JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

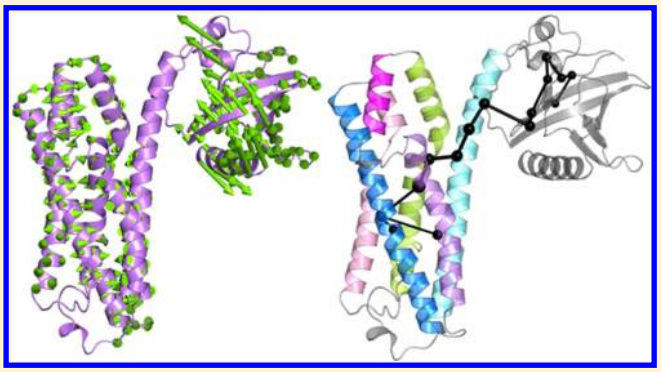
Catching Functional Modes and Structural Communication in Dbl Family Rho Guanine Nucleotide Exchange Factors

Francesco Raimondi,[†] Angelo Felline, and Francesca Fanelli*

Department of Life Sciences, University of Modena and Reggio Emilia, via Campi 103, 41125 Modena, Modena, Italy

Supporting Information

ABSTRACT: Computational approaches such as Principal Component Analysis (PCA) and Elastic Network Model-Normal Mode Analysis (ENM-NMA) are proving to be of great value in investigating relevant biological problems linked to slow motions with no demand in computer power. In this study, these approaches have been coupled to the graph theory-based Protein Structure Network (PSN) analysis to dissect functional dynamics and structural communication in the Dbl family of Rho Guanine Nucleotide Exchange Factors (RhoGEFs). They are multidomain proteins whose common structural feature is a DH-PH tandem domain deputed to the GEF activity that makes them play a central role in cell and cancer biology. While their common GEF action is accomplished by the DH domain, their regulatory mechanisms are highly variegate and depend on the PH and the additional domains as well as on interacting proteins. Major evolutionary-driven deformations as inferred from PCA concern the $\alpha 6$ helix of DH that dictates the orientation of the PH domain. Such deformations seem to depend on the mechanisms adopted by the GEF to prevent Rho binding, i.e. functional specialization linked to autoinhibition. In line with PCA, ENM-NMA indicates $\alpha 6$ and the linked PH domain as the portions of the tandem domain holding almost the totality of intrinsic and functional dynamics, with the $\alpha 6/\beta 1$ junction acting as a hinge point for the collective motions of PH. In contrast, the DH domain holds a static scaffolding and hub behavior, with structural communication playing a central role in the regulatory actions by other domains/proteins. Possible allosteric communication pathways involving essentially DH were indeed found in those RhoGEFs acting as effectors of small or heterotrimeric RasGTPases. The employed methodology is suitable for deciphering structure/dynamics relationships in large sets of homologous or analogous proteins.



INTRODUCTION

Biomolecular function depends on structural communication and the entangled dynamics, especially those motions that occur relatively slowly and require long time scale computations.[1] However, computational analyses using established techniques with well-chosen approximations have continually proven to be of great value in investigating interesting biophysical and biochemical problems with no demand in terms of computer power.[1−3] In this framework, we recently used approaches like Principal Component Analysis (PCA) and Elastic Network Model-Normal Mode Analysis (ENM-NMA) to identify the important structural flexibilities that enable proteins in the Ras superfamily to switch between their active and inactive states.[4] Such analysis, carried out on the crystal structures of representative members, led to a hypothesis regarding the evolutionary adaptation of structural deformations by the individual members of the superfamily to fulfill their specialized function. We have also coupled ENM-NMA with the analysis of Protein Structure Networks (PSN) in a so-called mixed PSN-ENM approach to infer structural communication pathways from single crystallographic structures.[5,6] In this study, all these single structure-based approaches have been merged to dissect functional dynamics and structural communication in the Dbl family of Rho Guanine Nucleotide Exchange Factors (RhoGEFs).

Rho (Ras homologue) proteins comprise a main family of the Ras superfamily of small GTPases.[7] They function as bimolecular switches by adopting different conformational states in response to binding GDP or GTP. In contrast to Rho-GDP, Rho-GTP actively transduces signals by interacting with downstream effectors, which mediate a range of different cellular functions including cell-cycle progression and gene transcription, adhesion and migration, phagocytosis, cytokinesis, neurite extension and retraction, cellular morphogenesis and polarization, growth and cell survival.[8] Cycling between GDP- and GTP-bound states is controlled primarily by two classes of regulatory molecules: GTPase activating proteins (GAPs), which enhance the relatively slow intrinsic GTPase activity of Rho proteins, and guanine nucleotide-exchange factors (GEFs), which catalyze the exchange of GDP for GTP. While GAPs suppress Rho activity, GEFs promote it.

Received: March 6, 2015
Published: August 31, 2015

DOI: 10.1021/acs.jcim.5b00122
J. Chem. Inf. Model. 2015, 55, 1878−1893

Evidence collected over the past decade has clearly established that Rho GTPases and RhoGEFs play important roles in many aspects of cancer development and tumor progression.[9−12]

Most of the RhoGEFs, 74 distinct members, share a catalytic domain homologous to that of the Dbl oncoprotein (DH domain).[13] In most Dbl family RhoGEFs, the DH domain is positioned immediately N-terminal to a pleckstrin homology (PH) domain. Structures of PH-DH tandem domains from several RhoGEFs have been determined either alone or in complex with their substrate GTPases, leading to key insights into the molecular mechanism that is used by Dbl-family RhoGEFs to promote the exchange of GDP for GTP. It has been hypothesized that they do this by turning the switch regions of the GTPases on themselves to eject GDP and Mg^{2+}.[14]

Early biochemical studies using oncogenic Dbl as a model system established that, whereas the DH domain is responsible for GEF catalytic activity, the adjacent PH domain is involved in assisting the binding of RhoGTPase for catalysis, stability, and intracellular targeting of the DH domain.[13,15−17] Although the PH-directed localization of the DH-PH module to the plasma membrane and the actin cytoskeleton conforms to the generalized view that PH domains function as regulated membrane-binding modules,[18] it also suggests a unique capability of the PH domain to associate with actin or actin-binding components. It is now clear that the DH domain, together with the PH domain, constitutes the minimum structural unit bearing transforming function.[15,16,19] Indeed, increasing evidence indicates that these linked domains cooperate to facilitate the activation of Rho GTPases, since DH-PH fragments show greater nucleotide exchange activity than the respective DH domains alone.[13,20,21]

The regulatory mechanism of Dbl proteins remains poorly understood and would rely on the action of the domains outside the DH-PH module as well as on the interaction between DH-PH and other proteins.[13] Lipid-binding may concur in the regulatory mechanisms as well. Many RhoGEFs are constitutively activated by the N-terminal truncation of sequences that lie upstream of the DH domain.[13] A clear example is represented by Vav1 (gene name: VAV1) whose PH-DH is N-terminally linked to a calponin homology (CH) domain and a successive α-helix (Acidic (AC) domain) that in basal conditions interact with the DH domain and, to a lesser extent, with the PH domain by preventing binding of Rac1.[22] The PH domain participates in such an autoinhibition exerted by CH. Phosphorylation by various receptor-associated tyrosine kinases, such as Lck, of Tyr174 and proximal tyrosines opens the DH domain to GTPases.[23,24] Another example of autoinhibition by the N-term is represented by the RGS-RhoGEF from Entamoeba Histolytica (EHI, gene name: EHI-010670) that holds a Regulator of G protein Signaling (RGS) binding domain linked to the DH domain by an α-helix that contributes to inhibit Rho binding.[25] Furthermore, Asef (gene name ARHGEF4), Collybistin (CB, gene name: ARHGEF9), and Intersectin (gene name: ITSN1) are autoinhibited by a SH3 domain N-terminal to the DH domain.[26−28] As for autoinhibitory C-terminal domains, FARP2 is autoinhibited by an additional PH domain (PH2) C-terminal to the primary PH domain.[29] Both additional SH3 and PH domains interact with the Rho binding regions of DH. Regulation of a number of Dbl RhoGEFs relies also on their being part of supramolecular complexes with scaffolding proteins like in the case of Tiam1 (gene name TIAM1) that participates in supramolecular assemblies through the scaffold proteins JIP2 to direct the activation of Rac.[30] Dbl proteins may act as effectors of heterotrimeric G proteins (e.g., $G\alpha_{12/13}$ and $G\alpha_q$) or small GTPases (e.g., RhoA). This is the case of the members of the Lbc-subfamily: A-Kinase Anchoring Protein-AKAP13 (AKAP-Lbc (Lbc), gene name: AKAP13), Leukemia-associated RhoGEF (LARG, gene name: ARHGEF12), p115-RhoGEF (p115, gene name: ARHGEF1), and PDZ-RhoGEF (gene name: ARHGEF11), which act as effectors of activated $G\alpha_{12/13}$ (i.e., only $G\alpha_{12}$ in the case of AKAP13)[31−36] or p63 (gene name: ARHGEF25), effector of $G\alpha_q$.[37] In the latter case, $G\alpha_q$ enhances the GEF activity of ARHGEF25 possibly by releasing the autoinhibition operated by the PH domain. An analogous case of PH-mediated autoinhibition is represented by AKAP13.[38,39] Differently from AKAP13 and ARHGEF25, the RhoGEFs ARHGEF1, ARHGEF11, and ARHGEF12 hold also a RGS homology (RH) domain through which they exert a GAP action that leads to deactivation of heterotrimeric $G\alpha_{13}$. Thus, the three AKAP13 subfamily members act both as effectors and GAPs for heterotrimeric G proteins, thus behaving as the main points of convergence between signaling cascades that are controlled by heterotrimeric G proteins and those that are regulated by RhoGTPases.[13] Collectively, whereas the GAP function requires the presence of an RH domain, the effector action may be shared by the DH-PH tandem domain (hereafter indicated as DH-PH domain) of different RhoGEFs and may rely on an allosteric communication between RhoA and heterotrimeric G protein binding sites. Other RhoGEFs like Tiam1 (gene name: TIAM1) act also as effectors of small GTPases. Indeed, N-terminal to the DH domain of TIAM1 is a Ras-binding domain (RBD).[40] Active Ras interacts specifically, albeit weakly, with the RBD of TIAM1, and this interaction is sufficient to stimulate the ability of the RhoGEF to activate Rac *in vivo*. Son of sevenless 1 (Sos1, gene name: SOS1) integrates the activation of Ras and Rac through distinct Cdc25-homology and DH domains, respectively.[41] Collectively these data depict a complex regulatory scenario, in which allosteric communications through the DH-PH domain is expected to play a central role at least for selected members, an important aspect that, together with evolutionary-driven deformations, has been addressed in this study.

The inferences from the whole set of crystallographic structures of Dbl-family RhoGEFs strengthens the power of the computational approaches employed in this study to unveil structure−function relationships within or among superfamilies of proteins.

EXPERIMENTAL SECTION

Principal Component Analysis of the DH-PH Domain of the Dbl Family RhoGEFs. Information on the evolutionary flexibility within the Dbl family of RhoGEFs was captured through PCA (by means of the Wordom software).[42] Resting on the assumption that the major collective modes dominate the functional dynamics of a system, information on such global motions can be inferred from the atomic fluctuations by means of the PCA. The latter allows the decomposition of the atomic fluctuations into a set of principal components (eigenvectors of the covariance matrix of positional fluctuations) that describe the concerted motions of these atoms (e.g., the Cα-atoms). The technique is based on the diagonalization of such covariance matrix producing a set of eigenvector and eigenvalue pairs in which the eigenvector and the eigenvalue describe,

DOI: 10.1021/acs.jcim.5b00122
J. Chem. Inf. Model. 2015, 55, 1878−1893

respectively, direction and amplitude of the concerted atomic motion (a mode). The atomic components of an eigenvector provide a quantitative measure of the participation of each Cα-atom to the collective motion described by the corresponding eigenvector. The subspace spanned by the major modes of collective fluctuations (e.g., a number of eigenvectors describing 90% of the total variance) is accordingly often referred to as "essential subspace (ES)". When applied on a pseudotrajectory made of a set of homologous structures as we have done in this study, eigenvectors from PCA may describe the "evolutionary-sampled" patterns of flexibility of the considered proteins, whereas the associated set of eigenvalues indicates the impact of each deformation on the total variance. In this framework, PCA has to be done on the structurally conserved regions shared by the homologous proteins, i.e. portions of the DH-PH domain in the case of Dbl RhoGEFs. Such portions were, hence, identified through a multiple structural alignment procedure by means of the Multiprot software.[43] All the subfamily representatives with a crystallographic structure containing both the DH and PH domains were considered. These concerned RhoGEFs either in their Rho-bound state (GEF^{bnd}) or in their free state (GEF^{unb}) (Table S1). The corresponding multiple sequence alignment (MSA), refined to improve secondary structure matches, is shown in Figure S1. The amino acid conservation profiles for the DH-PH structural core were calculated by using the identity histograms method within the Chimera software.[44]

The conserved core of the DH-PH domain includes 213 Cα-positions. Building of the data set for PCA followed DaliLite[45] searches by using the family representative structure of ARHGEF12 (PDB code: 1X86, chain A) as a query. A final set of 54 structures containing the conserved DH-PH structural core was obtained (Table S1) and subsequently subjected to PCA.

