Research Article

Molecular modelling study of the 3D structure of the biglycan core protein, using homology modelling techniques

Dimitrios Vlachakis[1], Spyridon Champeris Tsaniras[2], Christos Feidakis[1] and Sophia Kossida[1]

[1]Biomedical Research Foundation of the Academy of Athens, 11527, Athens, Greece
[2]Department of Physiology, Medical School, University of Patras, Rio, Patras, Greece

Received on February 28, 2013; Accepted on May 7, 2013; Published on June 22, 2013

Correspondence should be addressed to Sophia Kossida; Phone: +30 210 6597199, Fax: +30 210 6597545, Email: skossida@bioacademy.gr

Abstract

Herein we report the establishment of the 3D structure of the biglycan core protein, using conventional homology molecular modelling techniques. The 3D model has been structurally optimised via molecular dynamics. It was found that the final model of biglycan resembles in structure its template protein bearing a set of distinct parallel β-sheet structure patterns. The biglycan model bears a very hydrophobic amino acid region towards its inner cavity that acquires an arc-like structure. The external domain of the biglycan model is made up of hydrophilic residues that are exposed to the water solvent. It is those hydrophilic residues that are responsible for their interaction with polysaccharide polymers. Overall comparison of the model of biglycan to the recently determined x-ray structure of the same protein returns a very low Root Mean Square Deviation (RMSD), which confirms the viability of the model and its reliability as a platform for the study biglycan interactions.

Introduction

The Extracellular Matrix (ECM) plays a key role in the maintenance and function of connective tissue and consists predominantly of collagen, proteoglycans, glycosaminoglycans and other minor proteins (Hocking *et al.* 1998). Proteoglycans are macromolecules composed of a protein core and a carbohydrate glycosaminoglycan (GAG) side chain. GAGs are highly hydrophilic, negatively charged molecules located at specific sites around each collagen fibril (Scott & Haigh 1988). The core protein of a proteoglycan seems to be the most active part of the molecule, since it interacts with the collagen fibrils and the extracellular matrix at specific sites. According to Cintron (1989), the precise arrangement of PGs presumably reflects specific intermolecular interactions with collagens. Therefore, loss of PGs might cause alterations to the collagen organisation and hence to connective tissue stability.

Dermatan sulphate proteoglycans are small leucine rich proteins (SLRPs) with an "arc" shaped spatial conformation and tend to form dimers. The amino-acid sequence of these proteins is characterised by long arrays of leucine-rich repeat motifs of about 24 amino acids in length. A feature that is of great interest about SLRPs is that they contain an amphipathic consensus sequence, with leucine as the predominant hydrophobic residue placed in conserved positions (Hocking *et al.* 1998). This pattern is highly likely to be involved with "protein-protein" or "protein lipid" interactions (Krantz *et al.* 1991). With respect to collagen/proteoglycan interactions several theories have been proposed to date, with the majority of them supporting the idea that interactions with collagen are being established at the inner site of the "horseshoe" shaped molecule (Weber *et al.* 1996, Vesentini *et al.* 2005). Bearing in mind that the concave surface is involved in a high-affinity dimer interaction, this theory is now being debated (Scott et al. 2006a, Vesentini *et al.* 2006). However, it should be taken into account that there might be a dimer-to-monomer transition of proteoglycans in collagen/proteoglycan interactions. For example, dermatan sulphate proteoglycans bind collagen as a dimer and conservation analysis across class I SLRPs revealed a clustering of partially conserved residues on the sugar-free surface of leucine rich repeats (LRRs) IV-VI, a region that has been implicated in collagen binding. This theory though has yet to be confirmed by further biochemical studies

(Scott *et al.* 2004).

Biglycan is a dermatan sulphate proteoglycan and it receives its name because it is often substituted with two glycosaminoglycan chains at the N-terminus (Kresse *et al.* 1997, Naito 2005). It is expressed in a variety of epithelial tissues throughout the body, such as the lung, spleen, bone, liver, cartilage, tendon, skin, kidney, heart, sclera and cornea (Wegrowski *et al.* 1995, Hocking *et al.* 1998, Watson & Young 2004). Biglycan interacts with fibrillar collagen and based on immunohistochemical analysis, it has been shownn that the core protein of the proteoglycan interacts directly with the collagen fibril (Schonherr *et al.* 1995).

In this study the 3D structure of the core protein of biglycan was determined by means of homology based molecular modelling. Descent from a common ancestor, i.e. homology, can be hypothesised when similar properties are detected in biological objects (Grishin 2001). It is also widely accepted that statistically significant homology detected from the sequence alone reflects descent from a common ancestor (Aravind & Koonin 1999). Comparative homology-based modelling is based on the observation that sequence correlation above a certain proportion implies structural similarity. Thus, a protein of a known structure and sequence can be used as a template for the construction of a 3D model of another protein with a sequence similar to the template (D'Alfonso *et al.* 2001). Additionally, it should be taken into consideration that structure is conserved to a much greater extent than sequence and that there is a limited number of backbone motifs (Chothia 1992, Banerjee-Basu & Baxevanis 2002). A relative minimum percentage of homology between two protein sequences that can ensure the construction of a reliable model is 40% (MOE 2005). Template and model homology sequences identity that are greater than 50% normally have 90% or more of the individual structures within the common cores and an RMSD value smaller than 1Å, ensuring reliability of the proposed model (Chothia & Lesk 1986).

Materials and methods

Molecular Modelling

All calculations were performed on a dual Pentium 4 workstation running Linux and using the Molecular Operating Environment (MOE) 2005.03 software package developed by Chemical Computing Group (Montreal, Canada).

