

This article was downloaded by:

On: 14 January 2011

Access details: *Access Details: Free Access*

Publisher *Taylor & Francis*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Molecular Simulation

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713644482>

### Application of drug discovery software to the identification of heparin-binding sites on protein surfaces: a computational survey of the 4-helix cytokines

Barbara Mulloy<sup>a</sup>; M. J. Forster<sup>a</sup>

<sup>a</sup> National Institute for Biological Standards and Control, Potters Bar, Hertfordshire, UK

**To cite this Article** Mulloy, Barbara and Forster, M. J.(2008) 'Application of drug discovery software to the identification of heparin-binding sites on protein surfaces: a computational survey of the 4-helix cytokines', *Molecular Simulation*, 34: 4, 481 — 489

**To link to this Article:** DOI: 10.1080/08927020701784754

**URL:** <http://dx.doi.org/10.1080/08927020701784754>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

## Application of drug discovery software to the identification of heparin-binding sites on protein surfaces: a computational survey of the 4-helix cytokines

Barbara Mulloy<sup>a\*</sup> and M.J. Forster<sup>b</sup>

<sup>a</sup>National Institute for Biological Standards and Control, South Mimms, Potters Bar, Hertfordshire, UK; <sup>b</sup>Syngenta R&D IS, Jealotts Hill International Research Centre, Bracknell, Berkshire, UK

(Received 13 July 2007; final version received 30 October 2007)

Heparin, best known as a potent anticoagulant, also interacts with many other proteins for which the natural ligand is heparan sulfate (HS). The hope that HS would display specific sequences which would bind selectively to each of these proteins has not been fulfilled, but this may be at least in part due to the relationship between HS sequence and 3D structure. The example of FGF-1 is used to demonstrate that many different sequences can give rise to the 3D patterns of charge, which form binding motifs for proteins. Partly because of this redundancy in sequence-to-structure relationship, the application of conventional high-throughput drug discovery methods for the development of heparin or heparan based therapeutic agents is not yet practicable. However, it is possible to adapt systematic docking calculations to work in a moderately high throughput manner to screen protein structures from the Protein Data Bank (PDB) for predicted heparin-binding sites. A survey of protein structures in the Structural Classification of Proteins (SCOP) superfamily of 4-helical cytokines is presented.

**Keywords:** heparin; heparan sulfate; cytokine; drug discovery; docking

### 1. Introduction

The process of drug discovery may be generally defined as ‘the decision making process in the pharmaceutical industry by which compounds are identified and selected for clinical trial’ [1]. Within the last twenty years, the phrase has come increasingly to refer to a set of systematic, high throughput techniques for the screening of small molecule drug candidates. Such techniques include both experimental and theoretical screening methods, and there are many possible strategies for combining the two to yield agents which will interact effectively and selectively with a particular protein target, often an enzyme or a receptor.

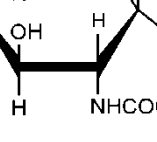
It is clear that not all useful drugs can be found in this way: heparin is a case in point. This highly heterogeneous polysaccharide (Figure 1), because of its variable patterns of sulfation and its domain structure, is capable of containing many different sequence motifs. It is a particularly highly sulfated example of the class of glycosaminoglycans (GAGs) known as heparan sulfate (HS; [2]), and is found in mast cell granules, where it is essential for the correct storage of proteases [3,4]. HS is present as the glycan side-chain of several proteoglycans, on the surfaces of most cells [2] and in the extracellular matrix [5], where it interacts with, and modulates the actions of, many of the cytokines, chemokines, growth factors, and other proteins which exert their effects in

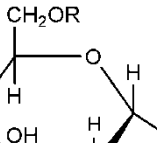
inflammation [6], reproduction [7], the immune system, and embryonic differentiation and development [8].

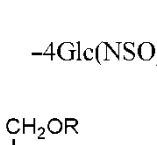
### 2. Sequence and structure in heparin and heparan sulfate

The most intensively studied and best understood sequence of monosaccharide residues in heparin is an unusual pentasaccharide, which is the minimum requirement for high affinity to antithrombin. It is this sequence which accounts for the high anticoagulant potency of heparin, and hence its use as an antithrombotic agent; the essential pentasaccharide has been prepared synthetically and is itself used as a drug [9]. When it became clear that heparin, as a model compound for HS, was capable of physiologically important interactions with other classes of protein, such as the fibroblast growth factors (FGFs) [10], the example of the antithrombin-binding sequence led to a search for other, equally specific sequences in either heparin or HS which would confer particular affinity for any given binding partner. This search for specificity of a high order has on the whole been unsuccessful, and a recent study of structures which are capable of potentiating FGF-mediated cell growth has concluded that HS fine structure may be less influential than has previously been supposed [11]. However, evidence from studies of model

