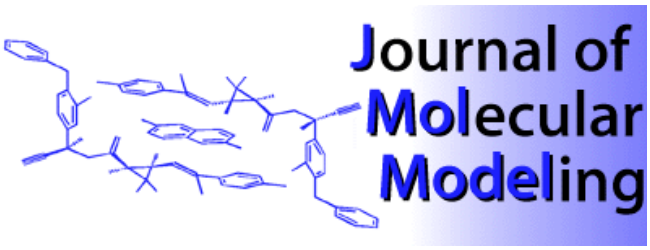
J. Mol. Model. **1998**, *4*, 310 – 322



FULL PAPER

Molecular Model for the C-type Lectin Domain of Human Thrombomodulin

Bruno O. Villoutreix and Björn Dahlbäck

Lund University, The Wallenberg Laboratory, Dept. of Clinical Chemistry, University Hospital, Malmö, S-205 02 Malmö, Sweden. Fax: (46) 40 33 70 44. E-mail: bruno.villoutreix@klkemi.mas.lu.se

Received: 20 August 1998 / Accepted: 18 September 1998 / Published: 1 October 1998

Abstract Thrombomodulin (TM) is a multi-modular membrane receptor (557 residues) present on the surface of endothelial cells. TM binds thrombin (T) and this complex promotes downregulation of the coagulation cascade via activation of protein C and delay fibrinolysis through activation of the thrombin-activatable fibrinolysis inhibitor (TAFI). The N-term region (155 residues) of TM possesses the signature of the C-type lectin domain. This module seems required for constitutive internalization of the T-TM complex, plays a role in the modulation of cell growth and may direct soluble forms of TM (or T-TM) to specific regions of the vasculature during inflammation and in a variety of vascular disorders. The understanding of this domain is however limited and structural information would contribute to the design of new experiments aiming at characterizing its functions. We have developed a 3D model for the lectin domain of TM using prediction-based threading and comparative model building. The X-ray structures of lithostathine (LIT), mannose-binding protein (MBP) and E-selectin (ESL) were used as initial templates. Despite a sequence identity of about 28 % between TM and LIT (best score) it is possible to build an accurate 3D model for TM. The TM lectin domain contains two α-helices, two β-sheets and a compact hydrophobic/aromatic core. The disulfide bridging pattern of TM has not been reported experimentally but the model proposes the formation of four disulfide bonds between C12-C17, C34-C149, C78-C115 and C119-C140. Based on the model, potential binding sites are proposed.

Keywords Coagulation, Thrombosis, Protein C, Lectin, Thrombomodulin, Protein modeling

Running Title 3D structure for the lectin domain of TM

Abbreviations PC, protein C; APC, activated protein C; T, thrombin; TAFI, thrombin-activatable fibrinolysis inhibitor; TM, thrombomodulin; T-TM, thrombin-thrombomodulin; EGF, epidermial growth factor; MBP, mannose-binding protein; LIT, lithostathine; ESL, E-selectin; r.m.s.d., root mean square deviation; 3D, three-dimensional.

Correspondence to : B.O. Villoutreix

Introduction

The relevance of the protein C (PC) anticoagulant system, to which thrombomodulin (TM) belongs, is illustrated by an increased risk of thrombosis in individuals with PC or protein S deficiency or with the R506Q amino acid substitution in factor V which results in resistance to activated protein C (APC) [reviewed in 1-3]. Thrombin, the final serine protease of the procoagulant pathway, binds to TM and the complex recruits the circulating protein C (PC) zymogen and catalyzes its rapid conversion to its activated form (APC). APC, then, inhibits further production of thrombin via proteolytic degradation of two essential cofactors (factors Va and VIIIa) in the presence of protein S [reviewed in 4-8].

Thrombomodulin is a multi-modular glycoprotein, essentially present at the surface of endothelial cell membrane, consisting of a N-terminal lectin-like domain [9] and hydrophobic segment, six EGF-like domains, a glycosylated serine/threonine-rich domain (with or without a chondroitin sulfate chain), a putative transmembrane domain and a C-terminal cytosolic tail [10-12]. TM EGF5 and EGF6 are essential for thrombin binding while the EGF6-EGF5-EGF4 segment is required for PC activation [13-15]. The previous segment together with TM EGF3 play a role in TAFI binding and activation by thrombin [16]. The functions of the lectin domain, of EGF1 and EGF2, of the transmembrane and cytoplasmic domains are not well defined.

The role of TM in hemostasis is well established as it markedly increases the ability of thrombin to activate PC and eradicates its procoagulant properties [reviewed in 6]. A new function of TM has also been demonstrated after a study in mice in which the TM gene had been deleted [17], causing embryonic lethality before the assembly of a functional cardiovascular system (indicating a crucial role of TM in early development). Recently, it was found that a single amino acid substitution in the mouse gene (E404P) leads to the dissociation of the developmental function of TM from its role in the regulation of blood coagulation [18]. In addition to its role in preventing coagulation, TM has been suggested to modulate angiogenesis, as the formation of the T-TM complex restrains human umbilical vein endothelial cell proliferation [19,20]. Loss of TM from the endothelial cell surface is thought to contribute to thrombosis as encountered in malignant and inflammatory diseases. Internalization or endocytosis of TM from the cell surface has been proposed to be one mechanism by which TM levels are reduced [21]. The lectin domain of TM and thrombin binding to TM have been implicated in the internalization process [22] but the data remain controversial [23]. Over-expression of TM decreases cell proliferation *in vitro* and tumor growth *in vivo*, and the TM lectin domain seems involved in the modulation of cell growth [24]. This mechanism is independent from thrombin and the thrombin receptor [24]. Possibly, a yet unknown ligand binds to the TM lectin domain and triggers signal transduction pathways dependent on the TM cytoplasmic domain [24]. Soluble forms of TM have been identified in plasma and urine [25]. A high level of soluble TM antigen in plasma was observed in diseases accompanying endothelium injury such as diabetes mellitus with microangiopathy, disseminated intravascular coagulation and systemic lupus erythematosus [26-28]. The role of soluble TM is not clearly understood but it may inhibit fibrinolysis by potentiating the activation of the thrombin-activatable fibrinolysis inhibitor, TAFI [29].

The TM lectin domain seems clearly important and structural information about this region would be valuable to investigate further the functions of this receptor. The N-term 155 residues of TM have been shown to belong to the C-type lectin family of module [9,30]. It is here important to underline that many investigators call the first 226 residues of TM a lectin domain while only the first 155 residues belong to the C-type lectin family. The C-type lectin domain or carbohydrate recognition domain, consists of about 130-140 residues. There are four cysteines which are strictly conserved and involved in disulfide bonds (see the PROSITE database, http://expasy.hcuge.ch/sprot/prosite.html; [31]). This lectin module appears in many proteins [reviewed in 32] and several experimentally determined 3D structures for this family of module [reviewed in 33] and molecules with a related 3D fold have been reported. For instance, the X-ray structures of the mannose-binding protein (MBP, 1.7 Å resolution) [34], E-selectin (ESL, 2.0 Å resolution) [35], the pancreatic inhibitor of stone formation, lithostathine (LIT, 1.55 Å resolution) [36], an angiogenesis inhibitor named endostatin (1.5 Å resolution) [37] and the NMR of the Link module [38] have been described. The structure of LIT is characterized by the presence of two long α-helices and six β-strands, the latter forming two antiparallel β-sheets (B1-B6-B2 and B3-B4-B5). The C-type lectin structures also contain several extended loops and stretches of non-regular secondary structure (about 60 % of the residues). Interestingly, this family of module adopts a similar fold in spite of a low to very low sequence identity. For example, the sequence identity between LIT and MBP or ESL is only 19 % with an r.m.s.d. for the 109 matched α carbon positions of about 1.6 Å [36]. The sequence identity between endostatin and ESL is only 9 % while 77 equivalent α carbons can be superimposed with an r.m.s.d. of 3.1 Å [37]. This observation indicates that it is possible to build 3D model for the C-type lectin module with a high degree of confidence as the fold is conserved.