Sequence alignment

The amino acid sequence of biglycan was obtained from the protein database from NCBI (http://www.ncbi.nlm.nih.gov/protein) (gi|30315664). The PSI-BLAST algorithm was used to identify homologous structures for biglycan by searching the structural

```
Rat      ---MRPLWLLTLLLALSQALPFEQKGFWDFTLDDGLLMMNDEEASGSDTTSGVPDLDSLT
Mouse    ---MCPLWLLTLLLALSQALPFEQKGFWDFTLDDGLLMMNDEEASGSDTTSGVPDLDSVT
Bovine   ---MWPLWPLAALLALSQALPFEQKAFWDFTLDDGLPMLNDEEASGAETTSGIPDLDSLP
Canine   ---MWPLWLVASLLALSQALPFEQKGFWDFTLDDGLPMLNDEEASGAETTSGVPDLDALT
Equine   MPTMWPLWLLASLLALSQALPFEQKGFWDFTLDDGLPMLNDEEASGADTTSGIPDLDSLT
Human    ---MWPLWRLVSLLALSQALPFEQRGFWDFTLDDGPFMMNDEEASGADT-SGVLDPDSVT
            * *** :. ************:.*********  *:********::* **: * *::.

Rat      PTFSAMCPFGCHCHLRVVQCSDLGLKTVPKEISPDTTLLDLQNNDISELRKDDFKGLQHL
Mouse    PTFSAMCPFGCHCHLRVVQCSDLGLKTVPKEISPDTTLLDLQNNDISELRKDDFKGLQHL
Bovine   PTYSAMCPFGCHCHLRVVQCSDLGLKAVPKEISPDTTLLDLQNNDISELRKDDFKGLQHL
Canine   PTYSAMCPFGCHCHLRVVQCSDLGLKAVPKEISPDTMLLDLQNNDISELRADDFKGLHHL
Equine   PTFSAMCPFGCHCHLRVVQCSDLGLKAVPKEISPDTTLLDLQNNEISELRKDDFKGLQHL
Human    PTYSAMCPFGCHCHLRVVQCSDLGLKSVPKEISPDTTLLDLQNNDISELRKDDFKGLQHL
         **:**********************:********* *******:***** ******:**

Rat      YALVLVNNKISKIHEKAFSPLRKLQKLYISKNHLVEIPPNLPSSLVELRIHDNRIRKVPK
Mouse    YALVLVNNKISKIHEKAFSPLRKLQKLYISKNHLVEIPPNLPSSLVELRIHDNRIRKVPK
Bovine   YALVLVNNKISKIHEKAFSPLRKLQKLYISKNHLCEIPPNLPSSLVELRIHDNRIRKVPK
Canine   YALVLVNNKISKIHEKAFSPLRKLQKLYISKNHLVEIPPNLPSSLVELRIHDNRIRKVPK
Equine   YALVLVNNKISKIHEKAFSPLRKLQKLYISKNHLVEIPPNLPSSLVELRIHDNRIRKVPK
Human    YALVLVNNKISKIHEKAFSPLRKLQKLYISKNHLVEIPPNLPSSLVELRIHDNRIRKVPK
         ********************************** ************************

Rat      GVFSGLRNMNCIEMGGNPLENSGFEPGAFDGLKLNYLRISEAKLTGIPKDLPETLNELHL
Mouse    GVFSGLRNMNCIEMGGNPLENSGFEPGAFDGLKLNYLRISEAKLTGIPKDLPETLNELHL
Bovine   GVFSGLRNMNCIEMGGNPLENSGFEPGAFDGLKLNYLRISEAKLTGIPKDLPETLNELHL
Canine   GVFSGLRNMNCIEMGGNPLENSGFEPGAFDGLKLNYLRISEAKLTGIPKDLPETLNELHL
Equine   GVFSGLRNMNCIEMGGNPLENSGFQPGAFDGLKLNYLRISEAKLTGIPKDLPETLNELHL
Human    GVFSGLRNMNCIEMGGNPLENSGFEPGAFDGLKLNYLRISEAKLTGIPKDLPETLNELHL
         ************************:********************************

Rat      DHNKIQAIELEDLLRYSKLYRLGLGHNQIRMIENGSLSFLPTLRELHLDNNKLSRVPAGL
Mouse    DHNKIQAIELEDLLRYSKLYRLGLGHNQIRMIENGSLSFLPTLRELHLDNNKLSRVPAGL
Bovine   DHNKIQAIELEDLLRYSKLYRLGLGHNQIRMIENGSLSFLPTLRELHLDNNKLSRVPAGL
Canine   DHNKIQAIELEDLLRYSKLYRLGLGHNQIRMIENGSLSFLPTLRELHLDNNKLSRVPSGL
Equine   DHNKIQAIELEDLLRYSKLYRLGLGHNQIRMIENGSLSFLPTLRELHLDNNKLSRVPAGL
Human    DHNKIQAIELEDLLRYSKLYRLGLGHNQIRMIENGSLSFLPTLRELHLDNNKLARVPSGL
         ************************************************:***:**

Rat      PDLKLLQVVYLHSNNITKVGINDFCPMGFGVKRAYYNGISLFNNPVPYWEVQPATFRCVT
Mouse    PDLKLLQVVYLHSNNITKVGINDFCPMGFGVKRAYYNGISLFNNPVPYWEVQPATFRCVT
Bovine   PDLKLLQVVYLHTNNITKVGVNDFCPVGFGVKRAYYNGISLFNNPVPYWEVQPATFACVT
Canine   PDLKLLQVVYLHTNNITKVGVNDFCPVGFGVKRAYYNGISLFNNPVPYWEVQPATFRCVT
Equine   PDLKLLQVVYLHTNNITKVGVNDFCPVGFGVKRAYYNGISLFNNPVPYWEVQPATFRCVT
Human    PDLKLLQVVYLHSNNITKVGVNDFCPMGFGVKRAYYNGISLFNNPVPYWEVQPATFRCVT
         ************:********:******:***************************** ***

Rat      DRLAIQFGNYKK
Mouse    DRLAIQFGNYKK
Bovine   DRLAIQFGNYKK
Canine   DRLAIQFGNYKK
Equine   DRLAIQFGNYKK
Human    DRLAIQFGNYKK
         ************
```

