Journal of Advanced Bioinformatics Applications and Research
ISSN 0976-2604.Online ISSN 2278–6007
Vol 5, Issue2, 2014, pp72-77
http://www.bipublication.com

IN SILICO SCREENING OF RECA-RECX DOCKED COMPLEXES AND THEIR INTERFACIAL SITE DETERMINATION

Subhra Mishra

Department of Chemistry, Alipurduar College,
North Bengal University, Pin-736122, West Bengal. India.
Corresponding author: E mail: subhra_apd@yahoo.com. Tel: 09474751733

[Received-07/03/2014, Accepted-07/04/2014]

ABSTRACT:

Understanding of interaction of two key macromolecular species is one of the major problems in structural and molecular biology. An understanding of protein – protein interactions depend upon knowledge of both the three dimensional structural details of the interactions and the chemical dynamics of the systems. An important protein-protein complex RecA-RecX has been taken for finding the interfacial site. The *E. coli* RecA protein plays a central role in homologous recombination and also important candidate in DNA repair and mutagenesis pathway. RecX is an important candidate in regulating the function of RecA. It is important to know the three dimentional structure of RecA-RecX complex to understand which residues are involved for interaction and how RecX regulate the function of RecA.

In this study, RecA-RecX complex and other oligomeric protein complexes are viewed from a network perspective to obtain new insights into protein association. The aim of this paper is to describe the computational approach to design the strategies to recognize the protein–protein interfaces in an automated, generalizable fashion. The successes suggest that these computational methods can be used to modulate, reengineer and design protein–protein interaction networks in living cells.

KEYWORDS: RecA, complex, RecX, filament, interaction.

INTRODUCTION:

DNA damage inside the cell is a well-known fact and in order to avoid genetic changes, cells have evolved mechanisms to repair this damage. DNA repair and recombination is intimately linked to transcription and cell cycle control. Recombination processes include homologous recombination site-specific recombination, and transposition. The proteins are the expressions of the information content of the nucleic acids. They in turn control the expression of information from nucleic acid. Such regulations are at the heart of many complex biological phenomena. An understanding of protein – protein interactions depend upon a knowledge of both the three dimensional structural details of the interactions and the chemical dynamics of the systems.

Protein–protein interactions are central to many processes within cells and organisms, ranging

from immune defence to cellular communication. For biological regulation, it is necessary to recognize their targets, and the networks responsible for interactions in macromolecular complexes [1]. So, it is important to know the three dimensional structure of the protein molecules as well as the protein-protein interface. But, the limited nature of Protein Data Bank, and further limited number of X-ray crystallographic structures of high resolution has been a major constraint in the previous studies. Recently, however, there has been a large increase in the number of known three-dimensional structures that contain protein-protein recognition sites and more high-resolution structures have been solved. The knowledge of those few structures guided us to determine the rules for general structural study. The effect of various physical and chemical parameters on the strength of the interaction can be determined by finding their correlation with the energy of complexation [2]. So finding the linear correlation between the different structural and chemical parameters can lead to the determination of those parameters, which play an important role in the determining the strength of the interaction.

It is necessary to discriminate correctly or near-correctly docked orientations from incorrectly docked ones, and developing a search method that will be able to `find' a near-correctly docked orientation with reasonable likelihood [3]. To use this, it is necessary to describe the surface shape of the protein. This may be done by discretising the molecule onto a grid in space and considering which cells are occupied, or by using some sort of 'surfacing algorithm', which calculates the solvent-accessible or solvent-excluded surface, and a point set that triangulates it [4]. In carrying out this calculation, many special cases of geometry need to be considered. The role of electrostatics in protein–protein interactions has been reviewed by Sheinerman [5], and was explored from a more physical point of view by Elcock [6]. In practice, then, the above considerations frequently lead to a two- or three-stage approach to docking, as outlined in. In the final stage, we deal explicitly with a model in full atomic detail and allow movement of the side chains and possibly backbone, minimizing an energy function. The second and third stages may be combined. The energy/score landscape is rough and so it is clearly desirable to make the search as effective as possible by the use of efficient optimization algorithms [7]. Many of these considerations apply to methods for docking small-molecule ligands to proteins and any developments will be mentioned if they may be relevant to protein–protein docking [8].

