Collège Doctoral du Languedoc-Roussillon

THÈSE

Pour obtenir le grade de Docteur

Délivré par l'**Université de Montpellier**

Préparée au sein de l'école doctorale
Sciences Chimiques Balard ED 459

Et de l'unité de recherche
Institut des Biomolécules Max Mousseron

Spécialité **Chimie**

Présentée par **Jean-Patrick Francoïa**

Le principe KISS appliqué à la conception de senseurs et à la veille bibliographique

Soutenue le 06/09/2016 devant le jury composé de :

Pr Julien LECLAIRE	Université de Lyon	Rapporteur
Dr Jean-François NIERENGARTEN	Université de Strasbourg	Rapporteur
Pr Christopher HUNTER	Université de Cambridge	Examinateur
Pr Michael SMIETANA	Université de Montpellier	Président du jury, examinateur
Dr Robert PASCAL	Université de Montpellier	Directeur de thèse
Dr Laurent VIAL	Université de Montpellier	Co-encadrant de thèse

UNIVERSITÉ MONTPELLIER

Table of Contents

List of abbreviations

3D . . . . . three-dimensional
AAAS . . . . . American Association for the Advancement of Science
ACS . . . . . American Chemical Society
ANSSI . . . . . Agence Nationale de la Sécurité des Systèmes d'Information
CF . . . . . Carboxyfluorescein
CPU . . . . . Central Processing Unit
DCM . . . . . Dichloromethane
DGL . . . . . DendriGraft of Lysine
DIC . . . . . Diisopropylcarbodiimide
DIEA . . . . . Diisopropylethylamine
DMF . . . . . Dimethylformamide
ESI . . . . . Electrospray Ionization
ELN . . . . . Electronic Lab Notebook
FI . . . . . Fluorescence Intensity
GAG . . . . . Glycosaminoglycan
GOx . . . . . Glucose Oxydase
GPU . . . . . Graphics Processing Unit
HPLC . . . . . High Performance Liquid Chromatography
LDA . . . . . Linear Discriminant Analysis
MD . . . . . Molecular Dynamic
MM . . . . . Molecular Mechanic
MS . . . . . Mass Spectrometer
NAS . . . . . National Academy of Sciences of the United States of America
NPG . . . . . Nature Publishing Group
PAMAM . . . . . PolyAMidoAMine
PEI . . . . . PolyEthylenImine
PCA . . . . . Principal Component Analysis
PLL . . . . . Poly-L-lysine
RMSD . . . . . Root Mean Squared Deviation
RSC . . . . . Royal Society of Chemistry
RSS . . . . . Rich Site Summary
RT . . . . . Room Temperature
SEM . . . . . Scanning Electron Microscopy
SVC . . . . . Support Vector Classification
TIPS . . . . . Triisopropylsilane
TF-IDF . . . . . Term Frequency–Inverse Document Frequency
TFA . . . . . Trifluoracetic Acid

TSA . . . . . Time Stampig Authority
UPLC . . . . . Ultra Performance Liquid Chromatography
USP . . . . . U.S. Pharmacopeial
XML . . . . . Extensible Markup Language

KISS:

KISS is an acronym for "Keep it Simple, Stupid" as a design principle noted by the U.S. Navy in 1960. The KISS principle states that most systems work best if they are kept simple rather than made complicated; therefore simplicity should be a key goal in design and unnecessary complexity should be avoided. Variations on the phrase include "Keep it Simple, Silly", "Keep it Short and Simple", "Keep it Simple and Straightforward" and "Keep it Small and Simple".

Source: Wikipedia

Part I

Introduction

Le professeur Jean Marie Lehn a défini la chimie supramoléculaire comme la chimie ”au-delà des molécules”[1]. Elle se focalise sur les assemblages moléculaires créés à partir d'interactions intermoléculaires. Ces assemblages supramoléculaires assurent, au sein de chaque organisme vivant, des fonctions biologiques primordiales. En ce sens, la Nature stupéfie par son ingéniosité, et nous fournit de nombreux exemples d'assemblages assurant des fonctions vitales: membranes bi-couches, ribosomes, hémoglobine, etc. Comprendre ces processus d'assemblage est essentiel pour le (bio)chimiste souhaitant reproduire ou améliorer ces systèmes. La tâche n'est cependant pas triviale car la Nature a masqué les caractéristiques réellement importantes de ses molécules à travers une complexité accumulée au fil de son évolution[2]. Le cas d'école très souvent évoqué est la haute reconnaissance moléculaire des enzymes naturelles[3–5]. Alors que celles-ci sont souvent spécifiques d'un substrat (ou d'un type de substrat), les systèmes synthétiques peinent à égaler cette faculté de reconnaissance.
Cependant, grâce à l'essor de la chimie supramoléculaire ces dernières décennies, des structures fascinantes ont été mises au point - telles que les cavitands, les nœuds, les hélicates, ou les anneaux borroméens[6]- avec des capacités de reconnaissance moléculaire égalant (parfois même surpassant) celles des récepteurs naturels. Par exemple le macrocycle de Vial *et al.*[7], sélectionné et synthétisé grâce à une approche de chimie combinatoire dynamique[8], se lie plus efficacement à la spermine que son récepteur naturel, la double hélice d'ADN (figure 1).

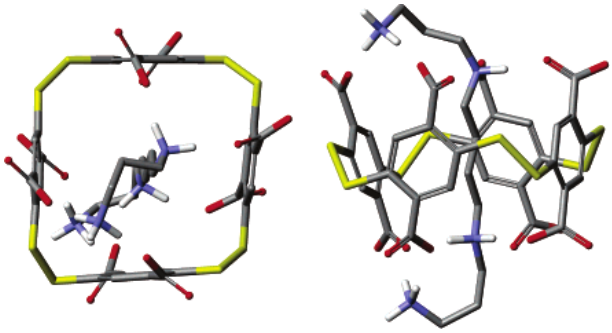


Figure 1: Vue de face et de profil du complexe macrocycle-spermine de Vial *et al.*[7]. Figure adaptée depuis la publication originale.

Koder *et al.*[9] ont quant à eux conçu une protéine de transport pour le dioxygène avec des propriétés surpassant celles de l'hémoglobine humaine (figure 2). En effet, ce récepteur présente une affinité diminuée envers le monoxyde de carbone (un gaz nocif), et une affinité augmentée envers le dioxygène.

Figure 2: Séquence et représentation schématique de la protéine transporteuse d'oxygène de Koder *et al.* [9]. Quatre hélices α sont reliées grâce à une structure "en chandelier". Figure adaptée depuis la publication originale.

Ces deux derniers exemples démontrent qu'il est tout à fait possible de concevoir des récepteurs simples, mais performants. La complexité n'est donc pas une caractéristique nécessaire des molécules biologiques, et les récepteurs synthétiques devraient être conçus le plus simplement possible.