A

SeqA	Name	Length(aa)	SeqB	Name	Length(aa)	Score
1	Bovine	369	2	Rat	369	95
1	Bovine	369	3	Mouse	369	94
1	Bovine	369	4	Canine	369	96
1	Bovine	369	5	Human	368	94
1	Bovine	369	6	Equine	372	97
2	Rat	369	3	Mouse	369	99
2	Rat	369	4	Canine	369	95
2	Rat	369	5	Human	368	95
2	Rat	369	6	Equine	372	96
3	Mouse	369	4	Canine	369	94
3	Mouse	369	5	Human	368	95
3	Mouse	369	6	Equine	372	96
4	Canine	369	5	Human	368	94
4	Canine	369	6	Equine	372	97
5	Human	368	6	Equine	372	95

B

Figure 1. Sequence alignments between biglycan core proteins of different species produced by the ClustalX program. (A) Homology scores from sequence alignments between the sequences of biglycan core protein from different mammal species. (B) Alignment scores were between 94-99% suggesting high homology and structural conservation of biglycan core protein throughout mammal species.

database of protein sequences in the Protein Data Bank (PDB) (Berman *et al.* 2002). The crystal structure of decorin (PDB code 1XKU) (resolution 1.5 Å) was selected as template structures for homology modelling of the biglycan protein. The selection of these structures was based on the highest sequence similarity (56.7% similarity). Sequence alignments were created with ClustalW (Thompson *et al.* 1994).

Homology modelling

The homology model was constructed using MOE version 2005.03. As indicated above, the homology model of biglycan was based on the crystal structure of 1XKU, which shares 56.7% similarity with the sequence of biglycan. The first requirement in the construction of a biglycan model structure is a multiple sequence alignment against its template. The sequence alignment is based on identifying structurally conserved regions (SCRs) common between the model and the template. An initial 3D structure of biglycan was obtained by transferring the backbone coordinates from the template residues to the corresponding residues of biglycan, except for several variable regions (LOOPs). To construct the structural variable regions, a loop-searching and generating algorithm over the databank of known crystal structures was used from within MOE. Through the procedure mentioned above, an initial 3D model was thus completed. The quality of this model was examined by WHATCHECK and ProCheck. The three-dimensional model was subjected to molecular mechanics energy minimization calculations using the updated MMFF94x and Amber algorithms implemented in MOE. The refinement of the homology model was carried out through energy minimization: 500 iterations of steepest descent (SD) calculation were performed and then the conjugated gradient (CG) calculation was carried out until achieving 0.1 kcal/mol $Å^{-1}$ of convergence on the gradient.

Molecular dynamics (MD) simulations

MD simulations were performed using MOE version 2005.03 and its built-in Molecular Dynamics module using an NVT ensemble. The protein was solvated in a cubic periodic box containing water molecules in order to perform simulations in an aqueous environment. Firstly, the model structure was refined by adding all hydrogen atoms and subsequently energy minimization was performed with the MMFF94x molecular mechanics module of MOE to ensure overall neutrality of the simulated system. The time step was 2 fs and all simulations were conducted at 300 K.

The model was first equilibrated for 100 ps keeping the whole protein fixed to allow the water molecules to relax. A subsequent 100 ps of equilibration with the protein backbone fixed was carried out. After the equilibration phase, we obtained 1 ns MD trajectory for the biglycan model. Additionally, in order to assess the structural stability of the protein, 500 pico-second snapshots of the molecular dynamics simulation trajectory were taken. This revealed 9 conformations that were superimposed together. Each snapshot conformation was coloured with a unique colour.

Finally, a conjugate gradient energy minimization of the full protein was performed until the root mean-square (rms) gradient energy was lower than 0.001 kcal/mol $Å^{-1}$. In this step, the quality of the initial model was improved. After the optimization procedure, the structure was checked again by ProCheck.

Docking

For the purposes of this study the docking suite ZDOCK (version 3.0) was used (Chen *et al.* 2003). Docking experiments were conducted on the models that had been energetically minimized and conformationaly optimized using molecular dynamics simulations. ZDOCK is a protein-protein docking suite that utilizes a grid-based representation of the molecular system involved. In order to efficiently explore the search space and docking positions of the molecules as rigid bodies, ZDOCK takes full advantage of a three-dimensional fast Fourier transformation algorithm. It

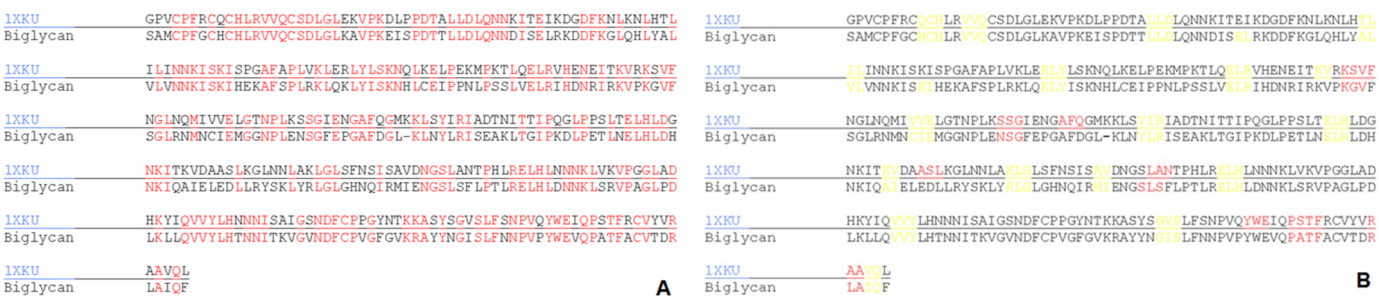

Figure 2. Sequence alignment between decorin (1XKU) and biglycan produced by the ClustalX program. Homology score between the two proteins was 56.7%. (A) Secondary structure features prediction. Red: α-helical secondary elements, Black: β-sheet pattern (B).