In the present study, we developed a 3D model for the lectin domain of TM using the LIT X-ray structure as initial framework while taking into account structural data from ESL and MBP. The LIT structure was selected to build TM as the sequence identity between the two proteins is higher (28 %) as compared to ESL (20 %) or MBP (23 %). Also, threading experiments [39] propose LIT as the most likely template for TM. We validate our model via structural analysis and by mapping known experimental data that possess atomic resolution (e.g., known glycosylated asparagine residues must be solvent exposed). We then screened the protein surface searching signals for potential binding sites.

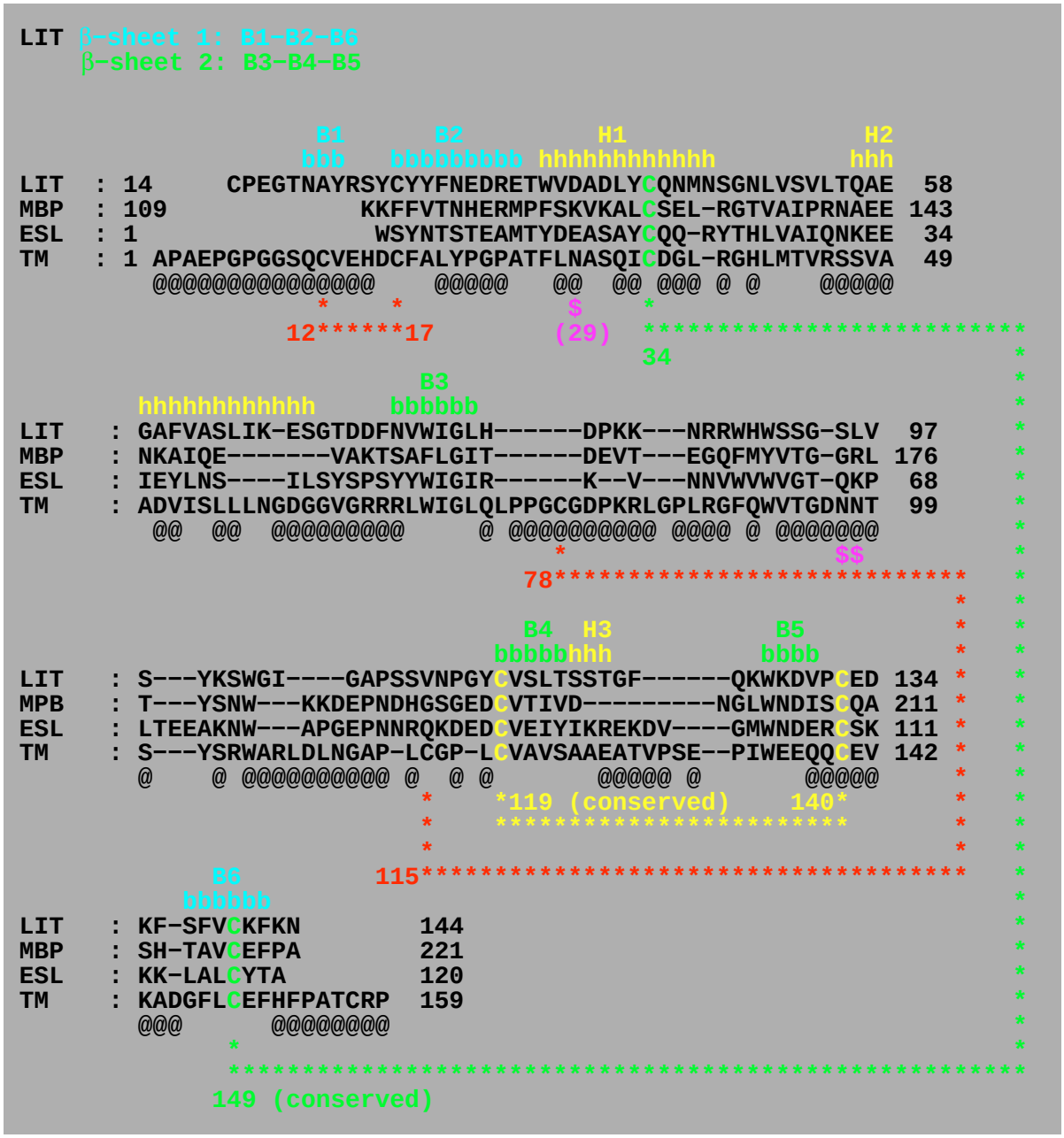


Figure 1 Sequence and structural alignments. *The structures of LIT and MBP and of LIT and ESL were aligned. The three amino acid sequences and TM were then aligned accordingly. The β-strands (b) and helices (h) present in the X-ray structure of LIT are reported above the sequences. These elements are essentially conserved in the TM model. The proposed disulfide bonds of TM are indicated by stars and the residue number is mentioned. The $ symbols indicate glycosylated asparagine residues in TM while the @ symbols indicate the TM residues which are solvent exposed. Clearly, the first N-term and last C-term residues are fully solvent exposed as they have been built in an extended conformation. Thus, some residues there may be buried in the native structure*

Methods

A Silicon Graphics Indigo2 (SGI) R10000 workstation was used for the construction of the model together with the InsightII/Discover program package (Biosym-MSI). Figures were essentially prepared within InsightII and edited with the SGI utility Showcase.

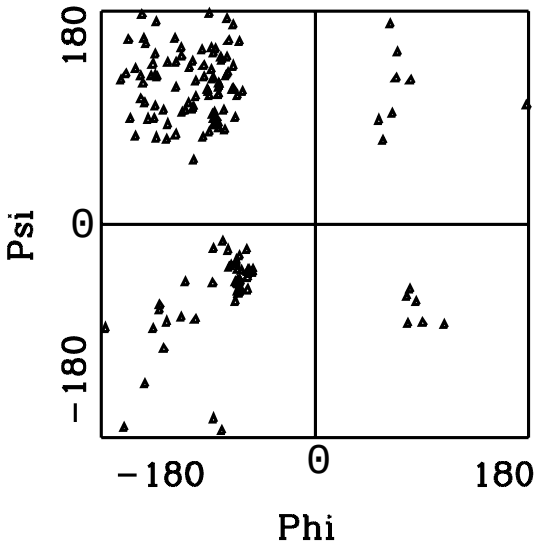


Figure 2 *Ramachandran plot for the TM model. This plot shows that most of the backbone dihedral angles are in the energetically favored regions*

Structural and sequence alignment

First, a multiple sequence alignment based on a structural alignment between LIT, MBP and ESL was used to align the lectin region of the TM sequence to LIT (Fig. 1). The sequence identity between the TM lectin domain and LIT, MBP and ESL is 28 %, 23 % and 20 %, respectively, with clear amino acid conservation within the central core. The UCLA-DOE fold recognition server [39] was also used. The top Z-score (= 10.64) identified LIT as most similar fold, followed by MBP (Z-score = 5.95) and ESL (Z-score = 3.04). With this method the confidence threshold is a Z-score of 4.8 ±1. Thus, as our Z-score is above this threshold, the structural prediction of TM using LIT as initial template is reliable. The sequence alignment between TM and LIT resulting from the threading experiments was similar to the structure-based multiple sequence alignment reported in figure 1. This last alignment was thus used to build the TM model.

