ISSN 0006-2979, Biochemistry (Moscow), 2012, Vol. 77, No. 5, pp. 435-443. © Pleiades Publishing, Ltd., 2012.
Original Russian Text © G. V. Novikov, V. S. Sivozhelezov, A. S. Shebanova, K. V. Shaitan, 2012, published in Biokhimiya, 2012, Vol. 77, No. 5, pp. 544-554.
Originally published in Biochemistry (Moscow) On-Line Papers in Press, as Manuscript BM11-325, March 18, 2012.

Classification of Rhodopsin Structures by Modern Methods of Structural Bioinformatics

G. V. Novikov[1], V. S. Sivozhelezov[1]*, A. S. Shebanova[2], and K. V. Shaitan[2]

[1]*Institute of Cell Biophysics, Russian Academy of Sciences, ul. Institutskaya 3, 142290 Pushchino, Moscow Region, Russia; fax: (4967) 330-509; E-mail: vsivo00@gmail.com; vsivo@icb.psn.ru*
[2]*Biological Faculty, Lomonosov Moscow State University, 119991 Moscow, Russia*

Received December 20, 2011
Revision received January 19, 2012

Abstract—We report a classification of the crystallographic structures of bovine and squid rhodopsins corresponding to different stages of their photocycles. Using the resource Protein (Structure) Comparison, Knowledge, Similarity, and Information server (ProCKSI, http://www.procksi.net/), selected spatial structures were compared on the basis of classification schemes (dendrograms). To compare the spatial structures of transmembrane proteins, optimal consensus was developed from methods implemented in ProCKSI. Structures were also clustered using principal component analysis, resulting in good agreement with the classification based on the ProCKSI consensus method. Analysis of the results revealed the basic movements of individual transmembrane domains of these proteins that we were able to relate to different stages of the photoactivation of rhodopsin. A combination of methods identified in this study can be used as an up-to-date analytical tool to study the conformational dynamics of membrane receptors.

DOI: 10.1134/S0006297912050033

Key words: receptor, rhodopsin, activation, conformation, photocycle, classification

G-protein coupled receptors (GPCRs) belong to the most extensive class of transmembrane receptors [1]. About 2000 genes from this class are now known, and they code about 900 different GPCRs. This class of membrane receptors is known to play an important role in different biological processes by mediating a response to such extracellular signals as hormones and neuromediators, odorous substances, and light. Therefore the therapeutic importance of GPCRs is evident, since they are primary targets of the majority of drugs [2-4].

GPCR proteins have pronounced structural dynamics; they can fold into different conformations depending on a number of external conditions [5]. Thus, a receptor molecule has the ability to interact with different ligands, which can eventually lead to changing of its conformation and initiation of specific biochemical cascades [6]. The mammalian photoreceptor rhodopsin (Fig. 1) is one of the best-studied seven-helical transmembrane proteins. The 11-*cis*-retinal chromophore is known to isomerize into the *trans*-form upon the absorption of a quantum of light. This reaction is considered to be the main precondition of minor conformational changes in individual transmembrane (TM) helices [7]. A few biophysical methods showed that the full cycle of rhodopsin photoactivation is accompanied by a significant shift in the intracellular end of the sixth helix by about 5-6 Å in the direction of the bundle plane of the α-helices. Moreover, a synchronous movement of the fifth as well as a small rotation of the seventh helix was observed (Fig. 2). As a result, the set of these conformational changes in the TM fragment lead to the opening of a cleft in the cytoplasmic area of this molecule, which a G-protein could subsequently enter.

Several structures of mammalian rhodopsin corresponding to different stages of receptor photoactivation have been reconstructed using X-ray crystallography. Comparative analysis of these experimental structures can be used to create a complete picture of rhodopsin photoactivation process, despite the possible lack of some intermediate structures in this cycle. Thus, studies of the conformational diversity of membrane receptors represented, for example, by existing experimental structures, will finally allow us to assess the relationship between the

Abbreviations: GPCR, G-protein coupled receptors; PCA, principal component analysis; PDB, Protein Data Bank; ProCKSI, "Protein (Structure) Comparison, Knowledge, Similarity and Information" Server; RMSD, root mean square deviation; TM, transmembrane.

* To whom correspondence should be addressed.

dynamics of their structure (motions of individual α-helices) and their function (signal transduction).

In this work, a structural and functional classification of rhodopsin experimental structures was created using modern methods of structural bioinformatics. It should be noted that the methods used in this paper had only been tested previously on globular (soluble) proteins. For example, a similar classification had already been constructed for a number of protein kinases [8]. It is known that soluble proteins generally have more diverse motifs of polypeptide chain packing as compared to the TM domain of membrane proteins. In the latter case, the hydrophobic membrane environment significantly limits possible packing variants for α-helices in the bundle of the TM domain. Thus, we had to test methods of structural comparison on topologically identical structures, only slightly differing in their spatial distribution of individual α-helices.

Using a particular combination of the above-mentioned methods, we were able to identify the main patterns of distribution of active and inactive forms of rhodopsin. Notably, the most reliable results were obtained using complex analysis of the data. The technique consists of (i) visual analysis of the resulting classification of experimental structures, (ii) search for answers in the literature to surfacing questions related to a specific crystallographic structure (hereinafter the original crystallographic work describing the specific 3D-structure is referred to as the primary source), and finally, (iii) our own verification of the primary source. The last stage was performed by comparing the root mean-square deviation (RMSD) between the structures in the study, as well as by visual study of 3D structures using the visualization program.

