions are mediated by direct contacts, such as the one that takes place between RGS or AGS and nucleotide binding sites [9], most are mediated allosterically. High resolution crystal structures of  $G\alpha i$  [12] and  $G\alpha t$  [13], as well as the heterotrimers [14,15], in their GDP- or GTP $\gamma$ S- (a nonhydrolysable analogue of GTP) bound forms have provided remarkable insight into the understanding of guanine nucleotide-dependent conformational changes in  $G\alpha$  and the mechanism of GTPase. However, the realization that conformational fluctuations may play an important role in mediating allosteric coupling [16,17] casts doubt on whether a complete picture of allosteric communication can emerge from static endpoint pictures provided by crystallographic studies. Indeed, despite the wealth of available data provided by biochemical, mutational, and structural studies, many questions regarding

the structural elements involved in allosteric communication remain to be answered. This is due to the fact that systematic screening of allosteric linkage, for example by using double-mutant-cycle strategies [18], is generally an intractable (if not impossible) experimental task, even for relatively small proteins.

In order to overcome these difficulties, at least in part, and to provide a structural map of a possible network of allosteric linkage in  $G\alpha$ , we adopted a strategy that utilizes two different computational approaches, which are based on independent and diverse assumptions and principles. We evaluated; 1) evolutionary data to draw information about statistical coupling between residues, and 2) a statistical thermodynamic model (i.e. COREX algorithm) of the native state ensemble of  $G\alpha$  to identify linkages in local folding free energies in  $G\alpha_i$  and  $G\alpha_t$ . We used currently available experimental information and the convergence of the results of the two independent analyses as a basis of judgment for the present results.

We show that the two approaches together reveal already known facts, such as the linkage between guanine nucleotide, receptor and effector binding sites, indicating that the predictions of the analyses can be viewed with a measure of confidence. Therefore, we present a set of residues and/or structural elements that may have potential functional importance for the allosteric communication in  $G\alpha$ , and that can be used as a guide to structural, spectroscopic, mutational, and theoretical studies on the allosteric network in  $G\alpha$  proteins, as well as to the interpretation of a vast amount of available experimental data.

## Methods

### Dependence of amino acid distributions in multiple sequence alignment

Correlated mutation analyses in multiple sequence alignments have long been considered as useful tools to identify sites involved in protein-protein contacts [19], or to obtain distance information for structural prediction and fold recognition [20,21]. However, Horowitz et al. [22] have observed that coordinated changes in amino acids at two different positions (i.e. compensatory mutations) in an evolutionarily sampled sequence could predict non-additivity in the corresponding double mutant cycle, which is indicative for allosteric coupling. Lockless and Ranganathan [23] later expanded on this idea and suggested that covariance information in a multiple sequence alignment (MSA) could also be used to identify long-range energetic couplings (i.e. allosteric linkage) in proteins. The reasoning was as follows: if proteins have evolved in such a way that a perturbation at one site can be allosterically transmitted to a distant site, then an evolutionary perturbation imparted by a random mutation at

one of these sites should affect the evolution of the other. This should result in a dependence of amino acid distributions at the relevant sites, which could be uncovered by covariance analyses in a well sampled MSA. Although this is equivalent to saying that distant sites in a protein may have been co-evolved under common functional or structural constraints, the subtle interpretation that such associations might indicate allosteric linkage between the sites has been proved to be useful [24]. However, the free energy-like measure that has originally been suggested by Lockless and Ranganathan for this statistical coupling has potentially serious disadvantages: It is based on a virtual perturbation experiment that yields an asymmetric measure of coupling, and it is extremely biased by the conservation of the sites. The bias is due to the properties of binomial probabilities that play a central role in the measure (see supplemental information in the Additional file 1 for detailed discussion). In general, the bias is so extensive that the measure gauges conservation rather than statistical coupling. Other disadvantages have also been discussed by others [25]. In the present analysis, we therefore used a modified measure of dependence as explained below.

We assessed the dependence of amino acid distributions by means of  $\chi^2$  value associated with relevant contingency tables. The tables were constructed as usual. The magnitude of  $\chi^2$  gauges the degree of departure from independence assumption. However, the absolute magnitude of  $\chi^2$  depends on the actual amino acid content of the sites, which can obviously be different for every pair of sites in a protein sequence, and therefore requires normalization in order to be comparable across different pairs of sites. We used the following normalization:

$${}_N\chi_{ij}^2 = 2 \frac{\chi_{ij}^2}{(\chi_{ii}^2 + \chi_{jj}^2)} \begin{cases} i = j \Rightarrow & {}_N\chi_{ij}^2 = 1 \\ i \neq j \Rightarrow & 0 \leq {}_N\chi_{ij}^2 < 1 \end{cases}$$

The normalizing factor is the average "self-coupling" of each site, which gives the upper limit of  $\chi^2$  for that site. For all possible pair of sites in the MSA,  ${}_N\chi_{ij}^2$  forms a correlation-like symmetric matrix with diagonal elements equal to 1. As a further refinement we made the following correction: we calculated average  ${}_N\chi_{ij}^2$  for each pair of sites by randomly permuting the amino acids at each site in the pair (this procedure affects only the joint distribution without changing the marginals). We considered this

value as the "background"  ${}_N\chi_{ij}^2$  for the sites  $i$  and  $j$ . For all possible pairs in the sequence, these values form a matrix of the same size as  ${}_N\chi^2$ . We used the difference between  ${}_N\chi^2$  and the background matrix as a final measure of dependence (or statistical coupling), which can be interpreted as the extra  ${}_N\chi^2$  due to the actual arrangement of amino acids in the MSA columns. Note that  $\chi^2$  in this procedure was not used in its usual statistical sense, but simply as a measure of normalized "distance" between two joint distributions. Obviously, physicochemical properties of amino acids were not considered in this analysis. A comparable strategy has been undertaken by Kass and Horovitz [26]. Finally, possible contributions of evolutionary noise to the covariance information [27] were disregarded in the present analysis, since the results of COREX analysis (as explained below) were used as a source of independent information to judge the relevance of covariance data. The entire procedure is available upon request as a Matlab script.

For the present analysis, we used a non-redundant multiple sequence alignment consisting of 233 samples of  $\alpha$ -subunits of heterotrimeric G protein sequences, obtained from the pfam database (please see Availability & requirements section below), from which redundancies were eliminated manually (the entire data set is available upon request). We excluded those sites for which the available number of amino acid samples was less than 90% of the size of the MSA. Highly conserved sites were also disregarded in the analysis since the covariance with or between such sites is statistically ambiguous as they possess little or no variance.

In order to identify the network of alike sites with respect to their coupling strength and pattern, we did a two-way hierarchical cluster analysis with the corrected  ${}_N\chi^2$  matrix after eliminating the sites that did not fulfill the sampling criterion mentioned above. We used complete linkage function and city-block distance metrics (sum of absolute differences) as the simplest and most natural choice [28]. We tested the efficiency of the entire procedure by using simulated multiple sequence alignments, in which different types of couplings with different strengths were imposed. Simulated patterns were perfectly recov-

ered by the above procedure, whereas the original measure of Lockless and Ranganathan failed [see Additional file 1].

#### Simulation of statistical ensemble

In thermodynamic equilibrium, the folded state of a protein comprises an ensemble of energetically close conformational states [29]. Each state in this ensemble emerges with a certain probability depending on its free energy. Equivalently, this ensemble can be viewed as the "state repertoire" of a single molecule that is continuously explored by the molecule with different mean dwell-times, the latter being eventually related to the state probabilities when the molecule is observed in the time scale of thermodynamic averaging. In any case, the native state of a protein for a given set of conditions can be specified by a probability distribution defined over this conformational space [30]. It can be theoretically shown that modification of this probability distribution by different perturbations made at distant sites on a protein results in energetic coupling of the perturbations when they are made simultaneously (e.g. coupling between two ligand binding processes at two distant binding sites) [31,32], which constitutes the microscopic basis of two equivalent linkage theories that have been formulated differently by Wyman [33] and Weber [34]. Thus, allosteric linkage between distant sites on a protein molecule can be analyzed by using such probability distributions [17,35], provided that the equilibrium probabilities associated with specific conformers in the native ensemble are available.

