Glycoconj J (2006) 23: 241–249
DOI 10.1007/s10719-006-7929-z

Structural studies on mannose-selective glycoprotein receptors using molecular modeling techniques

Madhumita Patra · Sujata Majumder · Chhabinath Mandal

Abstract Glycoproteins play important roles in various cellular events and their presence in appropriate locations in proper active conformations is essential for many biochemical functions. Recent evidences suggest that some glycoproteins may require sorting receptors for efficient exit from the endoplasmic reticulum. These receptors need the presence of calcium or other metal ions for their native activity. The three-dimensional structure of such a receptor, p58/ERGIC-53, has been recently solved by x-ray crystallography, which is a mannose-selective lectin and contains two Ca^{2+} ions. Homology search in the sequence databases indicates a large number of proteins which bear varying degrees of homology in a wide spectrum of species with this receptor. In this study we have systematically searched for such genes which are potential candidates for acting as mannose-mediated glycoprotein receptors in various species as initially inferred from their amino acid sequence homology. Structures of a number of proteins have been predicted using knowledge-based homology modeling, and their ability to act as the glycoprotein receptor has been explored by examining the nature of sugar-binding site. Tetramer of mannose was docked in the binding pockets of the modeled structures followed by energy minimization and molecular dynamics to obtain most probable structures of the complexes. Properties of these modeled complexes were studied to examine the nature of physicochemical forces involved in the complex formation and compared with p58/ERGIC-53-mannose complex.

M. Patra · S. Majumder · C. Mandal (✉)
Drug Design, Development and Molecular Modelling Division
Indian Institute of Chemical Biology, Jadavpur, Kolkata 700 032, India
e-mail: cnmandal@iicb.res.in

Keywords Mannose-binding glycoprotein receptors · Molecular modeling · Protein-sugar complex · Lectin

Abbreviations

BLAST	Basic local alignment search tool
CRD	carbohydrate recognition domain
AA	amino acid
MOLMOL	Molecule analysis and molecule display
PDB	protein data bank
RMSD	root mean square deviation

Introduction

Animal lectins, a family of sugar-binding proteins, participates in a number of different functions such as the endocytosis of glycoproteins, regulation of cell adhesion, cell migration, sorting and distribution of microsomal enzymes [1–9]. Mannose-binding lectins are thought to be able to recognize mannose groups on bacteria and mediate the destruction of theses pathogenic invaders. They appear to have a significant influence on the course of certain chronic diseases like rheumatoid arthritis and cystic fibrosis. They also have a role in providing some protection at an early stage of life since newborns have yet to develop a robust antibody system. Export of many proteins from endoplasmic reticulum(ER) in eukaryotic cells is a highly regulated and selective process [10]. Recent evidence suggests that some glycoproteins may require sorting receptor for efficient export from the ER.

p58/ERGIC-53, a calcium-dependent mannose-selective animal lectin, that recycles between the ER and Golgi complex, appears to act as a sorting receptor for a subset of soluble glycoproteins exported from the ER [11]. Recently, the structure of the carbohydrate recognition domain (CRD) of p58/ERGIC-53 in its calcium-bound form has been

determined [12]. It has a tetrameric structure of four identical subunits, each of them containing two calcium ions. The overall structure of the calcium-bound form of the CRD of p58/ERGIC-53 is quite similar to that of the calcium-free form [11] with a RMSD of 1.0 Å with respect to all the Cα atoms. Large conformational changes between the calcium-bound [12] and calcium-free [11] structures are limited to two regions of the protein that contain all the calcium ion-coordinating residues present in two calcium-binding loops. This large change in the conformations between the two forms highly modulates the carbohydrate binding ability of the receptor. The CRD comprises one concave and one convex β-sheet packed into a β-sandwich. p58/ERGIC-53, a highly conserved protein in the animal kingdom, mediates secretion of coagulation factors V and VIII and two additional glycoproteins, cathepsin C and cathepsin Z-like proteins [13–16]. Mutations in the gene coding for this protein cause a rare hereditary bleeding disorder in humans due to a combined deficiency of coagulation factors V and VIII in the circulation.