Structural parameters characterizing a protein-protein interface:

There are several parameters which can characterize a protein-protein interface like interface area, planarity, secondary structure, hydrogen bonds and hydrophobic and polar composition of the residues in the interface etc. The exposure of protein atoms to solvent can be obtained by calculating the surface area of atoms in contact with solvent molecule. The solvent accessible surface area is calculated by using Lee & Richards Algorithm [9]. The classification of secondary structure is based on the percentage frequency of alpha and beta secondary structures in the interface residues. The secondary structure composition of these segments was analyzed and was grouped into three different groups as: alpha (>80% alpha helix), beta (>80% beta sheet), coil (80 % coils), alpha/beta.

Interface residues are defined as those residues that possess an accessible surface area (ASA) that decreases by >1 angstrom squared on complexation [10]. It has been often been assumed that proteins associate with their hydrophobic patches but polar interaction between the interface is also important. It is therefore to explore the relative composition of

polar and nonpolar residues on the interface. The interface residue propensity is an indication of a particular residue to be in an interface.

An open reading frame, designated as *oraA*, is located directly downstream of *recA* and upstream of alas in *E.coli*. De Mot *et al* [11] reported the existence of this putative regulatory gene, *recX* conserved in gram positive and gram negative bacteria and found that 5 end of an ORF overlaps the *recA* gene which encodes a putative polypeptide presenting some homology to RecX [12]. *recX* genes are located downstream of *recA* or overlapping *recA* or sometimes in the chromosome like in Neisseria gonorrhoeae and Bacillus subtilis. More recently, Investigation of the role of RecX in *E. coli* in vitro as well as in vivo suggest that RecX of *E.coli* inhibits both DNA strand exchange and ATP hydrolysis of RecA [13]. RecX blocks RecA filament extension, thereby lead to filament disassembly. The interaction of RecX to RecA is also enhanced in presence of N-terminal His_6 tag on the RecX protein. So, RecX is an important candidate in regulating the function of RecA. It is important to know the 3D structure of RecA-RecX complex to understand which residues are involved for interaction and how RecA regulate the function of RecX. RecX is a modular protein assembled of repeated three-helix motifs. The relative arrangement of the repeats generates an elongated and curved shape that is well suited for binding within the helical groove of the RecA filament [14-15]. Dynamics study shows an improved model of regulation of RecA by RecX protein, where RecA filament elongation along ssDNA is blocked by RecX protein on the ssDNA region, located outside the filament [16]. In the present study, the mode of interaction and the interacting residues are to be finding out by some software package and by developing new algorithm. Then this has to be compared with some common protein complexes whose structure has already been solved.

MATERIALS AND METHODS:

The hexameric RecA has been generated using crystallographic 6_1 symmetry applied on the coordinates of the RecA monomer (2REB) deposited in the Protein Data Bank. The hexameric assembly and RecX were docked using GRAMM v1.03 to find out the binding mode of RecX to RecA. The parameters have been assigned according to Vakser, 1995[17].

But, docking at very low resolution may be used and will enable to find out the interface region and the correct relative orientation of the components was unlikely to be found in this case. Another method has to use the statistics of residue-residue contacts across the interfaces of complexes in the PDB, expressing how much more probable it was that residues would interact than would be expected merely from random contacts between residues with the observed global frequencies of occurrence.To evaluate the docked complexes obtained in the first stage of docking, the residue- residue contacts were measured using following algorithm.