Model building

The X-ray structures of LIT, MBP and ESL were obtained from the Protein Data Bank (PDB) [40] and LIT was used as initial framework to build the lectin domain of TM. Conservative side-chain replacements were modeled in conformations similar to the ones present in the LIT structure. Other residue replacements were modeled using the LIT coordinates as initial template but optimized, if needed, using low-energy rotamer conformations [41]. Low-energy rotamers were calculated using a 10 Å cutoff distance for non-bonded interactions. The insertion regions were built either from a search [42] among some of the high-resolution protein structures present at the PDB [43] or using the random tweak algorithm which generates a peptide fragment *de novo* by randomly searching the conformational space for a suitable backbone conformation while a screening for steric overlap violations and a set of distances between the two anchor residues are taken into account [44]. Indeed, both methods were tested and the fragment which accommodated the best with the remaining part of the structure was selected. For the loop search method [42], at least 2 residues prior (preflex) and after (postflex) the loop (flex) to be built were selected for the r.m.s. calculations (i.e., calculations between the α-carbon distance matrix database extracted from the PDB and the preflex and postflex residues of the segment to build). For the *de novo* loop generation, two backbone atoms, N and Cα, for the starting residue prior to the loop to build, and the Cα and CO atoms for the stop residue defined six distances that described the geometry of the base of the loop. The conformation of the base of the loop (preflex and postflex residues), like in the case of the loop search method, is expected to be very reliable as the structure is there conserved in the C-type lectin family. Thus, for the *de novo* loop generation, bond lengths and angles of the residues to introduce between the preflex and postflex residues must meet a certain criteria for the loop to be acceptably closed. The differences between the desired distances to build the loop and their current values are minimized using a linearized Lagrange multiplier method. After a series of iterations, all the loops were closed, indicating that the distance between the preflex and postflex is appropriate, since otherwise, the loops could not be closed. The resulting loop structures were investigated on the computer screen. These TM loops and connecting segments attempt to follow the overall conformation as observed not only in LIT, but also in MBP and ESL.

Deletions in TM when compared to LIT were effected by computationally removing the appropriate residues. TM residues 1-5 and 154-159 had no counterpart in LIT and were built in extended conformation. These two regions have thus to be considered with caution. Disulfide bonds were created interactively between the appropriate cysteine residues. As these residues were close in space, this procedure was straightforward.

Energy refinement

The TM model was then energy minimized using Discover (Biosym-MSI). All calculations were carried out using the CVFF force field parameters, a dielectric constant of 1 and a 20 Å cutoff distance for non-bonded interactions (repulsive and attractive Lennard-Jones and Coulombic terms) [reviewed in 45]. Other calculations were performed with a dielectric constant of 5, but no clear differences were noted between the resulting models. As simply cutting the nonbond interactions off at a given distance leads to discontinuities in the energy evaluation and its derivatives, a switching function as

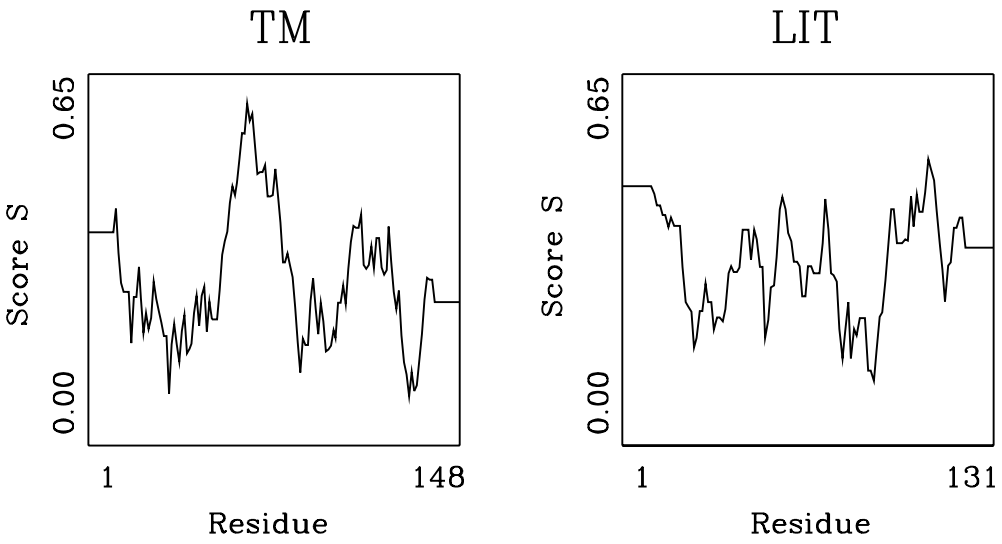


Figure 3 *Profile plot for the TM model and LIT X-ray structures. The compatibility score for residues in a 21-amino acid sliding window is presented. Such profile indicates that both proteins are correctly folded*

integrated in Discover was used. This function allows for the interaction energy and its first (and second) derivatives to be continuous. Hydrogen atoms were added to the model and partial charges were assigned to all atoms. However in an effort to account for the lack of solvent and ions, potentially charged residues (Arg,Lys,His,Glu,Asp) were given appropriate parameters to obtain electrostatic neutrality [45]. The model was refined in a stepwise fashion, initially the insertion/deletion regions were refined while the surrounding atoms were fixed. Then, all heavy atoms were slightly tethered (the force constant, *K*, added on all heavy atoms was 5 kcal·Å^{-2}) to their original positions and subsequently relaxed (K = 0) until the maximum Cartesian derivatives of the energy was less than 1.0 kcal·mol^{-1}·Å^{-1}. This convergence criteria was judged sufficient for optimal relaxation of covalent bonds or slightly overlapping (e.g., newly added hydrogens) atoms and for subsequent molecular dynamics run. At the beginning of the simulation a robust algorithm (steepest descents) was used but as soon as the derivatives were below 10.0 kcal/mol-Å, the refinement was performed with conjugate gradients.

Simulation of the insertion loops

A long molecular dynamics (MD) of the TM model in the presence of solvent would be computationally prohibitive and is beyond the scope of this study. Yet, in order to analyze and refine further the insertion regions, short MD calculations with annealing were performed. Thus TM segments 35-39; 75-87 and 113-118; 94-100; 105-110 were simulated. In all cases, only one loop was allowed to move during the simulation while the remaining part of the structure was fixed. However, loops 75-87 and 113-118 were not considered independent as they are direct topological neighbor. Thus both loops were simulated at the same time. The protocol used to probe the conformation of the TM insertion regions was as follow: a conformational search using a 20-picosecond (ps) high-temperature (500 K) molecular dynamics (MD) simulation was performed on the loop keeping the remaining part of the protein atoms held fixed. A temperature higher than 300 K was deemed appropriate in order to cross more easily energy barriers between various local minima and to reduce the length of the simulations. The coordinates saved in the 20-ps history file were collected at 1-ps intervals, annealed to 300 K for 1-ps and each 20 structures were energy minimized using 100 iterations of steepest descents algorithm and 100 iteration of conjugate gradients minimizer. The average structure for each loop was selected and the entire model briefly energy minimized. This last model was then used for subsequent analysis.

Electrostatic potential

Electrostatic potential calculations were carried out for the TM model structure. The Delphi package, which solves the Poisson-Boltzmann equation by a finite difference method [46], was used for the calculation of the 3D distribution of the electrostatic potential at physiological ionic strength and pH. The atomic coordinates of the model structure were mapped into a 3D grid with a resolution of about one grid point/Å. The grid was chosen to leave a 10 Å border between the protein and the grid edge. The dielectric constant was set to 5 for the protein interior and 80 for the surrounding solvent. Atomic radii definitions were taken from the Delphi

default parameters with the radii of hydrogen atoms set to 0. A standard set of formal charges was assigned to the titratable residues (e.g., Arg, +0.5e at the $N_{\eta 1}$ and $N_{\eta 2}$ nucleus). The N-terminal amino group (residue 1) and the C-terminal carboxy group (residue 157) were considered neutral as their exact position is not precisely defined. The resulting 3D isopotential contours were analyzed interactively within InsightII.