Fig. 1. Overall topology of seven-helix transmembrane receptor exemplified by mammalian rhodopsin (PDB code 3PQR). Gray scale shows the orientation of the chain, from extracellular N-terminus (light) to the intracellular C-terminus (dark). *Trans*-retinal ligand is shown as ball model.

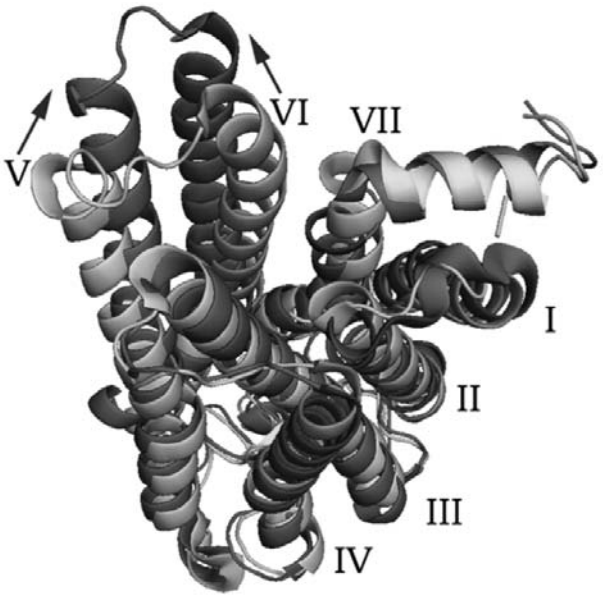

Fig. 2. Spatial superposition of inactive (dark) and activated (Meta-II*) state of mammalian rhodopsin. Structures are shown from the intracellular side. Light and dark shading show the inactive (PDB code 1GZM) and active (PDB code 3PQR) conformations of the receptor, respectively. Numerals indicate the number of transmembrane helices, and arrows show the direction of structural changes accompanying the transition from one form to another. The largest structural changes occur in the fifth and sixth helices. Simultaneously, the seventh helix also undergoes minor changes, rotating counterclockwise.

METHODS OF INVESTIGATION

Original data set. Three-dimensional structures of membrane receptors used in this paper were obtained from Protein Data Bank (http://www.rcsb.org/pdb). We studied all known crystallographic structures of rhodopsin available at the time of this study.

Protein (Structure) Comparison, Knowledge, Similarity, and Information server (ProCKSI). The main part of the work, including spatial alignment of the structures, construction and analysis of distance matrices, as well as clusterization of results, was carried out on an external server ProCKSI. ProCKSI server is a system for comparison and analysis of protein spatial structure with a number of internal algorithms [9]. It should be noted that in this study we decided to focus on spatial alignment algorithm

specifically, leaving methods of amino acid sequence alignment out of the scope of our research. This decision was due to the fact that different families of membrane receptors with low homology in amino acid sequence (typically 20-30%) have very high similarity in the spatial organization of the seven-helical transmembrane domain. As a result, we tested all the methods presented in ProCKSI, primarily based on pair-wise measuring RMSD for structures under study. This set included such techniques as DaliLite based on spatial alignment of distance matrices, TM-align based on dynamic programming, as well as the long-known method of combinatorial extension (CE) and, finally, the Vorolign technique based on Voronoi contacts. It is important to note that each of the methods presented, in addition to classical RMSD measurements, has specific estimation methods to assess the similarity of structures. Therefore, it is possible to determine in this study the most appropriate method for a specific type of structures. In the end, as output of ProCKSI, we can obtain classification distribution of the structures in a form similar to phylogenetic trees (dendrograms).

Evaluation of structural similarity of proteins. This section provides a brief description of the individual algorithms integrated into ProCKSI as well as their specific criteria for evaluation of spatial similarity of structures. The reader can obtain a more complete understanding of these methods from a detailed review.

The most traditional parameter to compare the spatial structures of proteins is RMSD [10, 11]. This value represents the average distance between the atoms (usually of the main chain only) of aligned proteins. It is known that when dealing with any dynamic system, one should take account of spatial deviations of the neighborhood under consideration. In this case, this indicator may as well include mean-square deviations of the system under study:

$$\text{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_i^2}\ , \qquad (1)$$

where δ_i is the distance between each of N identical atom pairs of the main chain of the protein (usually C_α atoms).

It should be noted that there are some more specific criteria for evaluation of similarity between the structures under study in addition to standard deviation estimation in each of the individual methods of ProCKSI. For example, the authors of TM-align algorithm [12], in addition to RMSD, use a specific estimate reflecting the overall degree of general topology similarity between the structures in the study – so-called Topological Score (TM-score):

$$\text{TM-score} = \max\left[\ 1/L_{\text{target}}\sum_{1}^{L_{\text{aligned}}}\frac{1}{1+d_i/d_0(L_{\text{target}})}\right]. \qquad (2)$$

In this equation, L_{target} and L_{aligned} are sequence lengths of the target protein and the aligned region, respectively; d_i is the distance between i-th residue pairs. Finally, the factor $d_0(L_{\text{target}})$ is a normalizing factor of the distance scale.

This criterion characterizes the similarity between two structures by the value in the (0,1) interval, where the value of 1 corresponds to two structures with absolutely identical spatial topology. For comparison, a value below 0.2 corresponds to randomly chosen structures, whereas higher than 0.5 result may indicate that these structures have common polypeptide chain packing in a rather rough approximation.