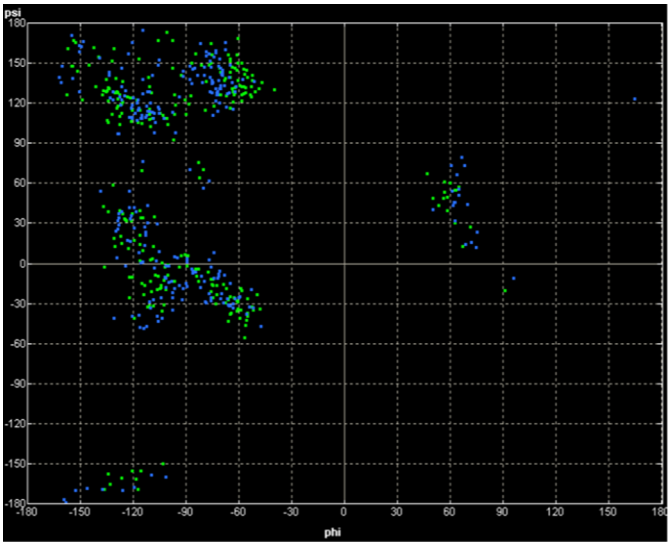

Figure 3. Ramachandran Plot. In both the template and the model, there are no residues in the disallowed regions of the ramachandran plot.

uses a scoring function that returns electrostatic, hydrophobic and desolvation energies as well as performing a fast pairwise shape complementarity evaluation. Moreover it uses the contact propensities of transient complexes of proteins to perform an evaluation of a pairwise atomic statistical potential for the docking molecular system. RDOCK was utilized to refine and quickly evaluate the results obtained by ZDOCK (Li *et al.* 2003). RDOCK performs a fast minimization step to the ZDOCK molecular complex outputs and re-ranks them according to their re-calculated binding free energies.

Molecular electrostatic potential (MEP)

Electrostatic potential surfaces were calculated by solving the nonlinear Poisson-Boltzmann equation using finite difference method as implemented in the PyMOL Software (Delano 2002). The potential was calculated on grid points per side (65, 65, 65) and the grid fill by solute parameter was set to 80%. The dielectric constants of the solvent and the solute were set to 80.0 and 2.0, respectively. An ionic exclusion radius of 2.0Å, a solvent radius of 1.4Å and a solvent ionic strength of 0.145 M were applied. Amber99 (Duan et al. 2003) charges and atomic radii were used for this calculation.

Model evaluation

Evaluation of the model quality and reliability in terms of its 3D structural conformation is critical for the viability of this study. Therefore, the produced models were initially evaluated within the Gromacs package by a residue packing quality function, which depends on the number of buried non-polar side chain groups and on hydrogen bonding. Moreover, the PROCHECK suite (Laskowski *et al.* 1996) was employed to further

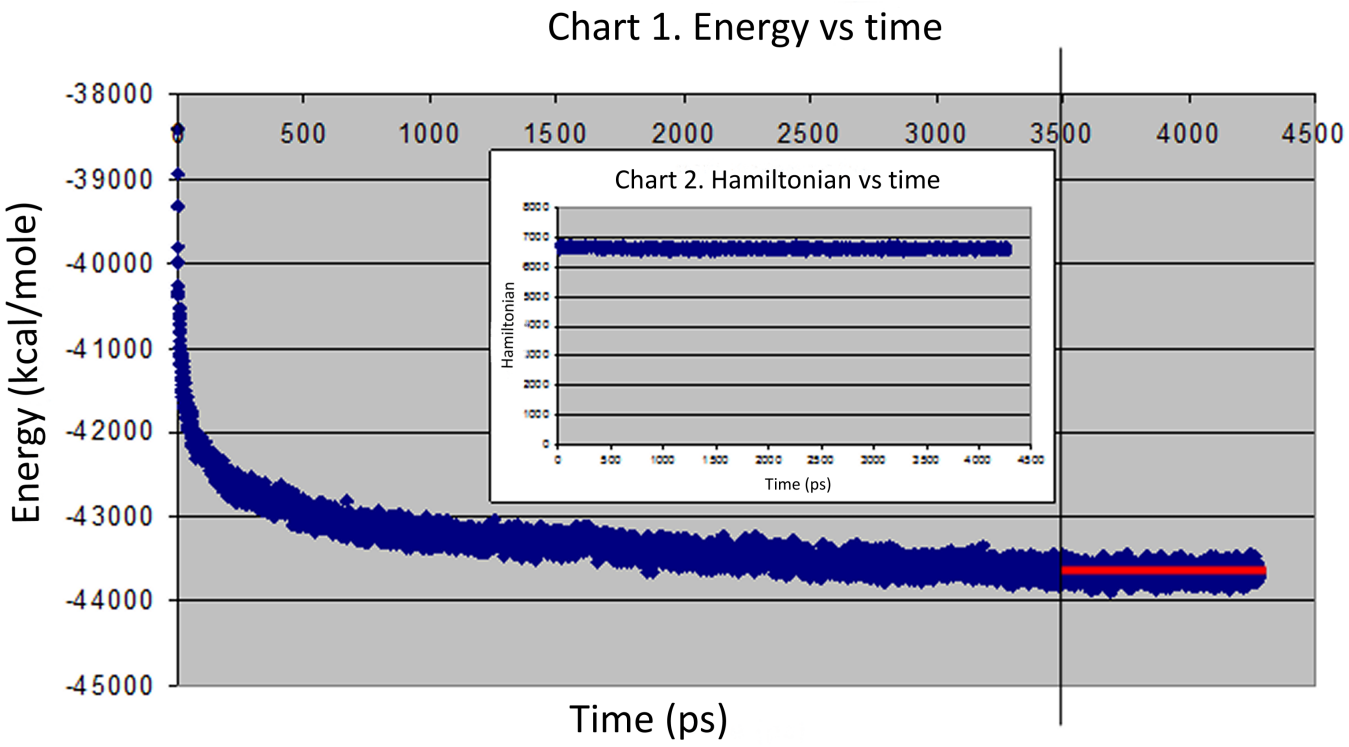

Figure 4. The model quickly reached a plateau, which is considered to be its global energy minimum (Chart 1). A constant Hamiltonian value revealed that the molecular system was stable throughout the course of the molecular dynamics simulation (Chart 2).

evaluate the quality of the produced model. Verify3D (Eisenberg *et al.* 1997) was also used to evaluate whether the model of Classical Swine Fever virus helicase is similar to known protein structures. Finally, the Molecular Operating Environment (MOE) suite was used to evaluate the 3D geometry of the models in terms of their Ramachandran plots, omega torsion profiles, phi/psi angles, planarity, C-beta torsion angles and rotamer strain energy profiles (Sellis *et al.* 2009, Vlachakis *et al.* 2012, 2013).