\*Corresponding author. Email: bmulloy@nibsc.ac.uk

(a)   
-4GlcNAc $\alpha$ 1-4Glc $\beta$ 1-

(b)   
-4Glc(NSO<sub>3</sub><sup>-</sup>) $\alpha$ 1-4Ido $\alpha$ 1-

(c)   
-4Glc(NSO<sub>3</sub><sup>-</sup>) $\alpha$ 1-4Glc $\beta$ 1-

R = H or SO<sub>3</sub><sup>-</sup>

HS is often represented and imagined, in terms of sequences, rather than 3D structures [15]. The crystal structure of FGF-1 (2axm.pdb) complexed with a heparin oligosaccharide (Figure 2; [16]) shows clearly that the pattern of sulfate groups interacting with the protein can be formed in two ways, involving clusters of sulfate and carboxylate substituents on either side of the polysaccharide chain. Two separate molecules of FGF-1, aligned in opposite directions along the heparin chain, each interact with a cluster of three sulfates, part of a second cluster, and the carboxylate between the two clusters. The charge-based interactions, between the acidic substituents on the polysaccharide and basic residues on the surface of the protein, usually dominate the interface, and the detailed nature of the sugar backbone carrying the substituents is much less important, so long as it presents the substituents in an appropriate 3D pattern. This ‘pseudo-symmetry’, in which the underlying asymmetry of the sugar backbone is hidden by the almost symmetrical arrangement of bulky and highly charged substituents, is a complicating factor in the interpretation of HS sequence requirements for affinity with different proteins. Another such factor is the finding that, for most interactions between heparin and proteins, substitution with additional sulfate groups does not decrease affinity. Bearing in mind both factors together, a simplistic calculation indicates 31 different pentasaccharide sequences (starting and ending with glucosamine) which will contain a single FGF-1 binding motif. Leaving the motif on one side of the molecule (Figure 2) undisturbed, and assuming that any or all of the four remaining sulfated positions may or may not be occupied,  $16\ (2^4)$  different possible compounds can be defined. Repeating this exercise for the second side gives 31 possible sequences in all (not counting the fully sulfated compound twice). Such a sequence is more likely to occur in highly sulphated regions of the polysaccharide. Potentiation of growth factor activity is more complex than simple affinity for the growth factor itself, and it is clear that the requirements for functional interaction with the growth factor/receptor complex are not the same as for the growth factor alone [17]; still, the search for a 3D pattern is more likely to be successful than for a specific sequence.

The use of 3D models in drug design is not a new concept [18]. Molecular modelling techniques, in particular those in which a small molecule is docked into its binding site, are frequently used in the design of new drugs.

A conventional application of theoretical techniques to the process of designing a new drug would be to take a particular protein, for example an enzyme, and to look at

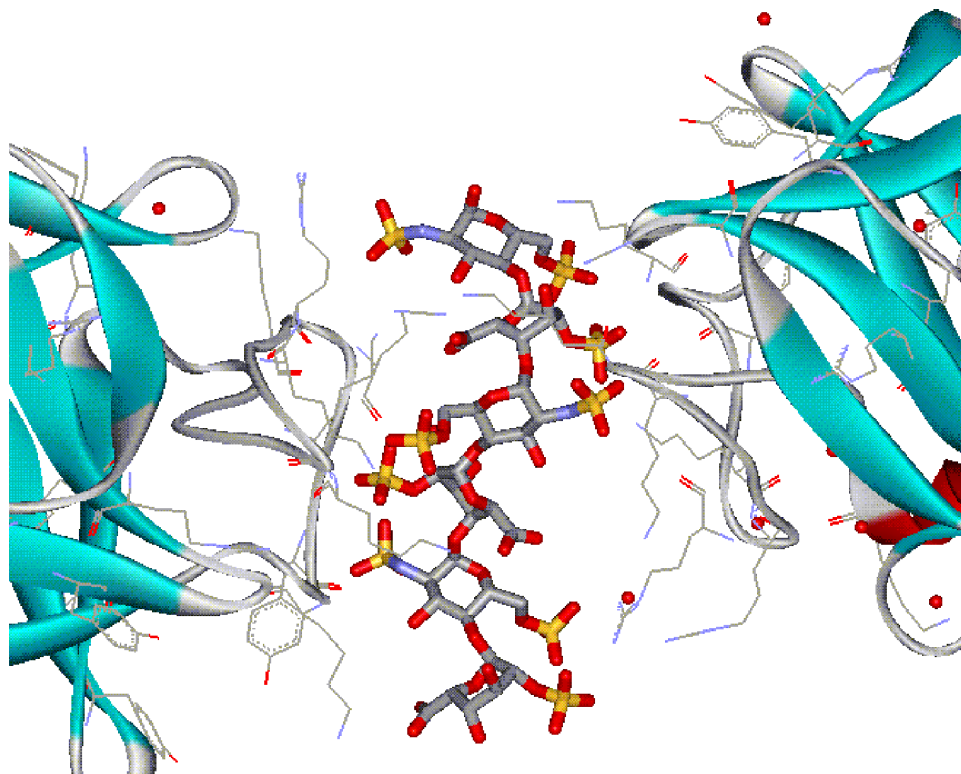


Figure 2. A detail from the crystal structure of FGF-1 complexed with a heparin hexasaccharide, with the protein shown in ribbon form and the saccharide in stick form (reproduced from [15] with permission). Two molecules of FGF-1 interact with the hexasaccharide, on opposite sides. Although the carbohydrate backbone is not symmetrical, the patterns of sulphate and carboxylate on either side are sufficiently similar as to bind to the same site on the protein.

the detailed experimental structure of a complex between the protein and its ligand, for example a substrate or inhibitor. On the basis of the structural details of the ligand and its binding site, new compounds may be proposed, perhaps with a view to increasing the affinity of an inhibitor. The compounds can be screened by molecular modelling, using the technique known as docking calculations, which explore many mutual orientations of protein and ligand to find out how best to accommodate the ligand in the binding site. Compounds with the best affinity for the protein will produce theoretical complexes with favourably low interaction energies. For glycosaminoglycan–protein interactions this course of action is not yet practicable. For one thing, the detailed modelling of GAG ligands with proteins is technically challenging [19], and secondly, the definition of a binding motif in heparin or HS in terms of its sequence is problematic, as described above. In addition, heparin and its fragments do not bind tightly to deep clefts in the protein surface, but to superficial sites, and the complexes are formed largely by electrostatic interactions. The molecular modelling of glycosaminoglycan sequences has therefore not yet reached the stage at which docking different sequences to one particular protein is useful; but

the reverse process, in which a generic heparin model ligand is used to find GAG binding sites on the surfaces of many proteins, is possible. Several groups have developed protocols for a systematic search of the entire surface of a protein to find the best heparin-binding site [20–22]. We have used a combination of predictive docking with experimental studies in the successful identification of several heparin-binding sites [23–25].