ENM Analyses of the Conserved DH-PH Domain. The observed robustness of global modes to details in atomic coordinates or specific interatomic interaction and their insensitivity to the specific energy functions and parameters that define the force field provided support to the development of simplified, i.e. coarse-grained (CG), descriptions of protein structures such as the ENM (reviewed in ref 2). The latter rely on the fact that the property that apparently dominates the shape of global modes is the network of inter-residue contacts, which is a purely geometric quantity defined by the overall shape or native contact topology of the protein. In recent years, ENM-based NMA (ENM-NMA), contributed significantly to improving our understanding of the collective dynamics of a number of allosteric proteins (reviewed in ref 2). In this study, the ENM-NMA implemented in the Wordom software[42] was instrumental in inferring the physical deformability patterns of the structurally conserved region in the most representative members of the Dbl family of RhoGEFs in either the GEF^{unb} or GEF^{bnd} forms.

As reported in a previous study, three ENM approaches differing for the degree of coarse graining have been implemented in the Wordom Software, the basic ENM, the Vibrational Subsystem Analysis (ENM-VSA),[46] and the Rotation Translation Block (ENM-RTB).[47,48] The difference between the basic ENM and the other two methods essentially resides in the strategies employed to reduce the dimensionality of the Hessian matrix for efficient diagonalization. According to the ENM-VSA, employed in this study, the total system is divided in two components: a) the subsystem, which is defined as the region of interest (i.e., part of the system that controls functionality) and is hence subjected to the vibrational analysis; and b) the environment, which consists of the less important remaining portions of the molecule, whose effects on the subsystem are implicitly taken into account.

For each ENM simulation, a set of eigenvectors of dimension $3N_{ss}$ (where N_{ss} is the number of Cα-atoms within the conserved DH-PH domain) was obtained. Structural portions not included in the DH-PH domain, e.g. additional domains or the RhoGTPases for the GEF^{bnd} states, were implicitly considered as environment during calculations.

Different conformational subspaces, defined by equal numbers of eigenvectors, either derived from ENM calculations or by PCA, can be compared through a number of similarity metrics, e.g. dot product, cumulative square overlap, Root Mean Square Inner Product (RMSIP).[4] As an example, the latter is defined by the following equation

$$\mathrm{RMSIP} = \sqrt{\frac{1}{s}\sum_{m=1}^{s}\sum_{n=1}^{s}(M_m \cdot N_n)^2} \tag{1}$$

where M_m is the m^{th} eigenvector from the reference ENM subspace, and N_n is the n^{th} eigenvector from the target subspace, either from ENM or PCA calculations, and s is an increasing number of eigenvectors up to the ES (in this study defined as the space described by the first 50 eigenvectors).

In order to characterize the conformational transition involved in the formation of the RhoGEF-RhoGTPase complex, we computed the involvement coefficients between ENM modes and the coordinate difference vectors obtained from the crystal structures of GEFs in their bound and unbound states using eq 2[49]

$$I_m = \frac{\sum_{k=1}^{3N} M_{mk}\Delta c_k}{\sum_{k=1}^{3N} M_{mk}^2 \sum_{k=1}^{3N} \Delta c_k^2} \tag{2}$$

where $\Delta c_k = c_k^{bound} - c_k^{unb}(c_k^{unb} - c_k^{bound})$ and $c_k^{bound(unb)}$ is the kth coordinate of the two conformers in the bound and unbound states, respectively.

If major motion of the protein consists of rigid structural domains moving relative to another, the deformation energy computed on all residue will likely distinguish between sites participating in a rigid body motion (small deformation energy) or in an internal structural change. The distribution of the deformation energy may hence serve to reveal flexible regions linking the rigid domains, such as the hinge sites.[50,51] The most likely hinge points have been investigated by computing the deformation energy (D_{mi}), directly obtained from interparticle distance fluctuations, for each particle i along each mode m from ENM, using eq 3 adapted from the works of[50,51]

$$D_{mi} = \sum_{j=1}^{p} \frac{1}{2} K(|\vec{R}_{ij}^{\,0} + \Delta\vec{R}_{mj} - \Delta\vec{R}_{mi}| - |\vec{R}_{ij}^{\,0}|)^2 / N\lambda_m^2 \tag{3}$$

where R_{ij}^0 is the equilibrium distance vector between particle i and j, $\Delta R_{mi(j)}$ is the displacement of atom i (or j) along the mth mode, p is the number of particles connected to residue i, K is the harmonic spring constant defined by eq 2, N is the total number of particles, and λ_m is the eigenvalue of the corresponding mth mode. It is possible to obtain the overall deformation energy D_i of a residue i by summing the contribution of each single eigenvector D_{mi} over the essential subspace.

DOI: 10.1021/acs.jcim.5b00122
J. Chem. Inf. Model. 2015, 55, 1878−1893

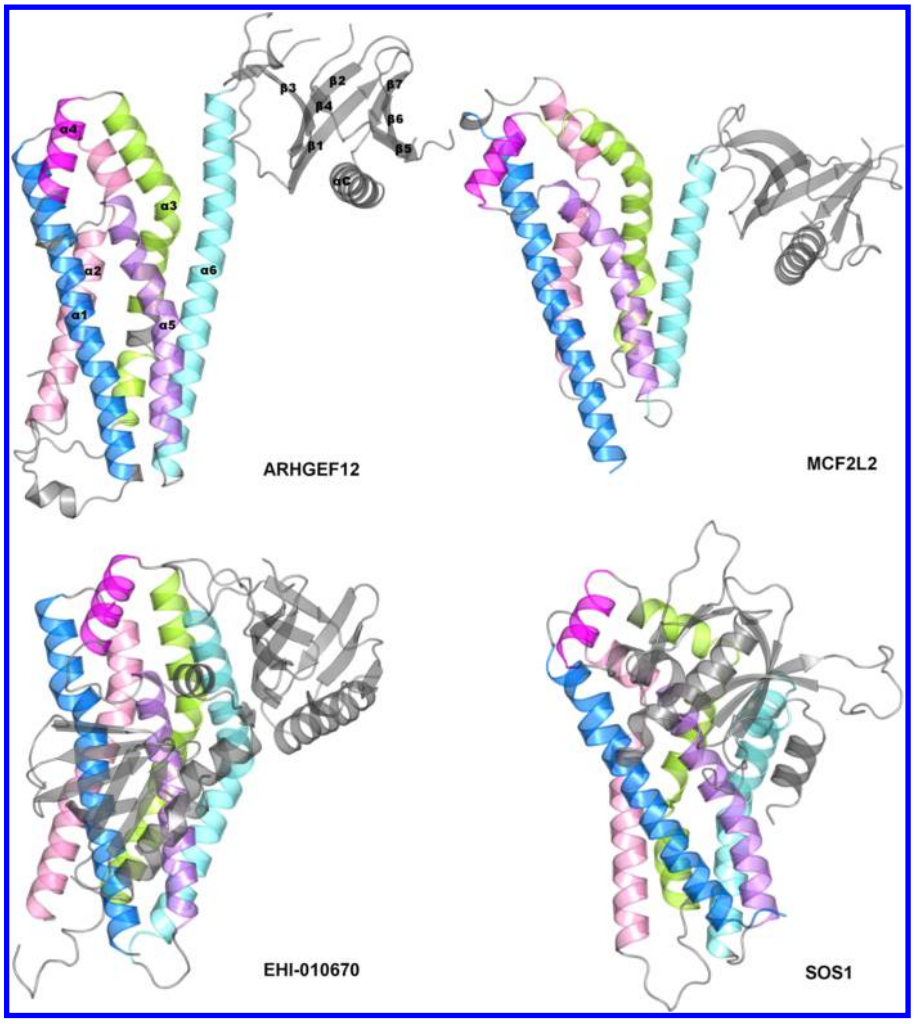


Figure 1. Crystal structures of selected RhoGEF representatives. The crystal structures of ARHGEF12, MCF2L2, EHI-010670, and SOS1 are shown in cartoons. All structures are in the GEF^{unb} state, EHI-010670 and SOS1 being in their autoinhibited form. The DH helices 1, 2, 3, 4, 5, and 6 are, respectively, blue, pink, green, magenta, violet, and cyan. The extra-DH domains are gray.

PSN Analysis. *Building of the PSG.* Building of the PSG was carried out by means of the PSN module implemented in the Wordom software.[42] PSN analysis is a product of graph theory applied to protein structures.[52] A graph is defined by a set of vertices (nodes) and connections (edges) between them. In a PSG, each amino acid residue is represented as a node, and these nodes are connected by edges based on the strength of noncovalent interactions between residues.[53] The strength of interaction between residues i and j (I_{ij}) is evaluated as a percentage given by eq 4

$$I_{ij} = \frac{n_{ij}}{\sqrt{N_i N_j}} \times 100 \quad (4)$$

where I_{ij} is the percentage interaction between residues i and j; n_{ij} is the number of atom−atom pairs between the side chains of residues i and j within a distance cutoff (4.5 Å); and N_i and N_j are normalization factors for residue types i and j, which account for the differences in size of the amino acid side chains and their propensity to make the maximum number of contacts with other amino acids in protein structures.

Thus, I_{ij} are calculated for all node pairs. At a given interaction strength cutoff I_{min}, any residue pair ij for which $I_{ij} \geq I_{min}$ is considered to be interacting and hence is connected in the PSG.

Residues making zero edges are termed as orphans, and those that make four or more edges are referred to as hubs at that particular I_{min}.

The following setting has been used for computation of the PSG: a) the distance cutoff for computing the interaction strength is 4.5 Å; and b) the minimal number of links a node must be involved in to be considered as a hub is ≥4. Such cutoff for hub definition relates to the intrinsic limit in the possible number of noncovalent connections made by an amino acid in protein structures due to steric constraints. The cutoff 4 is close to the upper limit. The majority of amino acid hubs indeed make from 4 to 6 links, with 4 being the most frequent value.

The PSGs concerning structural conserved regions of homologous proteins can be compared once assigned a position-dependent common numbering inferred from a structural sequence alignment (e.g., Figure S1). In this framework, the consensus network was computed as made of node and links occurring in ≥70% of all PSGs of the considered representatives.

Combining PSN with ENM To Search for the Shortest Communication Paths. The procedure has been previously described and validated.[5] The search for the shortest path(s) between pairs of nodes as implemented in Wordom relies on

