ISSN 1607-6729, Doklady Biochemistry and Biophysics, 2009, Vol. 427, pp. 215–217. © Pleiades Publishing, Ltd., 2009.
Original Russian Text © I.B. Kovalenko, A.M. Abaturova, G.Yu. Riznichenko, A.B. Rubin, 2009, published in Doklady Akademii Nauk, 2009, Vol. 427, No. 5, pp. 696–698.

BIOCHEMISTRY, BIOPHYSICS, AND MOLECULAR BIOLOGY

A Novel Approach to Computer Simulation of Protein–Protein Complex Formation

I. B. Kovalenko, A. M. Abaturova, G. Yu. Riznichenko, and Corresponding Member of the RAS A. B. Rubin

Received April 2, 2009

DOI: 10.1134/S1607672909040127

The majority of biochemical processes are associated with the functioning of protein molecules and their complexes in the reactions of enzymatic catalysis and cell signaling. Predicting the structure of protein complexes by their simulation is a complex problem that remains largely unsolved.

The factors that play the key role in complex formation are as follows: the rate of protein diffusion to the docking site; long-range electrostatic interactions between protein surfaces, geometric and chemical complementarity of binding areas; molecular mobility at the protein–protein interphase, hydrogen bonds, Van der Waals interactions, hydrophobic interactions, and salt bridges. It is known that different factors play different roles at different stages of complex formation [1]. Currently, there is no universal method for simulating protein complex formation that would make it possible to take into account all these factors and accurately predict the structure of protein complex [1, 2].

Molecular diffusion and long-range electrostatic interactions as well as molecule geometry play the decisive role in precomplex formation. The electrostatic interactions significantly accelerate the process of precomplex formation and thereby make it much more effective. If the geometrical correspondence of binding areas is established at the precomplex stage, this ensures the optimal relative position of two molecules prior to subsequent final complex formation. The hydrophobic interaction, hydrogen bonds, and molecular mobility, in turn, play the key role in the conversion of the precompex into the final complex [1].

In this work, we developed a new method for determination of binding areas in proteins and precomplex structure with allowance for the Brownian diffusion and electrostatic interactions of proteins that occur when proteins approach one another. This method significantly simplifies subsequent precise simulation and prediction of the final complex structure. The Brownian dynamics method, which can be used for predicting the structure of protein complexes, considers the interaction of only two molecules in solution [3–5]. A characteristic feature and novelty of our method, as is shown below, is the possibility to use it for studying interaction of several protein molecules simultaneously. This makes it possible to simulate the formation of a large number of complexes, which takes place in solution or cell compartments and to monitor the real-time kinetics of this process.

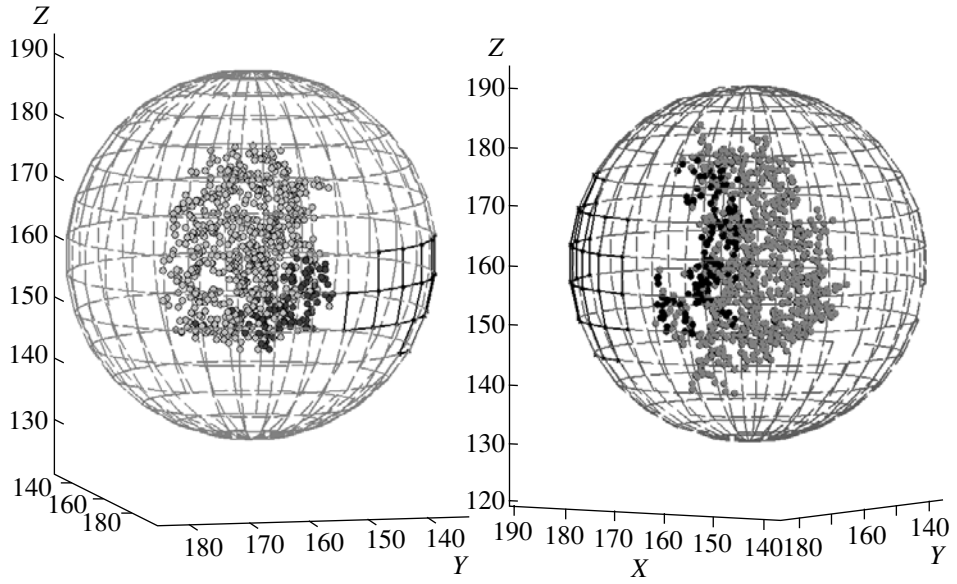
In our method, the process of protein complex formation is conditionally divided into several stages: (1) Brownian diffusion of proteins to the docking site, (2) their approach due to electrostatic attraction forces between molecules, relative spatial position of molecules, and precomplex formation; and (3) final complex formation. As is shown below, relative position of proteins in precomplexes, predicted on the basis of the computer model suggested, in most cases corresponds to their real orientation in final complexes.