Results

Sequence alignment

The sequence that was modelled into a 3D molecule in this study was the bovine biglycan core protein. A comparison between the sequences of this protein from various mammalian species revealed high homology (Figure 1B). Additionally, biglycan protein sequence is the same size in many mammalian species (Figure 1). Sequence alignment between decorin and biglycan core protein sequences revealed a high homology score (i.e. 56.7%) (Figure 2A).

Secondary features and model evaluation

The β-sheet pattern has been completely conserved between the model and the template. On the other hand, the template's α-helix distribution was not precisely assigned to the model. There are 3 α-helical secondary elements missing from the model, when compared to the template (Figure 2B). The ProCheck/ WHATCHECK evaluation of the model revealed that 95% of its residues were found in the core regions and another 5% in the allowed regions of the

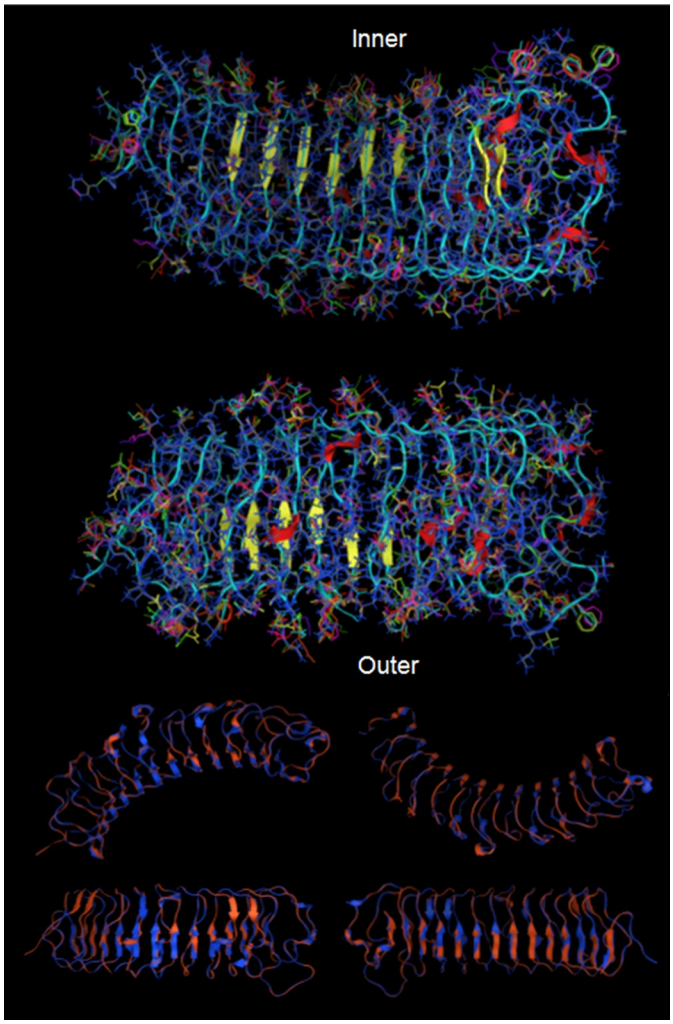

Figure 5. The structure of biglycan superimposed with its template. Reasonably high homology identity guaranteed the retention of the major secondary elements and shape. The above structure has been obtained after a 2 nanosecond molecular dynamics simulation.

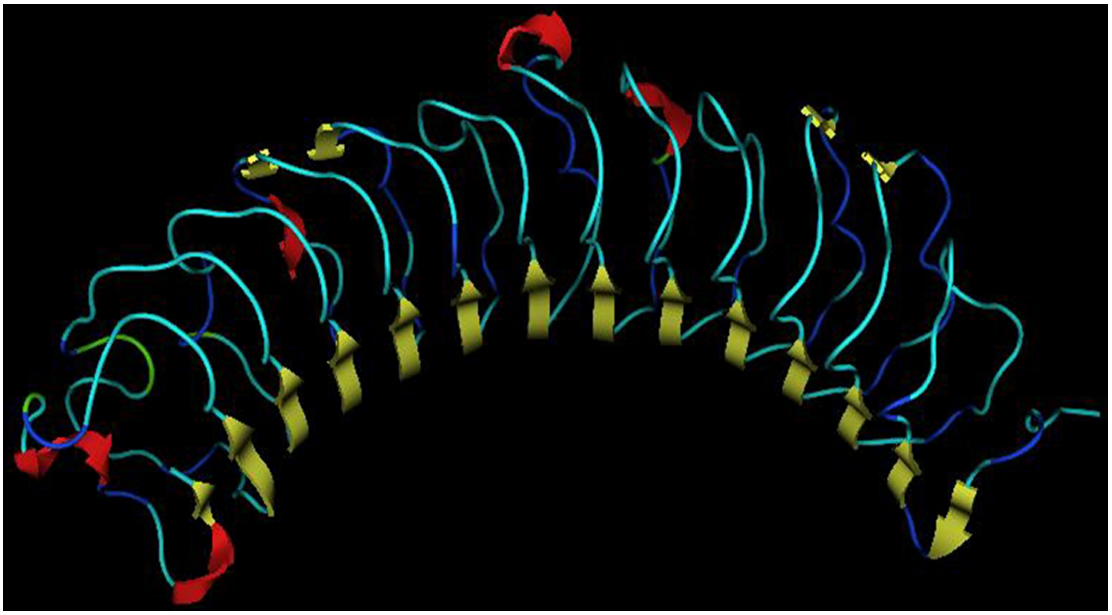

Figure 6. Proposed homology model for the structure of the biglycan core protein.