#### 4. Docking protocol for identification of heparin binding sites on protein surfaces

Our docking strategy was developed with the aim of discovering potential heparin-binding sites on protein surfaces, for use both in the illustration and rationalisation of experimental results such as NMR titrations [23] or the design of site-directed mutagenesis experiments [25]. As it is not necessarily the case that heparin structures will bind to proteins in a single, defined orientation (see for example the case of FGF-1 described in the previous section), no emphasis was placed on the detailed prediction of the geometry of the complexes, or accurate calculation of the interaction energy. Rather



than the conventional use of docking techniques, in which the geometry of a small ligand molecule in a known high-affinity binding site is optimised, we used a simple protocol to screen the entire surface of a small protein for clusters of basic residues which offer suitable shape and charge profiles complementary to the pattern of acidic substituents along the heparin chain.

Docking of several heparin oligosaccharide ligand models to protein structures was performed as previously described [26] using Autodock, version 2.4 [27], with partial charges for protein atoms taken from the AutoDock version of the AMBER force field. Co-ordinates for the heparin oligosaccharide ligands were derived from the NMR structure for the predominant repeating disaccharide of heparin 1hpn.pdb [28] with partial atomic charges from *ab initio* calculations using the Jaguar program (Schrodinger, Inc., Portland, Oregon, USA) on 1-OMe, 4-OMe substituted monosaccharides. Two pentasaccharide ligands were used, each with glucosamine at both reducing and non-reducing termini. For one of these, the two iduronate residues were both in the  $^1C_4$  conformation, and for the other they were both in the  $^2S_0$  conformation; in heparin both these forms are in equilibrium. All the exocyclic bonds in these pentasaccharide models were regarded as rotatable with the exception of the glycosidic linkages. Some calculations were also performed using a completely rigid endecasaccharide ligand model.

Docking was typically performed on a grid of  $120 \times 120 \times 120$  points, with the addition of a central grid point. The grid was centred on the mean of the coordinates of the protein. Grid spacing was  $0.7 \text{ \AA}$ , leading to a grid of  $84 \times 84 \times 84 \text{ \AA}$ . This determines the largest protein that can be studied by this protocol. Separate grids of van der Waals (VDW) interaction energy are then calculated for each atom type in the ligand (C, N, H, S, O) and an electrostatic interaction energy grid is computed for a single electron charge. These grids are used during the docking process to rapidly calculate the interaction energy of the ligand with the protein. This is achieved by finding the grid points surrounding each ligand atom and using an interpolation procedure to find the energy contribution at the current coordinates. The energies are then summed over all atoms in the ligand and the torsion energy terms added to the VDW and electrostatic energies. During the docking procedure the position, orientation and allowed torsion angles of the ligand structure are optimised by a Monte Carlo simulated annealing procedure. Initial simulation temperature (defined in RT units) of 1000 was used and a temperature reduction factor of 0.95 per cycle was used; typically 128 runs of 300 cycles were performed. As detailed analysis of the modes of binding was not an aim of the protocol, no clustering of the results was performed.

These parameters were selected in a validation study of the protocol, by performing simulations on a protein/heparin oligosaccharide complex of known crystal structure, that of FGF2 with a heparin hexasaccharide (1bfc.pdb; [29]), adjusting parameters to most reliably reproduce the known heparin-binding site. Docking was performed with a unit dielectric rather than a distant dependent dielectric, as this was found more reliable in reproducing known binding sites. Docking calculations typically required approximately 50 min on a 300 MHz SGI octane workstation. Docked ligand coordinates were extracted from output files using a set of in-house Perl scripts.

## 5. Use of docking calculations to screen the structure database for heparin-binding proteins

The docking protocol can readily be performed on a medium-high throughput basis, so that the possibility arises of a systematic survey of proteins with known 3D structures, in order to supplement the limited number of experimentally determined heparin-protein complexes [30]. There are over 41,000 structures in the Protein Data Bank (PDB), however, so an initial survey of a subset of solved protein structures is desirable. In the structural classification of proteins (SCOP) [31] system (<http://scop.mrc-lmb.cam.ac.uk/scop/>), members of the Superfamily of 4-helical cytokines (in the Class of all-alpha proteins) form a suitable group (Figure 3). They are small proteins, related in functional terms as well as by structure, and all performing their biological functions outside the cell, so that their environment is rich in GAGs.

We have performed semi-automated docking calculations using the program Autodock, as previously described [26], with co-ordinates for the heparin-based oligosaccharide ligands taken from the PDB file 1hpn.pdb [28] and co-ordinates for proteins derived from PDB files (<http://www.rcsb.org/pdb/>) corresponding to experimentally determined structures in the SCOP families of long- and short-chain 4-helical cytokines, and the interferon/interleukin-10 (IL-10) group. Only the structures of human cytokines were chosen. For all the proteins except for IL-5, IL-6 and IL-10 one monomer of the protein was used for docking; a single chain was edited from the pdb file where that file contained more than one chain. No attempts were made to replace missing co-ordinates in the pdb files by molecular modelling. Heparin-protein complexes calculated to have intermolecular interaction energies of less than  $-1000 \text{ kcal/mol}$  were regarded as predictions of capacity to bind heparin; those with interaction energies of more than zero were regarded as predicting no capacity to bind heparin. The energy units