DOI: 10.1021/acs.jcim.5b00122
J. Chem. Inf. Model. 2015, 55, 1878−1893

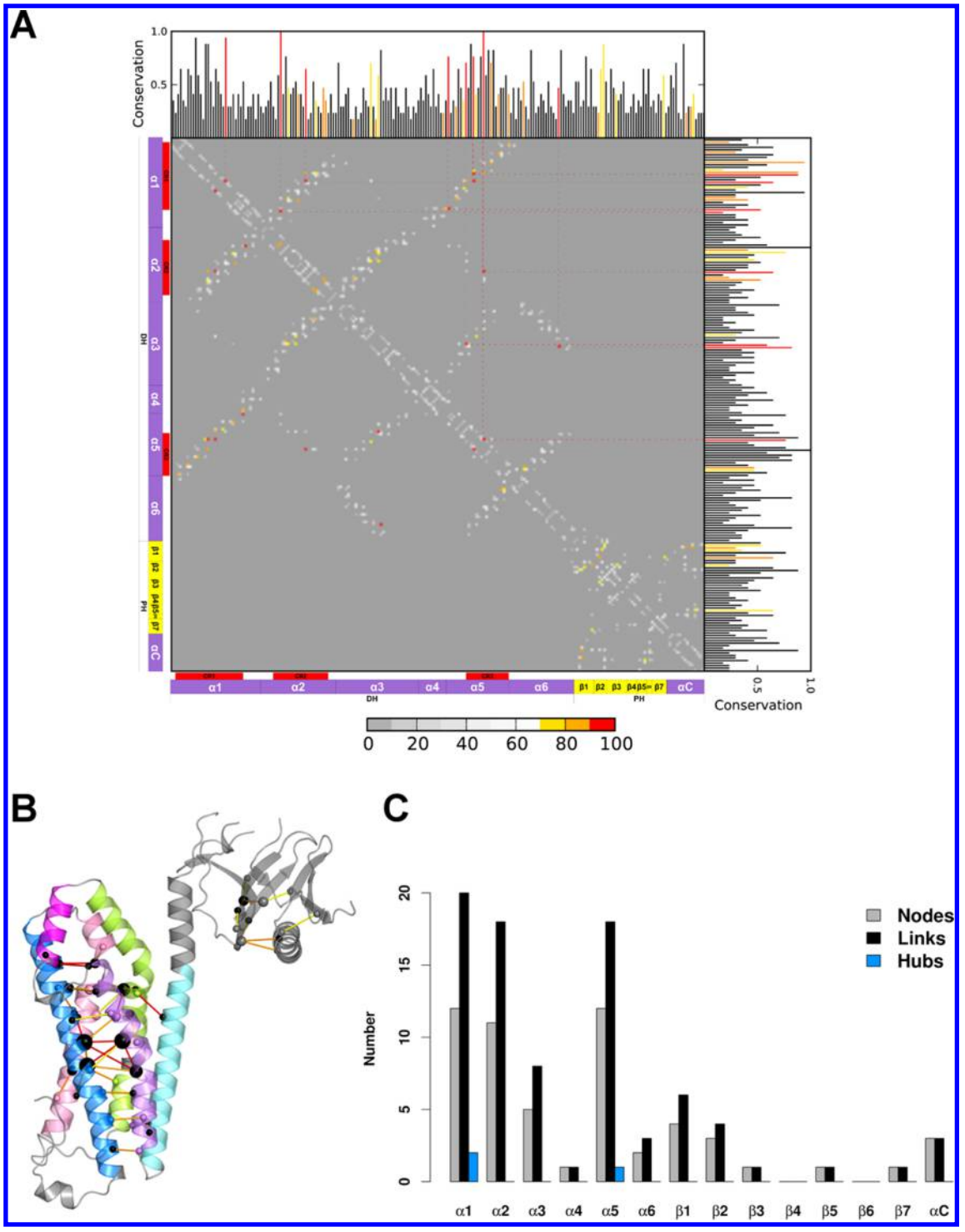


Figure 2. Relations between consensus structure network and primary sequences. A. The consensus PSN profiles and sequence conservation of the DH-PH domain are shown. The 2D matrix shows the link conservation profile from the consensus PSN of the DH-PH domain. The color scheme for the links with frequency ≥70% within the consensus network is the same as that in B. A gray scale is employed for links with lower frequency values. Top and right histograms indicate the amino acid identity profile from the multiple sequence alignment of the corresponding DH-PH portions (see Figure S1). Positions occupied by nodes involved in links with frequency ≥70% are colored from yellow to red according to the highest link frequency they are involved in (see below). B. The consensus network inferred from the networks of 15 representative members of the Dbl family in their unbound states and drawn on the crystal structure of ARHGEF12 (PDB ID: 1txd, chain A) is shown. Spheres centered on Cα-carbons represent nodes. The diameter of the spheres is proportional to the number of links made by the considered node. Link color refers to the recurrences of the link in the PSGs contributing to the consensus network. In this respect, yellow, orange, and red correspond, respectively, to the following frequency (F) ranges: 70% ≤ F ≤ 80%, 80% ≤ F ≤ 90%, and F ≥ 90%. Black spheres indicate the sites of mutations as inferred from the COSMIC database. C. Contributions of the secondary structure elements to the consensus network are shown. The network is expressed in terms of nodes, links, and hubs.

DOI: 10.1021/acs.jcim.5b00122
J. Chem. Inf. Model. 2015, 55, 1878−1893

Dijkstra's algorithm.[54] All possible residue pairs at a sequence distance greater than i, i+4 were set.

The matrix of correlated motions obtained from ENM-NMA intervenes in the filtering stage of the search for the shortest paths between pairs of nodes belonging to the same network cluster (i.e., a collection of nodes connected by at least one link). Thus, the ENM-based filtering stage consisted in retaining all those paths, in which at least one intermediate node holds correlated motions[5,55] with either one of the two extremities (i.e., the first and last amino acids in the path).

The relative number of residues holding correlated motions with either one of the two extremities is quantified by the correlation score.

Those paths that pass the filtering stage constitute the pool of paths of a system at given I_{min}, correlation coefficient, and minimum length cutoffs. The statistical analysis of such a pool of paths can lead to the building of global meta paths constituted by the most recurrent nodes and links in the pool.

The following parameters were set: a) minimum path length, excluding the two terminal nodes: 5; b) minimum correlation between a residue in the path and one of the apical residues for the path to be accepted: 0.6; c) sequence distance cutoff for two residues to be considered as path extremities: 5; d) node and link recurrence cutoff used in building meta paths: 20.

PSN analysis and search for communication pathways through the PSN-ENM method can now be performed through the WebPSN server (http://webpsn.hpc.unimore.it/).[6]

RESULTS

Insights into RhoGEF Structure through PSGs. The DH domain of RhoGEFs is characterized by six α-helices (i.e., α1-α6) organized in an up−down bundle architecture, that are roughly arranged along six main axes to form an oblong helical bundle that has been compared in appearance to a "chaise longue", with the seat back created by U-shaped arrangement of α-helices (Figure 1). The largest structural differences among different DH domains reside in the length and orientation of α6 (Figure S1). DH domains share three conserved regions (CR1-CR3), which pack to form the core. CR1, CR2, and CR3, located respectively on α1, α2, and α5, along with conserved residues within the C-terminus of the domain (α6), form a contiguous patch that constitutes the bulk of the GTPase binding surface. The PH domain is formed by a roll architecture made by seven antiparallel β-strands capped with a characteristic C-terminal helix (αC); the first part is a linker region, which contacts directly the α6 of the DH domain and presents different traits in random coil conformation (Figure S1).

Herein, the DH-PH conserved core of 15 representative members of the Dbl family was translated into structure graphs and inspected through the PSN analysis (see Experimental Section). RhoGEFs were basically taken in their GEF^{unb} state unless only the GEF^{bnd} state was available.

The secondary structure elements contributing the most to the structure network in terms of nodes, links, and hubs lay on the DH domain, whereas the PH domain contributes very little (Figure 2). As for the DH domain, the shortest helix, α4, gives the lowest contribution to the PSG. While the number of nodes from each helix is comparable among the different subfamilies, the number of links and hubs is rather variable. In particular, ARHGEF1, ARHGEF11, ARHGEF12, and Trio-N (gene name: TRIO) are characterized by a greater number of links and hubs on α1, α3, and, to a lesser extent, α2 compared to the other RhoGEFs (Figure S2). As for the PH domain, the αC helix and, to a lesser extent, the β1-β3 strands are the secondary structure portions that contribute the most.

The consensus network at 70% of recurrence (see Experimental Section) comprises 56 linked nodes and 42 interconnecting links (colored yellow, orange, and red in Figure 2, Table S2). 64% of such nodes (36 out of 56) are invariant (i.e., they recur in 100% of the networks). Helices α1, α2, and α5 hold 26 out of the 36 invariant nodes (i.e., present in the structure network of all the considered systems). Some of these nodes (nodes α1:15, α1:18, α5:11, where the Greek symbol is the secondary structure element, the number before the colon is the number of such secondary structure element, and the two digits after the colon indicate the amino acid position in that secondary structure element) behave also as hubs, suggesting a potential role in the stability core of the protein. The remaining invariant nodes are located on α3, α4, α6, β1, β2, β3, and β5 (Figure 2). On the other hand, only three links are shared across the family, i.e. α1:15-α5:11, α3:18-α6:20, and α5:11-α5:15. In line with the considerations above, less conserved portions of the core display low conservation in terms of structure networks. This is the case of α6, which is characterized by a variable contact profile with either α3 or α5.

Those residues highlighted as invariant by the analysis of primary sequence conservation profiles over the DH-PH domain are all present in the consensus PSN, and two of them, F(α2:8) and Y(α5:15), mediate invariant links with residues at positions α1:30 and both α5:11 (either K, R, or H) and α2:18 (either H, Q, N, or S), respectively (Figure 2).

Collectively, 73.1% (83.3% when considering only the DH domain) of the amino acids shared by ≥70% of the aligned sequences (Figures 2 and S1) participate as nodes in the consensus network. 27.6% (44.4% when considering only the DH) of them mediate invariant links, and 17.2% (27.7% when considering only the DH) behave as hubs. Thus, the few almost invariant positions in the sequence alignment make invariant links in the structure network and may hold hub behavior. They belong to the DH domain and are likely to participate in the stability core and/or in function retention. Relations between conservation in sequence and in structure network, however, concerns only the almost invariant positions. In fact, almost 59% of the amino acids involved in consensus links hold sequence conservation <50% and they fall essentially in α1, α2, α3, and α5 (Table S2 and Figure 2). Thus, sequence similarity does not correlate with conservation in intramolecular connections (i.e., accounted for by consensus links or consensus nodes). As an example, positions with low sequence conservation, such as α2:18, α3:18, α5:8, and α1:30, participate in the consensus network.

Remarkably, 81% of nodes from the consensus network in the DH domain belong to the three CRs. From the COSMIC (Catalogue of Somatic Mutations in Cancer; http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/)[56] database we annotated those sites found mutated in the human sequences of the RhoGEFs considered in this study. These COSMIC mutation sites constitute 63% of the nodes in the consensus network of DH (Figure 2); 77% of such mutation sites fall in the CRs, in particular CR1 and CR2. Thus, α1 and α2 participate more frequently in the consensus network and display a high concentration of COSMIC mutation sites (Figure 2).

DOI: 10.1021/acs.jcim.5b00122
J. Chem. Inf. Model. 2015, 55, 1878−1893

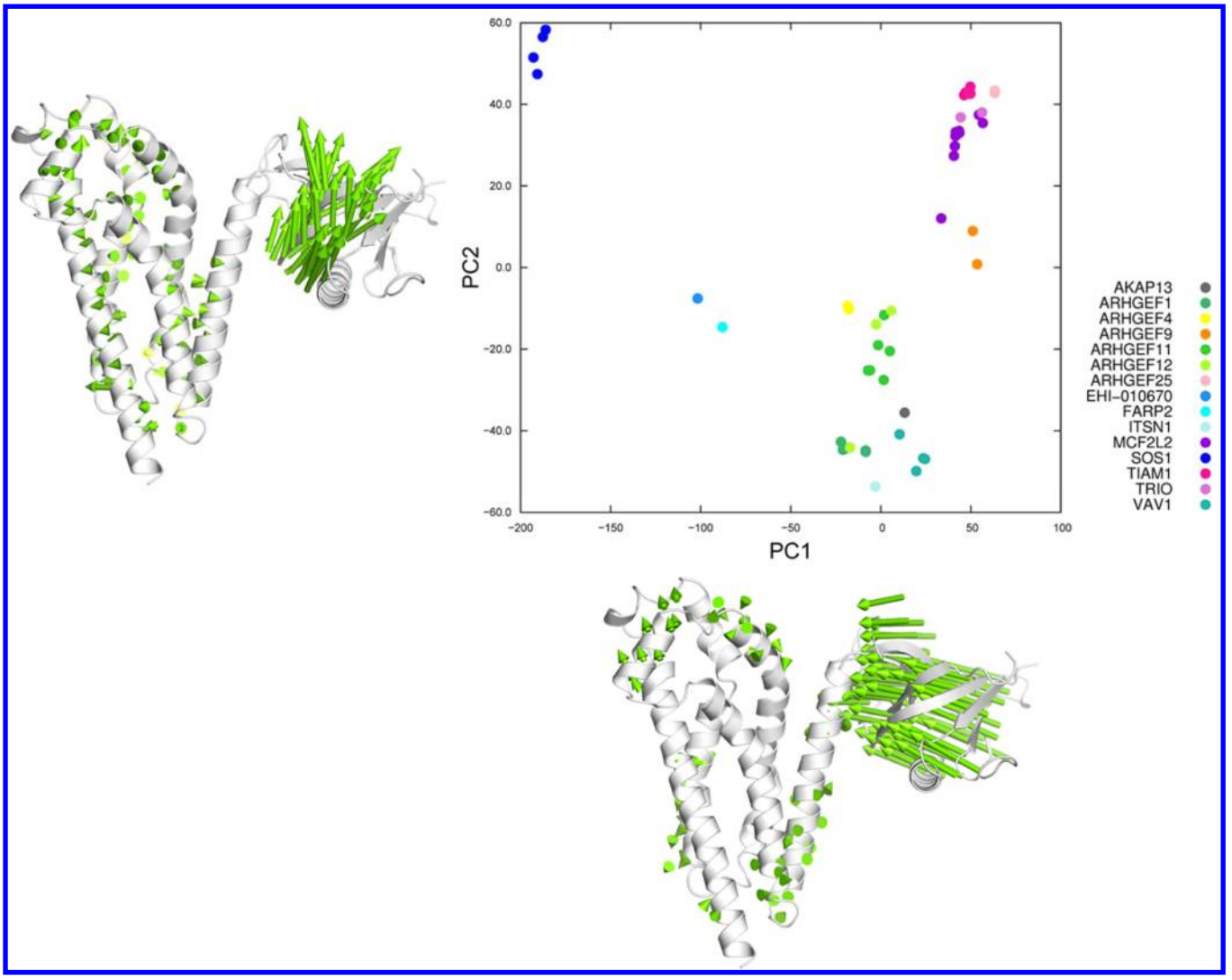


Figure 3. Collective motions from PCA of 54 representatives of the Dbl family. The Cα-atom displacements along PC1 and PC2 are shown represented as dots colored according to the 15 RhoGEF representatives. Color selection is such that it highlights RhoGEF separation in four clusters by PC1. In this respect, members of clusters 1, 2, 3, and 4 are colored according to green, red, light blue, and dark blue hues, respectively. The projections of the 213 Cα-atom along PC1 and PC2 are shown as well represented as green arrows holding the same direction of the eigenvector.