In a series of studies, Hilser, Freire and their co-workers have developed an effective strategy to model the native ensemble of folded states of proteins [30,36-38]. The procedure, known as the COREX algorithm, can be summarized as follows. The native ensemble of the protein is modeled as a collection of partially folded states of the protein: Each residue is considered as either folded (native-like as in the high-resolution structure) or unfolded (devoid of structure), and folding blocks of consecutive (usually 6 to 12) residues, in which residues are collectively folded or unfolded, are defined. The native ensemble is obtained by combinatorial folding and unfolding of these blocks. Each microstate in the ensemble thus generated is then assigned a Gibbs free energy by means of an empirically parameterized energy function [38]. This function, which is based on solvent accessible surface area and conformational entropy, has been previously calibrated and validated for globular non-membrane proteins [37]. Free energy distribution over the states ( $\Delta G_i$ ) gives the probability distribution of interest:

$$K_i = e^{-\Delta G_i / RT} \quad Q = \sum_{i=1}^n K_i \quad p_i = \frac{K_i}{Q}$$

$K_i$  is the Boltzman weight of state  $i$ ,  $Q$  is the partition function of the ensemble, and  $p_i$  is the probability of state  $i$ . The summation runs over the entire ensemble.

The probability distribution  $p$  can be used to estimate a descriptor of the residue-specific equilibrium; the residue-specific stability constant ( $k_f$ ). This quantity is the ratio of total probability of the states where residue  $j$  is in folded conformation to that of the states where residue  $j$  is in unfolded conformation:

$$k_{f,j} = \frac{\sum_{i=\text{index of states where } j \text{ is folded}} p_i}{\sum_{i=\text{index of states where } j \text{ is unfolded}} p_i}$$

It follows that the local folding free energy of residue  $j$  is  $-RT\ln(k_{f,j})$ . Although this quantity is designated as "local", it actually depends on all the other residues through the probability distribution, which is determined by the properties of all microstates in the ensemble to which every part of the protein contributes. Hence, in the context of a given ensemble,  $k_f$  is local in the sense that it is a true physical descriptor of the equilibrium state of a single residue. However, this does not mean that it is independent of the state of other residues in the protein. It has been shown by extensive analysis of various proteins that these residue-specific constants reproduce a number of experimental observables, suggesting that the actual ensemble is well-represented by the COREX ensemble [38].

Another interesting piece of information that can be drawn from the COREX analysis is the degree of dependence of local folding free energies among different sites of the protein, which can be interpreted as an additional sign of allosteric linkage within the molecule [17]. This information can be obtained by evaluating folding correlations between different sites in a subset of the ensemble which contains the most probable states (e.g. with 0.99 probability):

$$r_{ij} = \frac{\text{cov}(F_i, F_j)}{\sqrt{\text{var}(F_i) \text{var}(F_j)}}$$

$F_i$  and  $F_j$  indicate the folding state (i.e.  $F = 0$  or  $1$ ) of residue  $i$  and  $j$ , respectively, across the most-probable states. An alternative strategy to reveal such linkages is to calculate the propagation of thermodynamic mutations made at a specific residue to all other residues [36]. This is achieved by artificially increasing ( $\sim 1$  kcal/mol) the free energy of all states in which the residue of interest is folded in the representative ensemble, and then recalculating residue-specific stability constants. Such a virtual perturbation results in re-sorting of the probabilities asso-

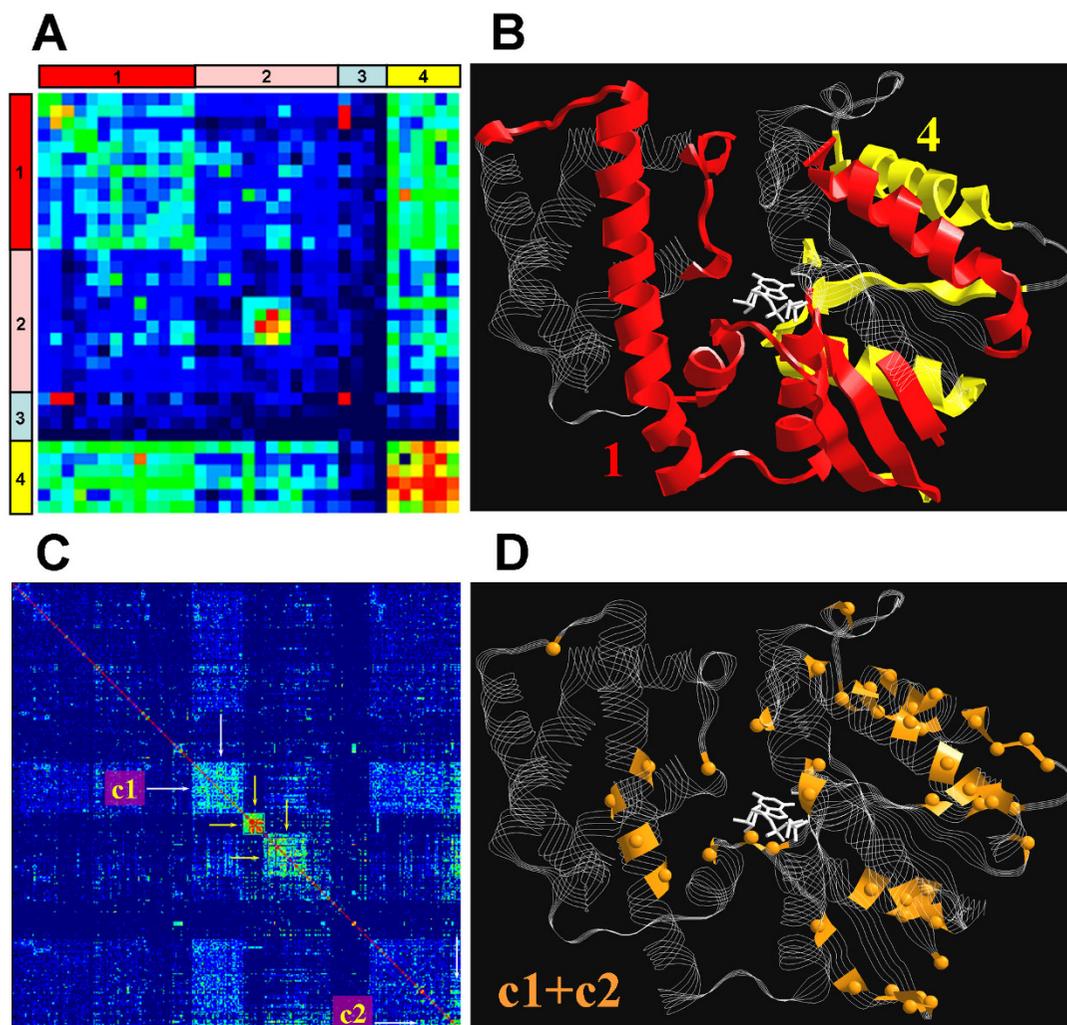
ciated with the states of the ensemble, and thus changes the value of residue-specific stability constants. Changes in the stability constants indicate how the perturbation made in a single residue propagates throughout the molecule. In principle, this information is similar to what the simple correlation analysis provides. Nevertheless, we performed both analyses, but provided only the results of the former as they indeed gave similar results.

In the present analysis we used high resolution structures of  $G_{\alpha i1}$  and  $G_{\alpha t}$ , which have been determined in their GDP- or GTP $\gamma$ S-bound forms (PDB ID: [1BOE](#), [1GIA](#) for  $G_{\alpha i1}$  and [1TAG](#)(A and B) for  $G_{\alpha t}$ ). We used the web server at UTMB (please see Availability & requirements section below) for Monte-Carlo sampling of COREX states (with their calculated free energies) of  $G_{\alpha t}$ . We performed high state density (i.e., exhaustive enumeration) COREX calculations only for  $G_{\alpha i1}$ . In both cases a window size of 8 residues per folding blocks was used. In high-resolution correlation calculations we evaluated the top 5000 states of  $G_{\alpha i1}$ .