Another specific criterion, V-score, is used in the Vorolign method as a normalized measure of similarity between the structures studied in the framework of this method [13]. In computational geometry, the Voronoi diagram is a special way of metric space expansion defined by the distances to the discrete set of selected objects in this space. In the simplest case there is a set of points S on the plane called the Voronoi sites. In this case each site S can be enclosed in the specific volume $V(S)$, called the Voronoi cell. In this volume, in turn, there is a set of points in the neighborhood of a given starting point S. To assess the degree of structural similarity between the studied proteins, we used C_β atoms of individual amino acid residues (C_α atoms for glycine) as the initial set of points in Euclidean space. The subsequent transformation of this space into the form of a convex polyhedron allows us to consider the original input data as a set of Voronoi cells. For example, for two compared proteins $X = x_1x_2...x_p$ and $Y = y_1y_2...y_q$, the set of nearest neighbors of the specified base x_i is defined as $N(x_i) = (x_{i1}, x_{i2}, ..., x_{in})$. Thus, to assess the similarity between two bases x_i and y_j, ultimately, it is necessary to calculate the similarity between adjacent sets (cells) of bases, which are determined by the corresponding functions $N(x_i)$ and $N(y_j)$. The similarity between two individual bases, located in neighboring cells $k = N(x_i)$ and $l = N(y_j)$ is defined as:

$$\text{Sim}\ (x_{i_k}, y_{j_l}) = \omega 1\ \text{AA}\ (x_{i_k}, y_{j_l}) + \omega 2\ \text{SSE}\ (x_{i_k}, y_{j_l}). \qquad (3)$$

In this expression, AA reflects the similarity between individual amino acid residues, and SSE is a measure of secondary structure element similarity. Both of these terms are determined in the process of calculating the values of the corresponding matrices. Finally, the estimation of similarity between two neighboring sets of cells is calculated with a dynamic programming method, with consideration of Eq. (3) as well as additional factor p_u, used to account for all non-aligned residue pairs. The value of $S(k,l)$ of dynamical matrix S is given as follows:

$$S(k,l) = \left\{\begin{matrix} S(k-1,\ l-1) + Sim(x_{i_k}, y_{j_l}) \\ S(k-1,\ l) - p_u \\ S(k, j-l) - p_u \end{matrix}\right\}. \qquad (4)$$

Finally, the value $Sim(N(x_i), N(y_j))$ corresponding to the last row and column of the dynamic matrix S is used to calculate the similarity between two neighborhoods of bases x_i and y_j.

ProCKSI algorithm for constructing a consensus. The main function of ProCKSI is to compare studied 3D structures through a series of integrated algorithms [7]. In the end, the results are converted into a form more convenient for subsequent analysis. We would like to specifically note that with the help of ProCKSI a unique opportunity appears to combine the results for a given data set into a single consensus. This operation produces the average of data obtained by different structural alignment methods. Thus, using this mechanism, we can significantly vary the initial set of data reflecting the similarity between the studied structures with different resultant consensuses. As a result, by addition or exclusion of individual components we can understand which consensus will reflect the most credible distribution of structures. It should be noted that this method has already been verified experimentally by structural comparison of the different families of protein kinases [8]. It was shown in the cited paper that the best distribution of these proteins was obtained by the consensus ProCKSI method, compared with the results of individual methods included in this consensus. Finally, the results of this work were tested with respect to the existing structural classification presented in the SCOP (Structural Classification of Proteins, http://scop.mrc-lmb.cam.ac.uk/scop/) database. Based on the results of this experiment, we decided to conduct our own review of all the methods presented in ProCKSI in order to determine which of these algorithms would be most suitable for our task. First of all, it was necessary to understand whether it makes sense to use these methods on the structures we studied (membrane receptors), because all these algorithms were originally developed for globular proteins.

In the first phase of our study we conducted a control experiment, which included a variety of proteins from bacteria and archaea (e.g. bacteriorhodopsin and sensory rhodopsins of archaea) along with GPCR structures. These structures were chosen primarily on the basis of structural similarities in the spatial organization of TM-domain of these proteins and GPCRs. In the case of adequacy of the methods, we expected to obtain the classification in which all the structures of GPCR receptors would be grouped separately from the structures of other bacterial proteins. As a result, we obtained the expected outcome.

In the next stage it was necessary to conduct finer tuning of the classification methods used in accordance with the expected tasks. It is important to note that these methods are highly accurate with respect to the tasks (namely, specific proteins) for which they were originally designed. For example, the high accuracy of the TM-align method was shown for water-soluble (globular) proteins [12, 14]. In these studies the authors have shown that TM-score was the best measure of similarity between studied 3D structures. Similar results were obtained for the Vorolign method [13], where the researchers had shown a high accuracy of V-score criterion, as compared with the classical measurement of RMSD. As to our research, it was empirically found that the measurements based on RMSD ultimately gave more correct distribution of experimental structures of membrane proteins. Finally, the best results were obtained using a combination of results provided by the TM-align and Vorolign methods.

Principal component analysis of mammalian rhodopsin conformational dynamics. Principal component analysis (PCA) is one of the main methods to reduce the dimensionality of input data with the least loss of significant information [15]. The calculation of principal components comes to the computation of eigenvectors and eigenvalues of input data covariance matrix. In addition, PCA is intensively used in bioinformatics to reduce the dimensionality of studied structures, to highlight the most relevant information, as well as during the visualization of results [15-18].