Grâce à leurs propriétés accrues, ces récepteurs synthétiques possèdent de nombreuses applications thérapeutiques. Citons par exemple le dendrimère peptidique (pour la définition d'un dendrimère, *vide infra*) de Zhang *et al.* [10] (figure 3). Cette molécule, riche en résidus tryptophane, forme des complexes supramoléculaires avec l'ADN contenu dans les cellules tumorales, et provoque l'apoptose de celles-ci.

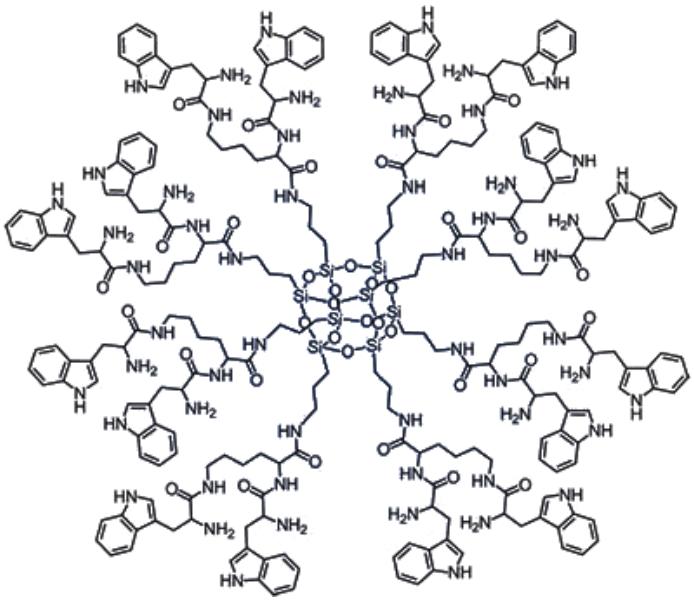


Figure 3: Dendrimère de peptide riche en résidus tryptophane. Figure adaptée de Zhang *et al.* [10].

Aussi, les assemblages supramoléculaires peuvent être utilisés pour la vectorisation de médicaments. Howe *et al.* [11] ont utilisé l'agent gelifiant 1,3:2,4-di(4-acylhydrazide)-benzylidene sorbitol (DBS – $CONHNH_2$, figure 4) pour former un hydrogel supramoléculaire. Une fois chargé

en principe actif, le gel permet la dissémination contrôlée du médicament de façon pH-dépendante. Trois anti-inflammatoires connus ont ainsi pu être vectorisés: l'Ibuprofen, le Naproxen, et la Mésalazine (figure 4). Grâce à cette dissémination sensible au pH, la Mésalazine - utilisée dans le traitement des maladies inflammatoires du colon - pourrait être délivrée préférentiellement au niveau des intestins, où le pH (environ 8) est voisin du pH de déstructuration du gel.

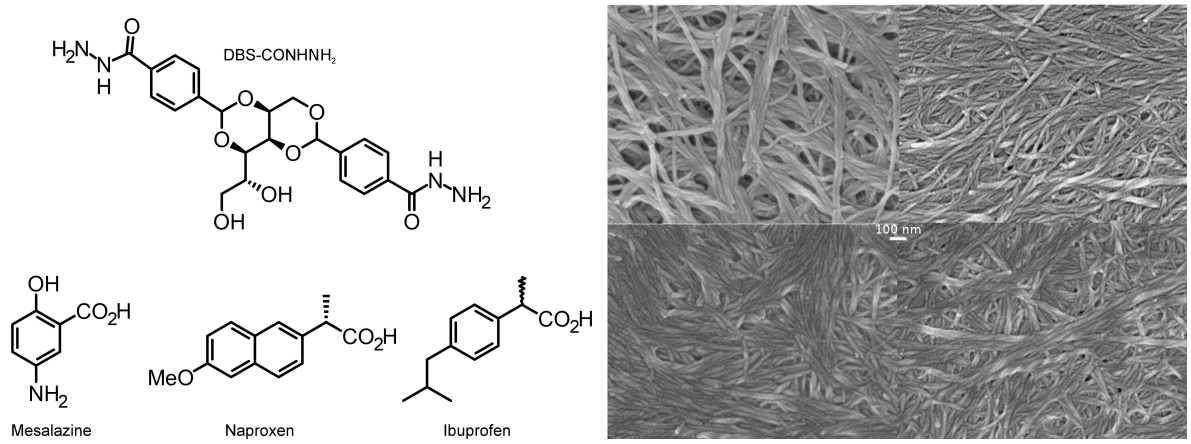


Figure 4: **Gauche**: structures chimiques de l'agent gelifiant DBS – $CONHNH_2$ et de trois anti-inflammatoires connus. **Droite**: images Scanning Electon Microscopy du gel sans additif (en haut à gauche), avec Ibuprofen (en haut à droite), avec Mésalazine (en bas à gauche) et avec Naproxen (en bas à droite). Figure adaptée de Howe *et al.* [11].

Outre les applications thérapeutiques, la reconnaissance moléculaire peut être utilisée à des fins de monitorage, ou *sensing* en anglais. Le senseur du glucose - que tout bon ouvrage traitant de senseurs se doit d'évoquer - en est un exemple célèbre (figure 5)[12]. Mis au point en 1962 par Clark et Lyons, il utilise l'enzyme *glucose oxydase* (GOx) comme récepteur, qui oxyde le glucose en acide gluconique et en peroxyde d'hydrogène (equation 1). En monitorant la quantité d'oxygène consommée grâce à une électrode à oxygène, il est possible de déterminer la glycémie d'un patient à partir de son sang total (cette technique a évolué depuis, et utilise maintenant une détection ampérométrique basée sur la libération de peroxyde d'hydrogène (équation 2). Pour plus de détails sur l'évolution du senseur du glucose, voir la revue de Wang [13]).

$$\text{glucose} + \text{oxygène} \xrightarrow{\text{GOx}} \text{acide gluconique} + \text{peroxyde d'hydrogène} \tag{1}$$

$$H_2O_2 \longrightarrow O_2 + 2\,H^+ + 2\,e^- \tag{2}$$

Aujourd'hui, ce cas d'école tend vers le système idéal: les volumes de sang nécessaires sont inférieurs au microlitre, la mesure est rapide (environ cinq secondes pour les appareils les plus

performants), l'appareillage se miniaturise et est accessible au plus grand nombre, facilitant ainsi la vie de nombreux patients atteints du diabète.

Figure 5: Appareillage typique utilisé pour mesurer le taux de glucose sanguin chez un patient. Source: Wikipédia, image reproduite sans modification. Image sous license Creative Commons.