#### Test statistics

In order to test whether an overlap between the two subsets of amino acid sites selected from the entire sequence was a result of pure coincidence, we used the following statistics (for the context of the problem, see results). The statistics is based on the probability distribution defined over the number of elements  $x$  in the intersection of two subsets selected randomly from a parent set. The null hypothesis is that the two subsets (whose sizes are given) are selected randomly. Given the size  $N$  of the parent set and the sizes of selected subsets  $K$  and  $M$ , the size of intersection  $x$  uniquely defines the partition as  $N = n+k+m+x$ , where  $k = K-x$ ,  $m = M-x$ ,  $n = N-(k+m+x)$ . The number of possible such partitions for a particular  $x$  is  $q_x = N! / (n!k!m!x!)$ , and for all allowed values of  $x$ , from 0 to  $\min(K,M)$ , provided that  $K+M \leq N$ , the total number of possible partitions is  $\Omega = \sum q_x \{x = 0 \dots \min(K,M)\}$ . Given the null hypothesis, all events in  $\Omega$  are equally likely. Hence, the probability of selecting  $K$  and  $M$  randomly with a particular number of overlaps  $x$  is  $p_x = q_x / \Omega$ . The statistics for the actually observed number of overlaps  $x_0$  then reads as  $p\{x \geq x_0\} = \sum p_x \{x = x_0 \dots \min(K,M)\}$ .

We used  $p\{x \geq x_0\} < 0.05$  as a one-tail rejection criterion for the null hypothesis that the observed overlap is coincidental (or equivalently, the observed overlap is not significantly different from the expected overlap under the null hypothesis). When the alternative hypothesis is the hypothesis of interest, the left-hand tail of the same distribution can obviously be used to test whether the observed lack of overlap is coincidental or not.



**Figure 2**

**Color coded statistical coupling matrices and mapping of selected sites on the structure of G $\alpha$ i.** (A) Pair wise

couplings of 35 secondary structure elements. Pixels indicate the values of statistical couplings ( $\chi^2_N$ ) on a rainbow color

scale (red: full coupling, dark blue: no coupling). The secondary structure elements are sorted according to the cluster analysis and cluster numbers are indicated on the sides of the matrix. (B) The elements in clusters 1 and 4 are shown on the structure of G $\alpha$ i (colors of the the numbers match the colors of the clusters in panel A). These clusters include following elements: N-terminal helix,  $\beta$ 1,  $\beta$ 2,  $\beta$ 3,  $\alpha$ 1, linker I,  $\alpha$ A,  $\alpha$ A- $\alpha$ B,  $\alpha$ D- $\alpha$ E,  $\alpha$ F, linker II (switch I),  $\alpha$ 3,  $\alpha$ 3- $\beta$ 5 (in #1),  $\beta$ 2- $\beta$ 3 hairpin tip,  $\beta$ 5- $\alpha$ G,  $\alpha$ 4,  $\beta$ 6,  $\beta$ 6- $\alpha$ 5,  $\alpha$ 5 (in #4).

(C) The original coupling matrix (361  $\times$  361) is sorted according to the cluster analysis. Pixels indicate the values of statistical couplings ( $\chi^2_N$ ) as in panel A. (D) Members of the two relevant clusters, indicated as c1 and c2

in panel C, are shown on the structure of G $\alpha$ i as orange strips and spheres corresponding to their  $\alpha$ -carbons. These sites correspond to (with SwissProt numbering of G $\alpha$ i1): V34, V72, S75, I78, S80, F95, N149, T177, K180, V185, H188, F191, K192, L194, H195, S206, C214, A235, M240, H244, L248, S252, C254, N256, T260, K271, T284, Y296, E298, A300, A301, Q304, Q306, L310, N311, K312, R313, E318, I319, T321, T327, F334, D337, A338, T340. Coordinates of GTP $\gamma$ S-bound G $\alpha$ i (PDB ID: [1GIA](#)) is used to render 3D-structures. RasTop program v2.0.3 (Philippe Valadon, 2003) is used for 3D visualization.

## Results

In this section we briefly summarize the findings. The results of the statistical coupling and corex analyses are given separately, and then a combined picture is provided at the end. The entire matrix of statistical couplings for further analysis is available [see Additional file 2].

### Statistical couplings in G protein family

Statistical coupling analysis in MSA provides a symmetrical matrix ( $361 \times 361$ ) of coupling values for all possible pairs of individual sites in G protein as described in the method section (actual number of couplings is  $361 \times 360/2$  due to the symmetry of the matrix). However, in order to acquire a general picture of statistical coupling, we first evaluated the couplings between secondary structure blocks by averaging the coupling of individual sites in a given secondary structure element. Overall, 35 such elements as helices, sheets, loops, linkers etc. are identifiable in the molecule (13  $\alpha$ -helices, 6  $\beta$ -sheets and 16 connecting loops and linkers). Hence, this procedure yielded a  $35 \times 35$  symmetric matrix, entries of which represent average pairwise coupling between the secondary structure blocks (net number of couplings is  $35 \times 34/2$ ). A hierarchical cluster analysis of the coupling matrix identified four clusters of elements with similar patterns and extents of couplings. The coupling matrix, rearranged according to these clusters, is shown in Figure 2A. Clusters #2 and #3 in Figure 2A consist of mainly uncoupled elements, although a mutually coupled spot is apparent in cluster #2. This spot, which is relatively isolated from the rest of the molecule in terms of statistical coupling, mainly consists of very conserved sites, whose functional importance is relatively well known (i.e.  $\alpha 2$ -helix,  $\beta 4$ -sheet, and  $\beta 4$ - $\alpha 3$  linker sequence, which correspond to switches II and III, respectively). Therefore, we did not further discuss these regions in the context of the present analysis. On the other hand, clusters #1 and #4 are formed by relatively less conserved sites. The elements in #4 are strongly, and those in #1 are moderately intra-coupled. The two clusters are also coupled to each other (see Figure 2A). Secondary structure elements that form these two clusters are shown on the structure of  $G\alpha_i$  in Figure 2A. They cover mostly those domains that are known or predicted to be involved in the interaction of  $G\alpha$  with its partners, namely  $G\beta\gamma$ , nucleotide, receptor, effector, RGS and AGS proteins (switches II and III are not present in this picture for reasons explained in the Methods section). Note that the sites in cluster 1 and 4 are located in the opposite faces of the molecule, including a part of the nucleotide binding cleft ( $\beta 6$ - $\alpha 5$ ,  $\beta 5$ - $\alpha G$ ). Therefore, the observation that the elements of cluster 1 are mutually coupled to those of cluster 4 implies a statistical coupling between the two faces of the molecule. Incidentally, a mechanical coupling between the two faces of the protein has also been revealed by molecular dynamics simulations [39]. These