As for PH, COSMIC mutation sites in the consensus network essentially concern nodes on the β2 strand and αC (Figure 2).

Evolutionary-Driven Deformations Linked to Functional Specialization. All the available crystal structures of RhoGEFs of the Dbl family (54 structures, Table S1) in their GEF^{unb} or GEF^{bnd} states provided the data set for the PCA aimed at inferring structural deformations driven by either evolution or by function. The analysis was performed on the 213 Cα-carbon atoms corresponding to the conserved DH-PH structural core (see Experimental Section). The total variance of the data set is 6846, and it is entirely described by the first principal components (PCs). Indeed, PC1, PCs1-3, and PCs1-10 account, respectively, for 58%, 88%, and 99% of the total variance. In contrast, when the analysis is performed on the sole DH domain (57 structures), the total variance is 440, and it is distributed over a higher number of PCs (PC1 = 35%, PCs1-3 = 68%, and PCs1-10 = 92%).

The structural deformations of the DH-PH described by the first two PCs mainly consist in roto-translational motions of the C-terminal half of α6 and PH with respect to DH. Displacements along PC1 cluster the data set essentially into four groups comprising the following: a) AKAP13, ARHGEF1, ARHGEF4, ARHGEF11, ARHGEF12, ITSN1, and VAV1 (cluster 1), b) ARHGEF9, Dbs (gene name: MCF2L2), ARHGEF25, TIAM1, and TRIO (cluster 2), c) FARP2 and EHI (cluster 3), and d) SOS1 representatives (cluster 4) (Figure 3).

The dramatic separation of cluster 4 from the other three clusters is due to a marked deformation in the architecture of the DH domain. Such deformation is localized on the C-terminal half of α6 that, together with the linked PH, occupies part of the Rho binding surface thus exerting autoinhibition (Figure 1). Another significant separation along PC1 concerns cluster 3 and, again, is linked to the bending of the C-terminal half of α6 toward α5 and the joint movement of the PH domain. Differently from the members of clusters 3 and 4, in the members of cluster 2, the C-terminal half of α6 is misfolded and bent away from the helix bundle, the PH moving jointly. Members of cluster 1, which include the Lbc-subfamily, hold

DOI: 10.1021/acs.jcim.5b00122
J. Chem. Inf. Model. 2015, 55, 1878−1893

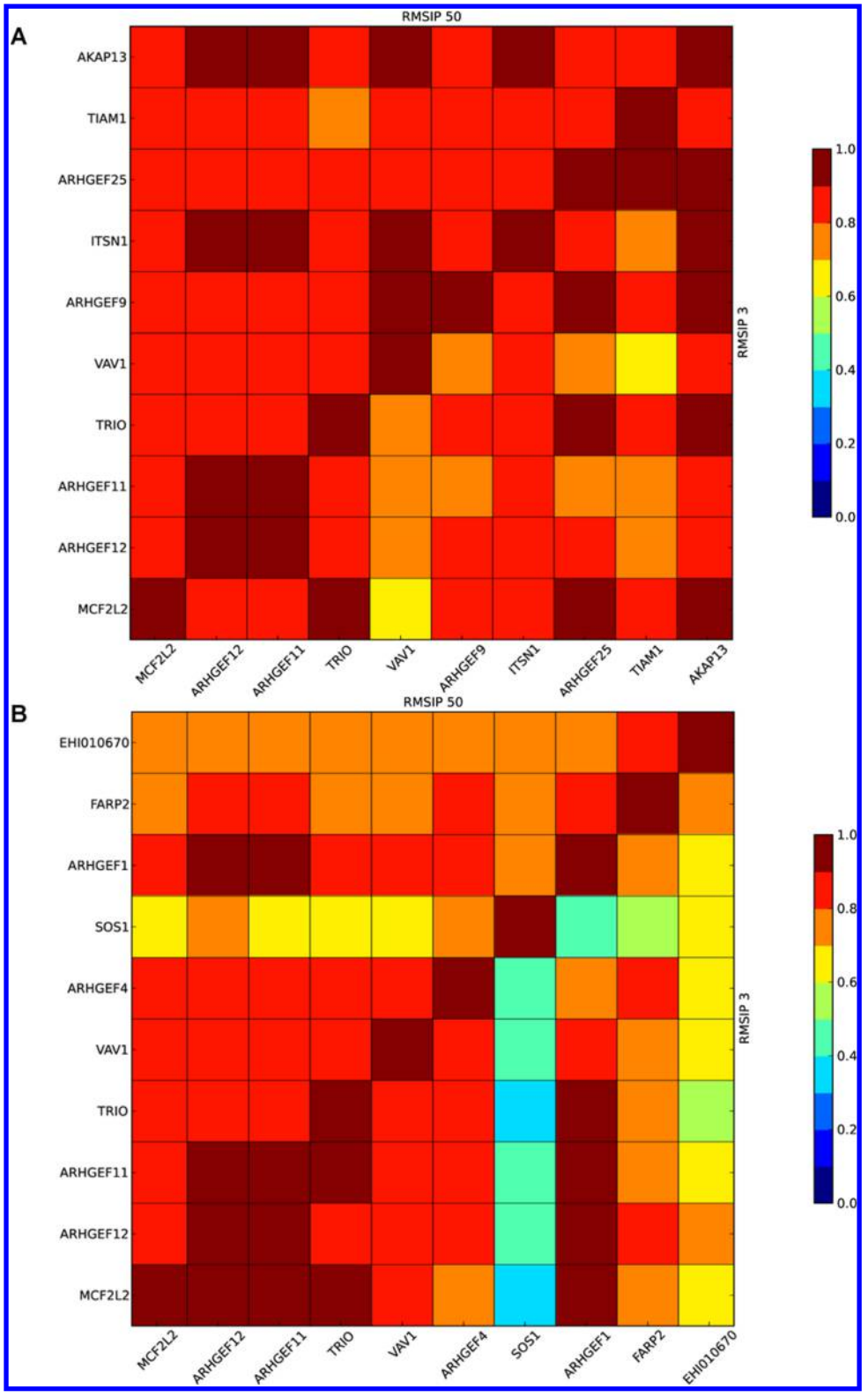


Figure 4. Similarities in the NMs of the considered RhoGEFs. The RMSIPs of the first 3 and first 50 normal modes (NMs) for the GEF^{bnd} (A) and GEF^{unb} forms (B) are shown. It is worth recalling that, according to PCA, the data set is essentially divided into four groups comprising the following: 1) AKAP13, ARHGEF1, ARHGEF4, ARHGEF11, ARHGEF12, ITSN1, and VAV1 (cluster 1), 2) ARHGEF9, MCF2L2, ARHGEF25, TIAM1, and TRIO (cluster 2), 3) FARP2 and EHI (cluster 3), and 4) SOS1 (cluster 4).

the lowest deformations of $\alpha 6$. Collectively, the members of clusters 3 and 4, on the left side of cluster 1, share a deformation of $\alpha 6$ toward the helix bundle, linked to autoinhibition, whereas the members of cluster 2, on the right side of cluster 1, tend to share deformations of $\alpha 6$ away from the helix bundle. As a consequence, members of cluster 2 show the maximal detachment of the PH domain from the GTPase binding site. Indeed, for some members of cluster 2,

DOI: 10.1021/acs.jcim.5b00122
J. Chem. Inf. Model. 2015, 55, 1878−1893

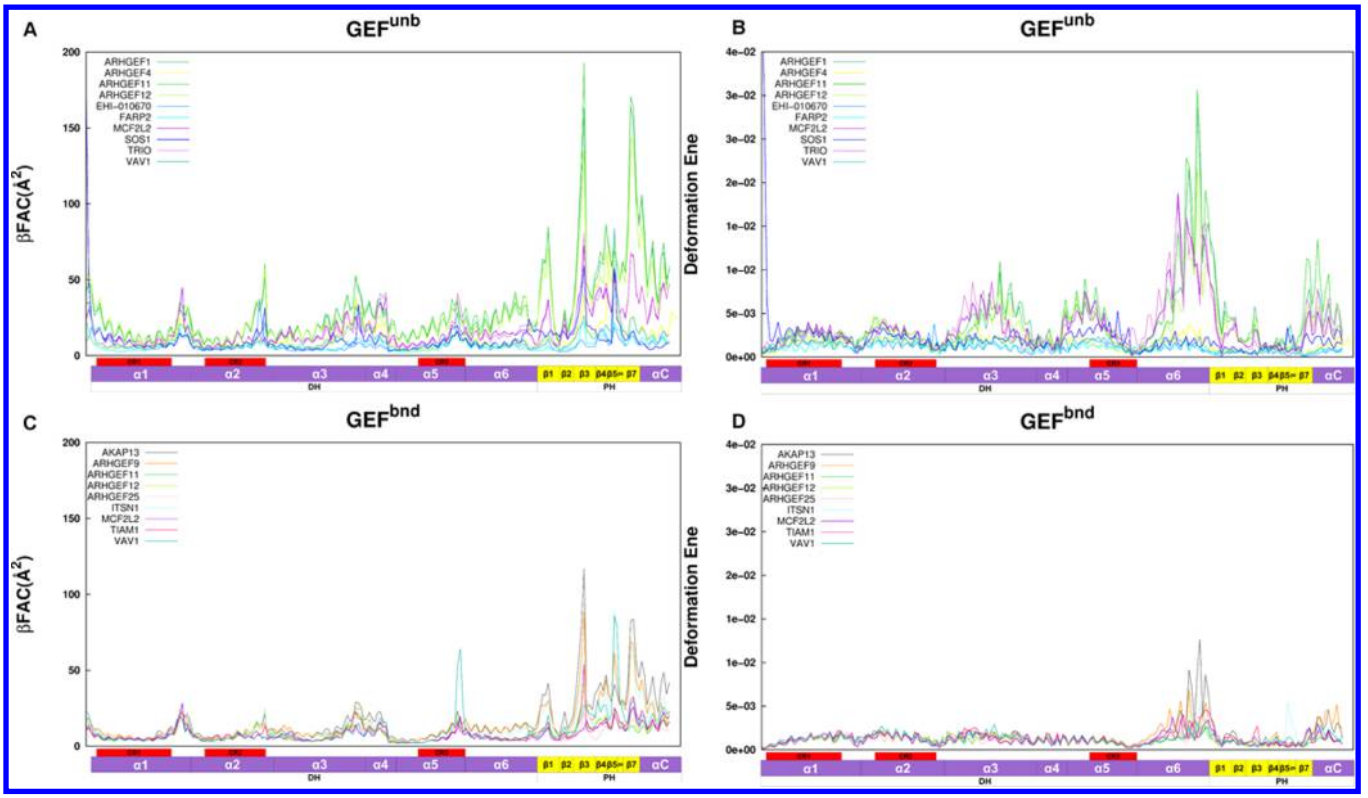


Figure 5. Flexibility patterns and mechanical properties of the considered RhoGEFs. The theoretical β-factors (A and C) and the deformation energies (B and D) are plotted for the GEF^{unb} (A and B) and GEF^{bnd} (C and D) states of 15 RhoGEF representatives. The color code of lines is the same as the one employed for distinguishing the RhoGEF representatives in the PCA plot according to their separation in four clusters (Figure 2). In this respect, members of clusters 1, 2, 3, and 4 are colored according to green, red, light blue, and dark blue hues, respectively.

two portions of the PH domain, the $\alpha6/\beta1$ linker and the $\beta2/\beta3$ turn, interact with regions of the DH domain (i.e., $\alpha2$, $\alpha3$, and the C-term of $\alpha1$) lying on an opposite face with respect to the Rho binding site (Figure 1). This may be instrumental in forming the binding site for interacting proteins, as in the case of ARHGEF25.[37]

PC2 does not segregate the proteins in clusters. It rather describes Z- and Y-rotations of the PH domain relative to the main axis of $\alpha6$, accounting in part for differences between GEF^{unb} and GEF^{bnd} states of the same RhoGEF (Figure 3).

Collectively, major deformations are described by PC1. Being essentially independent of the functional state of the RhoGEF, they are rather indicative of an evolutionary divergence of the DH domain essentially restricted to $\alpha6$ and, to a lower extent, $\alpha3$ and $\alpha4$. Evolutionary-driven deformations of $\alpha6$, which necessarily influence the orientation of the PH domain, are likely linked to functional specialization inside the Dbl family. Such specialization may relate to divergences in the modality to keep the inactive state, i.e. depending on whether the Rho binding site is masked by the C-terminal half of $\alpha6$ that carries the PH domain or by other domains or not masked at all. Evolution-driven deformations associated with the GEF^{bnd} state could be seen as well even if marginal compared to those linked to autoinhibition.