Ramachandran plot (Figure 3).

The amino acid sequences of the first 20 amino acids at the N-terminus of biglycan and decorin are different. In biglycan, two GAG chains are attached to Ser5 and Ser11 at the N-terminus whereas in decorin only one chain is attached to Ser4 (Scott *et al.* 2004).

Molecular dynamics

The Molecular Dynamics functions that were used in the present simulations solve the equations of motion for a molecular system and store the resulting trajectory information to a database (Balatsos *et al* 2012, Palaiomylitou *et al* 2008, Sellis *et al* 2012, Vangelatos *et al* 2009). These functions depend on the current state of the system, in particular, the force field, the

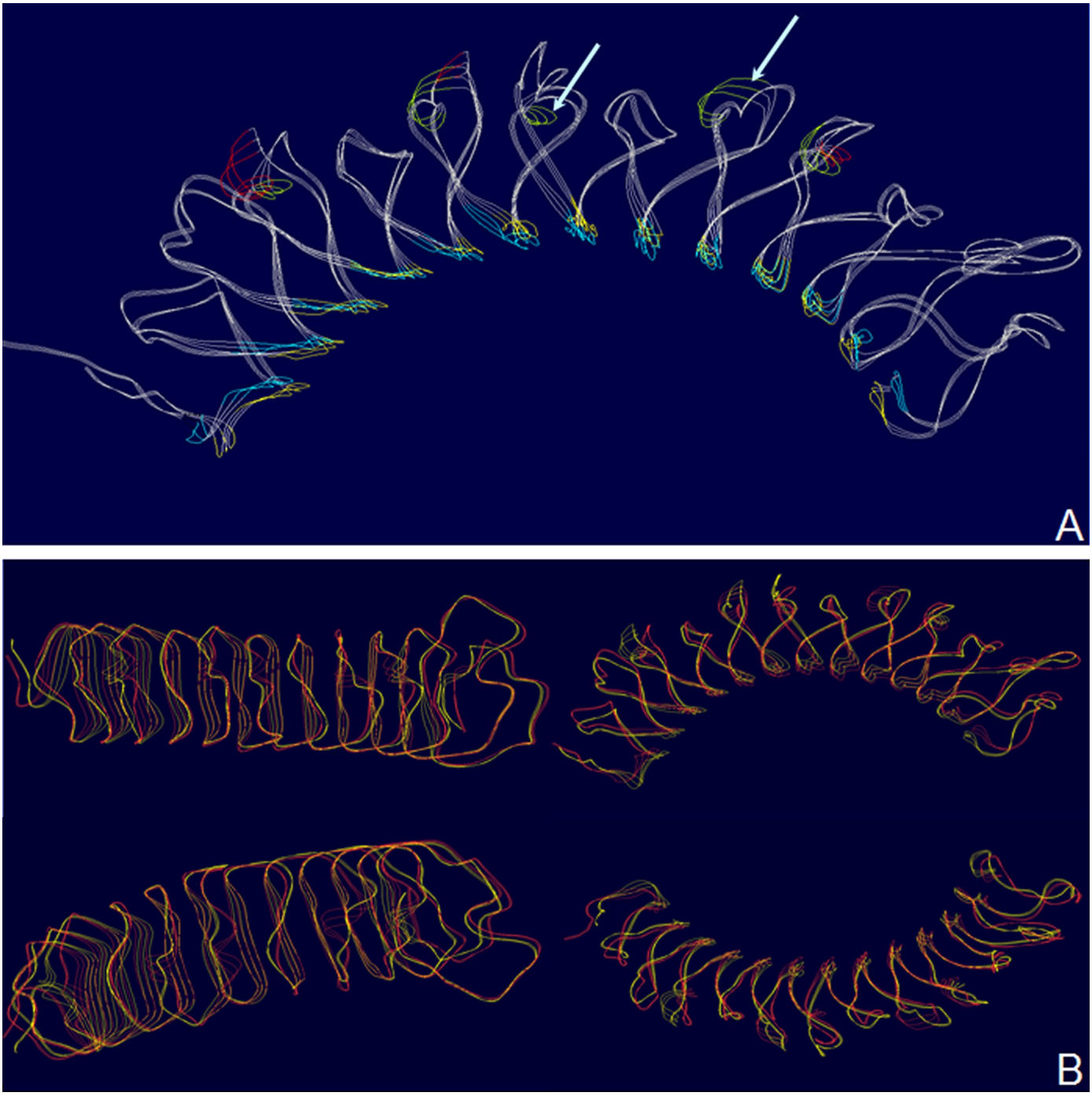

Figure 7: The structure of the homology model was very similar to the crystal structure. Superimposition of the two structures revealed the existence of two loops at the surface of the molecule (indicated by turquoise arrows). Key: Green and aqua → α-helix and β-sheet, respectively, for the homology model; Red and yellow → α-helix and β-sheet, respectively, for the crystal structure (A). The homology model of biglycan superimposed with the crystal structure of the same protein. Key: Yellow → X-ray structure, Red → homology model (B).

potential setup and the current restraint configuration (Vlachakis 2009).

The molecular dynamics simulation serves two purposes in the current study. Firstly, after 4 ns the protein is completely relaxed and close to its global energetic minimum (Figure 4, Chart 1). Secondly, the extended molecular dynamics simulation does impose a serious structural test to the folding stability of the model. It was proven that the energy of the system was reduced while its structure was conserved during the MD simulation, as shown by the energy and the Hamiltonian graphs (Figure 4, Chart 2). The Hamiltonian is the value of the full (extended) system at time *t*. In a properly functioning simulation, this quantity is conserved at all times (although fluctuations will be present). A drift in *H* is evidence of too large a time step.