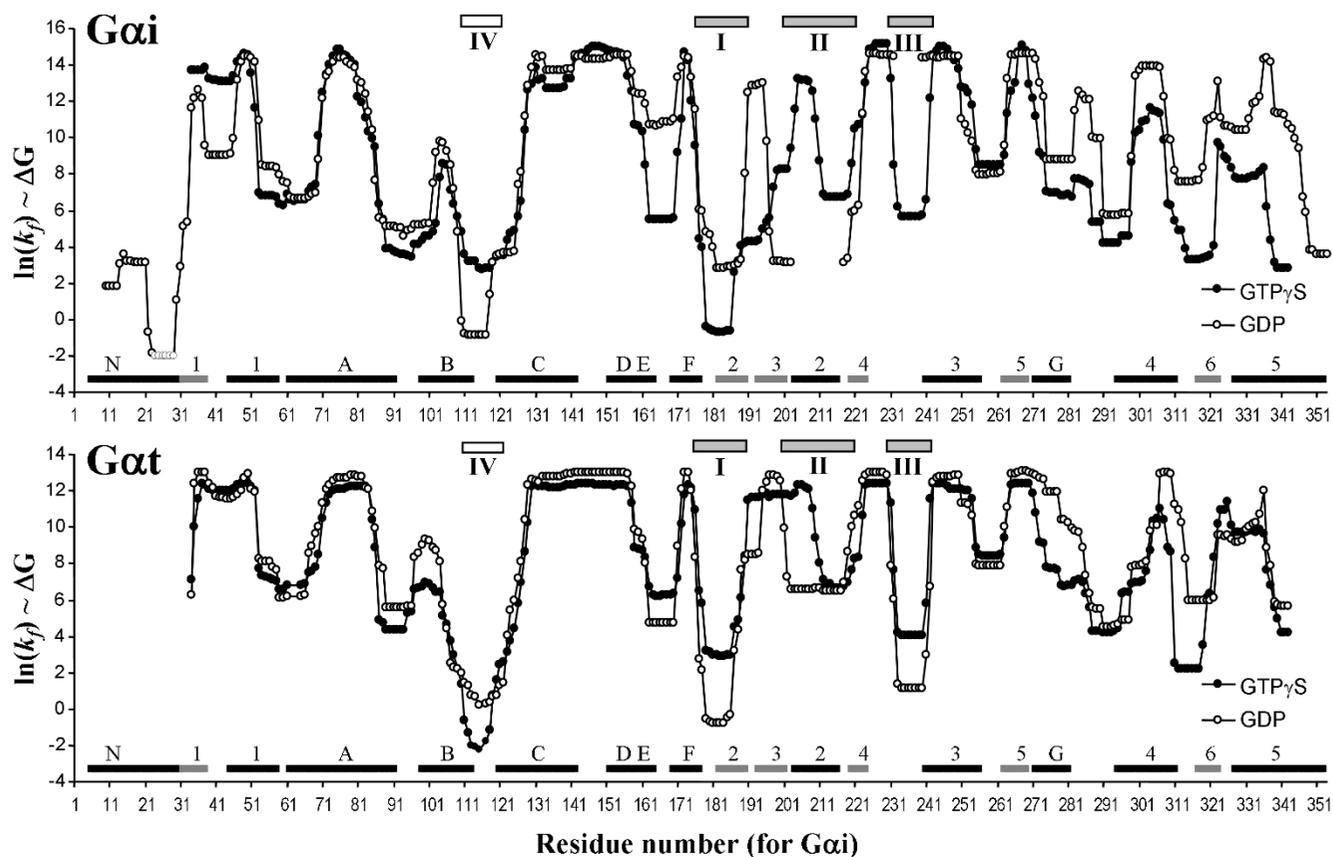
results imply and confirm that, on average, all the functional domains of  $G\alpha$  are statistically (or allosterically) intercoupled. Although this may seem to be trivial at first sight, a global intercoupling of these domains implies a more complex and flexible picture of allosteric regulation than a sequential interaction scheme would suggest. For example, statistical coupling between receptor, effector and  $G\beta\gamma$  binding sites raises the possibility of direct allosteric interactions between these sites which may by-pass the nucleotide exchange.

It is clear that secondary structure elements do not necessarily coincide with energetic or functional units in proteins [40]. Therefore, we further refined the picture given in Figure 2A by applying a cluster analysis to the coupling matrix of individual sites. The analysis identified 12 clusters of alike coupling patterns. Figure 2B shows the rearrangement of the original matrix according to these clusters. Out of twelve, two clusters designated as c1 and c2 in Figure 2C are of particular interest. These two clusters together consist of 50 moderately or poorly conserved sites (average conservation in the MSA is 40%, ranging from 19% to 86%). Moreover, these sites are slightly, albeit significantly, coupled to a considerable portion of the molecule. The sites that form cluster c1 and c2 are shown on the structure of  $G\alpha_i$  in Figure 2B. The pattern of coupling between secondary structural elements shown in Figure 2A is well represented by the distribution of the sites that belong to c1 and c2 (compare Figure 2A and Figure 2C). Two other highly coupled clusters in the coupling matrix are marked with yellow arrows in Figure 2C. Sites in these clusters mainly correspond to the highly inter-coupled spot in cluster #2 shown in Figure 2A (i.e. mainly those sites that form switches II and III).

In summary, cluster analysis of the coupling matrix revealed a set of individual sites that were all strongly coupled to each other and moderately coupled to the rest of the molecule in the statistical sense. The overall coupling between the two domains shown in Figure 2A is well represented by these sites. Therefore, we considered these sites as the core of allosteric interactions in  $G\alpha$  in the statistical sense. This core, which constitutes only about 15% of the entire molecule, is designated for brevity as the "allosteric core cluster".

### Analysis of statistical thermodynamic ensemble of $G\alpha_i$ and $G\alpha_t$

We evaluated equilibrium ensembles of  $G\alpha_i$  and  $G\alpha_t$  in their GDP- or GTP $\gamma$ S-bound forms that were modeled as partially folded states by COREX algorithm. Ensemble average of fractional folding of  $G\alpha_i$  or  $G\alpha_t$  showed that  $\sim 4\%$  of the residues were unfolded in equilibrium. However, GDP-bound forms of both  $G\alpha_i$  and  $G\alpha_t$  were slightly less folded compared to their GTP $\gamma$ S-bound forms (aver-



**Figure 3**  
**Distribution of local folding free energies over the residues of Gαi or Gαt.** Free energies are given as  $\ln(k_f)$  in the GDP or GTP $\gamma$ S bound forms of the proteins as indicated. Secondary structure elements are indicated as strips next to the x-axis (black for helices, grey for  $\beta$ -sheets) and named according to the convention given in figure 1. Known switch regions are also shown as grey strips numbered from I to IV on top of each picture. The two sequences are aligned and Swiss Prot residue numbering for Gαi is used in both cases.

age fractional unfolding was 5% vs. 3% in Gi, 4% vs. 3% in Gt for GDP- and GTP $\gamma$ S-bound forms, respectively). This was reflected by increased average conformational entropy and solvent accessible (polar and apolar) surface area in the GDP-bound form of the proteins, indicating that the GDP-bound form of Gα is slightly more "flexible" than the GTP $\gamma$ S-bound form in the thermodynamic sense.

Evaluation of folding states of individual residues in the ensemble (residue-specific stability constants) showed that the local folding free energy was not distributed uniformly over the residues of Gαi or Gαt (Figure 3). The pattern of local free energy distribution was roughly similar in Gαi and Gαt, as one might expect from the structural similarity of the two proteins. In Figure 3, it is worth noting that some of the low stability regions coincide with switches I, II and III (indicated in the picture), and with the  $\alpha$ B- $\alpha$ C loop, which has also been designated as a

switch region in Gαi (switch-IV) that assumed nucleotide-dependent conformations.

It is evident from Figure 3 that the local stability of many residues changes upon "nucleotide exchange", both in Gαi and Gαt. The patterns of these nucleotide-dependent changes ( $\Delta\Delta G$ ) in Gαi and Gαt are comparable. The present analysis can predict nucleotide-dependent changes in the local stability of the regions that have already been identified as switch regions in crystallographic studies (see Figure 3 for Gαt). Additional nucleotide-dependent changes that were not apparent in the crystal structure were also provided by the present analysis. The distribution of these nucleotide-dependent changes over the entire protein can be interpreted as the allosteric propagation of a free energy perturbation induced by nucleotide exchange at the guanine nucleotide binding site of Gα throughout the protein. Figure 4 shows



**Figure 4**  
**Changes in local folding free energy upon "nucleotide exchange" on  $G\alpha_i$ .** Changes in free energies upon nucleotide exchange in  $G\alpha_i$  are calculated from data given in figure 3 and mapped onto the structure of  $G\alpha_{i1}$ -GTP $\gamma$ S (PDB ID: 1GIA). Changes are shown on the color-coded trace-tube representation of the molecule (blue: no change – red: maximum change  $\sim 6$  kcal/mol). Trace-tube radius changes proportionally to the magnitude of free energy change to help visual perception. RasTop program v2.0.3 (Philippe Valadon, 2003) is used for 3D visualization.

the map of this propagation on the structure of  $G\alpha_i$ : Perturbation upon nucleotide exchange propagates nonuniformly to distant sites in the protein, and it covers a considerable part of the molecule (especially the GTPase domain). A qualitatively similar pattern of propagation was also observable in  $G\alpha_t$  (not shown). In addition to the switch regions, the following regions are also affected by nucleotide exchange (Figure 4): N-terminal of  $\alpha_G$ ,  $\alpha_G$ - $\alpha_4$  loop, almost the entire sequence of  $\alpha_4$ - $\beta_6$ - $\alpha_5$ ,  $\beta_2$ - $\beta_3$  hairpin, and obviously a large part of the nucleotide binding site. Among these regions,  $\alpha_G$ ,  $\alpha_G$ - $\alpha_4$  loop,  $\alpha_4$ ,  $\alpha_4$ - $\beta_6$  loop and  $\alpha_5$  have all been suggested to be involved in the formation of the binding interface between  $G\alpha$  and receptor [1,40,41].