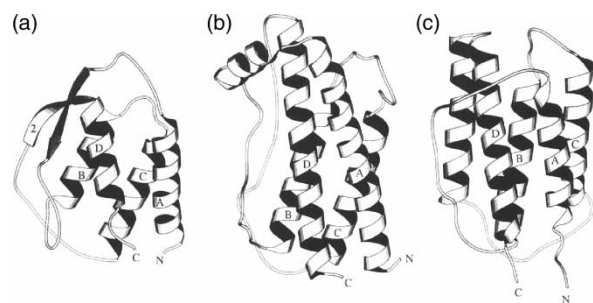


Figure 3. Ribbon diagrams showing structures of three proteins in the four helix bundle superfamily (reproduced from [46] with permission). The four helices are labelled A, B, C and D by convention. Long loops between helices A and B, and helices C and D, run the length of the molecule, so that A and B run parallel, as do C and D in the opposite direction. (a) A short-chain cytokine, showing a short two-stranded beta sheet. (b) A long-chain cytokine, with a short extra helix after A and no beta stands. (c) A member of the interferon/IL-10 family, with a helix in the C–D connection.

of the Autodock function are usually given as kcal/mol; the high figures in the tables are a consequence of the high weighting given to electrostatic terms in the forcefield by using a unit dielectric constant. Values given should be understood as the results of a ranking function with no significance in absolute terms. Co-ordinates were visualised and figures prepared using the programs InsightII and Weblab Viewer (Accelrys).

## 6. Long-chain cytokines

Docking calculations were possible for seven members of this family, listed in Table 1, out of a total of 9. IL-12 was omitted, as inspection of the crystal structure (1f45.pdb) showed that the 4-helical bundle is a minor part of this structure, which also includes a second subunit made up of three Immunoglobulin G (IgG) domains [32], and the co-ordinates for Oncostatin M (1evs.pdb) were too incomplete to give anything, other than an artefactual result. Four members of this group are strongly predicted to bind to heparin: leukaemia inhibitory factor (LIF), IL-6, erythropoietin (EPO) and thrombopoietin (TPO). Heparin is known to interact with IL-6 and a site involving the A and D helices has been proposed [33], which is entirely consistent with our predictions. This face of the bundle is closely involved with receptor interactions, and the proposed heparin-binding site lies close to, though not overlapping, the areas of IL-6 which interact with both its receptors; IL-6R and gp130. Though, it is possible to imagine heparin occupying the site at the same time as either of the two receptors, simple steric factors prevent heparin-binding to the complete signalling complex of IL-6 with both receptors. LIF has not so far been identified as a heparin-binding protein, but the docking results are highly indicative that an interaction is possible, though in contrast with IL-6 the proposed heparin-binding site lies along the length of helix A, extensively overlapping the gp130 binding site. Ciliary neurotrophic factor (CNTF) is

Table 1. Prediction of heparin binding to long-chain cytokines.

Cytokine (abbreviation)	PDB filename (monomer or dimer)	Intermolecular interaction energies for pentasaccharide ligands		Position of heparin binding site <sup>1</sup>
		IdoA <sup>1</sup> C <sub>4</sub>	IdoA <sup>2</sup> S <sub>0</sub>	
Leukemia inhibitory factor (LIF)	1emr.pdb (monomer)	– 1769	– 1785	Along helix A (N25, Q26, R28) and end of helix C (R133, K137)
IL-6 <sup>2</sup>	1alu.pdb (dimer)	– 1270	– 1262	Helices A (K27, R30, K66) and D (K171, Q175, R179, R182)
EPO	1buy.pdb (monomer)	– 1799	– 1844	Along helix D (R139, K140, R143, N147, R150, K154, R166)
TPO	1v7m.pdb (monomer)	– 1617	– 1617	Helices A (R10, K14, R17, H20, R25) and D (R136, R140)
Prolactin	1n9d.pdb (monomer)	– 627	– 625	
hGH <sup>2</sup>	3hhr.pdb (monomer)	– 307	– 277	
Leptin	1ax8.pdb (monomer)	+ 63	+ 81	
G-CSF <sup>2</sup>	1cd9.pdb (monomer)	– 154	– 169	
CNTF	1cnt.pdb (monomer)	– 508	– 520	

<sup>1</sup> Residues containing atoms within 3.5 Å of ligand in the five lowest energy complexes. <sup>2</sup> Reported to interact with heparin: see text.

involved in neuronal cell growth and survival; like IL-6 and LIF, it has gp130 as part of its receptor complex. CNTF also shares the LIF receptor, though unlike LIF, CNTF requires a specific, non-signalling receptor in addition [34]. Unlike IL-6 and LIF, CNTF has no plausible heparin binding site.

Both EPO and TPO are classified as long-chain cytokines with short-chain topology. Although, neither cytokine is recognised as a heparin-binding protein, the docking calculations indicate that a well-formed heparin-binding site is present on both. The crystal structure of the complex between EPO and two receptor molecules [35] clearly displays the same linear pattern of basic residues along helix D, which passes through the centre of one of the EPO receptor binding regions (Figure 4).

For the remaining long-chain cytokines, the results are as would be expected for a chance patch of basic residues on the protein surface rather than a functional heparin binding site, though heparin has been reported to interact with some of them (Granulocyte-colony stimulating factor (G-CSF) [36] and somatotropin (human growth hormone: hGH); [37]).