Validation of the model structure

The model was screened interactively and relevant experimental data were mapped onto the structure. The normalized static solvent accessibility was calculated according to the method of Lee & Richards [47]. The stereochemistry of the model and X-ray template was analyzed using ProStat (Biosym-MSI). Thus bond lengths, backbone Ω-angles, side chain χ1 and χ2 angles, chirality and values [48] for the phi-psi core region occupancy were investigated. The accuracy of the TM model structure was also assessed by inspection of its 3D profile [49] and the scores were compared with those computed for LIT. Secondary structure elements (e.g., helices and strands) were predicted according to the definition of Kabsch and Sander [50].

Results and discussion

Comparative model building: theory and background

Determining protein sequences is now routine in most laboratories and the number of sequence reported has increased very rapidly as compared to structural information resulting from X-ray crystallographic or NMR results. *A priori* folding rules have not yet been established but structural predictions based upon homologous reference proteins for which the 3D structures have been obtained experimentally represent an elegant and very efficient way of predicting the structure of a protein [51].

Thus comparative model building has been used successfully when the sequence identity between the protein to build and the references was around 40 % and above. Building 3D model when the sequence identity ranges from 15 % to 30 % has often been criticized essentially due to the lack of examples of proteins sharing a similar fold while having limited sequence identity. Although it is true that building 3D model for a protein that for example shares only 20 % with the template can be challenging, it is clear that very reliable models can be proposed [52]. Indeed, several powerful approaches have been developed to detect remote homologies (proteins or modules with a sequence identity between 10-20 % sharing the same fold and often having similar function), like for instance, the generation of profiles or the use of hidden Markov models analysis [reviewed in 53]. Then, with these methods in hand, a family of sequences can be aligned with precision and thus a 3D model can be developed in most cases by comparative model building. It is in fact known that fold recognition and structural prediction of remote homologues is achieved most of the time with the computational methods available at present (e.g., threading, secondary structure prediction, generalized profile or related methods) and a survey of the literature [54]. Errors can however occur (due to for instance local misalignment) but essentially cluster in the loop regions [52]. Even when such errors are present, the models are sufficiently accurate to predict binding sites or other functional features as well as to disqualify wrong interpretations of experimental data. Structural prediction of remote analogues (proteins with a sequence identity usually inferior to 10 % with no common ancestor and generally no common function, but similar 3D fold) is however still very difficult and the resulting 3D models have to be considered with cautions [54].

Modeling the N-term region of human thrombomodulin

Although the sequence identity between the C-type lectin modules is somewhat low, it is clear that the 3D structure of these modules is conserved. Because numerous experimentally determined structures have been reported, it is relatively easy to align the sequences and to introduce insertions and deletions in the correct regions of the proteins based upon their structural alignments. Yet, some residues can be aligned slightly differently in the loops and connecting segments depending on the criteria used (e.g., r.m.s.d., amino acid conservation). Such slight change(s) in the sequence alignment within these solvent exposed regions do not however modify significantly the overall output of the modeling process as they are regularized in the same fashion during energy refinement calculations. Thus, several C-type lectin models have been developed during these last few years and in all cases the structures shed light on some of the domain functions [55-57].

The distant homology between the first 155 residues of TM and the C-type lectin family is well established [9,30] and was used as background for the present study. Validation of the relationship between the N-term region of TM and the C-type lectin family based on previous sequence analyzes [9,30] is indeed supported by the threading investigation and the model building experiment reported in this article (see Methods). Detailed evaluation of the TM lectin model is reported below and structural analysis with the goal of identifying potential binding sites and function(s) of the module follows.

Validation of the TM model

Sequence identity evaluation and threading suggested that LIT is presently the best template to build the lectin domain of TM. Although a similar output would be achieved by using MBP as template, since the structure of LIT was available and that the resolution is very high (1.55 Å), it was clear that LIT was the appropriate template to build TM. Three

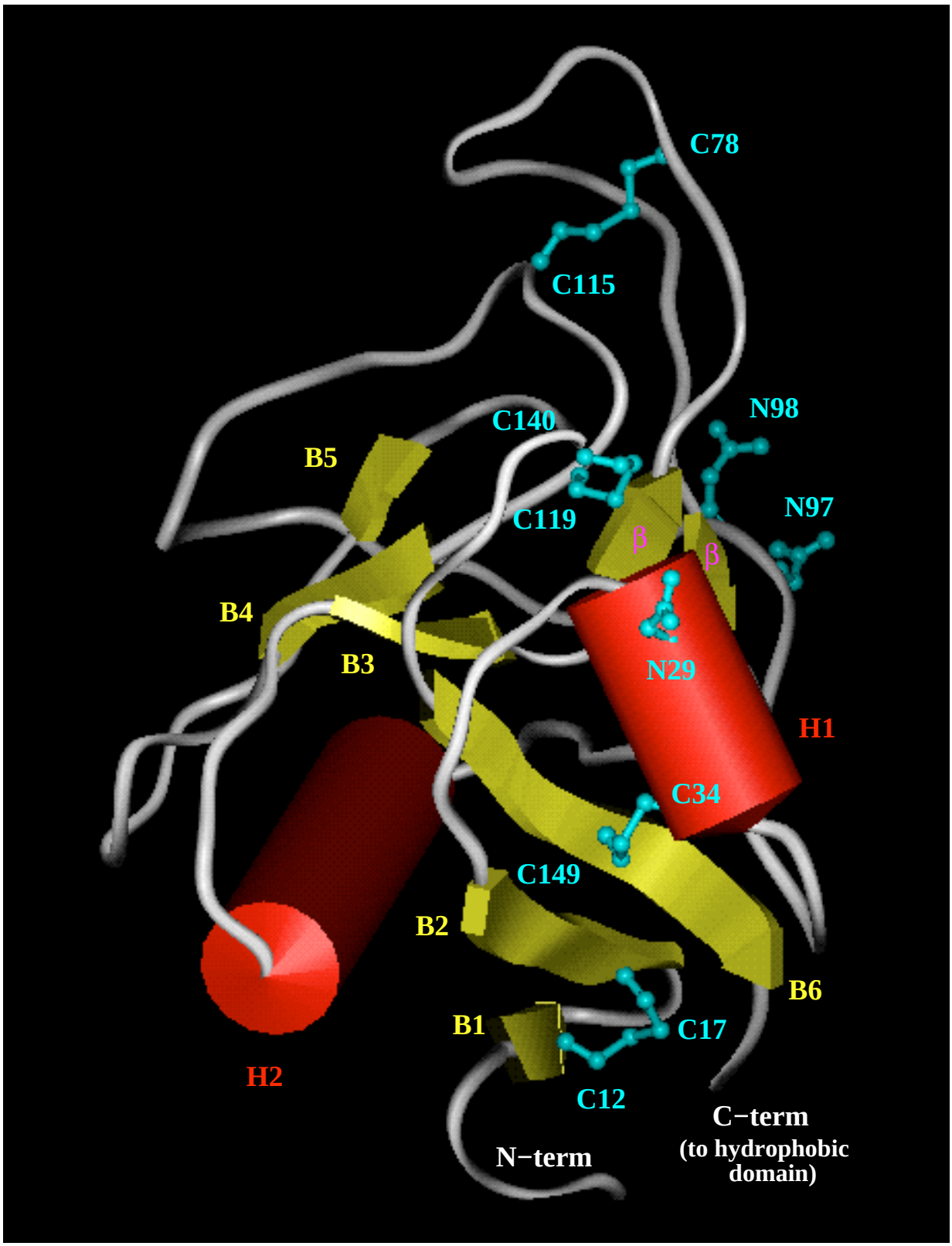


Figure 4 *Ribbon representation of the TM model. The trace of the TM model (residues 5-155) is in white with the β-strands in yellow and the two helices in red. The two proposed extra β-strands in TM as compared to LIT are highlighted with the symbol β (magenta). The disulfide bridges are in blue and the asparagine residues known to be glycosylated are also shown (blue)*

insertions in TM as compared to LIT are observed but are located in solvent exposed areas (Fig. 1). Thus, these regions could be built without disrupting the central core which is conserved in this family of module. The r.m.s.d. between the final TM model and LIT (for 83 equivalent Cα atoms, 59 % of the residues) is 2.5 Å. This r.m.s.d. is in the expected range and illustrates the need of some conformational changes within the LIT framework to accommodate the TM side chains.