The evaluation for the structural stability of the biglycan model showed that the inner cliff of the protein, where the parallel β-sheet motif is found, is much more stable than the rest of the protein. Visual investigation of the 9 conformations revealed that the inner region of the protein (β-sheet motif) is not as colourful as the rest of the protein (data not shown). This shows that the β-sheet motif of the biglycan protein is a very stable and conserved motif, whose residues are energetically and conformationally more stable than the rest residues of the protein.

Model evaluation

Superimposition of the biglycan proposed model to the decorin crystal structure revealed obvious similarities in the spatial arrangement as well as the structural motives of both proteins (Figure 5). Also, the proposed model for biglycan appears to obtain an "arc" shape spatial arrangement, which is consistent with the rest of the RSLPs (Figure 6). The inner cliff of the protein consists of parallel β-sheets made mainly of hydrophobic amino acids.

Comparison of theoretical model to its crystal structure

The crystal structure of biglycan has recently been released (Scott *et al.* 2006b) and a comparison of the homology model with the x-ray structure was performed. The secondary elements of the structure (i.e. α -helices and β-sheets) were found to be highly similar. The main differences between the model and the x-ray data lie in two extra α-helical loops the homology model has on its outer surface between residues 206-207 and 248-250, respectively (Figure 7A). The areas of interest, such as the inner cliff of the protein and N- and C- terminal regions were the same between the two molecules. Superimposition of the two models revealed an RMSD of 1.32Å (Figure 7B).

Discussion

The sequence of biglycan is highly conserved among higher mammalian species, a common characteristic of SLRPs. In addition to high homology scores (Figure 1B) biglycan sequences from different mammalian species have the same size (Figure 1). This observation suggests that the biglycan structure and hence function is very similar in the epithelial tissues of various mammals. Therefore, biglycan serves the same function to different species. Additionally, biglycan and decorin share a high homology between their sequences. These proteins belong to the same family of proteoglycans, type I Dermatan sulphate. Although their function differs structurally they share structural similarities. In previous studies it has been observed that decorin is involved with collagen fibrilogenesis, whereas biglycan is not. Additionally, decorin is released by keratocytes, whereas biglycan is the result of myofibroblast activity in epithelial tissues (Funderburgh *et al.* 2001, 2003).

It was also observed that the inner cavity of the "arc" shaped molecule consists mainly of hydrophobic amino acids, whereas the outer surface of the molecule consists of hydrophilic amino acids. This explains the arc shape of the molecule, as an adaptive response to "hide" these amino acids in aqueous biological environments. Additionally, parallel β-sheets are known to have a more tighten spatial arrangement than antiparallel β-sheets (Zubay 1998). This shape obtained for biglycan is consistent with the rest of the known SLRPs.

Findings confirm that a possible interaction site for the collagen fibrils is the inner cavity of the arc -shaped molecule. This is supported by three different arguments. Firstly, by structurally evaluating the model of biglycan, it appears that it has the correct dimensions to accommodate a collagen fibril. Secondly, the presence of hydrophobic residues implies they must be covered by a second molecule to avoid exposure to the aqueous phase. Thirdly, the inner cavity of the biglycan model is a beta-sheet motif, which is highly conserved among other collagen interacting proteins of the same families. That is also shown in the alignment between the template and the model.

Preliminary data investigation of the relevant literature combined with the findings of this study, pose important questions. How does biglycan interact with collagen fibrils? The proposed patterns of interaction can be investigated in a future protein-protein docking / interaction *in silico* investigation. Biglycan and collagen can be imported to the MOE docking algorithm and after the final interaction conformations are obtained, the stability and biological viability of

the system can be verified by means of molecular dynamics (MD). A separate MD simulation will be performed for each viable docking conformation and the various energies will be compared, until a mode of association and interaction can be suggested. However, technology means currently available are not powerful enough to simulate a whole collagen fibril.

Previous studies suggest a collagen binding domain in both the C- and N- terminal regions of the core protein of decorin. Data suggest a collagen binding site at the N-terminal half of the decorin core protein. The importance of amino acids 125-158 of the mature core protein is further supported by the finding that these amino acids are likely to be on the surface of the molecule according to computer modelling and antibody-binding studies (Scott 2003).

Conclusions

To summarize, the comparison between the proposed theoretical model for biglycan and the recently released X-ray structure for the same protein revealed high similarities between the two proteins. The low RMSD value confirms the accuracy of the theoretical model as well as the reliability of the homology methods and algorithms that were used for the current study. Future structural cross-validation with the X-ray crystal structure will yield invaluable insights for the molecular properties and biochemical role of biglycan.

References

Aravind L & Koonin EV 1999 Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. *J Mol Biol* **287** 1023-1040

Balatsos N, Vlachakis D, Chatzigeorgiou V, Manta S, Komiotis D, Vlassi M & Stathopoulos C 2012 Kinetic and in silico analysis of the slow-binding inhibition of human poly (A)-specific ribonuclease (PARN) by novel nucleoside analogues *Biochimie* **94** 214-221

Banerjee-Basu S & Baxevanis AD 2002 Predictive Methods Using Protein Sequences. In *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, edn 2. Eds AD Baxevanis & BFF Ouellette. New York, USA: John Wiley & Sons, Inc.

Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD & Zardecki C 2002 The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* **58** 899-907

Chen R, Li L & Weng Z 2003 ZDOCK: an initial-stage protein-docking algorithm. *Proteins* **52** 80-87.

Chothia C 1992 One thousand families for the molecular biologists *Nature* **357** 543-544

Chothia C & Lesk AM 1986 The relation between the divergence of sequence and structure in proteins *EMBO J* **5** 823-826

Cintron C 1989 The function of proteoglycans in normal healing cornea. In *Healing Processes in the Cornea.* Eds CC Beuerman RW, Crosson CE & Kaufman HE. Texas: The Portfolio Publishing Company.