Insights into the Intrinsic Flexibility of the DH-PH Domain. Insights into functional dynamics within the Dbl family were gained by the ENM-VSA approach (see Experimental Section). Intrinsic flexibility and mechanical properties were investigated on all family representatives, whereas the dynamics of transitions between GEF^{unb} and GEF^{bnd} states was investigated only for the RhoGEFs holding crystal structures of the DH-PH domain in both states, i.e. ARHGEF11, ARHGEF12, MCF2L2, TRIO, and VAV1. To allow for trans-member comparisons, only the structurally conserved regions of the DH-PH domain were explicitly considered.

The dynamics of the conserved core in RhoGEF family representatives is highly overlapping, either when the ES or only the first three NMs are taken into account (Figure 4). In fact, RMSIPs are above 0.8 for most of the considered representatives, similarity being significantly higher in the GEF^{bnd} state compared to the GEF^{unb} one, likely due to the stiffening caused by the interacting G protein.

For both functional states, the highest similarities are found within and between the PCA-derived clusters 1 and 2 (i.e., involving ARHGEF1, ARHGEF11, ARHGEF12, MCF2L2, and TRIO). This is more evident by considering the first 3 NMs than the first 50 ones (Figure 4A). On the contrary, in line with inferences from PCA, SOS1, and, to a lower extent, EHI and FARP2 display the lowest similarity, divergence being higher for the first three NMs, likely due to the peculiar autoinhibitory conformation of $\alpha6$. For the GEF^{bnd} states, the addition of AKAP13 and ITSN1 further expands the similarity in collective motions inside cluster 1 and between clusters 1 and 2. Remarkably, AKAP13 displays the most extended similarity among the homologues in the GEF^{bnd} states (Figure 4B).

The relations between intrinsic flexibility and mechanical properties of the conserved DH-PH tandem domain were inferred by comparing theoretical β-factors[4,5,57] and deformation energies (D_i). Dbl family RhoGEFs display similar patterns of flexibility. In general, flexible regions are preceded by hinge

DOI: 10.1021/acs.jcim.5b00122
J. Chem. Inf. Model. 2015, 55, 1878−1893

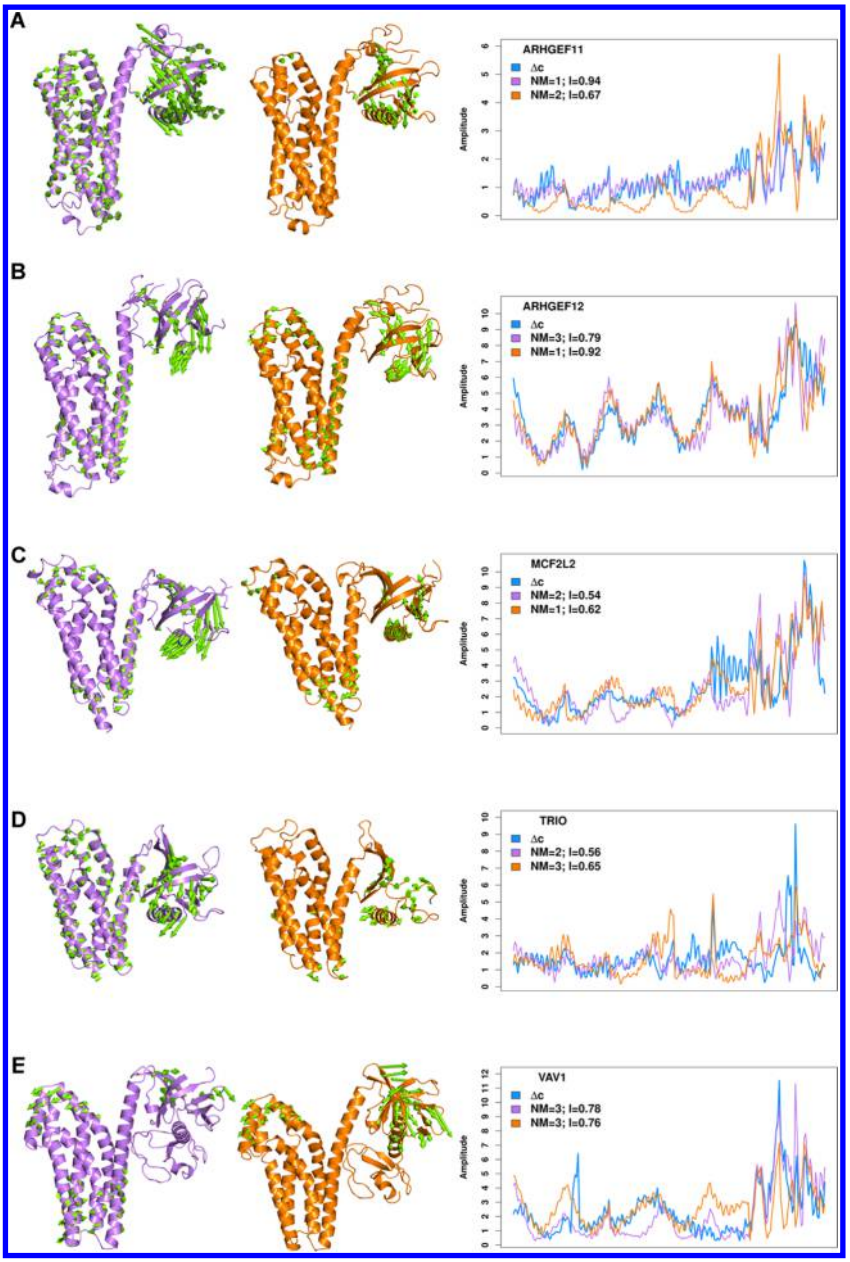


Figure 6. NMs inferred from ENM-VSA relative to the transition $GEF^{unb} \leftrightarrow GEF^{bnd}$. The C$\alpha$-atom projections of the NM showing the highest overlap (accounted for by the involvement coefficient I) with the coordinate difference vector (Δc) calculated between the GEF^{unb} and GEF^{bnd} states of ARHGEF11 (A), ARHGEF12 (B), MCF2L2 (C), TRIO (D), and VAV1 (E) are shown, represented as green arrows whose direction is the same as that of the NM. The arrows start from the Cα-atoms of the GEF^{unb} structure (violet) for the $GEF^{unb} \rightarrow GEF^{bnd}$ transition, whereas they start from the Cα-atom of the GEF^{bnd} state (orange) for the $GEF^{bnd} \rightarrow GEF^{unb}$ transition. The plots on the right display for each RhoGEF the amplitude of the NMs and of the coordinate difference vector (light blue).

points so that the higher the flexibility is the higher the deformation energy is.

The PH is generally more mobile than the DH domain, mobility in terms of rigid-body motions being higher in GEF^{unb} compared to the GEF^{bnd} states (Figure 5A and 5C). The most important hinge point (i.e., highest deformation energy) is located at the interdomain junction (i.e., C-terminal half of α6 from DH and β1 from PH; Figure 5B and 5D) and permits concerted motions of the PH domain with respect to the DH domain. Other minor hinge points can also be observed on α3 and α5 from the DH domain and on αC from PH (Figure 5B and 5D).

All the above considerations on flexibility and hinge points in the GEF^{unb} state count for all the considered RhoGEFs except ARHGEF4, EHI, FARP2, SOS1, and VAV1, which behave as rather rigid structures with no flexible regions nor hinge points (Figure 5), which is in line with their autoinhibited state. Indeed, in ARHGEF4 the DH-PH domain makes extensive contacts with the SH3 domain N-terminal to the DH.[26] In detail, the C-terminal half of α6 for EHI, FARP2, and SOS1, the N-terminal α-helix for EHI,[25] the PH2 domain for FARP2,[29] and the PH domain for SOS1[58] make contacts with the DH domain. Finally, in the GEF^{unb} state of VAV1, the mobility of the PH domain is restricted by the presence of the CH and AC domains N-terminal to the DH domain;[22] while

DOI: 10.1021/acs.jcim.5b00122
J. Chem. Inf. Model. 2015, 55, 1878−1893

the α-helix linker makes contacts with DH, CH makes contacts essentially with PH, by reducing significantly the intrinsic flexibility of the tandem domain. Indeed, if ENM calculations are performed by excluding those additional domains, fluctuations of the PH domain increase (Figure S3).

The ENM-VSA served also to infer the RhoGEF collective motions productive for recognizing the RhoGTPase. To this end, we measured the involvement coefficient (I) between the ENM-derived NMs and the coordinate difference vector (Δc) calculated between the GEF^{unb} and GEF^{bnd} states (see Experimental Section). For all five considered systems (i.e., those RhoGEFs with both states known at atomic resolution), the transition between the two states is described by one of the first three NMs when calculations are performed on the GEF^{bnd} state (Figures 6 and S4). Indeed, for that state, there is almost always one low-frequency NM that predominantly overlaps to the Δc with an I index $\geq$0.6 (Figures 6 and S4). In contrast, for the GEF^{unb} state, in two out of the five considered systems (i.e., MCF2L2 and TRIO) an overlap $\geq$0.6 with the transition vector could not be achieved by a single NM, suggesting that a more complex dynamics is required to transit from the GEF^{unb} state to the GEF^{bnd} one than vice versa. Collectively, the most significant NMs describe concerted motions of the PH domain with respect to the DH domain, the latter being quite rigid in line with theoretical β-factors. Incidentally, ARHGEF11 is the unique case, in which the highest overlap with Δc is reached by a NM (NM1) computed on the GEF^{unb} state and such overlap is significantly higher than the highest overlap computed on the GEF^{bnd} state. This peculiar behavior may be due, at least in part, to the fact that the GEF^{unb} DH-PH structure of ARHGEF11 holds GTP-RhoA bound to the PH, rather than being fully unbound as in the other cases.

The overlaps of each of the first 10 NMs highlight the high similarity in collective motions between ARHGEF11 and ARHGEF12 in line with their sequence and structure similarities; indeed, they are members of the Lbc-subfamily and of cluster 1 from PCA. As already seen for the RMSIPs computed on the first 3 or 50 NMs, the single-NM overlaps concerning the GEF^{bnd} states of ARHGEF11 and ARHGEF12 are higher than those concerning the GEF^{unb} ones (Figure S4). In fact, while the GEF^{unb} states reach overlaps $\geq$0.8 for seven out of the first ten NMs, the GEF^{bnd} states do so for all the first ten NMs. High single-NM overlaps are also reached by MCF2L2 and TRIO, both members of cluster 2 from PCA, which contrarily to ARHGEF11 and ARHGEF12 show higher overlaps in the GEF^{unb} state than the GEF^{bnd} one. In fact, whereas for GEF^{unb} all the first 10 NMs reach overlaps $\geq$0.8, in the GEF^{bnd} state only the first 3 NMs reach overlaps $\geq$0.8, whereas NMs 5−10 reach overlaps $\leq$0.6. Along this line, each of the first 4 NMs of GEF^{unb} MCF2L2 show high overlaps with the corresponding NMs of inactive ARHGEF11 and ARHGEF12 (Figure S4). Remarkably, low frequency collective motions of ARHGEF11, ARHGEF12, MCF2L2, and TRIO inferred from the same functional state are more overlapping between each other than the ones inferred from the GEF^{unb} and GEF^{bnd} states of the same RhoGEF. This suggests that Rho binding is associated with higher variability in the collective motions of the DH-PH domain than changes in primary sequence.

Selected RhoGEFs Display Allosteric Communication between the Binding Sites of Rho and Regulatory Proteins. As already remarked in the Introduction, the regulatory mechanism of Dbl RhoGEFs is quite complex also because it relies on the action of the domains outside the DH-PH as well as on the interaction between DH-PH and other proteins. Thus, depending on the regulatory mechanisms of Rho activation, selected RhoGEFs may act as allosteric proteins, whereas some others do not. In this study, the PSN-ENM method served to infer possible allosterism in representative RhoGEFs (Table S3). The analysis of structural communication relied on the computation of all possible shortest communication pathways according to a methodology already described.[5,6,59] For each considered protein, a coarse view of the global communication propensity was inferred through global meta paths, i.e. coarse paths made of the most recurrent nodes and links in the path pool. In this study, meta paths are made of nodes present in $\geq$20% of the path pool (i.e., "frequent nodes") and of links satisfying both conditions of being present in $\geq$20% of the paths and of connecting "frequent nodes".