One major task of structural bioinformatics is to explore the functions of various genes based on their structures predicted by theoretical tools of modern molecular modeling techniques. In the present study our aim is to identify mannose-binding environment in other proteins which may have structures similar to that of the p58/ERGIC-53 among the newly discovered gene sequences. Using BLAST search we have identified several sequences in a wide range of species which are homologous to this class of mannose-containing glycoprotein-sorting receptor implicating their potential function as receptors. Primary structure information points to their calcium- and mannose-binding ability, but 3-D tertiary structures can provide insight into their surface complimentarity to mannose-binding and proximity of relevant amino acid residues to form coordination to calcium ions. We have selected seven sequences from different species, namely, three from *Homo sapiens*, one from each of the species *Mus musculus, Newrospora crassa, Tetraodon nigroviridis* and *Anopheles gambiae* to examine in details the calcium binding loops and mannose-binding abilities in comparison with p58/ERGIC-53. Sequences with identities in the range from 30 to 50% were selected (except one with 89% identity from *Homo sapiens*) because higher identities would always lead to very similar structure. 3-D structures of these proteins have been predicted by homology modeling and the mannose molecule has been docked to arrive at the structures of the complexes using computer-aided modeling techniques. Analysis of these structures allowed us to examine the feasibility of the formation of calcium-binding loops and the nature of interaction of mannose with these proteins. We have also studied the binding affinities of the protein-sugar complexes in presence and absence of calcium in order to analyze the role of calcium in sugar binding. We calculated the electrostatic potentials of all the modeled proteins and examined their nature in the vicinity of the calcium- and sugar-binding pockets.

Methods

Protein structure predition by homolgy modeling

Three-dimentional structures of the mannose-binding glycoprotein receptors were predicted by knowledge-based homology modeling. The starting scaffold for the homology modeling was the x-ray crystallographically determined structure of p58/ERGIC-53 (PDB ID: 1R1Z). The initial structures were predicted using our in-house software package of ANALYN and MODELYN [17]. These structures were refined using the Insight II 2000.1 of Accelrys (San Diego, CA) equipped with DISCOVER as the energy minimization and molecular dynamics module. Distorted protein structures are produced at the places where insertion, deletion, and loop grafting were done during the automated structure generation from the starting scaffold and target sequences. These areas needed special attention in the process of structure refinement. Optimization of these regions involved the energy minimization (100 steps each of steepest descent and conjugate gradient methods) using cff91 force field followed by dynamics simulations. At the end of the dynamics simulation, the conformation with the lowest potential energy was picked for the next cycle of refinement. This combination of minimization and dynamics was repeated until satisfactory conformational parameters were obtained. Special care was taken in the structural zones where major insertion and deletions were made. Energy minimizations were performed with a convergence criterion of 0.001 Kcal/mol. Molecular dynamics simulations were carried out using a time step of 1 femto second for a total period of 1 pico second with 100 steps of equilibration and a conformation sampling interval of 1 in 10 during both equilibration and dynamics processes. During the regularizations of a selected segment, positional constraints were applied to the rest of the molecule in order to expedite the minimization and dynamics processes.

Prediction of structures of complexes

At present there is no crystal/NMR structure of the complex of the p58/ERGIC-53 with any carbohydrate, therefore, we applied molecular modeling techniques for predicting the structures of the complexes. Structures of the complexes of mono-, di-, tri- and tetra saccharides were built by scanning the complimentary surfaces of the x-ray structure of the p58/ERGIC-53 suitable for sugar binding. Structures of the complexes of the modeled proteins were obtained by the

superposition of the modeled protein structures with the structure of p58/ERGIC-53 docked with the appropriate oligomer of mannose followed by optimization with repeated energy minimization and dynamics simulations. Position constraints were applied to the atoms which were more than 10 Å away during energy minimization and molecular dynamics simulation; and also the distance between calcium ions and the atoms which are involved in co-ordination with the calcium ions was kept constant by applying 'Generic Distance Constraints' from DISCOVER module of InsightII.

Software and computation platforms

Our in-house molecular modeling software package of MODELYM and ANALYN was used for semi-automated structures prediction based on the existing homologous structures obtained from protein data bank (PDB). ANALYN was used for the analysis of the pre-aligned sequences of the target and scaffold proteins leading to the creation of auxiliary files suitable for MODELYN to use in the automated protein building mode and was run on an IBM compatible PC. MODELYN was used for automated prediction of the target structure and their structural analysis after refinements, and it was run on both an IBM compatible PC in the Windows environment and FUEL workstation of Silicon Graphics Inc. in the UNIX environment. InsightII was run on FUEL workstation of Silicon Graphics, Inc. in the UNIX environment. CLUSTALW [18] was run through the Internet for multiple alignment of the AA sequences. The electrostatic potential surfaces of the proteins were determined by MOLMOL [19]. PROCHECK [20] was used for checking the structural parameters. Both MOLMOL and PROCHECK were run on FUEL in the UNIX operating system. Hydrogen bonding patterns of the modeled and x-ray structures were obtained by adding hydrogen followed by optimization of the complex by energy minimization and molecular dynamics. Protein BLAST [21] was used through the Internet for finding homologous sequences.