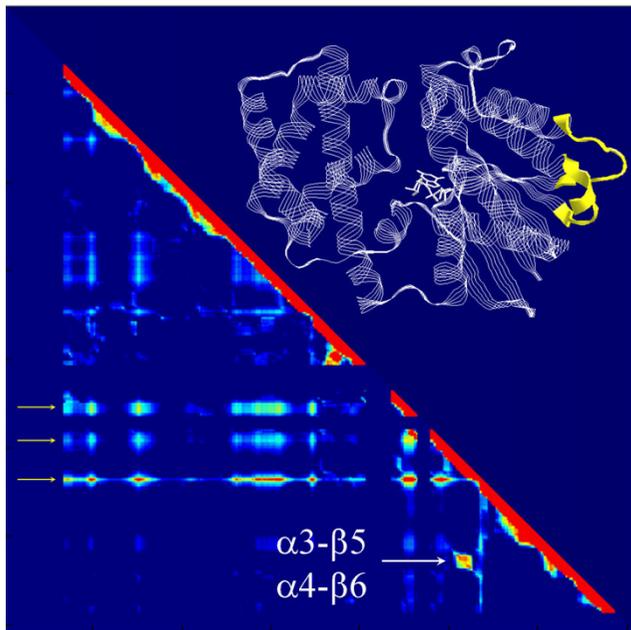
#### **Relationship between evolutionary and thermodynamic coupling**

We evaluated the relationship between the residues designated by the MSA analysis as "allosteric core cluster" and the residues identified by the Corex analysis to be involved in the propagation of nucleotide-induced thermodynamic effect. We used an arbitrary cut-off of 1.5 kcal/mol (comparable with an H-bond) for the local folding free energy difference ( $\Delta\Delta G$ ) between GDP- and GTP $\gamma$ S-bound  $G\alpha_i$  to construct the set of residues that were affected by nucleotide exchange. The overlap between the two sets was remarkable (ca. 65% of the

allosteric core intersected with the selected set of residues for which  $\Delta\Delta G > 1.5$  kcal/mol). However, the set of residues that changed local stability (above the cut-off) upon nucleotide exchange was quite large (about 1/3 of the entire molecule), which raised the possibility of a coincidental overlap between the two sets. We tested this possibility by using the test statistics described in the method section, which yielded a low probability of coincidence ( $p = 0.0086$  for  $N = 322$ ,  $K = 144$ ,  $M = 45$  and  $x = 28$ ). Therefore, we interpret the observed overlap as significant, and suggest that the residues in the "allosteric core cluster" are very likely to be mediating allosteric propagation of energetic perturbations in  $G\alpha$ .

Additionally, we searched for folding correlations as another potential indicator of allosteric coupling in the high resolution statistical ensemble generated by the COREX algorithm for  $G\alpha_i$ . For this analysis, 5000 partially folded states of GDP- $G\alpha_i$  or GTP $\gamma$ S- $G\alpha_i$  that possessed the highest probability of occurrence in the ensemble, and which accounted for a cumulative probability of greater than 99% were used. In this analysis nucleotide-dependent couplings in the molecule were disregarded, as the propagation of nucleotide-induced perturbation was already considered in the above analysis. In order to exclude the effect of nucleotide, only the correlations that are common to GDP- or GTP $\gamma$ S-bound forms of  $G\alpha_i$  were selected. Also excluded were high correlations between those residues that have a high level of stability in the representative ensemble. These correlations are statistically ambiguous for reasons similar to the correlations of highly conserved (invariant) residues in the statistical coupling analysis of MSA (as mentioned above).

In order to filter out the correlations that were not common in GDP- $G\alpha_i$  and GTP $\gamma$ S- $G\alpha_i$ , we constructed a correlation matrix whose entries were obtained through element-by-element multiplication of the two correlation matrices calculated for GDP- $G\alpha_i$  or GTP $\gamma$ S- $G\alpha_i$  (Figure 5). In this correlation matrix, an unambiguous correlation group that corresponds to residues forming  $\alpha_3$ - $\beta_5$  and  $\alpha_4$ - $\beta_6$  loops was identifiable (indicated with arrows in Figure 5). The same two loops were also represented in the "allosteric core cluster" by 8 residues, and the test statistics yielded a  $p$  value equal to 0.0018 (for  $N = 322$ ; total,  $K = 45$ ; allosteric core,  $M = 19$ ; correlation group,  $x = 8$ ; overlap) indicating that the intersection between the correlation group and the "allosteric core cluster" could not be explained by pure coincidence. This again confirms the convergence of the two independent analyses, as far as the allosteric coupling is concerned. Nucleotide-independent coupling between  $\alpha_3$ - $\beta_5$  and  $\alpha_4$ - $\beta_6$  loops may indicate a nucleotide-independent allosteric interaction between receptor and effector on  $G\alpha$ , as the two loops have been implicated to be involved in effector [10] and receptor



**Figure 5**  
**Common folding correlations in GDP- or GTP $\gamma$ S-bound G $\alpha$ i.** Correlation matrix is calculated from top 5000 partially folded COREX states of G $\alpha$ i-GDP and G $\alpha$ i-GTP $\gamma$ S and is shown on a rainbow color scale (blue to red correspond to 0 to 1). The matrix is constructed as explained in the text and the sites are sorted from N to C terminal (left to right, top to bottom). Only the lower triangle of the symmetric matrix is shown. Yellow arrows indicate regions of very high local stability. An unambiguous correlation group is indicated with white arrows and mapped on the structure of G $\alpha$ i (inset). RasTop program v2.0.3 (Philippe Valadon, 2003) is used for 3D visualization.

[1,41,42] interactions respectively. Such a nucleotide-independent coupling between effector and receptor sites on G $\alpha$  may explain the experimental observation that  $\beta$ 2-adrenoceptors are able to activate adenylate cyclase through Gs independently from nucleotide-exchange [43]. Using the statistical ensembles of GDP-G $\alpha$ i and GTP $\gamma$ S-G $\alpha$ i, we also performed a mutual perturbation response analysis based on thermodynamic "mutation" in the ensemble [37]. This analysis also identified the same coupling group as the one we found in the correlation analysis above (data not shown).