The final TM model bond lengths and angles were analyzed with ProStat (Biosym-MSI). No unusual structural features were noticed supporting the appropriateness of our modeling strategy. A Ramachandran plot is reported in figure 2 and it can be seen that most phi-psi angles are within the energetically favored regions. The value for percent of phi-psi core region occupancy for the TM model was about 65, indicating that the backbone conformation is mostly correct [48]. This score is consistent with the large amount (around 60 %) of residues adopting a non-regular secondary structure in the C-type lectin family.

When the compatibility scores of a misfolded proteins are plotted as a function of the amino acid sequence, the overall profile is often below 0.1 and can dip several times below zero [49]. We have run such calculations for the TM model (Fig. 3). The profile is always above zero and the self-compatibility scores are essentially above 0.1, supporting also the quality of our model structure. The profile for the TM model is equivalent to the one calculated for LIT (Fig. 3), further supporting the overall accuracy of our prediction.

The distribution of the hydrophobic and charged residues within the TM model is thus in agreement with what is known from structural analysis of correctly folded proteins. After interactive analysis of the model we noticed that a few charged residues were partially buried from the solvent (Fig. 1). However, they were all involved in salt bridges. D16-H152 (partially buried), H40-D35-R38, D51-R103-E132 (partially buried), R66-E126, D80-K82, R83-E137 (partially buried), D108-R67-E136 and E150-R45-D16 (partially buried) were found involved in electrostatic interactions. These ionic contacts are not conserved in LIT. It is however known that ionic interactions are not always conserved in a family of protein.

Another key point for the validation of our model was the ability to reproduce the lectin C-type disulfide bridging pattern. To our knowledge, the disulfide bonds (S-S bonds) have not been reported for the N-term region of TM. However, it is clear that disulfide bridges involve TM C34-C149 and C119-C140. These S-S bonds are conserved in the family (Figs. 1 and 4). It seems that TM C12 and C17 form a disulfide bridge by analogy with an almost equivalent bond in LIT (LIT C14 and C25). Definitively, TM C78 and C115 are far from C157 but are close in space. As it would be very surprising to find surface exposed cysteines not involved in disulfide bridging (free cysteines are almost always buried in a native protein), it is likely that C78 and C115 forms a S-S bond. Thus, C157, at the beginning of the hydrophobic region, should form a S-S bond with C206. Further validation of the TM model involved identification of asparagine residues known to be glycosylated (N29, N97, N98) [58]. Asparagines at position 29, 97 and 98 (in the mature protein numbering) are surface exposed and can thus be glycosylated (Figs. 1 and 4). Taken together, the above data indicate that our model structure is correct and can thus be used for the following structural analysis.

Description of the structure

The TM lectin domain displays the essential features of the C-type lectin modules. It folds into a globular domain that consists of two helices and six β-strands forming two antiparallel β-sheets. The short helix 3 of LIT is not present in the TM model (Fig. 1) while a possible extra antiparallel β-sheet is noted in TM involving residues 72-73 and 91-92 (Fig. 4). The first 13 residues of LIT are disordered in the crystal. This region of LIT could thus be flexible and may protrude into the solvent. In the TM model the first N-term residues were built in an extended conformation. This region has thus to be considered with caution. MD simulation for this segment indicates that it should be very flexible. Also, it is important to note that the N-term and C-term regions are close in space (Figs. 4 and 5) and that the conformation of these two areas may be depending on each other. The last C-term residues of TM have been built in an extended conformation. The conformation of these C-term residues (155 to 159) can not be predicted with high accuracy, but for the time being, our model clearly demonstrates that C157 can not be involved in a S-S bond with the cysteine residues present in the lectin domain region. The hydrophobic segment of TM could thus form a compact structure allowing for the formation of a disulfide bond between C157 and C206. No clear 3D templates were identified for the hydrophobic region of TM.

Surface features of TM

The structural comparison between LIT, MBP and ESL has already been reported and will not be described here [36]. Insertions of interest in the TM model as compared to LIT are shown in figure 5. The key insertions in TM are thus in the regions involved in calcium binding or areas that interact with carbohydrates (see below). MD simulation of the long TM insertion (residues 75 to 87) suggests the loop to be flexible. It could either protrude outside the molecular surface or have tight contact with the molecule as presented here. Surface exposed loops are most often involved in protein-protein or protein-membrane interaction. Thus the region surrounding this long insertion loop in TM and the loop itself probably have functional roles (see below).

C-type lectin domains usually bind two calcium ions and this is crucial for the function of the modules as sugar binds at the level of these calcium binding pockets (Figs. 5 and 6). Thus in most cases sugars are bound to the lectin domains via calcium coordination. ESL binds only one calcium ion via ESL E80, N82, N105, D106 [35]. MBP binds two calcium ions and the calcium ligands that are conserved be-