In the case of an ensemble of biological 3D structures, this method corresponds to orthogonal linear transformation of raw data from the Cartesian coordinate system to the new system, so-called collective coordinates. The main objective of PCA is to simplify the initial structural diversity in a data set by detection of the most prevalent directions of structural changes. In the resulting system of collective coordinates, the main variance of the source data is located along the first principal axis. Usually, PCA is applied to an ensemble of different conformers of the studied proteins, identified, for example, in the presence of various substrates. Another example is the analysis of various models corresponding to NMR structure under study. Finally, it is possible to use an ensemble of molecular dynamics snapshots as an input. There is experimental evidence of the reliability of the results obtained with PCA. For example, the suitability of this method for studies of functional dynamics has recently been demonstrated by the results of works carried out on ubiquitin structures [19]. In this case, the unique motion component responsible for ligand-recognizing ability of this protein was discovered.

Consider the trajectory of N atoms at time $F(X_i, Y_i, Z_i)$, where $i = 1, 2, \ldots N$. To analyze the ensemble of experimental protein structures, the first step is construction of covariance matrix C, consisting of $3N \times 3N$ elements. Covariance matrix C is calculated according to the formula:

$$\left\langle \Delta R \Delta R^T \right\rangle = m^{-1} \sum_A [\Delta R^{(A)} \Delta R^{(A)T}], \qquad (5)$$

where R is a set of coordinates of a trajectory point, the summation is over all m structures under study, and ΔR is

deviation of the specific structure A from ensemble average $<R>$. Index T denotes transposition of the original matrix. The subsequent decomposition of matrix C, according to transformation (6)

$$C = \sum_{i=1}^{m} \sigma_i p_i p_i^T , \quad (6)$$

results in calculation of principal components (eigenvectors) p_i, as well as the corresponding variances (eigenvalues) σ_i. Here, σ_i represents the largest variance of the system, and p_i ($3N$-dimensional vector) describes the displacement of N bases, represented in this study by their C_α atoms, along the most volatile mode, often called the first principal mode. Finally, root mean standard deviation between the investigated structures can be found as the trace (tr) of the C matrix:

$$<\text{RMSD}> = [\text{tr}(C)/N]^{1/2}. \quad (7)$$

Equations (6) and (7) make it obvious that the first principal component (p_1), which corresponds to the maximum variance σ_1, is a major contributor to mean standard deviation <RMSD>.

In this work the study of conformational dynamics of mammalian rhodopsins was conducted using PCA. During this analysis an ensemble of all known crystallographic structures of rhodopsin was constructed, reflecting the conformational behavior of this receptor. The first step consisted in spatial alignment of the investigated experimental structures to 1GZM structure [20]. This structure is an inactivated form of the receptor, obtained in a trigonal crystal form with a resolution of 2.85 Å. In the next step, we calculated the covariance matrix and obtained a projection of the ensemble of crystallographic structures on the plane of the principal coordinates. The final stage consisted of visualization of the resulting data. All of these operations were performed using the Python-based ProDy program [21] (http://www.csb.pitt.edu/prody/).

The resulting classification trees (dendrograms) of the structures were visualized using the Newick Utilities software package [22] running on Linux. Spatial superposition of individual structures was carried out using the PyMOL visualizer.

RESULTS AND DISCUSSION

Classification of rhodopsin experimental structures according to ProCKSI. Using the ProCKSI consensus method, we created a classification of all known experimental structures of bovine and squid rhodopsins. The choice of source animals was made primarily because quite a number of experimental structures of mammalian rhodopsin reflecting different stages of receptor photoactivation are available [23-25]. For comparison, a comparable number of crystallographic structures exist for β-adrenergic receptor, but in this case the vast majority of these structures represent the inactive form of the receptor corresponding to the ground state of rhodopsin. Only in 2011 two new crystallographic structures corresponding to the true active form of β-2-adrenergic receptor were obtained [26]. Notably, no intermediate forms of the receptor corresponding to known intermediates of rhodopsin were crystallized. As for all other GPCRs, only a small number of experimental structures of these receptors were available at the time of this study, and they represented mainly the inactive forms. Therefore, the experimental structure of bovine rhodopsin was the only statistically significant sample.

The general character of the calculated distribution of rhodopsin crystallographic structures is shown on the dendrogram (Fig. 3). The analysis of the classification showed the adequacy of the resulting distribution. Figure 3 shows that receptors from both rhodopsin families are grouped separately, suggesting the existence of structural differences between them [24, 27]. This observation is confirmed by a number of primary sources, in which features of the studied structures were discussed in detail. For example, the works of Murakami and Kouyama discussed the major structural differences between these two receptor types [24, 28]. The authors noted first of all the differences in the degree of protrusion of the fifth and the sixth helices in the cytoplasmic space (Fig. 4). Moreover, in the structure of the squid rhodopsin an additional ninth helix was found on the C-terminus, which is located after a short eighth helix present in mammalian rhodopsin as well. Finally, there is also a significant difference between the signaling pathways of the two receptors [27]. Bovine rhodopsin is known to transmit signals through a specific protein transducin (Gt), which interacts with phosphodiesterase. In the case of squid rhodopsin, signals are transmitted through another specific G-protein (Gq), which acts through phospholipase C. Thus, the functional features of the two rhodopsins are directly reflected in their spatial structures. It is important to note that subtle structural differences within the same spatial topology have been correctly recognized with the consensus ProCKSI method and are shown on dendrograms.