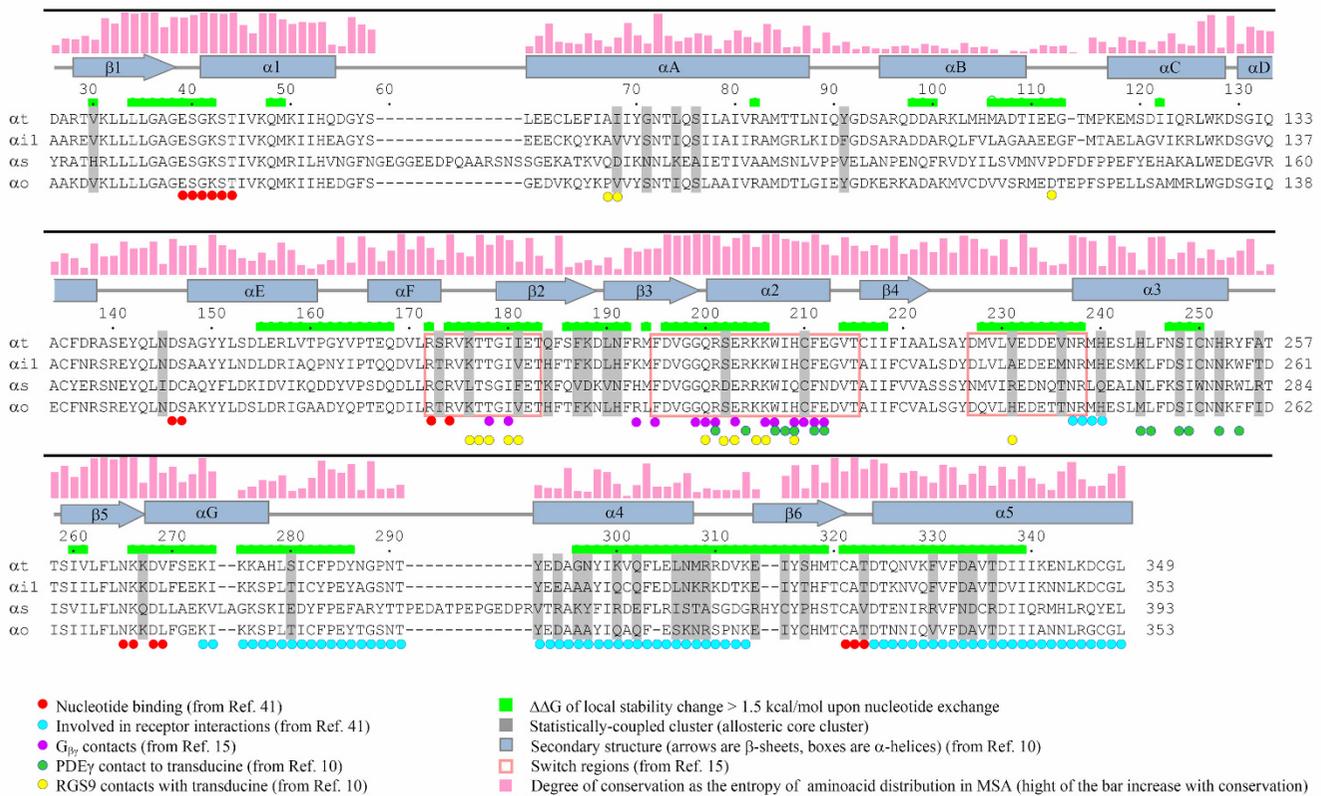
## Discussion

In the present study, we analyzed 1) statistical couplings in the MSA of heterotrimeric G protein family, and 2) statistical ensemble properties of G $\alpha$ i and G $\alpha$ t by using COREX algorithm to assess allosterically linked sites in this signal transducer family, which was not apparent in crystallographic studies. We showed that the residues captured by the two methods overlapped significantly. Con-

sidering that the information processed by the two methods was essentially different and independent, convergent results of the two analyses can be interpreted as they pointing (to a certain extent) to a common aspect of the protein sites (i.e. their involvement in allosteric communication in the protein). This is particularly important for the interpretation of statistical coupling in MSA, which is less obvious than the information provided by COREX analysis, as the covariance in MSA does not necessarily indicate allosteric coupling [44]. In this sense, the COREX algorithm seems to be a very useful tool to complement and confirm the covariance information used for this particular purpose.

The results of the entire analysis, along with information available from previous analyses concerning G protein structure and function, are summarized in Figure 6 using the aligned sequences of prototypical G proteins. It is evident from Figure 6 that the residues that change stability upon nucleotide exchange (green), those that are included in the "allosteric core" (gray) and those that are located in the known functional domains of G $\alpha$  (represented by different colors and symbols) show a considerable co-aggregation along the G $\alpha$  sequence. Highly conserved sites (especially those that are involved in nucleotide binding and those that form switch II) do not appear in the allosteric core due to the nature of the statistical coupling analysis as discussed above. This does not mean that these sites are not involved in allosteric coupling. Conserved sites are excluded simply due to the fact that statistical coupling analysis was not appropriate to diagnose the involvement of conserved sites in allosteric linkage. Nevertheless, it is evident from COREX analysis that most sites in the switch regions (including switch-II) are thermodynamically sensitive to the identity of bound nucleotide, which implies a thermodynamic coupling between the nucleotide binding site and the switch regions (see Figures 3 and 6). Together with the observation that the folding correlation between the  $\alpha$ 3- $\beta$ 5 and  $\alpha$ 4- $\beta$ 6 loops revealed by COREX analysis was also represented significantly in the allosteric core (not indicated in Figure 6), we interpret the general picture in Figure 6 as evidence for the existence of an allosteric network that connects all (known) functional domains of G $\alpha$  to each other in all possible ways (as the residues in the allosteric core are all mutually coupled). Sites in the allosteric core that have not been associated so far with an obvious function (such as those in the helical domain or in the  $\beta$ 2- $\beta$ 3 hairpin tip), may then be considered as potential candidates for the mediation of intramolecular allosteric communication in the G protein family.

Two interesting observations made in the present analyses may warrant further discussion. The first one is that residues that are thermodynamically affected by nucleotide



**Figure 6**  
**Sites in the allosteric core cluster and residues that change stability upon nucleotide exchange.** Sites that are found in MSA and COREX analyses, along with the sites involved in known functions, are shown together on the aligned sequences of indicated  $G\alpha$  proteins. Coding is shown on the picture.

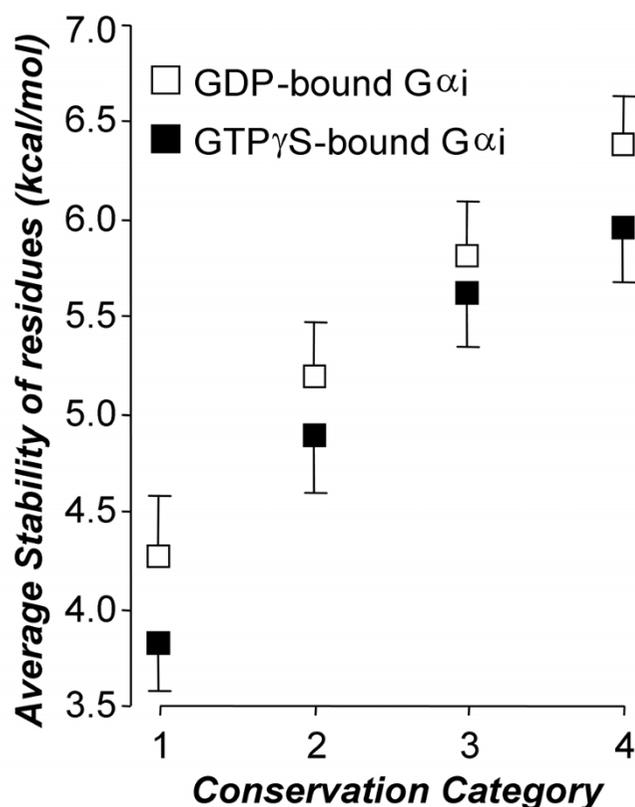
exchange and those that are represented in the allosteric core overlap much better in the GTPase domain than in the helical domain (see Figure 6). In the helical domain, COREX analysis revealed a nucleotide-induced local thermodynamic effect on the residues of switch IV and the  $\alpha E$ - $\alpha F$  sequence, whereas the statistical coupling analysis hit residues along the  $\alpha A$  helix. A trivial explanation for the observed mismatch can be that residues in the  $\alpha A$  helix have nothing to do with allosteric linkage, as statistical coupling analysis always has the potential of such bias when it is not supported by independent information. However, it is also possible that sites identified by COREX analysis as undergoing nucleotide-dependent changes in the helical domain (i.e. switch IV and the  $\alpha E$ - $\alpha F$  sequence) are not linked to other functional domains in the GTPase region as they are not covered by the allosteric core where all residues are mutually coupled. Only by experiments can these two possibilities be distinguished. The second observation is that there is a nucleotide-independent linkage between the  $\alpha 3$ - $\beta 5$  and  $\alpha 4$ - $\beta 6$  loops, which was revealed by both COREX and statistical coupling analyses. This suggests a direct (nucleotide-inde-

pendent) coupling between effector and receptor on  $G\alpha$ . This linkage may result in a nucleotide exchange-independent modulation of effector- $G\alpha$  interaction by the receptor. As mentioned above, experiments have shown that this may indeed be the case for  $\beta 2$ -adrenoceptor-induced adenylyl cyclase activation through  $G_{\alpha s}$  [43].