Results and discussions

Selection of target protein sequences

BLAST search with the sequence of the mannose selective animal lectin, p58/ERGIC-53, as query in the non-redundant (nr) data base gave 63 hits with a expect value below 1.0×10^{-23} and an amino acid (AA) identity above 30%. Most of the highly significant BLAST matches were with the proteins from higher mammals including human. Homologous proteins of higher eukaryotes with high amino acid identity ($> 50\%$) with the corresponding higher positive scores (67%) are expected to produce structures very similar to the mannose-selective receptor protein. However, proteins showing amino acid identities between 30–50% are expected to give rise to similar structures and may provide similar functions of mannose-selective glycoprotein transporters. Therefore, the selection of target sequences for structure prediction and analysis was confined mostly to this range of AA identity. There are at least 9 distinct human proteins with more than 30% AA identity. We selected three proteins from *Homo sapiens* (Accession No - AAH00347, AAQ89086, CAA5065,) with AA identities of 34%, 46%, and 89%, of which the highest was only the sequence with an AA identity greater than 50% used in this study, which served as a positive control in model building for examining the calcium- and mannose-binding environments akin to those of the mannose selective receptor, p58/ERGIC-53. In addition to these three human proteins, four additional sequences, one from each of the species, were selected from *Mus musculus* (Accession No - BAB29313), *Newrospora crassa* (Accession No - XP_326097), *Tetraodon nigroviridis* (Accession No - CAG12563) and *Anopheles gambiae* (Accession No - XP_313693) which showed reasonable homology with p58/ERGIC-53. Multiple alignment of all these sequences is shown in Fig. 1 (partial region from AA No 61 to 180 spanning calcium- and sugar-binding regions) along with the sequence of p58/ERGIC-53 using 1R1ZA-A as the sequence identifier.

Prediction of homologous structures

Initial three-dimensional structures of the target proteins were predicted by the combinations MODELYN and ANALYN. Most effected segments during insertion, deletion and loop grafting were regularized by energy minimization and molecular dynamics. General structural characteristics of the predicted models were checked by measuring all the bond distances and bond angles and calculating the deviation of these parameters from the standard values for appropriate types of bonds and angles. The quality of backbone conformations were checked by calculating the phi and psi dihedral angles and drawing Ramachandran's plots for all the structures. Table 1 presents the RMSD of bond lengths and bond angles of the predicted structures along with the percentages of backbone Phi-Psi angles in different areas of Ramachandran's plots obtained just after an initial prediction. RMSD of the bond lengths around 0.02 A and those of bond angles around 3 degrees from the respective standard values indicate good general structural parameters of the modeled structures. The overall values of above 95% Phi-Psi pairs in the core and allowed areas of Ramachandran's plot indicate good quality of the backbone conformations of the modeled structures. These Phi-Psi plots were calculated with the structures of targets just after the initial set of regularization; further refinement was done to reduce the number of backbone

Fig. 1 The Clustal W multiple alignment of the partial sequences taking the calcium and sugar binding regions (AA No 61 to 180) of the modeled proteins and the starting scaffold (1R1Z). AAs involved in the co-ordination with the two calcium atoms. The conserved set of AA in all the sequences are identified by a * and semi-conserved AAs are marked with : or . (last-but-one line above numbers)

conformations falling outside the core and allowed areas of the plot. These deviated backbone Phi-Psi pairs could be removed from almost all the structures after proper refinement.

All the modeled structures were structurally superposed with respect to the Cα atoms of the common core of the structures and shown in Fig. 2. The sugar-binding site of p58/ERGIC-53 is not known as the structure of any complex between this receptor with any carbohydrate or glycoprotein is yet to be solved. But it is known that the calcium-mediated sugar binding takes place near the calcium-binding loops so that the binding affinity is modulated by their conformational changes [22]. The location of the mannose tetramer in the modeled complex with the x-ray structure is shown in Fig. 2 in balls-and-sticks representation. It may be noted that the major variations of the target structures occur in the two calcium-binding loops. These structural variations may give rise to the altered specificity and affinity for sugar binding by different proteins of this homologous class.

Calcium binding environment

Both calcium ions in the x-ray structure of p58/ERGIC-53 are hepta co-ordinated [12]. Five co-ordination positions are satisfied from the residues of proteins and two from water molecules. Calcium atoms were added to our modeled structures by superposing them with the x-ray structure with respect to a set of Cα atoms which showed RMSD less than 0.3 Å followed by the transfer of the pair of calcium ions. Equivalent AAs were identified on the modeled structures and distant constraints were applied in order to obtain a co-ordination shell of similar geometry. Energy minimization and molecular dynamics were performed to get a

Table 1 General and backbone structural parameters of the modeled structure of the target sequences as well as the x-ray structure of the p58/ERGIC-53 in its calcium-bound form (PDB ID: 1RIZ)