## 7. Short-chain cytokines

Eight structures are listed in Table 2: a ninth, macrophage colony-stimulating factor, was omitted as co-ordinates (1hmc.pdb) were incomplete. Strongly positive predictions of heparin binding were made for interleukins IL-2, IL-4 and IL-13, and (less strongly) for IL-3; a positive prediction is also made for the intertwined dimer IL-5. Interactions have been found experimentally between heparin and IL-2 [38], IL-3 [39], IL-4[40], and IL-5 [41], but not reported so far for IL-13. Overlaid complexes for IL-2, IL-3, IL-4 and IL-13 are shown in Figure 5.

The NMR structure of IL-3 (1jli.pdb) has as many as 13 altered amino acids. The predicted heparin-binding site involves none of these, but the prediction should still be regarded with caution. Like Granulocyte-macrophage colony stimulating factor (GM-CSF), IL-3 is adsorbed by HS [39], but no detailed structural description of the interaction is available.

GM-CSF cannot be confidently predicted to bind heparin on the basis of the docking calculations. However,

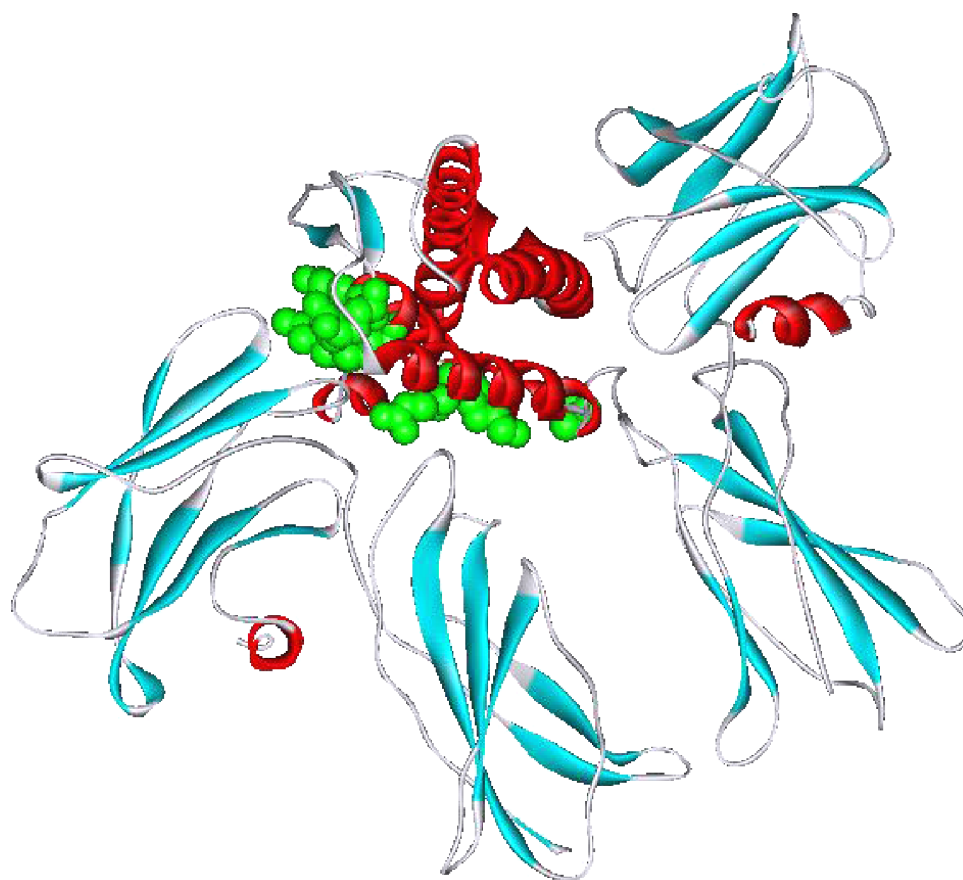


Figure 4. The predicted heparin binding site for EPO. EPO (the central bundle of helices, coloured red) and its receptor (largely blue-coloured strands) are represented as a ribbon diagram, with the basic residues on EPO predicted to bind heparin displayed as sticks and coloured green. The potential heparin-binding site runs along helix D and overlaps with the receptor binding site [35].

Table 2. Prediction of heparin binding to short-chain cytokines.

Cytokine (abbreviation)	PDB filename (monomer or dimer)	Intermolecular inter- action energies for pentasaccharide ligands		Position of heparin binding site <sup>1</sup>
		Ido A <sup>1</sup> C <sub>4</sub>	Ido A <sup>2</sup> S <sub>0</sub>	
IL-4 <sup>2</sup>	1hik.pdb (dimer)	– 2293	– 2303	Helix C (H74, K77, R81, K84, R85, R88)
IL-2 <sup>2</sup>	3ink.pdb (monomer)	– 1225	– 1252	AB (K32, K35, R38) and CD (Q74, K76, R81) loops
IL-13	1ijz.pdb (monomer)	– 1099	– 1120	AB loop (K25) and Helix D (K97, K104, K105, R108)
IL-3 <sup>2</sup>	1jli.pdb (monomer)	– 890	– 862	AB loop (K28, R29) and Helix D (R108, R109)
IL-5 <sup>2</sup>	1hul.pdb (intertwined dimer)	– 1201	– 1215	Helix C (K70, K77, K84, K85, R90) and N-terminus
GM-CSF	lcsg.pdb (monomer)	– 407	– 408	
Flt3 ligand	lete.pdb (monomer)	– 556	– 544	
SCF	lexz.pdb (monomer)	– 159	– 199	

<sup>1</sup>Residues containing atoms within 3.5 Å of ligand in the five lowest energy complexes. <sup>2</sup>Reported to interact with heparin: see text.

pH-dependent interaction between GM-CSF and heparin has been described, and three histidine residues identified as important for binding [42]. In the crystal structures, these residues are not exposed, but it is possible this is not the case for the solution structure at low pH.