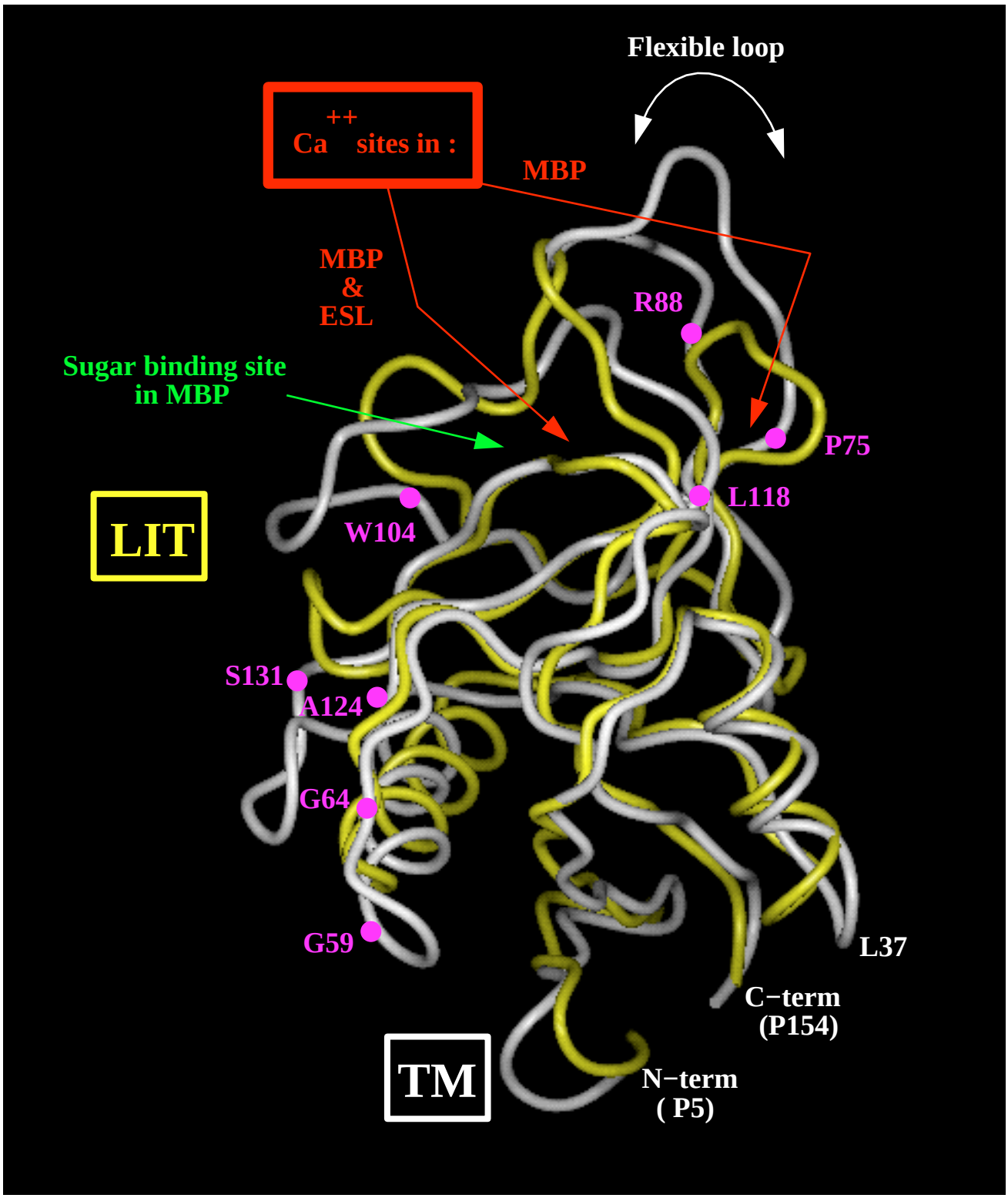


Figure 5 *Structural comparison: TM versus LIT. The TM model (white) is superimposed onto the LIT X-ray structure (yellow). Some key insertion loops in TM are mentioned. The calcium binding sites in MBP and ESL are shown. The carbohydrate binding site in MBP is also highlighted. The residues there are not conserved in LIT and TM but it is possible that these regions have a functional role in both proteins*

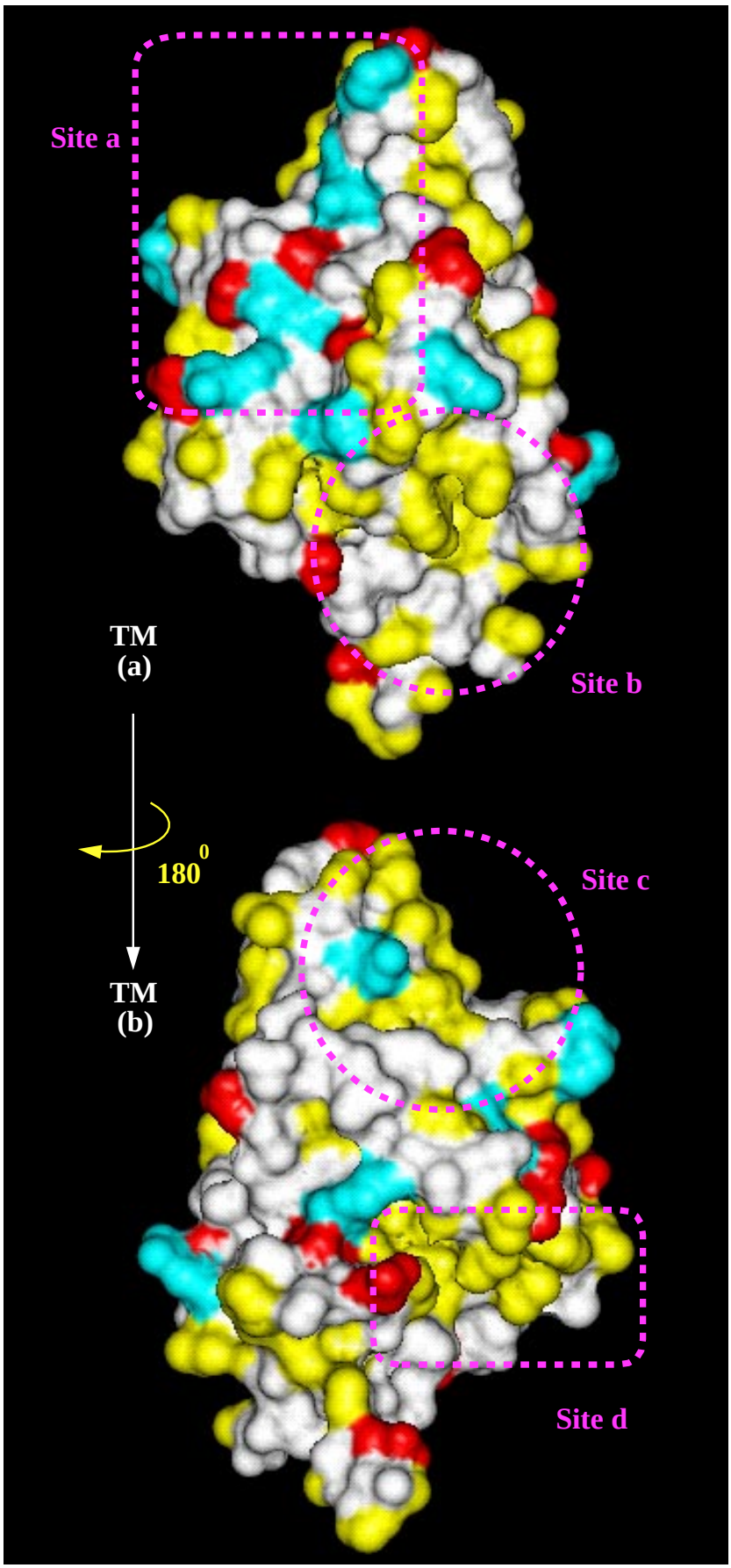


Figure 6 *Solvent accessible solid surface for the TM model. The backbone atoms and glycine residues, the side chains of polar residues (N, Q, T, S) are colored in white, side chains of positively charged amino-acids (K,R) are in blue, negatively charged (E, D) are in red and hydrophobic/aromatic residues (A,V,I,L,M, F,Y,W,P,C) in yellow. The top figure [TM(a)] is in an orientation similar to the one presented in figure 4. In the bottom figure [TM(b)], the molecule has been rotated 180+ according to the arrows. In this figure the model runs from residues 1 to 157. Four potential binding sites are listed (sites a, b, c, and d) and described in the text*