At this point, it may be worth mentioning the nature of allosteric linkage that was revealed by the present analyses. For practical purposes, the process of allosteric interactions can be seen from two different points of view: 1) The mechanistic approach, which sees the process as the propagation of a perturbation in one site to another via a series of structural distortions that extend from one site to the other, and 2) the thermodynamical or statistical mechanical approach, which considers the process as arising from the perturbation of the state distribution by ligand binding (or other effects such as mutations) in the equilibrium conformational ensemble of the protein. The latter view (to which the COREX approach belongs) can eventually represent the process only as an energetic linkage between different domains of the protein without

referring to any specific mechanism whatsoever. The information provided by statistical coupling analysis in MSA also belongs to that second category. From this point of view, the effects of perturbations in one part of a protein on all other parts are due to a redistribution of the conformational ensemble [17,35]. Whether a particular substructure is stabilized by the perturbation in a particular conformational state will depend on how those ensemble-averaged properties change as a result of the redistribution. Regions (or more specifically, particular conformations of a region) that are positively coupled will be stabilized as a result of the perturbation, and regions that are negatively coupled will be destabilized [35]. In consequence, the pattern of allosteric coupling in the protein is determined by the energetic hierarchy of states (i.e. which states are the most stable and what structural and functional attributes those states possess). This may also have implications for the interpretation of mutational screening data. For example, any mutation that causes a change in the energetic hierarchy of states in the ensemble can change the allosteric coupling between two sites, even if the network of structural elements that physically connect the two sites is unaffected, which complicates the interpretation of the effect of mutation from a mechanistic point of view. Hence, the ensemble view does not imply any mechanical coupling between energetically (or statistically) linked sites. The information summarized in Figure 6 should also be understood in this context.

Another practical implication of the present analysis concerns the functional (or structural) importance that has generally been attributed to evolutionarily conserved sites in proteins. Unconserved sites have attracted less attention than the conserved for obvious reasons. However, the local conservation of the sites in the allosteric core indicated as gray-shaded residues in Figure 6 was quite weak: Average frequency of most abundant amino acids at these sites was 40%, or the average entropy of their conservations was 0.33 on a normalized scale, where 0 and 1 correspond to none and full conservation, respectively. It follows that unconserved sites may play particularly important roles in the mediation of allosteric coupling. This assertion finds a firm basis when we consider the following observation made in the present analysis: We found that conservation of sites in the G protein family, on average, scales proportionally with the thermodynamic stability of these sites within  $G\alpha_i$  (Figure 7). In other words, there is (on average) a positive association between evolutionary and thermodynamic stability in G proteins. On the other hand, it has also been demonstrated that residues in unstable regions are important for mediating allosteric coupling [17,35]. Thus, it can be inferred that locally unconserved sites are indeed important for allosteric effects. This may explain why some



**Figure 7**  
**Association between thermodynamic and evolutionary stability.** Residues are classified into four (equally-spaced) categories according to their conservation entropy in  $G\alpha$ -MSA (1 to 4 correspond to weakest to strongest conservation), and plotted against average residue-specific stability constants in the categories. Stability constants are calculated in GDP- or GTP $\gamma$ S-bound forms of  $G\alpha_i$ . Bars are standard deviation of means. There are approximately equal numbers of sites in each category ( $n \approx 80$ ). Note that in average, there is a remarkable positive association between conservation of the residues and their local thermodynamic stabilities (local stabilities of the GDP-bound form are slightly but systematically higher than those of the GTP $\gamma$ S-bound form).

allosterically important sites elude attention when we consider the fact that the literature of screening by point mutations has been mostly concentrated on conserved sites. However, this should not be understood as an implication that the conserved sites are generally unimportant in allosteric coupling.

Besides the above discussion, we believe that the observed association between evolutionary and thermodynamic stabilities is interesting in itself from a theoretical point of view, as there is no obvious reason for such an associa-

tion: The former entity has to do with the variability of amino acid identity at a site in the G protein family, whereas the latter is associated with the folding stability of that site in a single member of the family (i.e. G $\alpha$ i). Although a detailed discussion of the issue is beyond the scope of the present work, we may be permitted to speculate that such an association is expected (on average) in the following sense: If the protein has such a structural, energetic and functional design that a thermodynamic flexibility (rather than a particular fold) is required at a site (which is observed as a low stability of the site), then evolution may permit any amino acid at that site (which results in an unconserved site), as the folding state of that site is determined by the entire design of the protein, but not by the actual identity of the amino acid that occupies that site. The thermodynamic instability of a site may well be a part of the functional design of the protein, as discussed above. Among many other possible mechanisms, this one alone can lead to a tendency of association between evolutionary and thermodynamic stabilities of the sites. It would be interesting to investigate whether such an association exists in other protein families.

## Conclusion

We propose that the sites indicated in Figure 6 (i.e. those that are linked to nucleotide binding sites and those that are statistically coupled to each other) constitute an allosteric network in the energetic sense within the G protein. Recent progress in G protein expression, purification and spectroscopy [45-47], together with molecular biological techniques, provide a potential tool to bring the information summarized in Figure 6 to experimental test, which may eventually contribute to a more comprehensive understanding of the functional design of G proteins.

## Abbreviations

GPCR: G protein coupled receptors; RGS: regulators of G protein signaling; AGS: activator of G protein signaling; MSA: multiple sequence alignment.

## Availability & requirements

Pfam database: <http://pfam.sanger.ac.uk>

BEST @ UTMD: <http://best.utmb.edu/BEST/>

## Authors' contributions

KS performed statistical analyses and participated in the interpretation of data, TL and VJH performed high resolution COREX calculations for G $\alpha$ , VJH helped in drafting the manuscript, ÖU and OO designed the study, participated in all statistical analyses, interpreted the data, and wrote the paper.

## Additional material

### Additional file 1

supporting information. A simulation test for the comparison of the measure of statistical coupling in MSA that was used in the present study with the one that has been originally suggested by Lockless and Ranganathan (ref 23) is provided as supporting information.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6807-8-23-S1.pdf>]

### Additional file 2

Statistical coupling matrix. Entire statistical coupling matrix for the G protein family is given as supporting data (as an MSExcel file) for further analysis.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6807-8-23-S2.xls>]

## Acknowledgements

The authors would like to thank to Tommaso Costa (ISS, Rome, Italy) and Francesca Fanelli (University of Modena, Italy) for their insightful comments and criticism, and Frederic Stark for editing the manuscript in terms of English. This work was supported, in part, by research grants from the Turkish Scientific and Technical Research Council 104s472 and 107s086.

## References

1. Oldham WM, Hamm HE: **Structural basis of function in heterotrimeric G proteins.** *Q Rev Biophys* 2006, **39**:117-166.
2. Hollinger S, Hepler JR: **Cellular Regulation of RGS proteins: Modulators and integrators of G protein signaling.** *Pharmacol Rev* 2002, **54**:527-559.
3. Lanier SM: **AGS proteins, GPR motifs and the signals processed.** *Biology of the Cell* 2004, **96**:369-372.
4. Noel JP, Hamm HE, Sigler PB: **The 2.2 Å crystal structure of transducin-alpha complexed with GTP $\gamma$ S.** *Nature* 1993, **366**:654-663.
5. Coleman DE, Berghuis AM, Lee E, Linder ME, Gilman AG, Sprang SR: **Structures of active conformations of G $\alpha$ i1 and the mechanism of GTP hydrolysis.** *Science* 1994, **269**:1405-1412.
6. Sunahara RK, Tesmer JJ, Gilman AG, Sprang SR: **Crystal structure of adenylyl cyclase activator G $\alpha$ s.** *Science* 1997, **278**:1943-1947.
7. Kreutz B, Yau DM, Nance MR, Tanabe S, Tesmer JJ, Kozasa T: **A new approach to producing functional G $\alpha$  subunits yields the activated and deactivated structures of G $\alpha$  (12/13) proteins.** *Biochemistry* 2006, **45**:167-174.
8. Lutz S, Shankaranarayanan A, Coco C, Ridilla M, Nance MR, Vettel C, Baltus D, Evelyn CR, Neubig RR, Wieland T, Tesmer JJ: **Structure of G $\alpha$ q-p63RhoGEF-RhoA complex reveals a pathway for the activation of RhoA by GPCRs.** *Science* 2007, **318**:1923-1927.
9. Kimple RJ, Kimple ME, Betts L, Sondak J, Siderovski DP: **Structural determinants for GoLoco induced inhibition of nucleotide release by G $\alpha$  subunits.** *Nature* 2002, **416**:878-881.
10. Slep KC, Kercher NA, He W, Cowan CW, Wensel TG, Sigler PB: **Structural determinants for regulation of phosphodiesterase by a G protein at 2.0 Å.** *Nature* 2001, **409**:1071-1077.
11. Grishina G, Berlot CH: **Mutations at the domain interface of G $\alpha$ s impair receptor-mediated activation by altering receptor and guanine nucleotide binding.** *J Biol Chem* 1998, **273**:15053-15060.
12. Coleman DE, Lee E, Mixon MB, Linder ME, Berghuis AM, Gilman AG, Sprang SR: **Crystalisation and preliminary crystallographic studies of G $\alpha$ i1 and mutants of G $\alpha$ i1 in the GTP and GDP bound states.** *J Mol Biol* 1994, **238**:630-634.