The Flt3-ligand (Flt-3l) and stem cell factor (SCF) have overlapping biological functions and share a receptor

type [43], but are dissimilar in their predicted behaviour towards heparin. While it is possible that Flt-3l may bind heparin, the prediction for SCF is strongly negative. The crystal structures of both Flt-3l (lete.pdb) and SCF (lexz.pdb) contain co-ordinated metal ions, which are not taken into account in our docking protocol. While for the acidic protein SCF this is unlikely to introduce any heparin binding capacity, it is conceivable that the presence of zinc might influence heparin binding to Flt-3.

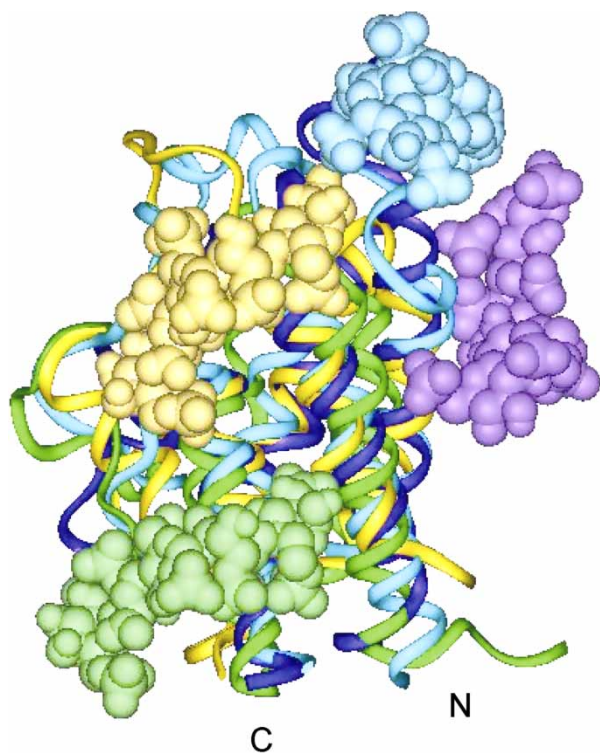


Figure 5. Four short-chain cytokines predicted to bind heparin (Table 1). The structures of IL-2 (3ink.pdb; pale blue), IL-3 (1jli.pdb; yellow) IL-4 (1hik.pdb; purple) and IL-13 (1ijz.pdb; green) shown as superimposed ribbon diagrams (in a similar orientation to the structures in Figure 3), each in complex with its lowest energy heparin pentasaccharide (Table 1). The four predicted heparin-binding sites do not overlap.

## 8. Interferons/interleukin 10 family

Members of this group have an extra short helix in one of the crossover connections; their heparin-binding predictions are listed in Table 3. Interferons  $\alpha$ -2A (1itf.pdb),  $\beta$  (1au1.pdb) and  $\gamma$  (1eku.pdb) were considered, and here the results of the docking calculations meet a common problem. A clear prediction was obtained only for interferon- $\beta$ , which has a strong potential binding site extending over helices A and D; a less clear prediction places a possible heparin binding site in the same area for interferon- $\alpha$ 2A. For interferon- $\gamma$ , post-calculation review of relevant literature indicated that prediction of a heparin binding site by this method is not valid, as the heparin binding site is known to be situated on an unstructured C-terminal tail [44]. It is not wise to place reliance on a prediction of this kind for the globular domain of a protein with N- or C-terminal unstructured sequences, without careful inspection of the unstructured sequences for concentrations of basic amino acids. Modelling of the complex would require an exhaustive conformation sampling scheme, followed by docking on each conformer generated. For an unstructured region of more than a few residues this procedure is not practicable.

Three IL structures also fall into this family; those for IL-10, IL-19 and IL-22. IL-10 is an intertwined dimer, whereas IL-19 and IL-22 are not. Predictions for both IL-19 and IL-22 are not clear; heparin may bind to IL-22



Table 3. Prediction of heparin binding to the interferon/IL10 group of 4-helical cytokines.

Cytokine (abbreviation)	PDB filename (monomer or dimer)	Intermolecular interaction energies for pentasaccharide ligands		Position of heparin binding site <sup>1</sup>
		Ido A <sup>1</sup> C <sub>4</sub>	Ido A <sup>2</sup> S <sub>0</sub>	
Interferon-β	lau1.pdb (monomer)	– 1501	– 1526	Helices A (K33, R35, N37) and D (R152, R159)
Interferon-γ <sup>2</sup>	leku.pdb (dimer-like single chain mutant)	– 858	– 892	
Interferon-α 2A	litf.pdb (monomer)	– 605	– 629	
IL-10 <sup>2</sup>	lilk.pdb (intertwined dimer)	– 1442	– 1463	Helices D (R102, R106, R107) and E (N116, K117, K119, K125)
IL-19	lnif.pdb (monomer)	– 899	– 911	C-D loop (R101, Q104, Q108, R112)
IL-22	lm4r.pdb (monomer)	– 809	– 820	Helices A (K44, Q48) and D (R175, N176)

<sup>1</sup> Residues containing atoms within 3.5 Å of ligand in the five lowest energy complexes. <sup>2</sup> Reported to interact with heparin: see text.

by the helix-A/helix-D face, and to IL-19 by the C-D loop and residues at the C-terminus. Heparin is strongly predicted to bind to IL-10, interacting with both bundles in the dimer, and heparin-IL/10 binding has been observed experimentally [45].