Species	Accession no	% of AA Identity	RMS deviation in		% of Phi-Psi pairs in the area			
			Bond (Angtrom)	Angle (Degree)	Core	Allowed	Generously allowed	Dis-allowed
Rattus norvegicus	PDB ID: 1R1Z	100	0.018	2.44	86.3	12.8	0.5	0.5
Homo sapiens	CAA50653	89	0.017	2.39	86.4	12.6	0.5	0.5
Homo sapiens	AAQ89086	46	0.017	3.02	86.1	12.8	0.5	0.5
Mus musculus	BAB29313	35	0.018	2.63	81.7	16.0	1.9	0.5
Anopheles gambiae	XP_313693	35	0.019	2.50	80.6	16.2	1.9	1.4
Homo sapiens	AAH00347	34	0.018	2.63	80.3	16.8	1.0	1.9
Newrospora crassa	XP_326097	33	0.018	2.53	78.0	17.7	2.4	1.9
Tetraodon nigroviridis	CAG12563	33	0.019	2.53	80.8	15.9	2.3	0.9

Fig. 2 Ribbon representations of superposed structures of p58/ERGIC-53(X-ray, red) and modeled proteins. Position of the bound tetramer of mannose is shown by balls-and-sticks and calcium ions by space-filling representations

regularized structure of the calcium-binding environment similar to the x-ray structure. The atoms of the residues involved in coordination are presented in Table 2. The structures of calcium co-ordination in p58/ERGIC-53 and one of the modeled proteins from *Homo sapiens* which have 46% AA identity with the starting scaffold are shown in Fig. 3. For most of the modeled proteins we found that five equivalent atoms of the protein are involved in ligand co-ordination of calcium ions. In addition, two water molecules could be placed in the similar positions as those of the x-ray structure.

The usual divalent ion binding site, named S1, present in leguminous lectins is absent in p58/ERGIC-53; but the other metal binding site, named S2, is found in this class of mannose-containing glycoprotein receptor. However, another calcium binding site was found at a different location 6.4 Å away from the S2 site which is unique for this class of receptors [12]. This distance is considerably greater than that between the S1 and S2 sites of the leguminous lectins (4.2–4.5 Å).

Complexes with carbohydrates

The crystal structure of the carbohydrate recognition domain of p58/ERGIC-53 does not provide the co-ordinate position for mannose or any other sugar molecule. It is known that this receptor is a calcium-dependent mannose-selective animal lectin and the carbohydrate-binding site is close to the Ca-binding loops which help to maintain proper conformation of the loops required for sugar binding. We have docked mannose oligomers near the Ca-binding site of p58/ERGIC-53 starting from monomer, dimer, trimer, and tetramer and selected the most probable structures, one for each oligomeric form, based on their empirical energies of binding (Fig. 4). The binding energies of complexes of momomer to tetramer of mannose along with their hydrogen bonding patterns are given in Table 3. It contains empirical energies, both van der Waals and electrical components, of the complex formation and the pattern of hydrogen bonding between the protein and the sugar. The most stable complex (−147.23 kcals/mole) in this series is formed with tetramer of mannose, and the least stable complex (−77.11 kcals/mole) is formed with monomer of mannose. The contribution of the van der Waals interactions for all the complexes range from −24 to −65 kcals/mole, and the electrical contributions show much variation ranging from −52 to −101 kcals/mole. The electrical energies are composed of hydrogen bonding as well as dipole-dipole interactions. In the complex with trimer of mannose higher numbers of hydrogen bonds are formed but tetramer forms the most stable complex due to much greater contribution in electrical energy (Table 3). Binding strength of the complexes increases from monomer to tetramer of mannose.

Tetramer of mannose has been docked into the binding site of the modeled proteins and the binding affinities are studied. The empirical energies of binding of the complexes of modeled proteins with tetramer of mannose along with their hydrogen bonding partners are given in Table 2. The contribution of the van der Waals interactions are very close for all the complexes ranging from −34 to −51 kcals/mole while the electrical contributions show much variation ranging from −66 to −101 kcals/mole. In terms of binding energy, the most stable complex (−130.19 kcals/mole), is formed between a hypothetical protein of *Homo sapiens* (Accession No. AAQ89086) and tetramer of mannose. In this case a higher number of hydrogen bonds is formed between the protein and sugar residue contributing more to the electrostatic interactions. The least stable complex (−109.17 kcals/mole) in this series is formed between a protein of *Newrospora crassa* and tetramer of mannose because it may contain only one calcium ion binding site. The binding strength of the other complexes falls in between.