Notons qu'en utilisant une enzyme naturelle comme récepteur, le senseur du glucose tombe dans la classe des *biosenseurs*, définis tels que[14]:

> **Biosensor:**
> *A device that uses specific biochemical reactions mediated by isolated enzymes, immunosystems, tissues, organelles or whole cells to detect chemical compounds usually by electrical, thermal or optical signals.*

La chimie supramoléculaire - depuis une vingtaine d'années environ - a su trouver sa place dans le domaine du sensing, et un nouveau domaine a vu le jour: la "chimie analytique supramoléculaire", nommée ainsi par le professeur Anslyn en 2007[15]. En pratique, ce domaine exploite des échanges dynamiques entre des structures synthétiques, résultant en la modulation d'un signal (optique, électrique, etc)[16]. L'Indicator Displacement Assay (IDA)[17] est un mécanisme qui répond à cette définition et qui permet la détection et la quantification d'un analyte cible (figure 6). En l'absence d'analyte (à gauche sur la figure), l'assemblage de l'indicateur et du récepteur résulte en un signal optique de base. La présence de l'analyte va entraîner le déplacement - l'échange - de l'indicateur, qui cédera sa place à l'analyte pour former un nouvel assemblage avec le récepteur. Cet échange entraîne une modulation du signal optique de base, qui permet à la fois la détection mais aussi la quantification de l'analyte.

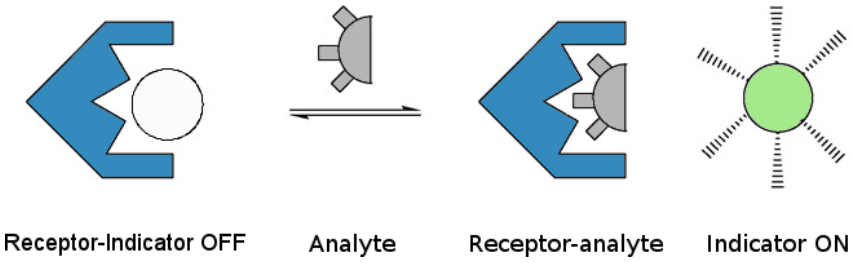


Figure 6: Le mécanisme IDA simplifié. Figure adaptée de Nguyen and Anslyn [17].

La figure 6 représente un cas de senseur dit *turn-on*, car l'interaction du senseur avec l'analyte entraîne une apparition/augmentation du signal optique. Ce type de senseur est largement préférable au type *turn-off*, pour lesquels la diminution du signal peut être causée par des interférents, ou bien par altération de l'indicateur. Zhong and Anslyn [18] ont néanmoins exploité ce concept pour créer un senseur turn-off de l'héparine, un anti-coagulant largement utilisé dans le domaine médical (figure 7).

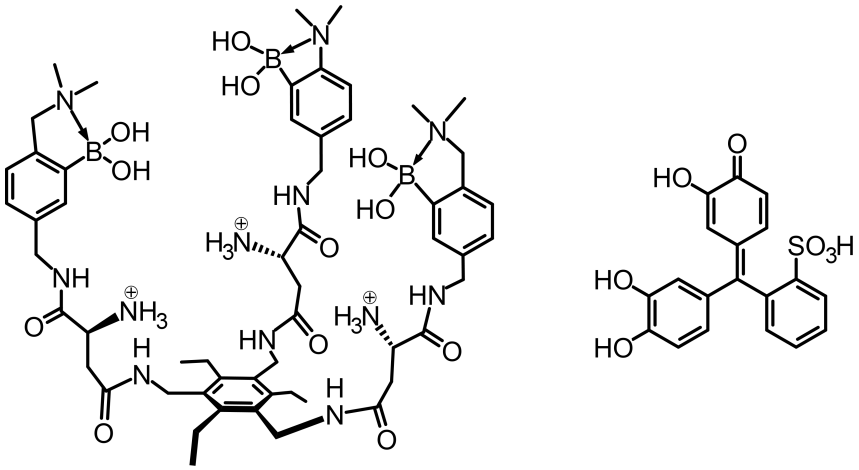


Figure 7: Récepteur (à gauche) et indicateur coloré (à droite, pyrochatécol violet) utilisés pour le senseur IDA à héparine de Zhong and Anslyn [18].

Bien évidemment, le concept de l'IDA repose sur l'affinité du récepteur pour l'analyte, qui doit être plus forte que celle du récepteur envers l'indicateur. De plus, de potentiels interférents ne doivent pas être capables de déplacer l'indicateur, car cela rendrait le senseur inutilisable. La difficulté réside donc dans la conception d'un système chimique respectant toutes ces conditions. Quand cela n'est pas possible, *plusieurs* systèmes moins spécifiques peuvent être utilisés pour l'identification d'un analyte. On parle alors d'un *sensor array* (ou barrette de senseurs en français, voir figure 8).

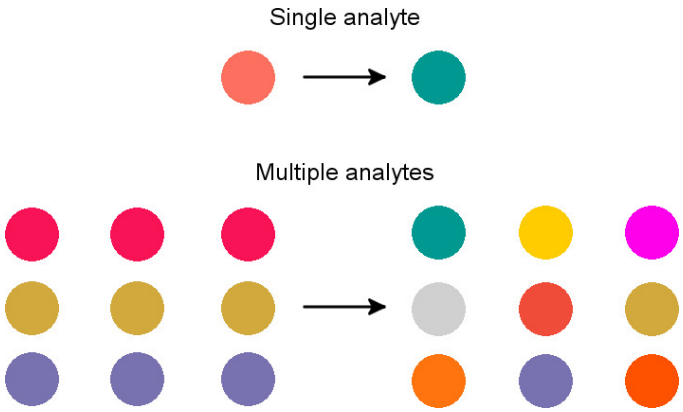


Figure 8: **Haut**: réponse unique d'un senseur spécifique à un analyte. **Bas**: réponse composite d'un sensor array pour trois analytes.

Un sensor array est constitué d'une combinaison de récepteurs et d'indicateurs différents. Pour chaque association récepteur/indicateur de l'array, un analyte donné va entraîner une réponse particulière. Toutes ces réponses vont constituer un motif particulier, propre à l'analyte (une sorte "d'empreinte digitale"). Les sensors arrays de Jagt *et al.*[19] et Elci *et al.*[20] (figures 9 et 10) sont par exemple capables de différencier plusieurs glycosaminoglycanes analogues à l'héparine. Leur utilisation en contrôle qualité pourrait faciliter la détection de lots d'héparine contaminés, et éviter des conséquences désastreuses chez certains patients[21]. Sur une note plus joyeuse, certains de ces sensor arrays sont aussi capables de différencier plusieurs variétés de vin, comme celui développé par Umali *et al.*[22].