After more detailed analysis of the classification of mammalian rhodopsin (Fig. 3), an even distribution of the crystallographic structures corresponding to different intermediate forms of this receptor can be highlighted. It is evident that the structures corresponding to the inactive (ground) state (PDB codes 1F88, 1GZM) are grouped separately from all other structures of this receptor. A more detailed examination of the dendrogram reveal a smooth transition from the cluster of inactive structures in the direction of the active forms corresponding to Meta-II* or constitutively-active (Opsin*) forms of the receptor (PDB codes 3DQB, 3PXO, 2X72, 3CAP). It

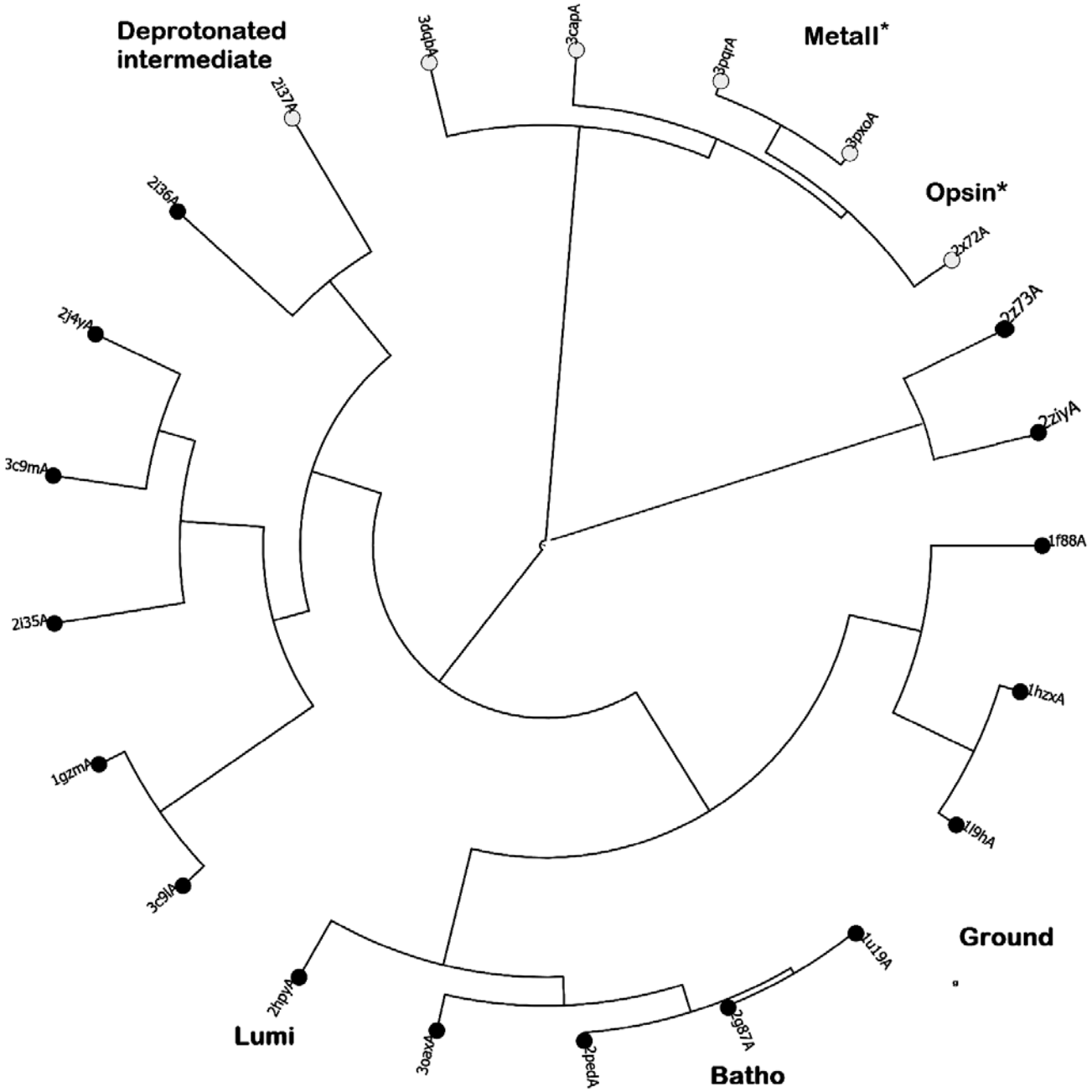

Fig. 3. Classification of experimental structures of the two rhodopsin families obtained with the consensus method of ProCKSI. Dark gray circles correspond to inactive (Ground, Batho, Lumi, Meta-I) crystallographic structures of mammalian rhodopsin, whereas light circles indicate active (Meta-II* or constitutively active opsin (Opsin*)) forms of the receptor. Black circles show the two structures of squid rhodopsin corresponding to its inactive (ground) state.

should be noted that one of the obvious advantages of mammalian rhodopsin model compared to other GPCRs is that all known intermediate forms of this receptor are described in the literature in sufficient detail. For example, a number of subtle structural differences were shown in a recent study, allowing separation of all inactivated intermediate forms of this receptor from its active forms [25]. Thus, tracking the dynamics of these changes, for example, comparing the structures corresponding to Batho- and Lumi-forms, with active Meta-II* structures, provides a holistic view of the nature of motion of separate TM receptor domains during the process of activation.