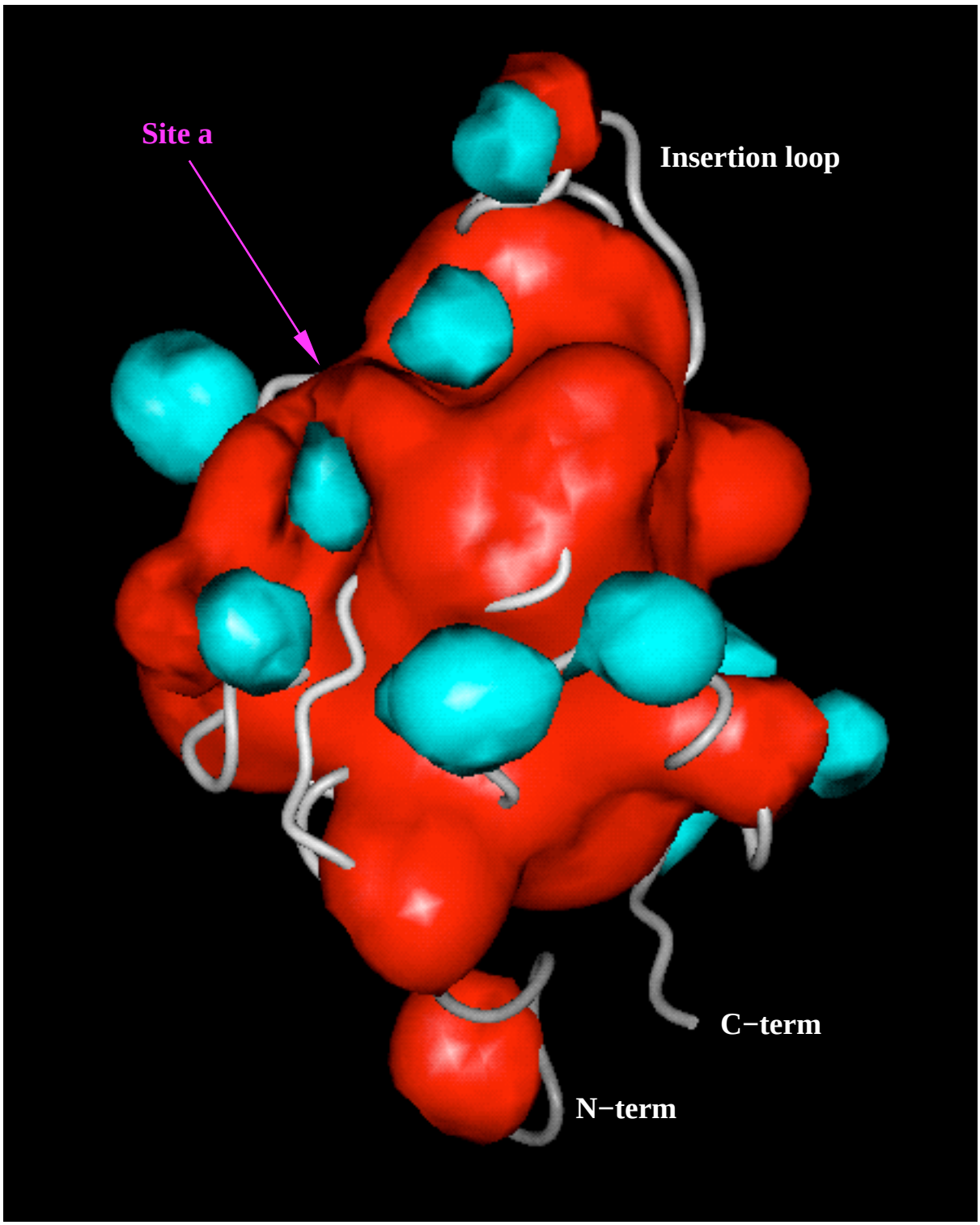


Figure 7 *Electrostatic surface potential of the TM model. The model is shown with the same orientation than the one used in figure 4. The net charge is -3.5e and residues 1 to 157 were taken into account in the calculations. The insertion loop underlines the region of residues 75-87. Site a refers to the* ***site a*** *area mentioned in figure 6 and in the text. The electrostatic isosurfaces are shown at a level of -1 (red) and +1 (blue) kcal/mol/e*

tween MBP and ESL, involves MBP E185, N187, E193, N205 and D206. This conserved site is directly involved in sugar binding. The other calcium binding site in MBP involves D161, E165, D188 and D194. These residues nor the appropriate structure are conserved in LIT in order to carry calcium-dependent oligosaccharide binding activity. This is thus consistent with the absence of carbohydrate binding to LIT [36]. Such loss of function is also noticed in other proteins having a C-type lectin fold, like endostatin [37]. These calcium ligands are not conserved in TM. Indeed, we do not notice any clear calcium binding pocket in the TM model as the negatively charged residues (and N or Q) that are close in space are partially counterbalanced by positively charged residues (Figs. 6 and 7). Yet, electrostatic calculations clearly show that an electronegative surface cover most of the central core of TM at -1 kcal/mol/e while the electropositive surface is almost missing at this energy level (Fig. 7). It is interesting to note that the neutralization of the negatively charged residues is not complete despite their overall involvement in salt bridges. This negative electric field is essentially due to D51, E126, E132, E136, E137, E141 and D145 (Fig. 7). However, for the time being, we suggest that this N-term domain in TM does not bind calcium but experimental work is needed to clarify further this point. Type II antifreeze proteins inhibits the growth of seed ice crystals in the blood of certain fishes [56]. They contain a C-type lectin domain homologous to MBP and have been studied by comparative model building. In this modeling study [56], the authors have suggested that the ice-binding site should be in the area involved in the calcium binding region that is also involved in sugars binding in MBP. Thus it is possible that the area around the sugar binding site has been selected in this family of modules during evolution as a key functional site, whether the domain binds sugars or not. This would avoid Nature the need of creating distinct and separate binding site. Possibly, this area is also important in TM (see below).

The highly polarized surfaces seen in LIT are not present in TM (Figs. 6 and 7). Thus the negatively charged LIT stretches expected to bind with the calcium carbonate crystal is only partially conserved in TM. Yet the central core of TM is mainly covered by an electronegative surface whose function remains to be characterized (Fig. 7). Endostatin inhibits angiogenesis and a key positive surface of the protein as been suggested to play a role in this process. This electropositive surface could bind heparan sulfate proteoglycans involved in growth factor signalling [37]. Although the structure of endostatin is not available yet, comparison of its molecular surface as presented in ref. [37] with the surface of TM as shown in figures 6 and 7 does not indicate that this segment of TM could have specific affinity with negatively charged compounds like heparin or related molecules. As many angiogenic inhibitors are fragments derived from larger proteins which by themselves are not regulating angiogenesis, it would be interesting to know if the lectin domain of TM is involved in such regulatory mechanism.

Surface exposed hydrophobic/aromatic or charged (polar) amino acid clusters are often involved in inter-molecular interactions. Residues potentially involved in protein-protein interactions tend also to be located in grooves. Thus we have color coded TM residues onto the solvent accessible surface with the goal of identifying potential binding sites (Fig. 6). Some exposed patches of potential functional interest are listed below. Such regions are thus good candidate for site directed mutagenesis experiments. Four main clusters are noted at the surface of TM. **Site a** (polar and charged) is in the area of the sugar binding site of MBP. It contains residues R66, K82, R83, D108, N110, E126, E136, E137, Q138, Q139 and E141 (see also Fig. 7). **Site b** (hydrophobic/aromatic) is close to the C-term region and could have contact with the following TM hydrophobic domain. It involves residues A19, P22, P24, A25, Y21, P22, I33, L37, A144, F147 and F151. **Site c** is essentially hydrophobic/aromatic and is located close to the long and most likely flexible insertion loop noted above. It contains L84, P86, L87, F90, W104, L109, A112, P113 and L114. **Site d** is essentially hydrophobic and involves residues V13, V48, A49, V52, L55, V129 and P130.

Conclusion

We provide an accurate 3D model for the lectin domain of human TM that will be valuable in assisting the design of experiments. The model provides a working hypothesis for the location of potential binding sites and proposes a disulfide bonding pattern which can be tested by site directed mutagenesis. For the time being we suggest that this region of TM does not bind calcium. This part of TM could thus still interact with sugar molecules like the cell surface receptor CD44 or TM looses its "lectin" property like lithostathine. Overall, our data are of importance to probe further the role of TM either as an anticoagulant protein, as a receptor of importance during early development or as an inhibitor of angiogenesis.

Acknowledgements This work was supported by grants from the Swedish Medical Research Council (projects n. 07143 and 11793) and research funds from the University Hospital, Malmö; the Louis Jeantet Fondation de Médecine and the Swedish National Network for Cardiovascular Research (NNCR).

References

1. Koeleman, B. P.; Reitsma, P. H.; Bertina, R. M. *Semin. Hematol.* **1997**, *34*, 256.
2. Dahlback, B. *Semin. Hematol.* **1997**, *34*, 217.
3. Aiach, M.; Borgel, D.; Gaussem, P.; Emmerich, J.; Alhenc Gelas, M.; Gandrille, S. *Semin. Hematol.* **1997**, *34*, 205.
4. Sadler, J. E. *Thromb. Haemost.* **1997**, *78*, 392.
5. Dahlback, B. *Thromb. Haemost.* **1991**, *66*, 49.
6. Di Cera, E.; Dang, Q. D.; Ayala, Y. M. *CMLS Cellular and Molecular Life Sciences* **1997**, *53*, 701.
7. Bird, P. *Hematol. Rev.* **1997**, *9*, 251.