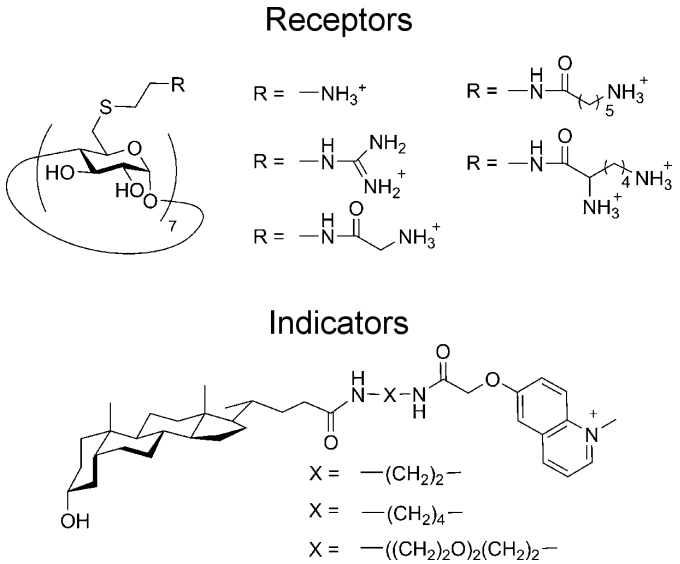


Figure 9: Sensor array de Jagt *et al.*[19], composé de 5 récepteurs et de 3 indicateurs. Figure adaptée de Jagt *et al.*[19].

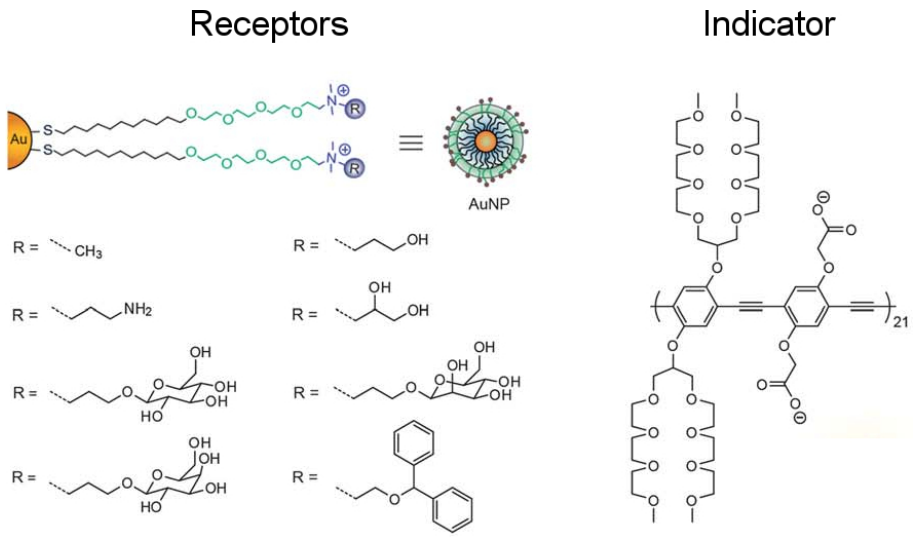


Figure 10: Sensor array de Elci *et al.* [20], composé de 8 récepteurs et d'un indicateur. Figure adaptée de Elci *et al.* [20].

Tout au long de ce manuscript, nous nous intéresserons en détail à la conception de senseurs utilisables pour la détection et la quantification de glycosaminoglycanes, et plus particulièrement de l'héparine. Fidèles au principe KISS, nous avons utilisé des dendrigrafts de poly-*L*-lysine (DGL)[23] (assimilables à des mimes de protéines) inventés dans notre laboratoire en 2006[24], et mis à notre disposition par l'entreprise Colcom (http://www.colcom.eu). Les dendrigrafts font partie des polymères dendritiques, qui sont divisés en trois catégories: les dendrimères, les polymères hyper-branchés, et les dendrigrafts (figure 11)[25–28].

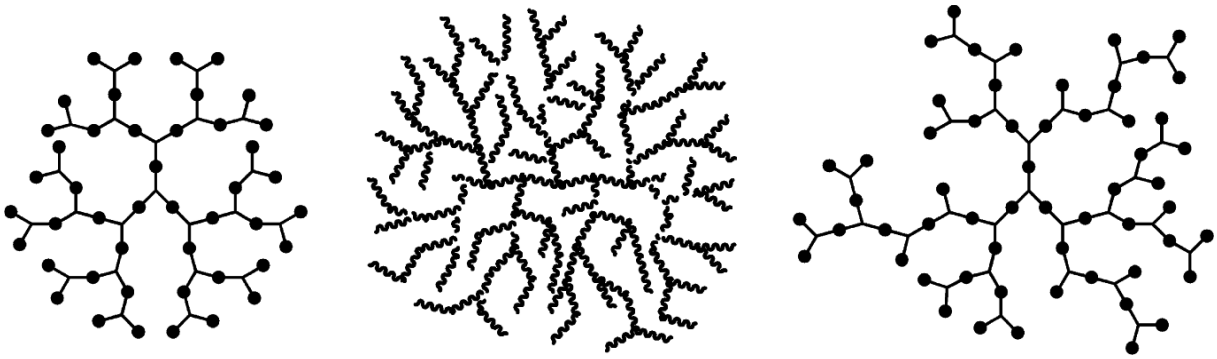


Figure 11: De gauche à droite, représentation schématique d'un dendrimère, d'un dendrigraft, et d'un polymère hyper-branché. Figure adaptée de Teertstra and Gauthier[28].

Les dendrimères (historiquement, la première catégorie des polymères dendritiques découverte) sont des objets moléculaires bien définis, obtenus après une synthèse dite *générationelle*. Cette synthèse implique des cycles de protection, condensation, et déprotection de blocs monomériques. Lorsqu'un monomère protégé est condensé avec un monomère dépro-

tégé, il ne peut plus réagir jusqu'à la prochaine déprotection. De fait, un dendrimère composé de n blocs monomériques ne peut avoir qu'une seule et unique structure. Les PAMAMS sont un exemple représentatif de la classe des dendrimères (pour plus de détails, voir la revue de Caminade *et al.* [25]). Les polymères hyper-branchés sont quant à eux structurellement à l'opposé des dendrimères. Il sont souvent obtenus par auto-condensation de monomères, en une synthèse dite *one-pot* (littéralement: un pot, ou synthèse en une seule étape). Leur structure est donc très peu contrôlée: pour un seul type de monomère avec deux points de branchement, il existe 2^{n-1} structures possibles pour un polymère hyper-branché constitué de n monomères (abstraction faite de toute opération de symétrie). Finalement, les dendrigrafts peuvent être qualifiés d'intermédiaires entre les deux catégories précédemment citées. Dans le cas des DGLs, la synthèse est générationelle, mais chaque nouveau monomère introduit n'est pas totalement protégé (seule la chaîne latérale de la lysine est protégée, voir figure 12). Ce bloc peut donc encore réagir grâce à son amine α, pour se condenser à un autre monomère, qui pourra lui même répéter le processus. Pour un DGL de génération 2 composé de 48 résidus lysine, 107.474.640 (soit environ 10^8) structures sont possibles. Un polymère hyper-branché de 48 monomères aurait quant à lui environ 10^{14} structures possibles. La protection de la chaîne latérale de la lysine introduit donc un biais, limitant le nombre de structures possibles pour les DGLs, et les empêchant de tomber dans la catégorie des polymères hyper-branchés. On peut dire qu'ils sont seulement super-branchés ! (pour plus de détails sur la structure des DGLs, voir la partie V).