Algorithm used for extracting the residues in respective distance Bins:

1. The required PDB file of docked complex (combine.pdb) was opened up.
2. RecA and RecX were separated.
3. The RecA file was split into its component subunits and the individual monomeric residues of RecA were identified.
4. Three distance bins of 0-2.5 Å, 2.5-5 Å, and 5-7 Å were generated.
5. The distance of atoms of RecX residues with those of RecA was calculated using the distance formula

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

6. The distance bins were identified and residue pairs were recorded in respective distance bins.

To study the nature of interaction, the individual interacting interfaces for both RecA (six monomer) and RecX, are fed into protein-protein interaction server. To evaluate the RecA-RecX complex with some solved protein complexes in PDB, a database has been generated.

The analysis was carried out by selecting the protein structures of which were predicted by methods other than X-Ray crystallography were filtered out. The Protein Quaternary Server was used to determine the biologically active structure of a given protein complex.

RESULTS AND DISCUSSION:

The hexameric RecA has been generated using crystallographic 6_1 symmetry. The central diameter of the hexamer RecA has been measured as 25 Å. The high-resolution docking has been used for accurate complex prediction.

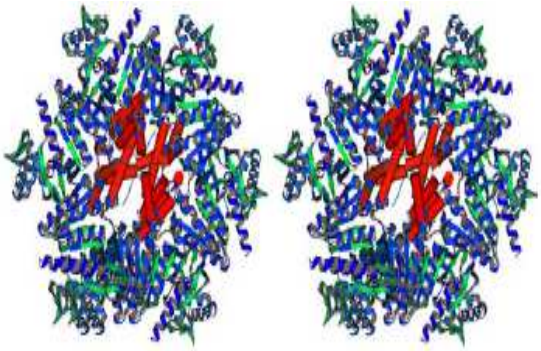

Fig: 1. Interaction of RecX with RecA [14]. RecX molecules are shown in red with the helices depicted in cylindrical form.

The docked structure (Fig. 1) indicates that RecX having a width of 24.98 Å enters the filalamentous RecA molecule along the spiral direction of the RecA molecule[14]. The docked structure suggest that three RecA molecule interact with one RecX molecule. Three molecules of RecA generated by crystallographic symmetry are depicted as A, B and C.

It has been found that besides other interacting residues the residues, which are close to the L2 disorder loop region of RecA, are suppose to interact with RecX. In, 'A' subunit of RecA, the region (192,193,194,210,211,212 and 213) interacts mainly with the N-terminal part of RecX (15, 18, 19, 22 and 26) (Fig. 2).

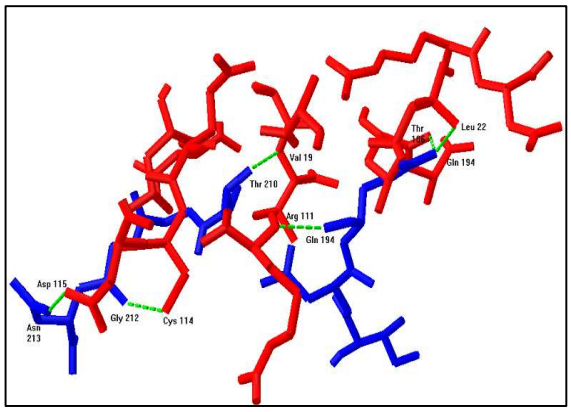

Fig: 2. Hydrogen bonding interaction between RecA (A subunit) with RecX.

It has been reported that GLN 194 [18] of RecA involves in DNA binding. This residue has been also found to interact with negatively charged residues like GLU 107 and ASP 26 of RecX. In 'B' subunit of RecA, besides other interaction the region (194,210,211,213 and 216) is supposed to be interacting with the residues (124,139,153 and162) of RecX. It has been found that Gln 194 of RecA is interacting with Asp 124 of RecX. Gln 194 being the basic residues makes electrostatic interactions with negatively charged component mimicking the phosphate group of DNA and helps in stabilizing the RecA-RecX interaction. Similarly, in case of 'C' subunit of RecA, the Gln 194 residue binds to the negatively charged Glu 140 of RecX (Fig. 3). Geometrical analysis shows that the positively charged residues of RecA (Gln 194, Asn 213, Lys 216, Arg 226 and Arg 227) interact with negatively charged residues of RecX Glu 106, Asp 124 and Glu140. Gln194 of RecA plays the major role in stabilizing the interaction.