The classification of mammalian rhodopsin is in good agreement with data obtained for some other experimental structures of this receptor from primary sources

[29, 30]. For example, structures of some constitutively active opsins (PDB codes 3DQB and 3CAP), according to the authors, could represent experimental models of the states of rhodopsin closest to the state after the full cycle of photoactivation. These studies showed that compared to the inactivated state of this receptor obtained in complex with 11-*cis*-retinal, in the apo-form of this protein (in opsin) significant structural changes were registered in regions responsible for binding with a G-protein. These changes included a deviation of the sixth receptor helix from the TM bundle plane and a lateral shift of the fifth helix. Additionally, a destabilization of two important contacts was detected: a disruption of the salt bridge between the third and the sixth helix and also a disruption of the polar contact between the seventh and the eighth receptor helix. Finally, emergence of two new stabilizing contacts was noted between the third and the fifth, as well as between the fifth and the sixth helices, respectively. It is important to note that this set of experimental data also reflected on our dendrograms in the form of a joint group of the above-mentioned structures in a single sub-branch with other active structures of bovine rhodopsin corresponding to Meta-II* state (Fig. 3). Finally, we would like to mention the experimental data about the recently crystallized structures of bovine rhodopsin (PDB codes 3OAX and 3PQR). They correspond to activated (Meta-II*) state of this receptor [23]. Similar conformational changes accompanying receptor photoactivation were registered in these structures by analogy with the two above-mentioned constitutively active opsin forms (PDB codes 2X72 and 3DQB). The set of this experimental data is also reflected on the dendrogram (Fig. 3).

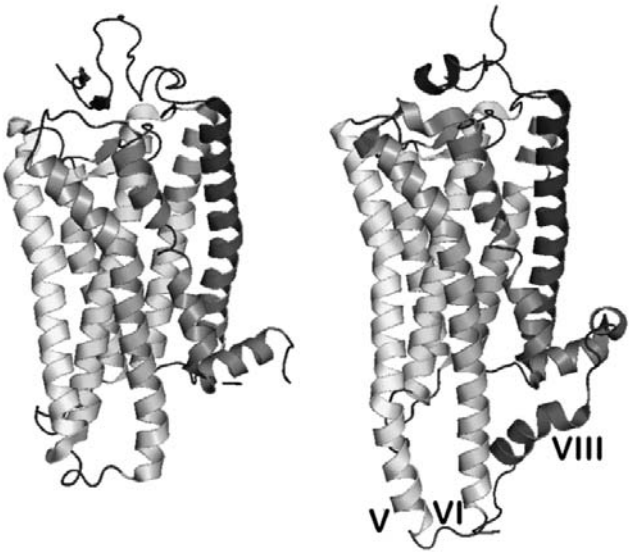

Fig. 4. Comparison of crystallographic structures of mammalian rhodopsin (left) and squid rhodopsin (right) in main (inactive) state (PDB codes 1GZM and 2Z73, respectively). Gray scale shows the orientation of the chain, from extracellular N-terminus (light) to intracellular C-terminus (dark). Squid rhodopsin is characterized by more pronounced immersion of the fifth and sixth helices of the receptor into the cytoplasmic space, as well as by the presence of an additional short eighth helix that lies perpendicular to the plane of the α-helical bundle.

The method of principal components. Projection of mammalian rhodopsin experimental structures into two main coordinates. In the next step we decided to check the classification of mammalian rhodopsin using PCA. The choice of this method was made primarily on the basis of published data demonstrating the successful application of this technique for studying protein dynamics [18]. In particular, in Bahar's laboratory a similar experiment was previously conducted with all existing at that time mammalian rhodopsin crystallographic structures [16]. The results of that work were generally in good agreement with the ProCKSI distribution obtained in our work. We decided to double-check the reliability of this experiment by repeating it for more mammalian rhodopsin experimental structures and comparing the results with ProCKSI structural trees. We would like to emphasize that at the time of our work we had a large number of bovine rhodopsin experimental structures compared to ones described in the pioneering study of Bahar [16]. In particular, we had several additional structures of the active receptor at high resolution.

Analysis of the results (Fig. 5) showed good correspondence of the obtained projection of rhodopsin crystallographic structures with the results of the earlier study by Bahar. As a result, the first principal component showed correct clusterization of all inactive (dark-, Batho- and Lumi-forms) and active structures of rhodopsin (opsin- and Meta-II* forms) into two different groups. Distribution along the second principal component, in turn, showed finer distribution of experimental structures within a cluster of all inactivated structures. The latter distribution, in turn, is a good representation of the dispersion of structures caused by structural differences in the individual loops, as well as in the terminal regions of the receptor. Finally, the most remarkable was the fact that the resulting distribution along two major components was in strict accordance with the rhodopsin structure classification created with ProCKSI (Figs. 3 and 5). This fact, in turn, once again proves the correctness of the chosen classification algorithm (ProCKSI consensus method). We would like also to note that due to the fact that in comparison with the work of Bahar, we operated with a somewhat larger set of experimental structures, and therefore the results of PCA analysis also had some differences. For example, a relatively uniform initial distribution along two principal coordinates obtained in Bahar's work, was in our case broken by the structure 2i37 (photoactivated deprotonated intermediate), which was also grouped between active and inactive forms of this receptor on ProCKSI dendrograms. In this case, it is obvious that the use of larger sets of experimental structures results in more accurate results of PCA