## 9. Discussion

A structural comparison among short-chain helical cytokines [46] has pointed out that, though these compounds are structurally and functionally related, their amino acid sequences are not strongly similar. This is true of the 4-helical cytokines in general, but in spite of poor sequence conservation, they are considered a monophyletic group, arising at some point in the distant past from a single ancestral gene [47]. Their receptors are however closely related, and shared subunits are common: for example the gp130 receptor subunit shared by several of the long-chain cytokines (such as IL-6 and LIF) and the common gamma receptor subunit shared by several short-chain cytokines. There are clearly elements of surface character in common between cytokines which share receptors, and it is reasonable to ask whether other surface features, such as heparin binding sites, are shared. However, the results of this brief survey do not indicate that the capacity to bind heparin, or the site at which heparin might bind, is conserved between the 4-helical cytokines, as is shown by the comparison of the four heparin binding sites of IL-2, IL-3, IL-4 and IL-13 illustrated in Figure 5. The heparin binding site for IL-2 is constructed from loops [38], whereas, those for the other three ILs lie across helices A and D, or A and C.

This study was undertaken in part to investigate the feasibility of high-throughput screening of structures in the PDB for the presence of heparin-binding sites, in

order to add to the limited number of such sites so far defined by crystallised complexes. Several problems arise for this type of exercise. First, there are entries in the PDB which are incomplete, and missing residues in loop or terminal unstructured regions may cause a heparin-binding protein to be missed. Secondly, it is not always clear which oligomeric form of the protein is most appropriate for the docking exercise. Other considerations include glycosylation and other post-translational modifications, which are not always clear in the PDB structure. However, it is relatively quick and simple to generate results from our protocol. Several of the predictions made will be investigated experimentally.

Detailed analysis of the predicted binding sites has been avoided as being premature in the absence of experimental data. Though, these molecular modelling exercises offer no more than informed predictions, they are in agreement with all the experimental data so far in failing to identify any single amino acid sequence motif or feature of secondary structure common to all heparin binding sites [15,26,48]. The significance and function (if any) of the capacity of cytokines to bind heparin or HS cannot be generalised; each case requires separate investigation. However, it seems unlikely that HS acts as a co-receptor for these cytokines by analogy with the FGFs, as so many of the predicted heparin binding sites overlap receptor binding sites. An ability to bind to GAGs in the extracellular matrix may reasonably be expected to affect the diffusion of cytokines, and it may be in the location, rather than the signalling, of cytokines that HS plays its part in their action.

## References

- [1] L.P. Schacter et al., *Drug discovery and development in the pharmaceutical-industry*, Semin. Oncol. 19 (1992), p. 613.