Method description. The proposed approach is based on direct computer simulation of diffusion and complex formation between mobile electron-transport proteins [6–8]. Simulation is performed in a virtual 3D cubical reaction volume containing randomly distributed protein molecules. Movement is described by the Langevin equation, which describes changes of each coordinate in time caused by random and outer forces:

$$\xi_x \frac{dx}{dt} = f_x(t) + F_x,$$

where x is the coordinate along which movement is considered; ξ_x is the factor of viscous friction along this coordinate; and $f_x(t)$ and F_x are the projections of a random and electrostatic forces, respectively, on the abscissa axis. The random force $f_x(t)$ is distributed normally with a zero mean and variance $\frac{2kT\xi_x}{\Delta t}$. Here, k is

Biological Faculty, Moscow State University, Moscow, 119991 Russia



The probability of barstar (on the left) and barnase (on the right) binding calculated on the basis of the binding probability model and represented as a probability distribution sphere. Dark circles correspond to the atoms belonging to the contacting amino acid residues revealed on the basis of X-ray data for the barnase–barstar complex. The sectors of the sphere surrounded by dark lines correspond to protein regions with a high binding probability (according to calculations for the model).

the Boltzmann constant, T is temperature, and Δt is the time increment (constant in our method). To calculate the viscous friction factors in the model, the protein molecule shape is approximated by an ellipsoid of revolution rather than by a sphere, which is commonly accepted in Brownian dynamics models [3–5].

Three-dimensional protein molecules are constructed on the basis of data extracted from the Protein Data Bank. To calculate protein collisions, the shape of proteins is described using a small number (10–100) of spheres, which sufficiently adequately reflects the molecule surface for calculating collisions with other molecules and simultaneously reduces the calculation time compared to Brownian dynamics methods, in which the atomic resolution of the protein surface is used. The electrostatic interactions between proteins are taken into account when proteins approach one another to distances smaller than 35 Å. The protein is represented as an area with the dielectric constant $\varepsilon = 2$ and spatially distributed partial charges; for the surrounding solution, $\varepsilon = 80$. The electrostatic field created by the charges on the protein surface was calculated using the Poisson–Boltzmann equation, which made it possible to take into account different values of dielectric permittivity of proteins and solution.

In this model, the movement and interaction of several hundreds of protein molecules is simulated, which provides an opportunity to directly monitor the protein interaction kinetics taking into account simultaneous interaction of several molecules and to study precomplex formation depending on the geometrical size and shape of the reaction volume. We divided the set of relative positions of two proteins into paired disjoint subsets corresponding to a 12° step in the angle of rotation of one molecule relative to the other one in a spherical coordinate system (figure). These sectors, obtained as a result of small-step division, contain only one or two amino acid residues located on the molecule surface. The objective of simulation is to find those sectors and corresponding amino acids that approach one another most closely as a result of diffusion and electrostatic interaction of proteins (i.e., to calculate the probability of approach of various amino acids and select those of them for which this probability is the highest). The movement and interaction of 100 pairs of protein molecules in 1-ms time interval at a constant step of 100 ps was simulated in all numerical experiments with the use of the computer model in a cubical reaction volume of $70 \times 70 \times 70$ nm.