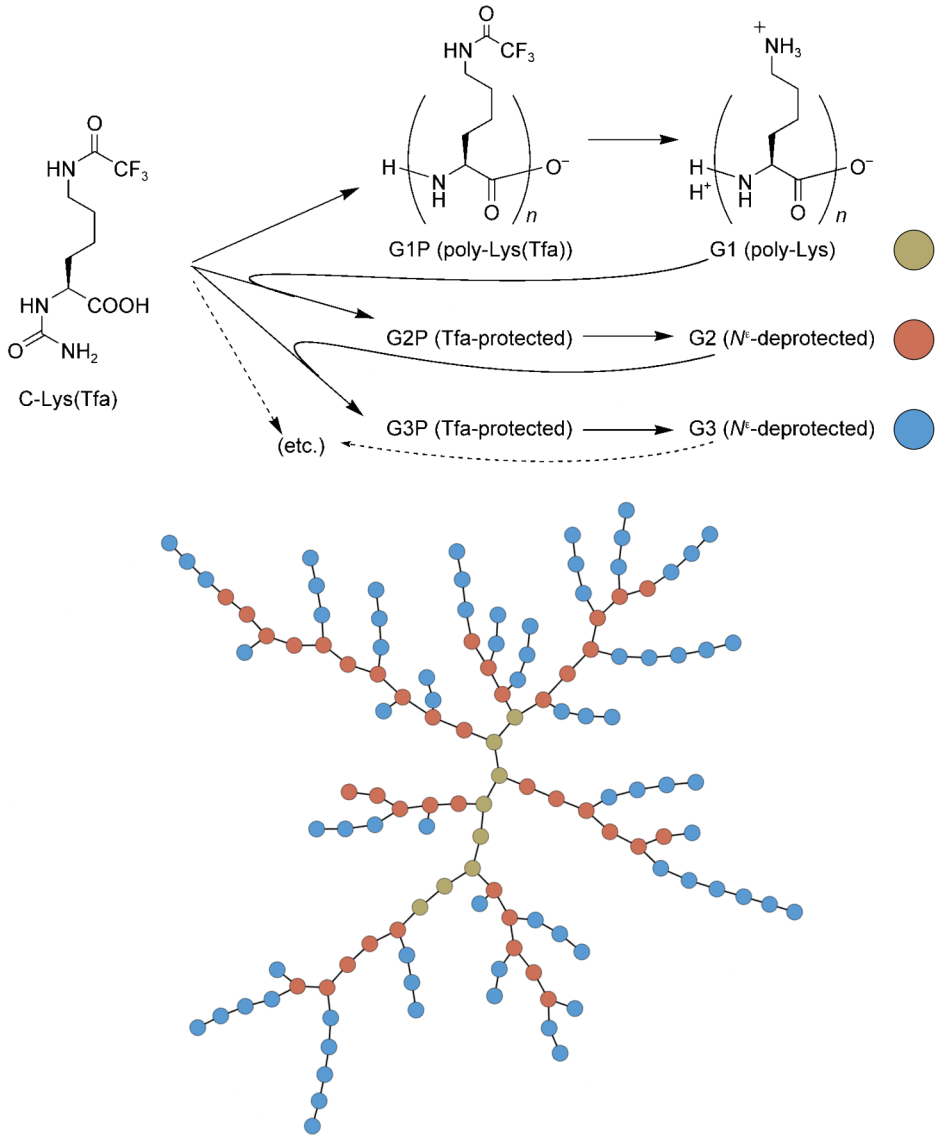


Figure 12: Schéma de la synthèse générationelle des DGLs. Figure adaptée de Collet *et al.* [23].

Les propriétés remarquables des DGLs, combinées à de courts peptides fluorescents (facilement accessibles par synthèse sur phase solide, philosophie KISS), nous ont permis de mettre au point le premier biosenseur fluorescent capable de quantifier l'héparine dans du sang humain total (voir partie III). Grâce à de très simples variations de nos systèmes, nous avons su transformer ce simple senseur en un *sensor array* qui permet l'identification formelle de cinq glycosaminoglycanes de structures analogues (voir partie IV). Nous finissons avec la partie V qui - grâce des techniques de mécanique moléculaire innovantes - apporte de nouvelles données concernant la structure et le comportement des DGLs, au niveau atomique.

Dans un tout autre registre, nous aborderons au travers de la partie II l'utilisation d'outils informatiques modernes pour réaliser une veille bibliographique efficace. Plus généralement, les résultats décrits ci-après n'auraient probablement pas pu être obtenus sans un usage intensif

de l'informatique et de la programmation, et la communauté de la chimie gagnerait à intégrer certains concepts et outils de l'open-source.

Commençons par le cahier de laboratoire, toujours très matérialisé en 2016. Il est évident que tout travail de recherche doit être documenté et décrit quotidiennement, mais il est curieux que les chimistes tiennent à garder leurs notes sur support physique. Aujourd'hui, les avantages de la dématérialisation compensent largement ses (potentiels) inconvénients. Le premier argument que le cahier de laboratoire électronique - Electronic Lab Notebook en anglais (ELN) - se voit opposé qu'il est incapable d'assurer l'authenticité des résultats qu'il contient. Cet argument est faux car, grâce à l'avènement des techniques de chiffrement, un document signé électroniquement possède une valeur légale bien supérieure à quelques notes écrites sur un cahier physique. Plusieurs logiciels de ELN proposent la fonctionnalité de *trusted timestamping* (horodatage certifié en français). Le principe est simple: à partir des données originales, un *hash* est généré (un hash est une empreinte digitale numérique, une chaîne de bits pratiquement impossible à reproduire à partir d'autres données). L'altération des données originales résultera en un hash complètement différent. Ce hash est ensuite envoyé à une tierce partie faisant office d'autorité (Time Stamping Authority, TSA), comme l'Agence Nationale de la Sécurité des Systèmes d'information (ANSSI), une organisation nationale française (la procédure peut être répétée auprès de plusieurs TSA pour augmenter la fiabilité du processus). Le TSA va ensuite concaténer le hash à un timestamp (un timestamp est un nombre entier indiquant le nombre de secondes écoulées depuis le premier janvier 1970, il représente tout simplement une date à la seconde près), et calculer le hash de cette concaténation. Finalement, ce dernier hash sera signé avec la clé privée du TSA et renvoyé au propriétaire original des données, avec le timestamp datant la procédure (voir figure 13). Évidemment, le processus repose sur la crédibilité des TSA, et sur leur capacité à garder secrète leur clé privée, mais cette procédure d'authentification reste sûre, et est valable légalement.