In each case, the Gln 194 residue is involved in interaction with negatively charged Glu and Asp

of RecX suggesting that RecX might replace the DNA and binds to the DNA binding site of RecA.

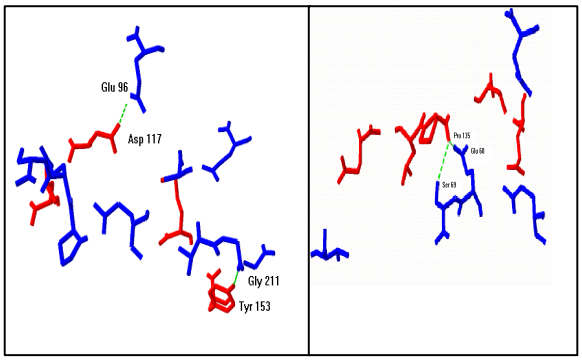

Fig: 3. Hydrogen bonding interaction between RecA (C and D subunit) with RecX

The docked complex of RecA-RecX has been analysed to determine the residues involved in surface interactions. Different geometrical values have been calculated to compare with that of known complexes. Interface residue propensity calculation of different complexes shows that there is high propensity of Phe, which is followed by Trp, Leu and Ile in the interface of dimeric complexes [19]. It has also been found that in case of RecA-RecX complex there are four Leu residues and 2 Ile residues at the interface corresponding to the propensities of Leu and Ile residues seen at dimeric interfaces.

Comparison of the structural parameters (Table 1) indicates that the planarity value (3.14-4.16) obtained for RecA-RecX (for three RecA chain) complex closely match with the same for dimeric complexes (3.93) [20]. The percentage of polar and non-polar atoms in interface also supports the existance of RecA-RecX dimer. It has been found that the values of ASA and interface residue segment are close to the calculated values for solved dimeric complexes (Table 1).

Structural parameters	**Dimer**	**RecA-RecX complex**
Accessible surface area (Å^2)	2093.952 (mean)	RecA(A chain) - RecX =1480.08 RecA(B chain) - RecX = 662.77 RecA(C chain) - RecX = 345.10
Planarity	3.93 (mean)	RecA(A chain) - RecX = 3.14 RecA(B chain) - RecX = 3.15 RecA(C chain) - RecX = 4.16
% Polar Atoms in Interface	37.808 (mean)	RecA(A chain) - RecX = 46.35 RecA(B chain) - RecX = 46.12 RecA(C chain) - RecX = 24.81
% Non polar Atoms in Interface	62.148 (mean)	RecA(A chain) - RecX = 53.60 RecA(B chain) - RecX = 53.80 RecA(C chain) - RecX = 75.10
Interface Residue Segments	6.488 (mean)	RecA(A chain)- RecX = 8 RecA (B chain)- RecX = 7.5 RecA(C chain) – RecX = 6.5

Table: 1. Comparison of the structural parameters of RecA-RecX complex with solved dimeric protein complex

The overall analysis of the docked structure indicates high probability of interactions involved between RecA and RecX. Hydrophobic as well electrostatic interactions are involved in stabilizing the complex. Electron microscopy has shown that RecX forms a block across the deep helical groove of the RecA filament, where strand exchange occurs [21]. This binding of RecX inhibits the strand exchange activity of RecA, a view confirmed by this proposed model as well.

CONCLUSION:

The construction and analysis of oligomeric protein structure networks and their comparison with monomeric protein structure networks provide insights into protein association. We believe this analysis will significantly enhance our knowledge of the principles behind protein association and also aid in protein design.

ACKNOWLEDGEMENT: I acknowledge with thanks, to my research supervisor Dr Amit Kr. Das, Professor, Department of Biotechnology, IIT, Kharagpur for providing me all sorts of facilities and invaluable suggestions during this work.