Independently of the functional state, i.e. GEF^{unb} or GEF^{bnd}, and considering all together the global meta paths inferred from all the considered structures, α1, α5, and α6 from DH contribute the most to those paths in absolute terms (Figures 7 and 8, Table S3). Relative to the helix length, α5 contributes with the highest number of positions to the communication pathways (80% for α5 and <50% for the other helices). This highlights the centrality of α5 in the structure communication through the DH domain. The three CRs contribute significantly as well to meta paths. In fact, 48% of the DH nodes found in meta paths come from the CRs and represent 60% of the CR positions. The aromatic amino acids at positions α1:18, α2:18, α3:17, and α5:15 are among the most recurrent nodes in the meta paths of the considered structures (Table S3). Remarkably, α1:18 and α2:18, together with the fully conserved Y(α5:15) and the highly conserved E/D(α1:15), participate in the interface among the three helices.

As for the GEF^{unb} forms, the structures corresponding to an autoinhibited state tend to hold an extended communication that in some cases like ARHGEF4 involves DH-PH and extra domains whereas in other cases is essentially localized on the DH-PH domain (Table S3). As for ARHGEF4, major communication pathways involve DH portions participating in the interface with the SH3 and PH domains, i.e. α1, α5, and α6. Such interface is overlapping with that of the Rho GTPase. Passing through α5, the communication pathways extend also to the site opposite to that of SH3, i.e. positions α3:12, α3:15, and α3:18, all occupied by aromatic amino acids. The main part of the meta path extends inside and between the PH and SH3 domains. If calculations are made by excluding the SH3 domain, meta paths are almost entirely lost. This suggests that the DH-PH domain of ARHGEF4 acquires communication pathways because of the interactions with the SH3 domain, likely linked to autoinhibition. Surprisingly, in a recently released crystal structure of "closed" autoinhibited ARHGEF9 very similar to the corresponding functional state of ARHGEF4,[27] no meta paths could be found at the same recurrence cutoff as ARHGEF4. This may be due or not to the low resolution of the crystal structure. Similarly to ARHGEF4, the "open" structure of SH3-deprived nonautoinhibited ARHGEF9 does not hold any meta path as well. Collectively, the unbound forms of ARHGEF9, analogously to TRIO, seem not to hold relevant structural communication. In those cases like EHI, FARP2, and SOS1, in which autoinhibition is essentially operated by the peculiar C-terminal half of α6 that makes extensive contacts with DH, communication pathways

DOI: 10.1021/acs.jcim.5b00122
J. Chem. Inf. Model. 2015, 55, 1878−1893

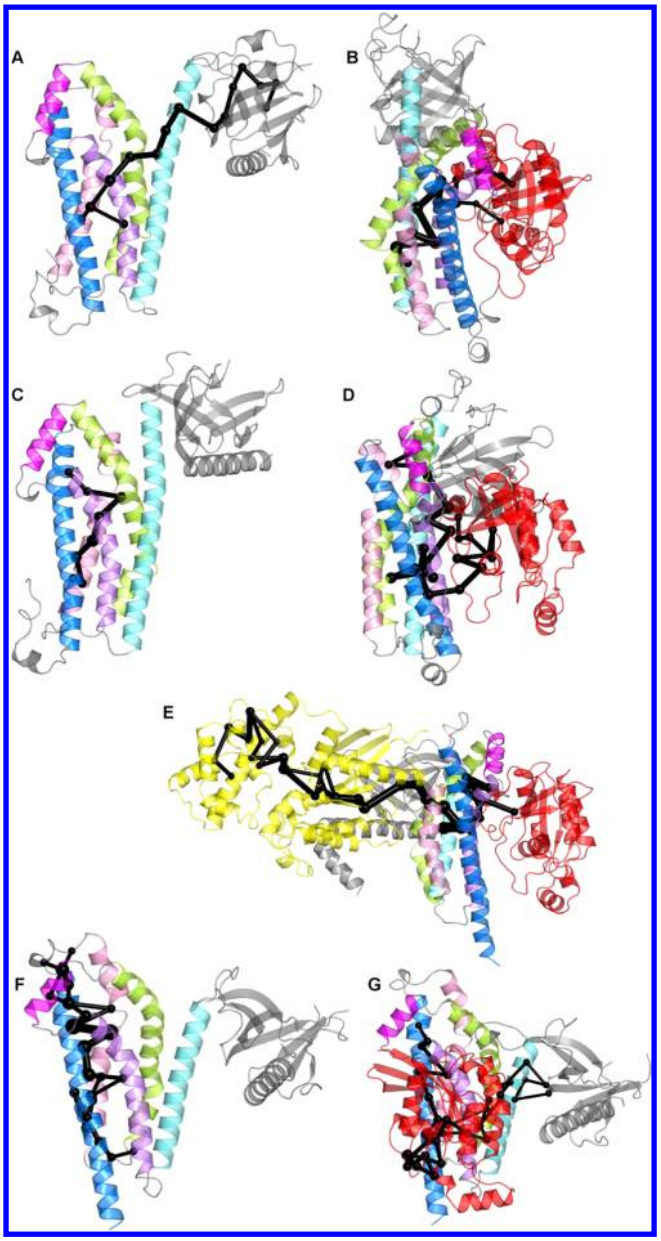


Figure 7. Selected meta paths. The meta paths of unbound and RhoA-bound ARHGEF11 (A and B, respectively), unbound and RhoA-bound ARHGEF12 (C and D, respectively), RhoA and $G\alpha_{q/i}$-bound ARHGEF25 (E), and unbound and RhoA-bound MCF2L2 (F and G) are shown. The DH is colored according to the helices, the PH domain is gray, RhoA is red, and $G\alpha_{q/i}$ is yellow. Meta paths are black; the radius of the spheres and the width of links are proportional to the recurrence of nodes and links, respectively, in the path pool.

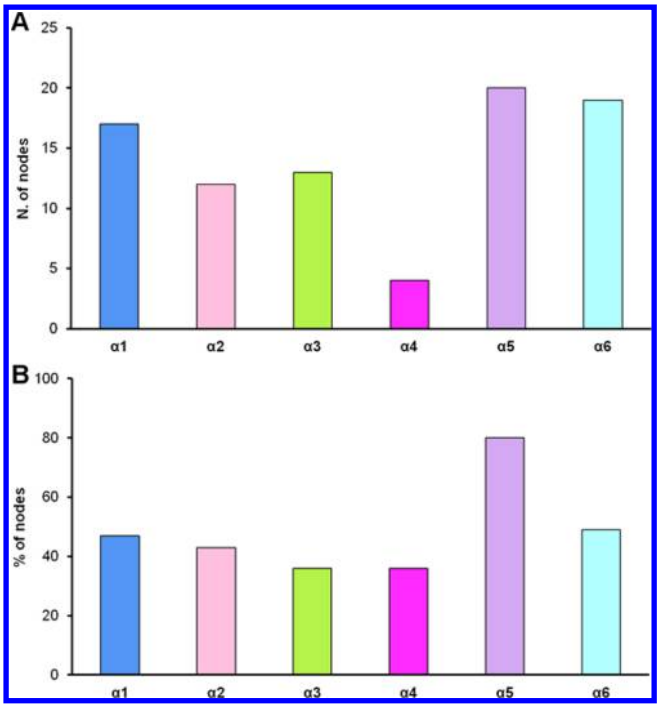


Figure 8. DH helix positions contributing to meta paths. The number (A) and percentage (B) of sites in the DH helices contributing to meta paths in all considered RhoGEFs is shown.

tend to extend along the main axis of the helix bundle and heavily involve the DH domain. In fact, they remain almost the same after removal of the additional domains, RGS and α-helical linker for EHI and PH2 for FARP2. Meta paths are, however, different in these three RhoGEFs. While in EHI the relevant meta path extends from α1 to the RGS domain (i.e., α5-α1-α3-α6-RGS), in FARP2 three major meta paths could be found made of α5-α6, α5-PH2, and α6-PH links, and in SOS1 the meta path involves the α1-α5, α2-α3, and α3-α5 interfaces (Figure 7 and Table S3). As for the last autoinhibited form considered in this study, VAV1, major communication pathways extend horizontally from α1 to the PH domain passing essentially through α3, α5, and α6. The N-terminal CH domain, which occupies the binding site of Rac1, is not involved in meta path links with the DH domain but only in one link with the PH domain. Indeed, removal of the CH domain does not change significantly the meta path on the DH domain and at the DH-PH interface, suggesting a marginal implication of such a path in autoinhibition.

Nonautoinhibited GEF^{unb} states are essentially characterized by a structural communication between α1 and α2 that, similarly to a number of autoinhibited forms, involves nodes distributed along the main helix axes. This is particularly true for MCF2L2 (Figure 7 and Table S3). The communication does not involve the PH domain in all these RhoGEFs except for ARHGEF11, in which the major meta path shows an α1-α2-α5-α3-α6-PH communication (Figure 7 and Table S3). This may be linked to the fact that the PH domain of the considered structure is bound to an activated RhoA molecule. In this respect, the peculiar extension of the meta path to the PH domain may be linked to the autoregulatory mechanism that characterizes this RhoGEF and that involves its role as an effector of GTP-RhoA. In line with their high homology, ARHGEF11 and ARHGEF12 share an entire portion of the meta path that holds conserved amino acids in the two RhoGEFs: E(α1:15)-H(α2:18)-E/Q(α5:8)-F(α3:17)-I(α5:5) (Table S3).

In the GEF^{bnd} states, the communication essentially involves the interface between the DH domain and switches 1 and 2 (sw1 and sw2, respectively) of Rho (Table S3). This implies a more marked involvement of nodes on α1, α5, and α6, compared to the GEF^{unb} state. This is evident for MCF2L2, in which the most recurrent communication pathways extend from α1 to the PH domain via the interface with RhoA. The involvement of the RhoGEF-Rho interface is even stressed in CBII, in which the major meta path includes one link between α3 (position α3:14) and the sw2 of Cdc42. Indeed, in this

DOI: 10.1021/acs.jcim.5b00122
J. Chem. Inf. Model. 2015, 55, 1878−1893

RhoGEF, the protrusion of α3 in between α5 and α6 makes α3 participating in the G protein binding site. In the case of CBII, major communication pathways seem to be instrumental in maintaining intermolecular interactions between DH domain and Cdc42 in line with the highly extended surface exposed to the G protein. Such surface is hypothesized to serve also to recognize gephyrin in competition with Cdc42.[60]

Remarkably, in the GEF^{bnd} state of the RH domain-holding RhoGEFs, ARHGEF11 and ARHGEF12, the meta path is characterized by the involvement of F(α3:8) linked to the fully conserved Y(α5:15) and to C(α6:18). Position α3:8 precedes the α-helical insertion typical of Lbc-subfamily RhoGEFs and is occupied by an aromatic amino acid only in the members of such subfamily. Incidentally, in AKAP13 this situation is visible upon reducing the recurrence cutoff for links and nodes in the meta paths. A peculiarity of ARHGEF11 and ARHGEF12 is that they act both as RGS and effectors of heterotrimeric G$\alpha_{12/13}$. Their meta paths describe a communication between the DH-RhoA interface and the site of the helix bundle opposite to the RhoA binding site, i.e. the N-term of α3 that, in addition to the RH domain, is likely to participate in the binding site of the heterotrimeric G protein. This hypothesis is supported by the meta paths computed on the crystal structure of ARHGEF25 in complex with AlF_4^--activated G$\alpha_{i/q}$ and RhoA, in which G$\alpha_{i/q}$ engages both the DH surface opposite to the RhoA binding site and portions of the PH domain of ARHGEF25 (Figure 7). Global meta paths computed on such a complex show the presence of a communication between RhoA and G$\alpha_{i/q}$ mediated by nodes in α5, α6, α2, and α3. Analogously to G$\alpha_{12/13}$-coupled RhoGEFs, α3 contributes with an aromatic amino acid in its N-terminus, F(α3:14). Position α3:14 follows the α-helical insertion and holds an aromatic amino acid in many RhoGEFs but not in the members of the Lbc-subfamily (Figure 7).

Another RhoGEF considered in this study and acting as an effector of Ras GTPases is MCF2L2, whose Rac1 docking site resides on the PH domain.[61] The extension of the major meta path from the DH-RhoA interface to the PH domain may be related to a possible regulatory mechanism operated by Rac1 on the MCF2L2 GEF activity toward RhoA.

DISCUSSION

Dbl family RhoGEFs play a central role in cell biology, and many of them are oncoproteins whose overexpression is often associated with highly malignant forms of cancers.[10]

They are multidomain proteins whose common structural feature is a DH-PH domain deputed to the GEF activity. While their common GEF action is accomplished by the DH domain, their regulatory mechanisms are highly variegated and depend on the PH and the additional domains as well as on interacting proteins.

In this study, information on the intrinsic flexibility and structural communication features of the DH-PH domain were inferred from the analysis of almost all the crystallographic structures released so far (i.e., 57 structures) and correlated with functionality.