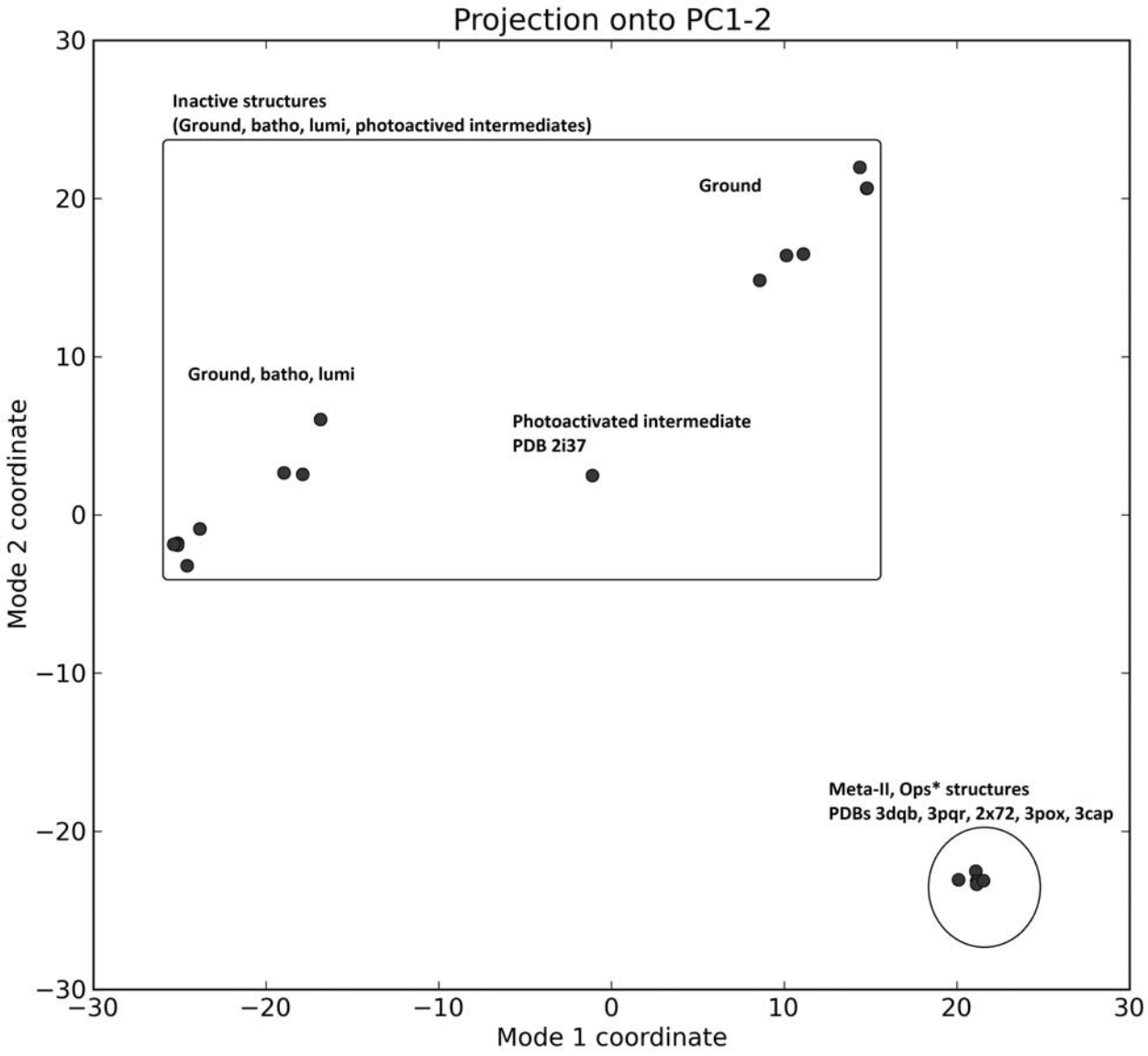

Fig. 5. Results of PCA analysis of 20 mammalian rhodopsin crystallographic structures. Projection of the ensemble of structures on the two main components is shown. According to the resulting distribution, the first PCA mode correctly clusters all inactive (wider distribution – marked by rectangle) and active (narrow distribution in a circle) Meta-II* and Opsin* structures into two distinct groups. The second mode traces further differentiation of the structures in a cluster, showing conformational variance (e.g. spatial differences in the organization of the loops).

analysis. Thus, a statistically significant set of the different crystallographic structures of the studied protein will eventually form a true conception of the conformational behavior of the protein during photoactivation.

As a result of this work, we created a classification of crystallographic structures of rhodopsins. The analysis of the resulting distribution elucidated a number of structural and functional features of these proteins. For example, analysis of a set of structural changes that are present in a number of intermediate structures of mammalian rhodopsin relates the different stages of photoactivation to the movements of individual transmembrane domains of these proteins. As a result, an approach was developed that yields visual schema reflecting the different stages of membrane receptor activation. It is important to note that this approach can be later applied to other seven TM receptors that have a greater therapeutic significance. Thus, in addition to significant theoretical importance of these data it may also, for example, facilitate the development of new highly effective drugs.

We should also note that the experimental distribution of GPCR structures obtained in this study is in good agreement with the source data. It is important that various structural and functional features that are present in

the resulting classification can be easily explained on the basis of the relevant literature. In particular, the clearest distribution was obtained using the ProCKSI consensus method, giving much more correct classification of the structures than the individual methods included in this consensus alone. Finally, the resulting distribution of the experimental rhodopsin structures was confirmed by the results of PCA. In the latter case, the distribution pattern of the investigated structures was identical to that obtained by the ProCKSI consensus method. This, in turn, further supports adequacy of the methods used in relation to the investigated structures of membrane receptors.