Electrostatic potential surfaces

Finally we have calculated the electrostatic potential surfaces of the modeled proteins as well as those of the starting scaffold. The electrostatic potential environment of the calcium binding site is extensively negative due to the presence of acidic residues required for calcium-binding except for a hypothetical protein of *Newrospora crassa*, which contains coordinating residues only for one calcium ion. Electrostatic potential surfaces of p58/ERGIC-53 and a few modeled

Table 2 Empirical binding energies of some complexes of proteins with tetramer of mannose along with their hydrogen bonding partners and residues involved in co-ordination with calcium ions

Species	Accession No	% Identity	Residues involved in Co-ordination with		Binding Energy (Kcals/mol)	Residues involved in H-bonding
			210 Calcium	215 Calcium	VdW Elect Total	
p58/ERGIC-53	PDB ID: 1R1Z	100	Asp-129(OD1) Asp-129(OD2) Phe-131(O) Asn-133(OD1) Asp-158(OD2)	Asp-132(OD1) Asp-134(OD1) Asn-138(OD1) Asn-139(OD1) Asp-158(OD1)	−46.07 −101.15 −147.23	Asp-129(OD1)–C3H:H6 Asp-129(OD2)–C2H:O2(H) Asp-160(OD2)–C3H:O3(H)
Homo sapiens	CAA50653	89	Asp-129(OD1) Asp-129(OD2) Phe-131(O) Asn-133(OD1) Asp-158(OD2)	Asp-132(OD1) Asp-134(OD1) Asn-138(OD1) Asn-139(OD1) Asp-158(OD1)	−47.02 −66.20 −113.22	Asp-98(OD1)–C3H:H7 Asp-98(OD2)–C2H:O4(H) Phe-115(NH)–C3H:O2 Asp-129(OD1)–C3H:O3(H) Asn-133(OD1)–C3H:O4(H)
Homo sapiens	AAQ89086	46	Asp-129(OD1) Asp-129(OD2) Pro-131(O) Glu-133 (OE1) Asp-158(OD2)	Asp-134(OD1) Asp-134(OD2) Asp-138(OD1) Gln-136(OE1) Asp-158(OD1)	−34.05 −96.15 −130.19	Leu-115(NH)–C3H:O3 Glu-133(OE2)–C3H:O4(H) Glu-133(OE2)–C3H:H7 Glu-154(OE2)–C1H:O3(H) Glu-154(OE2)–C1H:O4(H) Tyr-229(NH)–C3H:O6 Tyr-229(OG1)–C4H:O3(H)
Mus musculus	BAB29313	35	Asp-129(OD1) Asp-129(OD2) Tyr-131(O) Asn-133(OD1) Asp-158(OD2)	Asp-134(OD1) Asp-134(OD2) Glu-136(OE1) Glu-136(OE2) Asp-158(OD1)	−51.11 −66.36 −117.46	Arg-64(HE)–C4H:O3 Arg-64(HH21)–C4H:O3,O4 Glu-67(OE1)–C1H:O1(H) Glu-67(OE2)–C1H:O1(H) Asp-98(OD1)–C3H:H7 Phe-115(NH)–C3H:O2 Gly-116(NH)–C2H:O2 His-155(ND1)–C3H:O3 His-155(ND1)–C3H:O3(H)
Anopheles gambiae	XP_313693	35	Asp-129(OD1) Asp-129(OD2) Tyr-131(O) Asn-133(OD1) Asp-158(OD2)	Asn-135(OD1) Asn-136(OD1) Gln-138(OE1) Asp-158(OD1)	−47.30 −76.74 −124.04	Asn-67(HD21)–C2H:O5 Asp-98(OD1)–C2H:O4(H) Asp-98(OD1)–C3H:H7 Asn-133(OD1)–C3H:O4(H) Asn-133(OD1)–C3H:O3(H) Asp-229(OD1)–C4H:H7
Homo sapiens	AAH00347	34	Asp-129(OD1) Asp-129(OD2) Tyr-131(O) Asn-133(OD1) Asp-158(OD2)	Glu-134(OE1) Glu-135(OE1) Gln-138(OE1) Glu-139(OE1) Asp-158(OD1)	−45.96 −73.49 −119.46	Asp-98(OD1)–C3H:O4(H) Asp-98(OD1)–C3H:H7 Asp-98(OD2)–C3H:H7 Pro-113(O)–C3H:O2(H) Asn-133(ND2)–C3H:O4 His-155(ND1)–C1H:O3(H) His-155(ND1)–C1H:O3
Newrospora crassa	XP_326097	33	Asp-129(OD1) Asp-129(OD2) Tyr-131(O) Asn-133(OD1) Asp-158(OD2)	Asn-134(OD1) Asp-158(OD1)	−35.90 −73.27 −109.17	Asp-98(OD2–C3H:O3(H) Lys-155(NZ1)–C3H:O2 Glu-226(OE2)–C2H:O2(H) Glu-229(OE1)–C2H:O4(H) Glu-229(OE1)–C4H:O4(H)
Tetradon nigroviridis	CAG12563	33	Asp-129(OD1) Asp-129(OD2) Tyr-131(O) Asn-133(OD1) Asp-158(OD2)	Asp-135(OD1) Asp-139(OD1) Asp-158(OD1)	−47.51 −80.19 −127.71	Asp-98(OD2)–C3H:O2(H) Phe-115(NH)–C3H:O3 Tyr-153(OH)–C3H:O4(H) Asp-229(NH)–C4H:O6 Asp-229(OD2)–C4H:O4(H) Asp-229(OD2)–C4H:H7