Autoinhibition turned out to be the main target of functional specialization by Dbl family RhoGEFs. Indeed, major evolutionary-driven deformations of the structurally conserved DH-PH domain are linked to the mechanisms adopted by the RhoGEF to prevent Rho binding. In fact, collective motions described by the first eigenvector separate those RhoGEFs, like SOS1, EH1, and FARP2 where the Rho binding site is masked by tertiary contacts with the C-terminal half of α6 and the linked PH, from those RhoGEFs in which the Rho binding site is masked by contacts with other domains or not masked at all. Evolution-driven deformations associated with the GEF^{bnd} state could be seen as well even if they are marginal compared to those linked to autoinhibition. Thus, apart from the autoinhibited forms highlighted above, grouping by the first principal component is indicative of a general structural/dynamics homogeneity independently of the functional state. This is in line with the high RMSIPs reached by the first three normal modes for most RhoGEFs except SOS1 and, to a lower extent, FARP2 and EHI.

The analysis of theoretical β-factors, in line with PCA and NMA, highlights the α6 of DH and the linked PH domain as the portions of the tandem domain holding almost the totality of intrinsic functional dynamics, with the α6/β1 junction acting as a hinge point for the collective motions of PH that undergoes rigid-body motions. The set of representatives with both functional states resolved at the atomic detail is too small for allowing conclusions on state-transition dynamics. Whereas for some RhoGEFs like ARHGEF11 no major motions seem to be associated with RhoA binding,[62] for those RhoGEFs in which the C-terminal half of α6 undergoes bending and kinking for autoinhibition, more complex and larger collective motions of α6 and PH are expected to be required for Rho binding. This statement is supported by the very recent low resolution structure of autoinhibited ARHGEF9[27] (not considered in our analyses because of the resolution) whose DH-PH deviates from the GEF^{bnd} state by 8.5 Å Cα-RMSD.

Apart from α6, the DH domain does not undergo collective motions to bind Rho and to exert the RhoGEF function, which is rather regulated by PH and other proteins.[20,62−65] Thus, the GEF activity by the DH domain appears to be essentially linked to a static scaffolding and hub behavior, in which structural communication is expected to play a central role in the regulatory actions by other domains/proteins. Indeed, almost the entire consensus structure network and communication pathways involve the DH. In RhoGEFs, the relations between conservation in sequence and conservation in structure network essentially concern the almost invariant positions of the CRs. However, the gross of the consensus network relies on less conserved amino acids in the CRs and in α3 in line with the concept that folds keep better than sequences. In line with a recent PSN-based study on Rossmann fold-holding proteins,[66] we expect the lack of correlation between conservation in interaction pairs and conservation in primary sequence being ever emphasized if, instead of comparing homologous proteins as in this study, we would have compared analogous proteins sharing only the fold. Collectively, in Dbl family RhoGEFs, high and low conserved amino acids in α1, α2, α3, and α5 make almost the entirety of the conserved structure network, suggesting that those regions participate in the stability core of the DH domain, ultimately instrumental in GEF function. This is supported by the evidence that amino acid substitutions within CRs typically adversely affect the nucleotide exchange activity.[13] Thus, amino acids important for folding and for function coincide or crowd in the same portions of the DH domain, in line with the inferences that for the Rossmann fold nucleation and functional residues would coincide.[66]

Differently from the PSG analysis that was oriented to infer a consensus network across the Dbl family, the analyses of the shortest communication pathways were instead aimed at highlighting functional specialization within the family. Coarse

DOI: 10.1021/acs.jcim.5b00122
J. Chem. Inf. Model. 2015, 55, 1878−1893

representations of such pathways, i.e. the meta paths, were used as fingerprints of RhoGEF-specific communication. Meta paths catch the most crucial information flow in the network as they are made of the most frequent nodes and edges in all shortest communication pathways. Thus, meta path nodes and edges resemble nodes and edges with the highest degree of betweenness, as defined in a recent study.[66] The analysis of RhoGEF-specific meta paths highlighted a highly variegated scenario, which is in line with the complex regulatory mechanisms of these proteins. Autoinhibited forms generally share extended pathways almost always involving the Rho binding regions of DH and the portions participating in autoinhibition, which are highly variable within the Dbl family of RhoGEFs. Autoinhibition can indeed be made by additional domains at the N-terminus of DH,[13,25−28] at the C-terminus of PH,[29] by the PH domain,[37−39] or by both an additional N-terminal domain and PH.[41] This suggests that in the GEF^{unb} state of autoinhibited forms, structural communication is instrumental in preventing Rho binding.

In a number of RhoGEFs, like ARHGEF11 and MCF2L2, whose GEF activity is regulated by allosteric binding of RhoA and Rac1 to the PH domain, meta paths extend from the DH to the PH domain in either GEF^{unb} or GEF^{bnd} states, being hallmarks of allosteric regulatory mechanisms. Allosteric communication emerged also for those RhoGEFs acting as effectors of heterotrimeric G proteins. Indeed, for those RhoGEFs like ARHGEF11 and ARHGEF12 that act both as GAPs and effectors for heterotrimeric G proteins, the analysis of the GEF^{bnd} state highlighted a peculiar communication between RhoA binding region and an opposite site on $\alpha 3$ holding an aromatic amino acid. This finding is in line with evidence from NMR determinations indicating a communication between the Rho binding site and both $\alpha 2$ and $\alpha 3$.[62] Accordingly, in ARHGEF25, an effector of activated $G\alpha_q$, an aromatic amino acid in the N-terminal half of $\alpha 3$ is found as well to participate in the communication between RhoA and $G\alpha_q$ in the quaternary complex.[37] On these bases, we postulate that the N-terminal half of $\alpha 3$ acts as a binding region for the heterotrimeric G protein in all RhoGEFs acting as effectors of $G\alpha$ proteins. The involvement of the N-terminal half of $\alpha 3$ in $G\alpha$ binding is also supported by site-directed mutagenesis data showing that mutation of W(α3:9) in ARHGEF1 abrogates $G\alpha_{13}$-dependent Rho activation.[36]

CONCLUSIONS

The extensive dissection of the crystal structures of Dbl-family RhoGEFs indicated evolutionary-driven functional specialization as essentially linked to the variegate ways the PH and/or other domains autoinhibit GEF function. The study also highlighted the centrality of the DH domain in the structure networks and communication pathways, likely linked to a role as a scaffold and a hub domain. These features of the DH domain associate with lack of intrinsic dynamics, which is relegated to the C-terminal half of $\alpha 6$ whose collective motions are linked to those of PH and are expected to play a regulatory role in Rho binding and activation. Both the autoinhibition and the action as effectors of other GTPases pass through communication pathways that may respectively serve to keep the inactive state or to modulate GEF function via allosteric communication with interacting proteins.

The approach employed in this study being robust and not computational time-consuming is quite powerful to infer structure−function relationships in superfamilies or families of proteins.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.5b00122.

Tables S1−S3 and Figures S1−S4 (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: fanelli@unimo.it.

Present Address

†University of Hidelberg, Im Neuenheimer Feld 267 69120 Heidelberg, Germany

.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding

This study was supported by an Airc-Italy grant [IG14811] to F.F.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The employment of the PyMOL 0.99rc6 software for the realization of all drawings is acknowledged.

REFERENCES

(1) Van Wynsberghe, A. W.; Cui, Q. Conservation and Variation of Structural Flexibility in Protein Families. *Structure* **2010**, *18*, 281−283.

(2) Bahar, I.; Lezon, T. R.; Bakan, A.; Shrivastava, I. H. Normal Mode Analysis of Biomolecular Structures: Functional Mechanisms of Membrane Proteins. *Chem. Rev.* **2010**, *110*, 1463−1497.

(3) Kitao, A.; Go, N. Investigating Protein Dynamics in Collective Coordinate Space. *Curr. Opin. Struct. Biol.* **1999**, *9*, 164−169.

(4) Raimondi, F.; Orozco, M.; Fanelli, F. Deciphering the Deformation Modes Associated with Function Retention and Specialization in Members of the Ras Superfamily. *Structure* **2010**, *18*, 402−414.

(5) Raimondi, F.; Felline, A.; Seeber, M.; Mariani, S.; Fanelli, F. A Mixed Protein Structure Network and Elastic Network Model Approach to Predict the Structural Communication in Biomolecular Systems: The PDZ2 Domain from Tyrosine Phosphatase 1E As a Case Study. *J. Chem. Theory Comput.* **2013**, *9*, 2504−2518.

(6) Seeber, M.; Felline, A.; Raimondi, F.; Mariani, S.; Fanelli, F. WebPSN: a Web Server for High-throughput Investigation of Structural Communication in Biomacromolecules. *Bioinformatics* **2015**, *31*, 779−781.

(7) Colicelli, J. Human RAS Superfamily Proteins and Related GTPases. *Sci. Signaling* **2004**, *2004*, RE13.

(8) Jaffe, A. B.; Hall, A. Rho GTPases: Biochemistry and Biology. *Annu. Rev. Cell Dev. Biol.* **2005**, *21*, 247−269.

(9) Karlsson, R.; Pedersen, E. D.; Wang, Z.; Brakebusch, C. Rho GTPase Function in Tumorigenesis. *Biochim. Biophys. Acta, Rev. Cancer* **2009**, *1796*, 91−98.

(10) Cook, D. R.; Rossman, K. L.; Der, C. J. Rho Guanine Nucleotide Exchange Factors: Regulators of Rho GTPase Activity in Development and Disease. *Oncogene* **2014**, *33*, 4021−4035.

(11) Wirtenberger, M.; Tchatchou, S.; Hemminki, K.; Klaes, R.; Schmutzler, R. K.; Bermejo, J. L.; Chen, B.; Wappenschmidt, B.; Meindl, A.; Bartram, C. R.; Burwinkel, B. Association of Genetic