D'Alfonso G, Tramontano A & Lahm A 2001 Structural conservation in single-domain proteins: implications for homology modeling *J Struct Biol* **134** 246-256

DeLano WL 2002 The PyMOL User's Manual. CA, USA: DeLano Scientific

Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J & Kollman P 2003 A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem* **24** 1999-2012

Eisenberg D, Lüthy R & Bowie JU 1997 VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol* **277** 396-404

Funderburgh JL, Funderburgh ML, Mann MM, Corpuz L & Roth MR 2001 Proteoglycan expression during transforming growth factor beta -induced keratocyte-myofibroblast transdifferentiation. *J Biol Chem* **276** 44173-44178

Funderburgh JL, Mann MM & Funderburgh ML 2003 Keratocyte phenotype mediates proteoglycan structure: a role for fibroblasts in corneal fibrosis *J Biol Chem* **278** 45629-45637

Grishin NV 2001 Fold Change in Evolution of Protein Structures *J Struct Biol* **134** 167-185

Hocking AM, Shinomura T & McQuillan DJ 1998 Leucine-Rich Repeat Glycoproteins of the Extracllular Matrix *Matrix Biol* **17** 1-19

Krantz DD, Zidovetzki R, Kagan BL & Zipursky SL 1991 Amphipathic beta structure of a leucine-rich repeat peptide *J Biol Chem* **266** 16801-16807

Kresse H, Liszio C, Schonherr E & Fisher LW 1997 Critical Role of Glutamate in a Central Leucine-rich Repeat of Decorin for Interaction with Type I Collagen *J Biol Chem* **272** 18404-18410

Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R & Thornton JM 1996 AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* **8** 477-486

Li L, Chen R & Weng Z 2003 RDOCK: refinement of rigid-body protein docking predictions. Proteins **53** 693-707.

Naito Z 2005 Role of the small leucine-rich proteoglycan (SLRP) family in pathological lesions and cancer

cell growth *J Nippon Med Sch* **72** 137-145
Palaiomylitou M, Tartas A, Vlachakis D, Tzamarias D & Vlassi M 2008 Investigating the structural stability of the Tup1-interaction domain of Ssn6: Evidence for a conformational change on the complex *Proteins* **70** 72-82
Sellis D, Drosou V, Vlachakis D, Voukkalis N, Giannakouros T & Vlassi M 2012 Phosphorylation of the arginine/serine repeats of lamin B receptor by SRPK1—Insights from molecular dynamics simulations. *Biochim Biophys Acta* **1820** 44-55
Sellis D, Vlachakis D & Vlassi M 2009 Gromita: a fully integrated graphical user interface to gromacs 4 *Bioinform Biol Insights* **3** 99-102
Schonherr E, Hausser H, Beavan L & Kresse H 1995 Decorin-type I collagen interaction. Presence of separate core protein-binding domains *J Biol Chem* **270** 8877-8883.
Scott JE 2003 Elasticity in extracellular matrix 'shape modules' of tendon, cartilage, etc. A sliding proteoglycan-filament model. *J Physiol* **553** 335-343
Scott JE & Haigh M 1988 Identification of specific binding sites for keratan sulphate proteoglycans and chondroitin-dermatan sulphate proteoglycans on collagen fibrils in cornea by the use of cupromeronic blue in 'critical-electrolyte-concentration' techniques. *Biochem J* **253** 607-610
Scott PG, Bishop PN & Bella J 2006a On the calculation of the binding force between decorin and collagen. *J Biomech* **39** 1159-1160
Scott PG, Dodd CM, Bergmann EM, Sheehan JK & Bishop PN 2006b Crystal structure of the biglycan dimer and evidence that dimerisation is essential for folding and stability of class I small leucine rich repeat proteoglycans. *J Biol Chem* **281** 13324-13332
Scott PG, McEwan PA, Dodd CM, Bergmann EM, Bishop PN & Bella J 2004 Crystal structure of the dimeric protein core of decorin, the archetypal small leucine-rich repeat proteoglycan Proc Natl Acad Sci U S A **101** 15633-15638
Thompson JD, Higgins DG & Gibson TJ 1994 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
Vesentini S, Montevecchi FM & Redaelli A 2006 Response to letter to the editor: On the calculation of the binding force between decorin and collagen *J Biomech* **39** 1160-1162
Vesentini S, Redaelli A & Montevecchi FM 2005 Estimation of the binding force of the collagen molecule-decorin core protein complex in collagen fibril *J Biomech* **38** 433-443
Vangelatos I, Vlachakis D, Sophianopoulou V & Diallinas G 2009 Modelling and mutational evidence identify the substrate binding site and functional elements in APC amino acid transporters. *Mol Membr Biol* **26** 356-370
Vlachakis D 2009 Theoretical study of the Usutu virus helicase 3D structure by means of computer-aided homology modelling. *Theor Biol Med Model* **6** 9
Vlachakis D, Pavlopoulou A, Tsiliki G, Komiotis D, Stathopoulos C, Balatsos NA & Kossida S 2012 An integrated in silico approach to design specific inhibitors targeting human poly(a)-specific ribonuclease. *PLoS One* **7** e51113
Vlachakis D, Tsagrasoulis D, Megalooikonomou V & Kossida S 2013 Introducing Drugster: a comprehensive and fully integrated drug design, lead and structure optimization toolkit. *Bioinformatics* **29** 126-128
Watson PG & Young RD 2004 Scleral structure, organisation and disease. A review *Exp Eye Res* **78** 609-623
Weber IT, Harrison RW & Iozzo RV 1996 Model structure of decorin and implications for collagen fibrillogenesis *J Biol Chem* **271** 31767-31770
Wegrowski Y, Paltot V, Gillery P, Kalis B, Randoux A & Maquart FX 1995 Stimulation of sulphated glycosaminoglycan and decorin production in adult dermal fibroblasts by recombinant human interleukin-4. *Biochem J* **307** 673-678
Zubay G 1998 Biochemistry. Boston: McGraw Hill