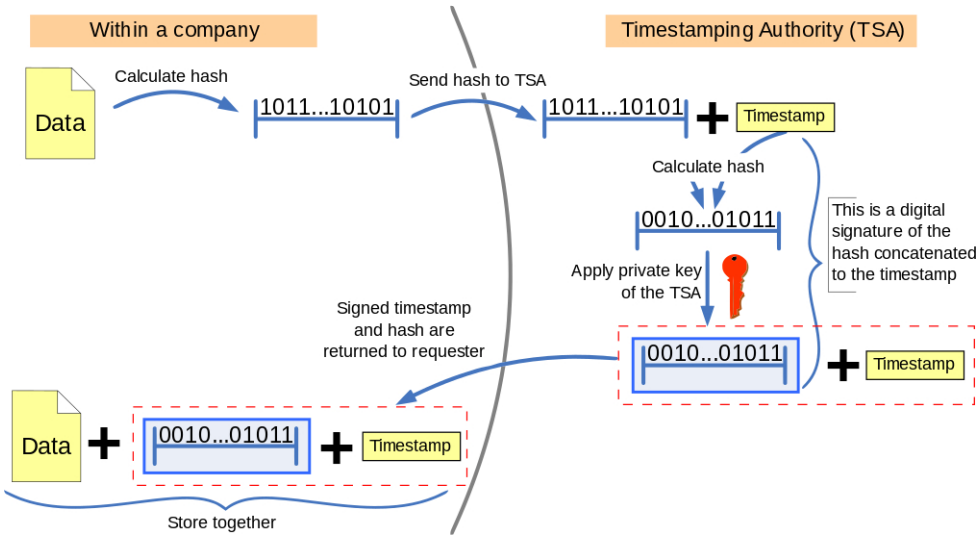


Figure 13: Schéma détaillant la création d'un horodatage certifié. Source: Wikipédia, image reproduite sans modification. Image sous license Creative Commons.

L'argument de l'authentification écarté, intéressons nous aux avantages pratiques d'un cahier de laboratoire électronique:

- Toute sorte de fichiers peuvent être rattachés à une expérience: photos, spectres RMN, classeurs Excel, etc, et tout est stocké au même endroit. Cela diminue les risques de perte et de confusion dans les données
- Des fonctions de recherche avancée peuvent être utilisées. Il est par exemple très simple de rechercher un mot-clé parmi les données accumulées tout au long d'un doctorat. La même chose serait plus compliquée avec des cahiers de laboratoire physiques
- La documentation des résultats est clairement séparée de l'expérimentation. Aucun cahier (souvent pollué par des produits chimiques) ne fait la navette entre la paillasse et le bureau de l'expérimentateur
- La sauvegarde des données est très simple et peut être automatisée. Il n'y a plus de risque de perdre plusieurs mois de recherche si un cahier de laboratoire physique est perdu ou détruit
- Selon le serveur utilisé (si l'ELN est hébergé sur un serveur) et le degré de sécurité désiré, l'ELN peut être accessible depuis n'importe où, et par plusieurs personnes. Un chef d'équipe peut ainsi suivre l'avancement des travaux de ses collaborateurs en temps réel

Bien sûr, comme évoqué plus haut, la sécurité des données est un point qui doit être analysé avec précaution, surtout si l'ELN est voué à être hébergé sur un serveur. Idéalement, aucune entreprise tierce ne doit pouvoir accéder aux données d'une équipe de recherche. Il est donc

impératif d'avoir la maîtrise de ce serveur. De même, aucun logiciel d'ELN qui ne serait pas open-source ne peut être utilisé.

Pour répondre à ce cahier des charges, notre choix s'est porté sur le Raspberry Pi (https://www.raspberrypi.org/, voir figure 14), un ordinateur coûtant approximativement 30€, pas plus gros qu'une carte de crédit, que nous utilisons en tant que serveur. Accessible seulement sur le réseau local de notre laboratoire, il est à l'abri de la plupart des attaques informatiques extérieures, ne prend que très peu de place, et ne consomme que 3W environ. Nous avons ensuite installé le logiciel open-source eLabFTW (http://www.elabftw.net/) sur le Raspberry Pi, qui présente tous les avantages d'un ELN listés ci-dessus, et plus encore. Déjà adopté par l'institut Curie et d'autres groupes, ce logiciel est simple d'utilisation et adapté au chercheur. Nous l'utilisons quotidiennement depuis deux ans sans avoir jamais rencontré de problèmes. De plus, le code source de eLabFTW étant accessible sur GitHub (https://github.com/elabftw/elabftw), tout chercheur peut contribuer au projet, et demander de nouvelles fonctionnalités (nous avons effectivement fait ajouter un éditeur de molécule à eLabFTW).

Figure 14: *Gauche:* un Raspberry Pi assemblé et branché. *Droite:* un Raspberry Pi, son boîtier et sa carte SD.

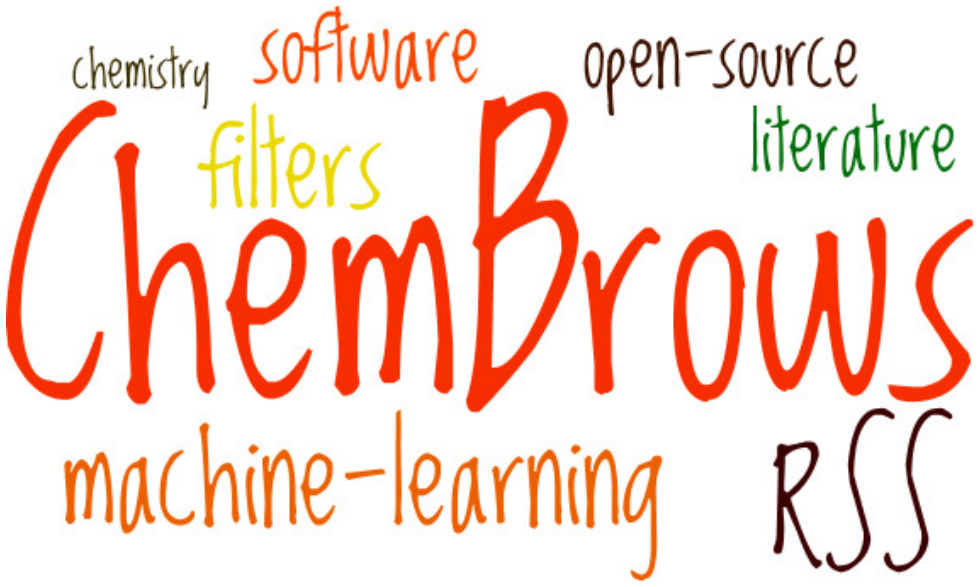
Nous pensons que la philosophie de l'open-source peut beaucoup apporter à la recherche en chimie. Il est par exemple facile de se former à la programmation grâce à de nombreux tutoriels disponibles gratuitement en ligne. Alliées à des projets open-source comme Numpy[29] et scikit-learn[30], quelques compétences en programmation peuvent faire économiser des sommes d'argent substantielles à une équipe de recherche, car l'achat de licenses pour certains logiciels propriétaires n'est plus nécessaire. De plus, en utilisant et en contribuant à des logiciels open-source, de nouvelles connaissances et outils sont mis à la disposition du plus grand nombre.