13. Lambright DG, Noel JP, Hamm HE, Sigler PB: **Structural determinants for activation of the alpha-subunit of a heterotrimeric G protein.** *Nature* 1994, **369**:621-628.
14. Wall MA, Coleman DE, Lee E, Iniguez-Llhui JA, Posner BA, Gilman AG, Sprang SE: **The structure of G protein heterotrimer G<sub>α1β1γ2</sub>.** *Cell* 1995, **83**:1047-1058.
15. Lambright DG, Sonddek J, Bohm A, Skiba NP, Hamm HE, Sigler PB: **The 2.0 Å crystal structure of a heterotrimeric G protein.** *Nature* 1996, **379**:311-319.
16. Cooper A, Dryden DTF: **Allosterism without conformational change.** *Eur Biophys J* 1982, **11**:103-109.
17. Pan H, Lee JC, Hilser VJ: **Binding sites in E. Coli dihydrofolate reductase communicate by modulating the conformational ensemble.** *Proc Natl Acad Sci USA* 2000, **97**:12020-12025.
18. Horovitz A, Fersht AR: **Strategy for analysing the co-operativity of intramolecular interactions in peptides and proteins.** *J Mol Biol* 1990, **214**:613-617.
19. Göbel U, Sander C, Schneider R, Valenzia A: **Correlated mutations and residue contacts in proteins.** *Proteins* 1994, **18**:309-317.
20. Shindyalov IN, Kolchanov NA, Sander C: **Can three-dimensional contacts in protein structures be predicted by analysis of correlated mutations?** *Protein Eng* 1994, **7**:349-358.
21. Olmea O, Rost B, Valenzia A: **Effective use of sequence correlation and conservation in fold recognition.** *J Mol Biol* 1999, **293**:1221-1239.
22. Horovitz A, Bochkareva ES, Yifrach O, Girshovich AS: **Prediction of an inter-residue interaction in the chaperonin GroEL from multiple sequence alignment is confirmed by double-mutant cycle analysis.** *J Mol Biol* 1994, **238**:133-138.
23. Lockless SV, Ranganathan R: **Evolutionarily conserved pathways of energetic connectivity in protein families.** *Science* 1999, **286**:295-298.
24. Russ WP, Lowery DM, Mishra P, Yaffe MB, Ranganathan R: **Natural-like function in WW domains.** *Nature* 2005, **437**:579-583.
25. Dekker JP, Fodor A, Aldrich RW, Yellen G: **A perturbation-based method for calculating explicit likelihood of evolutionary covariance in multiple sequence alignments.** *Bioinformatics* 2004, **20**:1565-1572.
26. Kass I, Horovitz A: **Mapping pathways of allosteric communication in GroEL by analysis of correlated mutations.** *Proteins* 2002, **48**:611-617.
27. Noivirtl O, Eisenstein M, Horovitz A: **Detection and reduction of evolutionary noise in correlated mutation analysis.** *Protein Engineering, Design & Selection* 2005, **18**:247-253.
28. Süel GM, Lockless SV, Wall MA, Ranganathan R: **A perturbation-based method for calculating explicit likelihood of evolutionary co-variance in multiple sequence alignments.** *Nat Struct Biol* 2003, **10**:59-69.
29. Frauenfelder H, Parak F, Yang RD: **Conformational substates in proteins.** *Annu Rev Biophys Biophys Chem* 1988, **17**:451-479.
30. Hilser VJ, Freire E: **Structure-based calculation of the equilibrium folding pathway of proteins: correlation with hydrogen exchange protection factors.** *J Mol Biol* 1996, **262**:756-772.
31. Onaran HO, Costa T: **Agonist efficacy and allosteric models of receptor action.** *Ann NY Acad Sci* 1997, **812**:98-115.
32. Onaran HO, Scheer A, Cotecchia S, Costa T: **A look into receptor efficacy. From the signalling network of the cell to the intramolecular motion of the receptor.** *Handbook of Experimental Pharmacology* 2000, **148**:217-259.
33. Wyman J, Gill SJ: *Binding and Linkage* Mill Valley, California:University Science Books; 1990.
34. Weber G: *Protein interactions* New York, London: Chapman and Hall; 1992.
35. Liu T, Witten ST, Hilser VJ: **Ensemble-based Signatures of Energy Propagation in Proteins. A New View of an Old Phenomenon.** *Proteins: Struct Func Bioinform* 2006, **62**:728-738.
36. Hilser VJ, Freire E: **Predicting the equilibrium protein folding pathway: structure-based analysis of staphylococcal nuclease.** *Protein Struct Funct Genet* 1997, **27**:171-183.
37. Hilser VJ, Dowdy D, Oas TG, Freire E: **The structural distribution of cooperative interactions in proteins: analysis of the native state ensemble.** *Proc Natl Acad Sci USA* 1998, **95**:9903-9908.
38. Hilser VJ, Garcia-Moreno EB, Oas TG, Kapp G, Whitten ST: **A statistical thermodynamic model of the protein ensemble.** *Chem Rev* 2006, **106**:1545-1558.
39. Ceruso MA, Periole X, Weinstein H: **Molecular dynamics simulations of transducin: interdomain and front to back communication in activation and nucleotide exchange.** *J Mol Biol* 2004, **338**:469-481.
40. Larson SA, Hilser VJ: **Analysis of the thermodynamic information content of a homo sapiens structural database reveals hierarchical thermodynamic organization.** *Prot Sci* 2004, **13**:1787-1801.
41. Fanelli F, Menziani C, Scheer A, Cotecchia S, De Benedetti PG: **Theoretical study of the electrostatically driven step of receptor-G protein recognition.** *Proteins* 1999, **37**:145-156.
42. Fanelli F, Dell'Orco D: **Rhodopsin activation follows precoupling with transducin: inferences from computational analysis.** *Biochemistry* 2005, **44**:14695-14700.
43. Uğur O, Oner SS, Molinari P, Ambrosio C, Sayar K, Onaran HO: **Guanine Nucleotide Exchange-Independent Activation of Gs Protein by β2-Adrenoceptor.** *Mol Pharmacol* 2005, **68**:720-728.
44. Fodor AA, Aldrich RW: **On Evolutionary Conservation of Thermodynamic Coupling in Proteins.** *J Biol Chem* 2004, **279**:19046-19050.
45. Ridge KD, Abdulaev NG, Zhang C, Ngo T, Brabazon DM, Marino JP: **Conformational changes associated with receptors stimulated guanine nucleotide exchange in a heterotrimeric G-protein α-subunit. NMR analysis of GTPγS-bound states.** *J Biol Chem* 2006, **281**:7635-7648.
46. Van Eps N, Oldham WM, Hamm HE, Hubbell WL: **Structural and dynamical changes in an α-subunit of a heterotrimeric G protein along the activation pathway.** *Proc Natl Acad Sci USA* 2006, **103**:16194-16199.
47. Oldham WM, Van Eps N, Preininger AM, Hubbell WL, Hamm HE: **Mapping allosteric connections from the receptor to the nucleotide-binding pocket of heterotrimeric G proteins.** *Proc Natl Acad Sci U S A* 2007, **104**:7927-7932.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