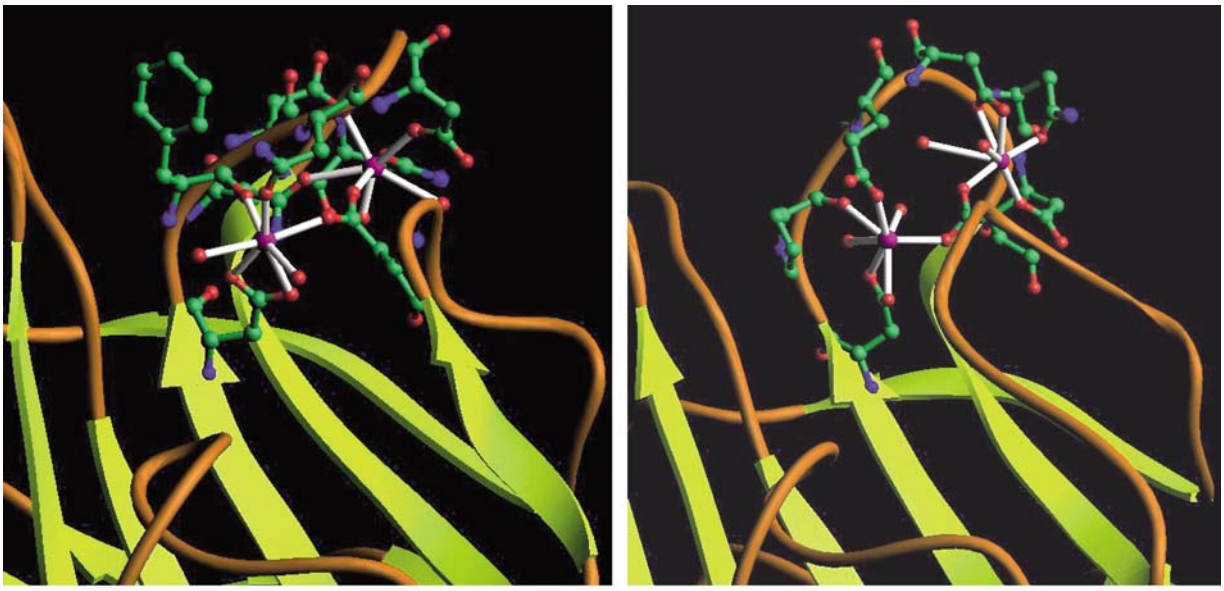

Fig. 3 Residues of protein involved in co-ordination with calcium ions are depicted in balls-and-sticks representation and calcium ions in pink spheres. (A) p58/ ERGIC-53 [calcium ions are hepta co-ordinated: five from the residues of protein and two from water molecule]. (B) a hypothetical protein of Homo sapiens [same pattern of calcium co-ordination with the starting scaffold]

proteins are shown in the Fig. 5. The mannose-binding site is mostly negatively charged with a small patch of a positively charged region in case of modeled proteins from *Homo sapiens* (Accession No. CAA50653), *Newrospora crassa*, and p58/ERGIC-53. The mannose-binding site is totally negatively charged in case of modeled protein from *Homo sapiens* (Accession No. AAQ89086). The highly negative potential surface around the positively charged calcium ion-binding site is necessary for the neutralization of the positive charge as is observed in other calcium-binding lectins. The mannose-binding site being very close to the calcium-binding sites also has the high negatively charged environment.

Table 3 Empirical binding energies of the complexes of p58/ERGIC-53 with monomer to tetramer of mannose along with their hydrogen bonding partners

	Bindings Energies in Kcals/mol*			
Structure	VdW	Elect	Total	Residues involved in H-bonding
p58/ERGIC-53 in complex with monomer	−24.74	−52.37	−77.11	Asp-129(OD2)–O2(H) Phe-146(NH)–O3 Asp-160(OD1)–O4(H) Asp-160(OD2)–H6
p58/ERGIC-53 in complex with dimer	−45.45	−52.54	−97.99	Asp-160(OD1)–C1H:O2(H) Asp-164(OD1)–C2H:O1(H) Asp-185(O)–C2H:O4(H)
p58/ ERGIC-53 in complex with trimer	−65.50	−61.60	−124.10	Asp-129(OD1)–C1H:O4(H) Asn-169(HD21)–C3H:O6 Asn-170(ND2)–C2H:O1(H) His-186(HE2)–C1H:O6 Asp-189(O)–C2H:O2(H) Asp-189(OD2)–C1H:H6 Ser-197(HG)–C3H:O2 Ser-197(HG)–C3H:O3
p58/ ERGIC-53 in complex with tetramer	−46.07	−101.15	−147.23	Asp-129(OD1)–C3H:H6 Asp-129(OD2)–C2H:O2(H) Asp-160(OD2)–C3H:O3(H)

*VwD = van der Waals & Elect = electrical energies.