To test the proposed method, we used pairs of proteins for which the structures of complexes are known from experiments. If a precomplex is not electrostatically optimal, the final complex determined in the model will significantly differ from the final complex of the test proteins in solution. In total, we studied eight pairs of proteins. For seven of these pairs (barnase–barstar, colicin E9DNase–immune protein 9, acetylcholinesterase–fasciculin 2, thrombin–thrombomodulin, erythropoietin–erythropoietin-binding protein, interleukin 4–interleukin 4 receptor, and plastocyanin–cytochrome *f*), the binding areas predicted by simulation corresponded to the experimentally determine areas. Only for one pair (colicin E3 RNase–immune protein 3), these areas did not coincide. The results of simulation for barnase and barstar molecules are shown in the figure, in which the frequently and rarely

approaching areas of the proteins are shown in dark and light gray, respectively. As seen in the figure, the areas with a high binding probability, calculated on the basis of the model (the sectors contoured with dark lines) are located opposite to the experimentally determined binding areas (dark circles designating atoms), i.e., correspond to the experimental data. In the model, the predicted precomplex structures form as a result of electrostatic interactions and geometrical complementarity of binding areas. Apparently, in the vast majority of cases, the electrostatic interactions between approaching proteins with the highest probability ensure the relative position of molecules that is most advantageous for their subsequent binding and formation of a procomplex that is then converted into the final complex.

Thus, we have developed a procedure for computer simulation of protein–protein complex formation, which takes into account simultaneous diffusion and interaction of several hundreds of protein molecules. The surface of the protein molecule is approximated by a set of spheres; the shape, by the ellipsoid of revolution. The electric field of surface charges is calculated with allowance for different values of the dielectric constant for proteins and water. Using eight pairs of proteins as an example, we showed that this method makes it possible to predict with a sufficient accuracy the binding areas and structure of the precomplex of protein molecules and is sufficiently precise to simulate the process of complex formation. The constructed model makes it possible to simulate protein interaction at different distribution of charges on proteins and different ionic strength and pH of medium and, therefore, to predict the binding areas for various proteins at various conditions.

ACKNOWLEDGMENTS

We are grateful to G. Smith and J. Smith for discussing the structural peculiarities of the protein complexes studied. This study was supported by the Russian Foundation for Basic Research (project nos. 07-04-00375 and 08-04-00354).

REFERENCES

1. Kleanthous, C., *Protein–Protein Recognition*, New York: Oxford Univ. Press, 2000.
2. Tramontano, A., *The Ten Most Wanted Solutions in Protein Bioinformatics*, Boca Raton: CRC, 2005.
3. Northrup, S.H., Luton, J.A., Boles, J.O., and Reynolds, J.C.L., *J. Computer-Aided Mol. Design*, 1987, vol. 1, pp. 291–311.
4. Gross, E.L. and Pearson, D.C., Jr., *Biophys. J.*, 2003, vol. 85, pp. 2055–2068.
5. Spaar, A., Dammer, C., Gabdoulline, R., et al., *Biophys. J.*, 2006, vol. 90, no. 6, pp. 1913–1924.
6. Kovalenko, I.B., Abaturova, A.M., Gromov, P.A., et al., *Phys. Biol.*, 2006, vol. 3, pp. 121–129.
7. Kovalenko, I.B., Abaturova, A.M., Grachev, E.A., et al., *Biofizika*, 2008, vol. 53, no. 2, pp. 261–270.
8. Kovalenko, I.B., Diakonova, A.N., Abaturova, A.M., and Riznichenko, G.Yu., in *Proc. Flavins and Flavoproteins Int. Symp.*, Prensas Univ. de Zaragoza, 2008, pp. 437–442.