Fidèles à cette approche, nous avons tenté de répondre à un besoin non satisfait en recherche: comment rester "à jour" de la bibliographie ? Partants du constat que chaque jour plus de 600

nouveaux articles sont publiés en chimie, nous avons décidé de mettre au point un outil simple (philosophie KISS) qui se chargerait de sélectionner les articles les plus pertinents pour chaque chercheur, selon ses préférences: ChemBrows était né.

Part II

ChemBrows: An Open-Source Application Software To Keep Up to Date with the Current Literature



Résumé en français

Fonctionnant comme un lecteur de flux RSS amélioré intégrant des filtres par mots-clés et un moteur de recommandation basé sur l'apprentissage machine, ChemBrows est un logiciel qui assiste les scientifiques/enseignants/étudiants dans leur veille bibliographique. Développé en Python, ce logiciel est open-source, libre, gratuit, et est disponible sur toutes les plateformes. ChemBrows présente de manière agréable et centralisée les abstracts et graphical abstracts des articles récemment publiés en chimie, et permet surtout le tri des articles par "hot paperness", un score caractérisant la pertinence de chaque article déterminé par apprentissage machine. Un tri des articles par date est aussi possible, permettant le parcours exhaustif de la littérature récente.

1 Introduction

To stay up-to-date with the latest research results as a scientist/teacher/student has clearly become a burden. According to Thomson Reuters Web of Science, the number of publications increased linearly over the last two decades, with in 2014 a stunning figure of +600 papers published every single day in the sole field of chemistry (figure 15. Data were generated from http://apps.webofknowledge.com. Typical search input: databases = all, year published = 2014, refined by research areas = chemistry). Following this growing rate, the milestone of a quarter million publications per year should be reached this year.

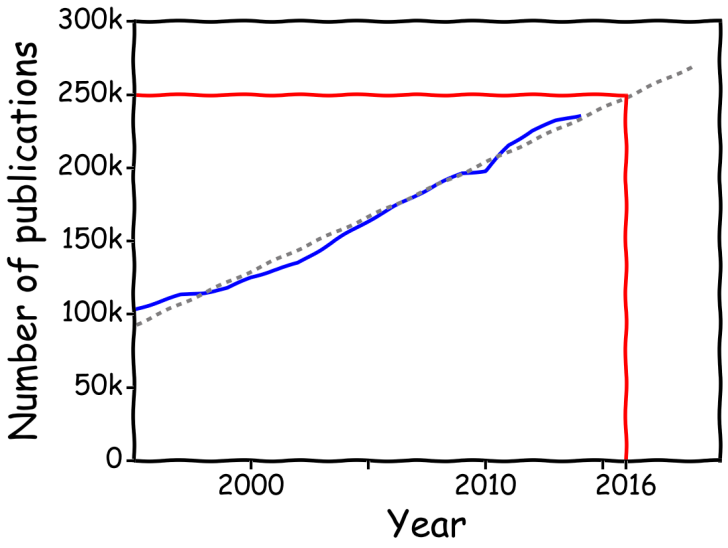


Figure 15: Number of publications in the chemical sciences over the last two decades.

Multiple factors may contribute to this rise in research output, which includes the growth of universities internationally, the emergence of new areas of research, and the infamous pressure to "publish or perish". Though the two former factors are good news, the latter one cer-

tainly participates to the avalanche of tedious – even fraudulent – papers in literature[31]. As a consequence, finding new, relevant and inspiring research works is looking for a needle in a haystack.

2 Common methodologies toward the survey of the literature

Various methodologies can be applied to tackle this issue. Among them, the old-fashioned literature routine involves the regular browsing of publishers' Web sites or RSS feeds to pick out fresh papers of interest. Although this methodology is nowadays time-consuming in regard to the flood of publications, it still has the advantage to be the most comprehensive one. In addition, this routine leads naturally to the browsing of papers from various research fields, which may stimulate unexpected connections and bracing ideas. Less fastidious solutions are available for literature surveying.

By using keywords-based alerts as provided for instance by PubMed, it is possible to significantly reduce the input of papers. Such an approach is almost mandatory in order to efficiently follow its own topic(s) of interest[32]. On the other hand, a survey based on keywords potentially introduces blinkers to what is scientifically possible.

More recently, literature-recommendation services have emerged (e.g. ReadCube and Mendeley, to name but two), relying on algorithm-based engines that learn from their users' interests to suggest relevant contents[33]. These tools often provide with unnoticed and pertinent research articles, but – as a balance – also narrows the flood of literature to a trickle. To summarize, these previous methodologies toward the survey of literature come with a distinct time-saving/comprehensiveness ratio, each methodology having its own advantages and being indeed complementary to the others (figure 16).

Herein, we introduce ChemBrows, an in-house piece of software that proposes an integrated solution to stay up-to-date with the literature.

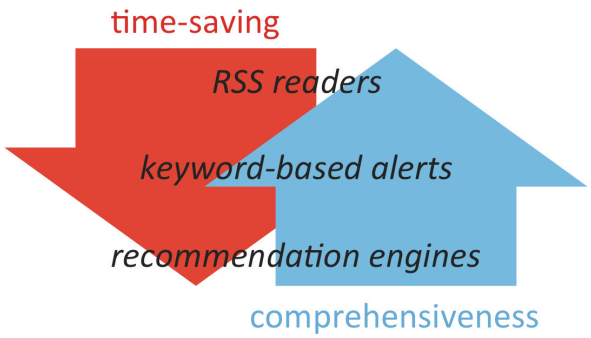


Figure 16: Available approaches toward the survey of the literature and their time-saving/comprehensiveness trends.

3 An all-in-one solution with ChemBrows

Working primary as a RSS reader, ChemBrows feeds the user with the latest papers from a customizable selection of journals (figure 17).