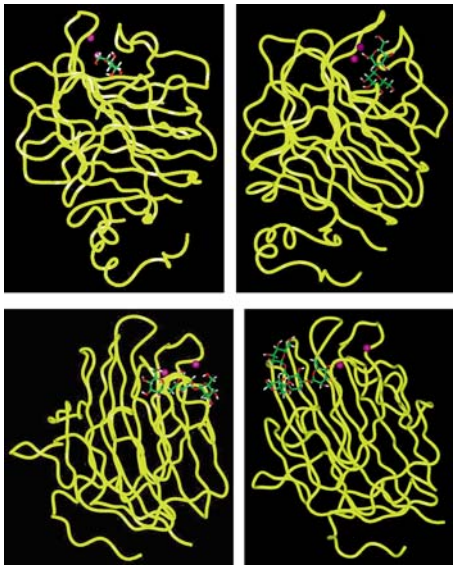

Fig. 4 Ribbon representation of the structure of p58/ERGIC-53 (yellow) in complex with monomer to tetramer of mannose. Calcium ions are denoted as pink spheres and oligomers of mannose are depicted in stick representation. Top left: monomer; Top right: dimer; Bottom left: trimer; Bottom right: tetramer

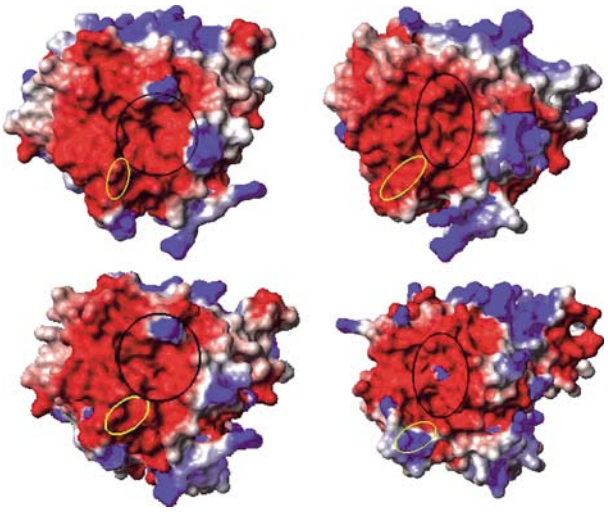

Fig. 5 Electrostatic potential surfaces of a few modeled proteins. Blue region represents positively charged environment, red for negatively charged, and white for hydrophobic surroundings. The calcium-binding site is shown by yellow bands and mannose-binding site by black bands. Top left: p58/ERGIC-53; Top right: a hypothetical protein of *Homo sapiens* (46%); Bottom left: a hypothetical protein of *Homo sapiens* (89%); Bottom right: a hypothetical protein of *Newrospora crassa*

Conclusions

We have modeled the structures of some proteins from various species based on the experimental structure of p58/ERGIC-53, and mannose (from monomer to tetramer) was docked into the sugar-binding sites of these structures to study the physico-chemical forces involved in such binding. Calculation of empirical energies of binding of the complexes of p58/ERGIC-53 with different oligomers of mannose indicates the binding strength increases monomer to tetramer. Similar type of homology modeling and docking studies was done by Wang *et al.* [23] for selective Histone deacetylase inhibitor design, who estimated the free energy of binding in the docked structure and found to agree with the experimental observation. Electrostatic potential environment of the calcium-binding site is extensively negative due to the presence of acidic residues required for calcium binding except for a hypothetical protein of *Newrospora crassa*, which contains coordinating residues only for the 1st calcium ion. Empirical binding energy calculation of the complexes of tetramer of mannose with these proteins revealed that a protein of *Homo sapiens* (Accession No. AAQ89086) would bind most strongly (−130 Kcals/mol) due to the formation of a large number of hydrogen bonds between the protein and mannose. The least binding energy (−109.17 Kcals/mol) was found with the protein of *Newrospora crassa* because it may contains only one calcium ion-binding site. In summary it may be stated that three-dimensional structure prediction of putative receptors identified by homology search using molecular modeling techniques could identify the calcium- and mannose-binding sites allowing structural analysis at the molecular level of this structurally unique class of glycoprotein receptors.

Acknowledgments This work was financially supported by the Mission Mode Project No CMM0017 of CSIR, India.