REFERENCES:

1. Brinda K.V. and Vishveshwara S., (2005). Oligomeric protein structure networks: insights into protein-protein interactions, BMC Bioinformatics., 6, 296.
2. Blundell T.L. and Srinivasan N., (1996). Symmetry, stability, and dynamics of multidomain and multicomponent protein systems, Proc. Natl Acad. Sci.USA., 93**,** 14243–14248.

3. Veith M., Hirst J.D., Koilinski A. and Brooks C.L., (1998). Assesing energy functios for flexible docking, J. Comp Chem., 19(14), 1612-1622.
4. Nooren I.M.A. and Thornton J.M., (2003). Structural characterization and functional significance of transient protein-protein interactions, J Mol Biol., 325, 991–1018.
5. Sheinerman, F.B., Norel, R., and Honig, B. (2000).electrostatic aspects of protein-protein interactions. Curr Opin Struct Biol. 10 (2). 153-9.
6. Elcock A.H., Gabdoulline R.R., Wade R.C. and McCammon J.A. (1999). Computer simulation of protein-protein association kinetics: acetylcholinesterase-fasciculin, J MolBiol., 291(1), 149-62.
7. Bahadur R.P., Rodier F. and Janin J., (2007). A dissection of the protein-protein interfaces in icosahedral virus capsids. J.Mol. Biol., 367, 574–590.
8. Jones S. and Thornton J.M., (1996). Principles of protein–protein interactions, Proc. Natl Acad. Sci., 93, 13–20.
9. Lee, B and Richards, F.M. (1971). The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol*, 55: 379-400.
10. Miller S., Lesk A.M., Janin J and Chothia C. (1987). The accessible surface area and stability of oligomeric proteins. *Nature*, 328:834–836.
11. R. De Mot, G. Schoofs, J. Vanderleyden. (1994). A putative regulatory gene downstream of recA is conserved in Gram-negative and Gram-positive bacteria. Nucleic Acids Res. 22. 1313-1314.
12. Vierling, S., Weber T., Wohlleben, W., and Muth, G. (2000). Transcriptional and mutational analysis of the Streptomyces lividans recX gene and its interference with RecA activity. J. Bacteriol. 182. 4005-4011.
13. Ragone, S., Maman, J, D., Furnham, N., and Pellegrini, L. (2008). Structural basis for inhibition of homologous recombination by the RecX protein. EMBO. 27(16). 2259–2269.
14. Mishra S, Mazumdar PA, Dey J, Das AK. (2003). Molecular modeling of RecX reveals its mode of interaction with RecA. Biochem Biophys Res Commun. 312(3).615-22.
15. Bakhlanova IV, Dudkina AV, Baitin DM, Knight KL, Cox MM, Lanzov VA. (2010). Modulating cellular recombination potential through alterations in RecA structure and regulation. Mol Microbiol. 78(6). 523-38.
16. Dudkina AV, Shvetsov AV, Bakhlanova IV, Baĭtin DM. (2011). Changing of filamentation dynamics of RecA protein, induced by D112R amino acid substitution or ATP to dATP replacement, results in filament steadiness TO THE RecX protein action**.** Mol Biol (Mosk). 45(3). 546-53.
17. Vakser, I.A. (1995). Protein docking for low-resolution structures. Protein Eng. 8 pp. 371–377.
18. Story, R.M. Weber, I.T. Steitz, T.A. (1992). The structure of the E. coli recA protein monomer and polymer, Nature, 355. 318–324.
19. Levy E.D., Pereira-Leal J.B., Chothia C and Teichmann SA. (2006). 3D complex: a structural classification of protein complexes. *PLoS Comput. Biol*, 2:e155.
20. Mishra S. (2012). Characterization of protein interfaces to infer protein-protein interaction. Res. J. Chem. Sci. 2(7).36-40.
21. VanLoock MS, Yu X, Yang S, Galkin VE, Huang H, Rajan SS, Anderson WF, Stohl EA, Seifert HS, Egelman EH. (2003). Complexes of RecA with LexA and RecX differentiate between active and inactive RecA nucleoprotein filaments. J Mol Biol. 333(2).345-54.