8. Cines, D.B.; Pollak, E.S.; Buck, C.A.; Loscalzo, J.; Zimmerman, G.A.; McEver, R.P.; Pober, J.S.; Wick, T.M.; Konkle, B.A.; Schwartz, B.S.; Barnathan, E.S.; McCrae, K.R.; Hug, B.A.; Schmidt, A-M.; Stern, D.A. *Blood* **1998**, *91*, 3527.
9. Petersen, T. E. *FEBS Lett.* **1988**, *231*, 51.
10. Jackman, R. W.; Beeler, D. L.; VanDeWater, L.; Rosenberg, R. D. *Proc. Natl. Acad. Sci. USA.* **1986**, *83*, 8834.
11. Wen, D. Z.; Dittman, W. A.; Ye, R. D.; Deaven, L. L.; Majerus, P. W.; Sadler, J. E. *Biochemistry* **1987**, *26*, 4350.
12. Suzuki, K.; Kusumoto, H.; Deyashiki, Y.; Nishioka, J.; Maruyama, I.; Zushi, M.; Kawahara, S.; Honda, G.; Yamamoto, S.; Horiguchi, S. *EMBO. J.* **1987**, *6*, 1891.
13. Hayashi, T.; Zushi, M.; Yamamoto, S.; Suzuki, K. *J. Biol. Chem.* **1990**, *265*, 20156.
14. Ye, J.; Liu, L. W.; Esmon, C. T.; Johnson, A. E. *J. Biol. Chem.* **1992**, *267*, 11023.
15. Tsiang, M.; Lentz, S. R.; Sadler, J. E. *J. Biol. Chem.* **1992**, *267*, 6164.
16. Kokame, K.; Zheng, X.; Sadler, J. E. *J. Biol. Chem.* **1998**, *273*, 12135.
17. Healy, A. M.; Rayburn, H. B.; Rosenberg, R. D.; Weiler, H. *Proc. Natl. Acad. Sci. USA.* **1995**, *92*, 850.
18. Weiler-Guettler, P. D. C.; Beeler, D.L.; Healy, A.M.; Hancook, W.W.; Rayburn, H.; Edelberg, J.M.; Robert, D.; Rosenberg, A. *J. Clin. Invest.* **1998**, *101*, 1983.
19. Grinnell, B. W.; Berg, D. T. *Am. J. Physiol.* **1996**, *270*, H603.
20. Lafay, R. L.; Le Bonniec B.F.; Lasne, D.; Aiach, M.; Rendu, F. *Thromb. Haemost.* **1998**, *79*, 848.
21. Maruyama, I.; Majerus, P.W. *Blood* **1987**, *69*, 1481.
22. Conway, E.M.; Pollefeyt, S.; Collen, D.; Steiner-Mosonyi, M. *Blood* **1997**, *89*, 652.
23. Chu, M.; Bird, C.H.; Teasdale, M.; Bird, P.I. *Thromb. Haemost.* **1998**, *80*, 119.
24. Zhang, Y.; Weiler-Guettler, H.; Chen, J.; Wilhelm, O.; Deng, Y.; Qiu, F.; Nakagawa, K.; Klevesath, M.; Wilhelm, S.; Bohrer, H.; Nakagawa, M.; Graeff, H.; Martin, E.; Stern, D.M.; Rosenberg, R.D.; Ziegler, R.; Nawroth, P.P. *J. Clin. Invest.* **1998**, *101*, 1301.
25. Ishii, H.; Majerus, P.W. *J. Clin. Invest.* **1985**, *76*, 2178.
26. Tanaka, A.; Ishii, H.; Hiraishi, S.; Kazama, M.; Maezawa, H. *Clin. Chem.* **1991**, *37*, 269.
27. Kazama, M. *Acta Hematol.* **1988**, *51*, 1387.
28. Takano, S.; Kimura, S.; Ohdama, S.; Aoki, N. *Blood* **1990**, *76*, 2024.
29. Bajzar, L.; Nesheim, M.; Morser, J.; Tracy, P.B. *J. Biol. Chem.* **1998**, *273*, 2792.
30. Patthy, L. *J. Mol. Biol.* **1988**, *202*, 689.
31. Bairoch, A.; Bucher, P.; Hofmann, K. *Nucleic Acids Res.* **1997**, *25*, 217.
32. Drickamer, K. *Curr. Opin. Struct. Biol.* **1993**, *3*, 393.
33. Rini, J.M. *Curr. Opin. Struct. Biol.* **1995**, *5*, 617.
34. Weis, W.I.; Drickamer, K.; Hendrickson, W.A. *Nature* **1992**, *360*, 127.
35. Graves, B.J.; Crowther, R.L.; Chandran, C.; Rumberger, J.M.; Li, S.; Huang, K.S.; Presky, D.H.; Familetti, P.C.; Wolitzky, B.A.; Burns, D.K. *Nature* **1994**, *367*, 532.
36. Bertrand, J.A.; Pignol, D.; Bernard, J-P.; Verdier, J-M.; Dagorn, J-C.; Fontecilla-Camps, J.C. *EMBO. J.* **1996**, *15*, 2678.
37. Hohenester, E.; Sasaki, T.; Olsen, B.R.; Timpl, R. *EMBO. J.* **1998**, *17*, 1656.
38. Kohda, D.; Morton, C.J.; Parkar, A.A.; Hatanaka, H.; Inagaki, F.M.; Campbell, I.D.; Day, A.J. *Cell* **1996**, *86*, 767.
39. Fisher, D.; Eisenberg, D. *Protein Sci.* **1996**, *5*, 947.
40. Bernstein, F.C.; Koetzle, T.F.; Williams, G.J.B.; Meyer, E.F.Jr.; Brice, M.D.; Rodgers, J.R.; Kennard, O.; Shimanouchi, T.; Tasmui, M. *J. Mol. Biol.* **1977**, *112*, 535.
41. Ponder, J.W.; Richards, F.M. *J. Mol. Biol.* **1987**, *193*, 775.
42. Jones, T.; Thirup, S. *EMBO J.* **1986**, *5*, 819.
43. Hobohm, U.; Sander, C. *Protein Sci.*, **1994**, *3*, 522.
44. Shenkin, P.S.; Yarmush, D.L.; Fine, R.M.; Wang, H.; Levinthal, C. *Biopolymers*, **1987**, *26*, 2053.
45. Mackay, D. H. L., Cross, A. J., Hagler, A. T. In: *Prediction of protein structure and the principles of protein conformation*; Fasman, G.D., Ed.; Plenum Publishing Corporation: New York and London, 1989, pp 317-358.
46. Sharp, K., Honig, B. *Annu Rev Biophys Chem.* **1990**, *19*, 301.
47. Lee, B.; Richards, F.M. *J. Mol. Biol.* **1971**, *55*, 379.
48. MacArthur, M.W.; Thornton, J.M. *Proteins* **1993**, *17*, 232.
49. Lüthy, R.; Bowie, J.U.; Eisenberg, D. *Nature* **1992**, *356*, 83.
50. Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.
51. Martin, A.C.; MacArthur, M.W.; Thornton, J.M. *Proteins* **1997**, *Suppl 1*, 14.
52. Bajorath, J.; Linsley, P.S.; Metzler, W.J. *J. Mol. Model.* **1997**, *3*, 287.
53. Bucher, P.; Karplus, K.; Moeri, N.; Hofmann, K. *Comput. Chem.* **1996**, *20*, 3.
54. Russel, B.R.; Saqi, M.A.S.; Bates, P.A.; Sayle, R.A.; Sternberg, M.J.E. *Prot. Engng.* **1998**, *11*, 1.
55. Bajorath, J.; Aruffo, A. *Protein Sci.* **1996**, *5*, 240.
56. Sönnichsen, F.D.; Sykes, B.D.; Davies, P.L. *Protein Sci.* **1995**, *4*, 460.
57. Mills, A. *FEBS Lett.* **1994**, *319*, 5.
58. Parkinson, J.F.; Vlahos, C.J.; Yan, S.C.B.; Bang, N.U. *Biochem. J.* **1992**, *283*, 151.