References

1. Caron, M., Bladier, D., Joubert, R.: Soluble galactoside-binding vertebrate lectins: a protein family with common properties. Int. J. Biochem. **22**, 1379–1385 (1990)
2. Barondes, S.H., Cooper, D.N.W., Gitt, M.A., Leffler, H.: Galectins, structure and functions of a large family of animal lectins. J. Biol. Chem. **269**, 20807–20810 (1994)
3. Weis, W.I., Drickamer, K., Hendrickson, W.A.: Structure of a C-type mannose-binding protein complexed with an oligosaccharide. Nature **360**, 127–134 (1992)
4. Iobst, S.T., Wormand, M.R., Weis, W.I., Dwek, R.A., Drickamer, K.: Binding of sugar ligand to Ca^{2+}-dependent animal lectins. Analysis of mannose binding by site-directed mutagenesis and NMR. J. Biol. Chem. **269**, 15505–15511 (1994)
5. Kolatkar, A.R., Weis, W.: Structural basis of galactose recognition by C-type animal lectins. J. Biol. Chem. **271**, 6679–6685 (1996)
6. Gabius, H.J.: Non-carbohydrate binding partners/domain of animal lectins. Int. J. Biochem. **26**, 469–477 (1994)
7. Gabius, H.J.: Detection and functions of mammalian lectins with emphasis on membrane lectins. Biochim. Biophys. Acta. **1071**, 1–18 (1991)
8. Gabius, H.J.: Animal lectins. Eur. J. Biochem. **243**, 543–576 (1997)

9. Barondes, S.H.: Lectins: their multiple endogenous cellular functions. Annu. Rev. Biochem. **50**, 207–231 (1981)
10. Neve, E.P.A., Svensson, K., Fuxe, J., Pettersson, R.F.: VIPL, a VIP36-like membrane protein with a putative function in the export of glycoproteins from the endoplasmic reticulum. Experimental Cell Research **288**, 70–83 (2003)
11. Velloso, L.M., Svensson, K., Pettersson, R.F., Lindqvist, Y.: Crystal structure of the carbohydrate recognition domain of p58/ERGIC-53, a protein involved in glycoprotein export from the endoplasmic reticulum. J. Biol. Chem. **277**, 15979–84 (2002)
12. Velloso, L.M., Svensson, K., Pettersson, R.F., Lindqvist, Y.: The Crystal structure of the carbohydrate-recognition domain of the glycoprotein sorting receptor p58/ERGIC-53 reveals an unpredicted metal-binding site and conformational changes associated with calcium ion binding. J. Mol. Biol. **334**, 845–51 (2003)
13. Nichols, W.C., Seligsohn, U., Zivelin, A., Terry, V.H., Hertel, C.E., Wheatley, M.A., Moussalli, M.J.: Hauri, H.P., Ciavarella, N., Kaufman, R.J., Ginsburg, D.: Mutations in the ER–Golgi intermediate compartment protein ERGIC-53 cause combined deficiency of coagulation factors V and VIII. Cell **93**, 61–70 (1998)
14. Moussalli, M., Pipe, S.W., Hauri, H.P., Nichols, W.C., Ginsburg, D., Kaufman, R.J.: Mannose-dependent endoplasmic reticulum (ER)–Golgi intermediate compartment-53-mediated ER to Golgi trafficking of coagulation factors V and VIII, J. Biol. Chem. **274**, 32539–32542 (1999)
15. Vollenweider, F., Kappeler, F., Itin, C., Hauri, H.P.: Mistargeting of the lectin ERGIC-53 to the endoplasmic reticulum of HeLa cells impairs the secretion of a lysosomal enzyme. J. Cell Biol. **142**, 377–389 (1998)
16. Appenzeller, C., Andersson, H., Kappeler, F., Hauri, H.P.: The lectin ERGIC-53 is a cargo transport receptor for glycoproteins. Nat. Cell. Biol. **1**, 330–334 (1999)
17. Mandal, C.: MODELYN-A molecular modelling program version PC-1.0 Indian Copyright No 9/98, (1998)
18. Thompson, J.D., Higgins, D.G., Gibson, T.J.: CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic. Acids. Res. **22**, 4673–4680 (1994)
19. Koradi, R., Billeter, M., Wuthrich, K.: Related Articles, MOLMOL: A program for display and analysis of macromolecular structures. J. Mol. Graph. **14**, 51–5 (1996)
20. Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M.: PROCHECK: A program to check the stereochemical quality of protein structures. J. Appl. Cryst. **26**, 283–91 (1993)
21. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J.: Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic. Acids. Res. **25**, 3389–02 (1997)
22. Mitchell, E., Houles, C., Sudakevitz, D., Wimmerova, M., Gautier, C., Perez, S., Wu, A.M., Gilboa-Garber, N., Imberty, A.: Structural basis for oligosaccharide-mediated adhesion of *Pseudomonas aeruginosa* in the lungs of cystic fibrosis patients. Nat. Struct. Biol. **9**, 918–21 (2002)
23. Wang, D.F., Helquist, P., Wiech, L.N., Wiest, O.: Toward slective histone deacetylase Inhibitor design: homology modeling, docking studies, and molecular dynamics simulations of human class I histone deacetylases. J. Med. Chem. **48**, 6936–47 (2005)