DOI: 10.1021/acs.jcim.5b00122
J. Chem. Inf. Model. 2015, 55, 1878−1893

Variants in the Rho Guanine Nucleotide Exchange Factor AKAP13 with Familial Breast Cancer. *Carcinogenesis* **2006**, *27*, 593−598.
(12) Feher, L. Z.; Pocsay, G.; Krenacs, L.; Zvara, A.; Bagdi, E.; Pocsay, R.; Lukacs, G.; Gyory, F.; Gazdag, A.; Tarko, E.; Puskas, L. G. Amplification of Thymosin Beta 10 and AKAP13 Genes in Metastatic and Aggressive Papillary Thyroid Carcinomas. *Pathol. Oncol. Res.* **2012**, *18*, 449−458.
(13) Rossman, K. L.; Der, C. J.; Sondek, J. GEF means go: Turning on Rho GTPases with Guanine Nucleotide-Exchange Factors. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 167−180.
(14) Snyder, J. T.; Worthylake, D. K.; Rossman, K. L.; Betts, L.; Pruitt, W. M.; Siderovski, D. P.; Der, C. J.; Sondek, J. Structural Basis for the Selective Activation of Rho GTPases by Dbl Exchange Factors. *Nat. Struct. Biol.* **2002**, *9*, 468−475.
(15) Cerione, R. A.; Zheng, Y. The Dbl Family of Oncogenes. *Curr. Opin. Cell Biol.* **1996**, *8*, 216−222.
(16) Whitehead, I. P.; Campbell, S.; Rossman, K. L.; Der, C. J. Dbl Family Proteins. *Biochim. Biophys. Acta, Rev. Cancer* **1997**, *1332*, F1−23.
(17) Erickson, J. W.; Cerione, R. A. Structural Elements, Mechanism, and Evolutionary Convergence of Rho Protein-Guanine Nucleotide Exchange Factor Complexes. *Biochemistry* **2004**, *43*, 837−842.
(18) Blomberg, N.; Baraldi, E.; Nilges, M.; Saraste, M. The PH Superfold: a Structural Scaffold for Multiple Functions. *Trends Biochem. Sci.* **1999**, *24*, 441−445.
(19) Zheng, Y. Dbl Family Guanine Nucleotide Exchange Factors. *Trends Biochem. Sci.* **2001**, *26*, 724−732.
(20) Liu, X.; Wang, H.; Eberstadt, M.; Schnuchel, A.; Olejniczak, E. T.; Meadows, R. P.; Schkeryantz, J. M.; Janowick, D. A.; Harlan, J. E.; Harris, E. A.; Staunton, D. E.; Fesik, S. W. NMR Structure and Mutagenesis of the N-terminal Dbl Homology Domain of the Nucleotide Exchange Factor Trio. *Cell* **1998**, *95*, 269−277.
(21) Rossman, K. L.; Campbell, S. L. Bacterial Expressed DH and DH/PH Domains. *Methods Enzymol.* **2000**, *325*, 25−38.
(22) Yu, B.; Martins, I. R.; Li, P.; Amarasinghe, G. K.; Umetani, J.; Fernandez-Zapico, M. E.; Billadeau, D. D.; Machius, M.; Tomchick, D. R.; Rosen, M. K. Structural and Energetic Mechanisms of Sooperative Autoinhibition and Activation of Vav1. *Cell* **2010**, *140*, 246−256.
(23) Aghazadeh, B.; Lowry, W. E.; Huang, X. Y.; Rosen, M. K. Structural Basis for Relief of Autoinhibition of the Dbl Homology Domain of Proto-oncogene Vav by Tyrosine Phosphorylation. *Cell* **2000**, *102*, 625−633.
(24) Crespo, P.; Schuebel, K. E.; Ostrom, A. A.; Gutkind, J. S.; Bustelo, X. R. Phosphotyrosine-Dependent Activation of Rac-1 GDP/GTP Exchange by the Vav Proto-Oncogene Product. *Nature* **1997**, *385*, 169−172.
(25) Bosch, D. E.; Kimple, A. J.; Manning, A. J.; Muller, R. E.; Willard, F. S.; Machius, M.; Rogers, S. L.; Siderovski, D. P. Structural Determinants of RGS-RhoGEF Signaling Critical to Entamoeba Histolytica Pathogenesis. *Structure* **2013**, *21*, 65−75.
(26) Murayama, K.; Shirouzu, M.; Kawasaki, Y.; Kato-Murayama, M.; Hanawa-Suetsugu, K.; Sakamoto, A.; Katsura, Y.; Suenaga, A.; Toyama, M.; Terada, T.; Taiji, M.; Akiyama, T.; Yokoyama, S. Crystal Structure of the Rac Activator, Asef, Reveals its Autoinhibitory Mechanism. *J. Biol. Chem.* **2007**, *282*, 4238−4242.
(27) Soykan, T.; Schneeberger, D.; Tria, G.; Buechner, C.; Bader, N.; Svergun, D.; Tessmer, I.; Poulopoulos, A.; Papadopoulos, T.; Varoqueaux, F.; Schindelin, H.; Brose, N. A Conformational Switch in Collybistin Determines the Differentiation of Inhibitory Post-synapses. *EMBO J.* **2014**, *33*, 2113−2133.
(28) Ahmad, K. F.; Lim, W. A. The Minimal Autoinhibited Unit of the Guanine Nucleotide Exchange Factor Intersectin. *PLoS One* **2010**, *5*, e11291.
(29) He, X.; Kuo, Y. C.; Rosche, T. J.; Zhang, X. Structural Basis for Autoinhibition of the Guanine Nucleotide Exchange Factor FARP2. *Structure* **2013**, *21*, 355−364.
(30) Buchsbaum, R. J.; Connolly, B. A.; Feig, L. A. Interaction of Rac Exchange Factors Tiam1 and Ras-GRF1 with a Scaffold for the p38 Mitogen-Activated Protein Kkinase Cascade. *Mol. Cell. Biol.* **2002**, *22*, 4073−4085.
(31) Diviani, D.; Soderling, J.; Scott, J. D. AKAP-Lbc Anchors Protein Kinase A and Nucleates Galpha 12-Selective Rho-Mediated Stress Fiber Formation. *J. Biol. Chem.* **2001**, *276*, 44247−44257.
(32) Fukuhara, S.; Chikumi, H.; Gutkind, J. S. Leukemia-Associated Rho Guanine Nucleotide Exchange Factor (LARG) Links Hetero-trimeric G proteins of the G(12) Family to Rho. *FEBS Lett.* **2000**, *485*, 183−188.
(33) Fukuhara, S.; Murga, C.; Zohar, M.; Igishi, T.; Gutkind, J. S. A Novel PDZ Domain Containing Guanine Nucleotide Exchange Factor Links Heterotrimeric G Proteins to Rho. *J. Biol. Chem.* **1999**, *274*, 5868−5879.
(34) Hart, M. J.; Jiang, X.; Kozasa, T.; Roscoe, W.; Singer, W. D.; Gilman, A. G.; Sternweis, P. C.; Bollag, G. Direct Stimulation of the Guanine Nucleotide Exchange Aactivity of p115 RhoGEF by Galpha13. *Science* **1998**, *280*, 2112−2114.
(35) Aittaleb, M.; Boguth, C. A.; Tesmer, J. J. Structure and Function of Heterotrimeric G Protein-Regulated Rho Guanine Nucleotide Exchange Factors. *Mol. Pharmacol.* **2010**, *77*, 111−125.
(36) Chen, Z.; Guo, L.; Hadas, J.; Gutowski, S.; Sprang, S. R.; Sternweis, P. C. Activation of p115-RhoGEF Requires Direct Association of Galpha13 and the Dbl Homology Domain. *J. Biol. Chem.* **2012**, *287*, 25490−25500.
(37) Lutz, S.; Shankaranarayanan, A.; Coco, C.; Ridilla, M.; Nance, M. R.; Vettel, C.; Baltus, D.; Evelyn, C. R.; Neubig, R. R.; Wieland, T.; Tesmer, J. J. Structure of Galphaq-p63RhoGEF-RhoA Complex Reveals a Pathway for the Activation of RhoA by GPCRs. *Science* **2007**, *318*, 1923−1927.
(38) Lenoir, M.; Sugawara, M.; Kaur, J.; Ball, L. J.; Overduin, M. Structural Insights into the Activation of the RhoA GTPase by the Lymphoid Blast Crisis (Lbc) Oncoprotein. *J. Biol. Chem.* **2014**, *289*, 23992−24004.
(39) Abdul Azeez, K. R.; Knapp, S.; Fernandes, J. M.; Klussmann, E.; Elkins, J. M. The Crystal Structure of the RhoA-AKAP-Lbc DH-PH Domain Complex. *Biochem. J.* **2014**, *464*, 231−239.
(40) Lambert, J. M.; Lambert, Q. T.; Reuther, G. W.; Malliri, A.; Siderovski, D. P.; Sondek, J.; Collard, J. G.; Der, C. J. Tiam1 Mediates Ras Activation of Rac by a PI(3)K-Independent Mechanism. *Nat. Cell Biol.* **2002**, *4*, 621−625.
(41) Innocenti, M.; Tenca, P.; Frittoli, E.; Faretta, M.; Tocchetti, A.; Di Fiore, P. P.; Scita, G. Mechanisms Through Which Sos-1 Coordinates the Activation of Ras and Rac. *J. Cell Biol.* **2002**, *156*, 125−136.
(42) Seeber, M.; Felline, A.; Raimondi, F.; Muff, S.; Friedman, R.; Rao, F.; Caflisch, A.; Fanelli, F. Wordom: A User-Friendly Program for the Analysis of Molecular Structures, Trajectories, and Free Energy Surfaces. *J. Comput. Chem.* **2011**, *32*, 1183−1194.
(43) Shatsky, M.; Nussinov, R.; Wolfson, H. J. A Method for Simultaneous Alignment of Multiple Protein Structures. *Proteins: Struct., Funct., Genet.* **2004**, *56*, 143−156.
(44) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera–a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.
(45) Holm, L.; Kaariainen, S.; Rosenstrom, P.; Schenkel, A. Searching Protein Structure Databases with DaliLite v.3. *Bioinformatics* **2008**, *24*, 2780−2781.
(46) Zheng, W.; Brooks, B. R. Probing the Local Dynamics of Nucleotide-Binding Pocket Coupled to the Global Dynamics: Myosin Versus Kinesin. *Biophys. J.* **2005**, *89*, 167−178.
(47) Tama, F.; Gadea, F. X.; Marques, O.; Sanejouand, Y. H. Building-Block Approach for Determining Low-Frequency Normal Modes of Macromolecules. *Proteins: Struct., Funct., Genet.* **2000**, *41*, 1−7.
(48) Durand, P.; Trinquier, G.; Sanejouand, Y.-H. A New Approach for Determining Low-Frequency Normal Modes in Macromolecules. *Biopolymers* **1994**, *34*, 759−771.

DOI: 10.1021/acs.jcim.5b00122
J. Chem. Inf. Model. 2015, 55, 1878−1893

(49) Marques, O.; Sanejouand, Y. H. Hinge-Bending Motion in Citrate Synthase Arising from Normal Mode Calculations. *Proteins: Struct., Funct., Genet.* **1995**, *23*, 557−560.
(50) Hinsen, K. Analysis of Domain Motions by Approximate Nnormal Mode Calculations. *Proteins: Struct., Funct., Genet.* **1998**, *33*, 417−429.
(51) Wang, Y.; Rader, A. J.; Bahar, I.; Jernigan, R. L. Global Ribosome Motions Revealed with Elastic Network Model. *J. Struct. Biol.* **2004**, *147*, 302−314.
(52) Vishveshwara, S.; Brinda, K. V.; Kannan, N. Protein Structure: Insights From Graph Theory. *J. Theor. Comput. Chem.* **2002**, *1*, 187−211.
(53) Vishveshwara, S.; Ghosh, A.; Hansia, P. Intra and Inter-Molecular Communications through Protein Structure Network. *Curr. Protein Pept. Sci.* **2009**, *10*, 146−160.
(54) Dijkstra, E. W. A Note on Two Problems in Connexion with Graphs. *Numer. Math.* **1959**, *1*, 269−271.
(55) Van Wynsberghe, A. W.; Cui, Q. Interpreting Correlated Motions Using Normal Mode Analysis. *Structure* **2006**, *14*, 1647−1653.
(56) Forbes, S. A.; Beare, D.; Gunasekaran, P.; Leung, K.; Bindal, N.; Boutselakis, H.; Ding, M.; Bamford, S.; Cole, C.; Ward, S.; Kok, C. Y.; Jia, M.; De, T.; Teague, J. W.; Stratton, M. R.; McDermott, U.; Campbell, P. J. COSMIC: Exploring the World's Knowledge of Somatic Mutations in Human Cancer. *Nucleic Acids Res.* **2015**, *43*, D805−811.
(57) Bahar, I.; Atilgan, A. R.; Erman, B. Direct Evaluation of Thermal Fluctuations in Proteins Using a Single-Parameter Harmonic Potential. *Folding Des.* **1997**, *2*, 173−181.
(58) Soisson, S. M.; Nimnual, A. S.; Uy, M.; Bar-Sagi, D.; Kuriyan, J. Crystal Structure of the Dbl and Pleckstrin Homology Domains from the human Son of Sevenless Protein. *Cell* **1998**, *95*, 259−268.
(59) Raimondi, F.; Felline, A.; Portella, G.; Orozco, M.; Fanelli, F. Light on the Structural Communication in Ras GTPases. *J. Biomol. Struct. Dyn.* **2013**, *31*, 142−157.
(60) Xiang, S.; Kim, E. Y.; Connelly, J. J.; Nassar, N.; Kirsch, J.; Winking, J.; Schwarz, G.; Schindelin, H. The Crystal Structure of Cdc42 in Complex with Collybistin II, a Gephyrin-Interacting Guanine Nucleotide Exchange Factor. *J. Mol. Biol.* **2006**, *359*, 35−46.
(61) Cheng, L.; Mahon, G. M.; Kostenko, E. V.; Whitehead, I. P. Pleckstrin Homology Domain-Mediated Activation of the Rho-Specific Guanine Nucleotide Exchange Factor Dbs by Rac1. *J. Biol. Chem.* **2004**, *279*, 12786−12793.
(62) Cierpicki, T.; Bielnicki, J.; Zheng, M.; Gruszczyk, J.; Kasterka, M.; Petoukhov, M.; Zhang, A.; Fernandez, E. J.; Svergun, D. I.; Derewenda, U.; Bushweller, J. H.; Derewenda, Z. S. The Solution Structure and Dynamics of the DH-PH Module of PDZRhoGEF in Iisolation and in Complex with Nucleotide-Free RhoA. *Protein Sci.* **2009**, *18*, 2067−2079.
(63) Chhatriwala, M. K.; Betts, L.; Worthylake, D. K.; Sondek, J. The DH and PH Domains of Trio Coordinately Engage Rho GTPases for their Efficient Activation. *J. Mol. Biol.* **2007**, *368*, 1307−1320.
(64) Oleksy, A.; Opalinski, L.; Derewenda, U.; Derewenda, Z. S.; Otlewski, J. The Molecular Basis of RhoA Specificity in the Guanine Nucleotide Exchange factor PDZ-RhoGEF. *J. Biol. Chem.* **2006**, *281*, 32891−32897.
(65) Jaiswal, M.; Gremer, L.; Dvorsky, R.; Haeusler, L. C.; Cirstea, I. C.; Uhlenbrock, K.; Ahmadian, M. R. Mechanistic Insights into Specificity, Activity, and Regulatory Elements of the Regulator of G-protein Signaling (RGS)-containing Rho-Specific Guanine Nucleotide Exchange Factors (GEFs) p115, PDZ-RhoGEF (PRG), and Leukemia-Associated RhoGEF (LARG). *J. Biol. Chem.* **2011**, *286*, 18202−18212.
(66) Bhattacharyya, M.; Upadhyay, R.; Vishveshwara, S. Interaction Signatures Stabilizing the NAD(P)-Binding Rossmann Dold: a Structure Network Approach. *PLoS One* **2012**, *7*, e51676.

DOI: 10.1021/acs.jcim.5b00122
J. Chem. Inf. Model. 2015, 55, 1878−1893