- [2] D.L. Rabenstein, *Heparin and heparan sulfate: Structure and function*, Nat. Prod. Rep. 19 (2002), p. 312.
- [3] E. Forsberg et al., *Abnormal mast cells in mice deficient in a heparin-synthesizing enzyme*, Nature 400 (1999), p. 773.
- [4] D.E. Humphries et al., *Heparin is essential for the storage of specific granule proteases in mast cells*, Nature 400 (1999), p. 769.
- [5] S.M. Knox and J.M. Whitelock, *Perlecan: How does one molecule do so many things?*, Cell. Mol. Life Sci. 63 (2006), p. 2435.
- [6] C.R. Parish, *The role of heparan sulphate in inflammation*, Nat. Rev. Immunol. 6 (2006), p. 633.
- [7] A.I. de Agostini, *An unexpected role for anticoagulant heparan sulfate proteoglycans in reproduction*, Swiss Med. Wkly. 136 (2006), p. 583.
- [8] S.M. Cool and V. Nurcombe, *Heparan sulfate regulation of progenitor cell fate*, J. Cell. Biochem. 99 (2006), p. 1040.
- [9] M. Petitou and C.A.A. van Boeckel, *A synthetic antithrombin III binding pentasaccharide is now a drug! What comes next?*, Angew. Chem. Int. Edit. 43 (2004), p. 3118.
- [10] M. Mohammadi, S.K. Olsen, and R. Goetz, *A protein canyon in the FGF–FGF receptor dimer selects from an a la carte menu of heparan sulfate motifs*, Curr. Opin. Struct. Biol. 15 (2005), p. 506.
- [11] J. Kreuger et al., *Interactions between heparan sulfate and proteins: The concept of specificity*, J. Cell Biol. 174 (2006), p. 323.
- [12] J.D. Esko and S.B. Selleck, *Order out of chaos: Assembly of ligand binding sites in heparan sulfate*, Annu. Rev. Biochem. 71 (2002), p. 435.
- [13] M.S.G. Pavao et al., *A unique dermatan sulfate-like glycosaminoglycan from ascidian – its structure and the effect of its unusual sulfation pattern on anticoagulant activity*, J. Biol. Chem. 270 (1995), p. 31027.
- [14] A.P. Alves et al., *Sulfated polysaccharides from the egg jelly layer are species-specific inducers of acrosomal reaction in sperms of sea urchins*, J. Biol. Chem. 272 (1997), p. 6965.
- [15] B. Mulloy, *The specificity of interactions between proteins and sulfated polysaccharides*, Anais. Acad. Bras. Cienc. 77 (2005), p. 651.
- [16] D.J. Stauber, A.D. DiGabriele, and W.A. Hendrickson, *Structural interactions of fibroblast growth factor receptor with its ligands*, Proc. Natl Acad. Sci. USA 97 (2000), p. 49.
- [17] O. Ostrovsky et al., *Differential effects of heparin saccharides on the formation of specific fibroblast growth factor (FGF) and FGF receptor complexes*, J. Biol. Chem. 277 (2002), p. 2444.
- [18] P. Gund et al., *Three-dimensional molecular modeling and drug design*, Science 208 (1980), p. 1425.
- [19] H. Verli and J.A. Guimaraes, *Insights into the induced fit mechanism in antithrombin–heparin interaction using molecular dynamics simulations*, J. Mol. Graph. Model 24 (2005), p. 203.
- [20] W. Bitomsky and R.C. Wade, *Docking of glycosaminoglycans to heparin-binding proteins: Validation for aFGF, bFGF, and antithrombin and application to IL-8*, J. Am. Chem. Soc. 121 (1999), p. 3004.
- [21] S. Ricard-Blum et al., *Characterization of endostatin binding to heparin and heparan sulfate by surface plasmon resonance and molecular modeling – role of divalent cations*, J. Biol. Chem. 279 (2004), p. 2927.
- [22] E. Krieger et al., *A structural and dynamic model for the interaction of IL-8 and glycosaminoglycans: Support from isothermal fluorescence titrations*, Proteins 54 (2004), p. 768.
- [23] Sachchidanand et al., *Mapping the heparin-binding site on the (13–14)F3 fragment of fibronectin*, J. Biol. Chem. 277 (2002), p. 50629.
- [24] S.E. Stringer et al., *Characterization of the binding site on heparan sulfate for macrophage inflammatory protein 1 alpha*, Blood 100 (2002), p. 1543.
- [25] D.J. Mahoney et al., *Characterization of the interaction between tumor necrosis factor-stimulated gene-6 and heparin – implications for the inhibition of plasmin in extracellular matrix microenvironments*, J. Biol. Chem. 280 (2005), p. 27044.
- [26] M.J. Forster and B. Mulloy, *Computational approaches to the identification of heparin-binding sites on the surfaces of proteins*, Biochem. Soc. Trans. 34 (2006), p. 431.
- [27] G.M. Morris et al., *Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock 2.4*, J. Comput. Aided Mol. Des. 10 (1996), p. 293.
- [28] B. Mulloy et al., *N.m.r. and molecular-modelling studies of the solution conformation of heparin*, Biochem. J. 293 (1993), p. 849.
- [29] B. Mulloy and M.J. Forster, *Conformation and dynamics of heparin and heparan sulfate*, Glycobiology 10 (2000), p. 1147.
- [30] A. Imberty, H. Lortat-Jacob and S. Perez, *Structural view of glycosaminoglycan–protein interactions*, Carbohydr. Res. 342 (2007), p. 430.
- [31] A.G. Murzin, S.E. Brenner, T. Hubbard and C. Chothia, *SCOP – A structural classification of proteins database for the investigation of sequences and structures*, J. Mol. Biol. 247 (1995), p. 536.
- [32] C. Yoon et al., *Charged residues dominate a unique interlocking topography in the heterodimeric cytokine interleukin-12*, EMBO J. 19 (2000), p. 3530.
- [33] R.S. Mummery and C.C. Rider, *Characterization of the heparin-binding properties of IL-6*, J. Immunol. 165 (2000), p. 5671.
- [34] W. He et al., *The N-terminal cytokine binding domain of LIFR is required for CNTF binding and signaling*, FEBS Lett. 579 (2005), p. 4317.
- [35] R.S. Syed et al., *Efficiency of signalling through cytokine receptors depends critically on receptor orientation*, Nature 395 (1998), p. 511.
- [36] A.Y. Liang et al., *Separation, identification and interaction of heparin oligosaccharides with granulocyte-colony stimulating factor using capillary electrophoresis and mass spectrometry*, Electrophoresis 26 (2005), p. 3460.
- [37] C. Zamiri and M.J. Groves, *Stabilization of somatropin by heparin*, J. Pharm. Pharmacol. 57 (2005), p. 555.
- [38] S. Najjam et al., *Further characterization of the binding of human recombinant interleukin 2 to heparin and identification of putative binding sites*, Glycobiology 8 (1998), p. 509.
- [39] R. Roberts et al., *Heparan sulphate bound growth factors: A mechanism for stromal cell mediated haemopoiesis*, Nature 332 (1988), p. 376.
- [40] H. Lortat-Jacob et al., *Human interleukin 4 is a glycosaminoglycan-binding protein*, Cytokine 9 (1997), p. 101.
- [41] R.J. Lipscombe et al., *Interleukin-5 binds to heparin heparan sulfate. A model for an interaction with extracellular matrix*, J. Leuk. Biol. 63 (1998), p. 342.
- [42] A. Sebollela et al., *Heparin-binding sites in granulocyte–macrophage colony-stimulating factor – Localization and regulation by histidine ionization*, J. Biol. Chem. 280 (2005), p. 31949.
- [43] S.N. Savvides, T. Boone and P.A. Karplus, *Flt3 ligand structure and unexpected commonalities of helical bundles and cystine knots*, Nat. Struct. Biol. 7 (2000), p. 486.
- [44] H. Lortat-Jacob and J.A. Grimaud, *Interferon–gamma binds to heparan sulfate by a cluster of amino acids located in the C-terminal part of the molecule*, FEBS Lett. 280 (1991), p. 152.
- [45] S. Salek-Ardakani et al., *Heparin and heparan sulfate bind interleukin-10 and modulate its activity*, Blood 96 (2000), p. 1879.
- [46] D.A. Rozwarski et al., *Structural comparisons among the short-chain helical cytokines*, Structure 2 (1994), p. 159.
- [47] M.O. Huising, C.P. Kruiswijk, and G. Flik, *Phylogeny and evolution of class-I helical cytokines*, J. Endocrinol. 189 (2006), p. 1.
- [48] B. Mulloy and R.J. Linhardt, *Order out of complexity – protein structures that interact with heparin*, Curr. Opin. Struct. Biol. 11 (2001), p. 623.