Therefore the resulting classification of the spatial structures of rhodopsin has made it possible not only to describe the structural changes occurring upon the activation of these proteins, but also to get closer to the description of the dynamics of the observed changes. A more detailed understanding of dynamic aspect of the structural changes accompanying membrane receptor activation can be achieved either with appearance of new experimental structures of these proteins or by using computer simulation techniques.

REFERENCES

1. Fredriksson, R., Lagerstrom, M. C., Lundin, L. G., and Schioth, H. B. (2003) *Mol. Pharmacol.*, **63**, 1256-1272.
2. Congreve, M., and Marshall, F. (2010) *Brit. J. Pharmacol.*, **159**, 986-996.
3. Hill, S. J., Williams, C., and May, L. T. (2010) *Brit. J. Pharmacol.*, **161**, 1266-1275.
4. Kenakin, T. (2004) *Trends Pharmacol. Sci.*, **25**, 186-192.
5. Kenakin, T., and Miller, L. J. (2010) *Pharmacol. Rev.*, **62**, 265-304.
6. Rajagopal, S., Rajagopal, K., and Lefkowitz, R. J. (2010) *Nature Rev. Drug Discov.*, **9**, 373-386.
7. Nygaard, R., Frimurer, T. M., Holst, B., Rosenkilde, M. M., and Schwartz, T. W. (2009) *Trends Pharmacol. Sci.*, **30**, 249-259.
8. Hanks, S. K., and Hunter, T. (1995) *FASEB J.*, **9**, 576-596.
9. Barthel, D., Hirst, J. D., Blazewicz, J., Burke, E. K., and Krasnogor, N. (2007) *BMC Bioinform.*, **8**, 416.
10. Damm, K. L., and Carlson, H. A. (2006) *Biophys. J.*, **90**, 4558-4573.
11. Maiorov, V. N., and Crippen, G. M. (1994) *J. Mol. Biol.*, **235**, 625-634.
12. Zhang, Y., and Skolnick, J. (2005) *Nucleic Acids Res.*, **33**, 2302-2309.
13. Birzele, F., Gewehr, J. E., Csaba, G., and Zimmer, R. (2007) *Bioinformatics*, **23**, e205-e211.
14. Pandit, S. B., and Skolnick, J. (2008) *BMC Bioinform.*, **9**, 531.
15. Ma, S., and Dai, Y. (2011) *Brief. Bioinform.*, **12**, 714-722.
16. Bahar, I. (2010) *J. Gen. Physiol.*, **135**, 563-573.
17. Maisuradze, G. G., Liwo, A., and Scheraga, H. A. (2009) *J. Mol. Biol.*, **385**, 312-329.
18. Yang, L. W., Eyal, E., Bahar, I., and Kitao, A. (2009) *Bioinformatics*, **25**, 606-614.
19. Lange, O. F., Lakomek, N. A., Fares, C., Schroder, G. F., Walter, K. F., Becker, S., Meiler, J., Grubmuller, H., Griesinger, C., and de Groot, B. L. (2008) *Science*, **320**, 1471-1475.
20. Li, J., Edwards, P. C., Burghammer, M., Villa, C., and Schertler, G. F. (2004) *J. Mol. Biol.*, **343**, 1409-1438.
21. Bakan, A., Meireles, L. M., and Bahar, I. (2011) *Bioinformatics*, **27**, 1575-1577.
22. Junier, T., and Zdobnov, E. M. (2010) *Bioinformatics*, **26**, 1669-1670.
23. Choe, H. W., Kim, Y. J., Park, J. H., Morizumi, T., Pai, E. F., Krauss, N., Hofmann, K. P., Scheerer, P., and Ernst, O. P. (2011) *Nature*, **471**, 651-655.
24. Kouyama, T., and Murakami, M. (2010) *Photochem. Photobiol. Sci.*, **9**, 1458-1465.
25. Lodowski, D. T., Angel, T. E., and Palczewski, K. (2009) *Photochem. Photobiol.*, **85**, 425-430.
26. Rasmussen, S. G., Choi, H. J., Fung, J. J., Pardon, E., Casarosa, P., Chae, P. S., Devree, B. T., Rosenbaum, D. M., Thian, F. S., Kobilka, T. S., Schnapp, A., Konetzki, I., Sunahara, R. K., Gellman, S. H., Pautsch, A., Steyaert, J., Weis, W. I., and Kobilka, B. K. (2011) *Nature*, **469**, 175-180.
27. Sugihara, M., Fujibuchi, W., and Suwa, M. (2011) *J. Phys. Chem. B*, **115**, 6172-6179.
28. Murakami, M., Kitahara, R., Gotoh, T., and Kouyama, T. (2007) *Acta Crystallogr. Sec. F*, **63**, 475-479.
29. Park, J. H., Scheerer, P., Hofmann, K. P., Choe, H. W., and Ernst, O. P. (2008) *Nature*, **454**, 183-187.
30. Scheerer, P., Park, J. H., Hildebrand, P. W., Kim, Y. J., Krauss, N., Choe, H. W., Hofmann, K. P., and Ernst, O. P. (2008) *Nature*, **455**, 497-502.