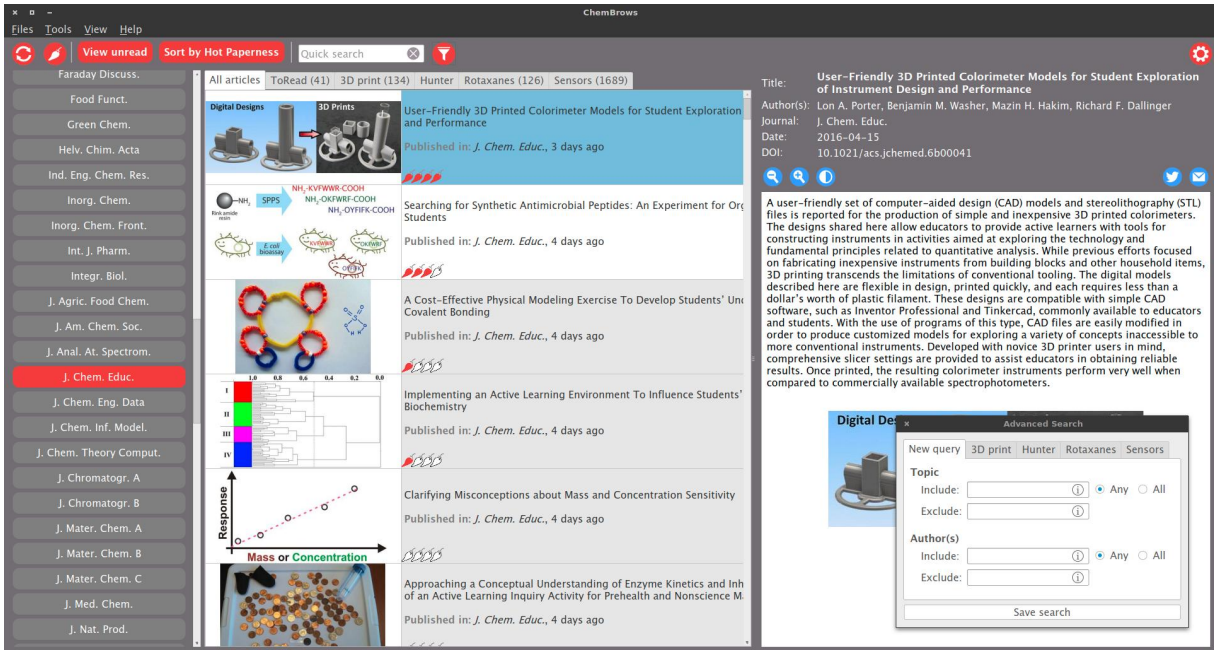


Figure 17: The interface of ChemBrows.

A catalogue of 122 journals is included so far, but virtually any title can be further added, including from other disciplines (new titles will be added upon requests from users). A dock appears upon mouse hovering on the left-side of the interface, allowing to display/hide titles (figure 17, *left side of the screenshot*). Because the browsing of papers is significantly facilitated by the presence of graphics, ChemBrows proposes – unlike classical readers – the graphical abstract embedded within the entries (figure 17, *middle of the screenshot*). Once an entry is selected, the corresponding untruncated abstract is displayed (figure 17, *right side of the screenshot*), and double-clicking on an entry opens the corresponding article's landing page in the user's default web browser. Users can also bookmark entries that will be saved in a separate "ToRead" tab.

In addition to providing scientists/teachers/students with the latest literature from a personal selection of journals, ChemBrows allows the creation of savable keyword-based filters in order to follow favorite topics and/or authors (figure 17, *dialogue box in the bottom-right corner of the screenshot*). It is obviously possible to perform multiple keywords searches that involve Boolean operators.

Because it also includes a machine-learning algorithm, ChemBrows can classify the entries by "hot paperness" (figure 17, *hot peppers in the screenshot*), which is a sorting based on a support-vector network[34]. To put it simply, when the read entries are – or are not – liked, the

algorithm extracts the words from the abstracts and gives them a score. Further incoming abstracts are evaluated with respect to its words' score, and are subsequently labeled with one to four "hot peppers" as a prediction of its significance for the user. Obviously, the more the user feeds ChemBrows, the more ChemBrows feeds back the user with pertinent articles.
The program is written in Python 3 and can be run on multiple platforms (*i.e.*, Linux, Mac OS X, and Windows). Information is parsed from the free content of the publishers' Web sites (*i.e.*, RSS feeds, and possibly landing pages if graphical abstracts and/or full abstracts are missing) and then stored locally in a SQLite database, which can be exported or imported from one computer to another. Upon the first boot, the software requires an e-mail address – the only personal information collected, that is only used to deliver important news/update notifications. The graphical user interface was designed to be intuitive, a short embedded tutorial allowing the easy mastering of every ChemBrows' features.

4 Algorithms

4.1 Fetching the data

The process of gathering the data is based on the RSS (Rich Site Summary) feeds provided on the publishers web sites. RSS feeds are simple web pages under the XML (Extensible Markup Language) format, which is a human and machine readable markup language. Each journal has its own RSS feed, which contains the data of the latest published articles (abstract, graphical abstract, DOI, authors' names, etc). The number of articles on a RSS page as well as its update frequency depends on the journal policy. Usually, such pages contain 20 to 100 articles, and are refreshed every week. However, publishers often do not provide complete RSS feeds, and the data can be truncated (missing graphical abstracts, incomplete abstracts, etc), which makes traditional RSS readers not fully efficient for staying up-to-date with the latest literature.
Addressing this issue is the main purpose of ChemBrows, and led us to a modular programming strategy. As publishers tend to adopt the same data structure for all of their journals (*e.g.* all of the ACS journals' RSS feeds have the same structure), specific treatments are applied to the articles with respect to its publisher. For example, the publisher ACS provides RSS feeds with a graphical abstract for each article, but the abstracts are missing. In this case, ChemBrows fetches and parses the abstracts directly from the web pages of the articles (figure 18). To allow a fast data fetching, ChemBrows starts a pool of n concurrent threads (n is the number of cores of the computer's processor), and each thread can start up to 40 asynchronous requests toward publishers' servers. For a typical first use of ChemBrows, it takes approximately 30 minutes to gather 6000 entries. When all the required data for an article have been acquired, ChemBrows writes them in a SQLite database. As the data fetching process is not limited by the Central Processing Unit (CPU), its execution time mainly depends on both the response time of the servers and the network speed. Also, ChemBrows will request the web page of an article only

if:

- The article is in the RSS feed
- The article (and its data) is not already in the user's database
- The article's data are missing from the RSS feed

These conditions drastically restrain the number of requests on the publisers' servers, and the bandwidth consumption should be inferior to a visit to the journals' Web sites through a common web browser.

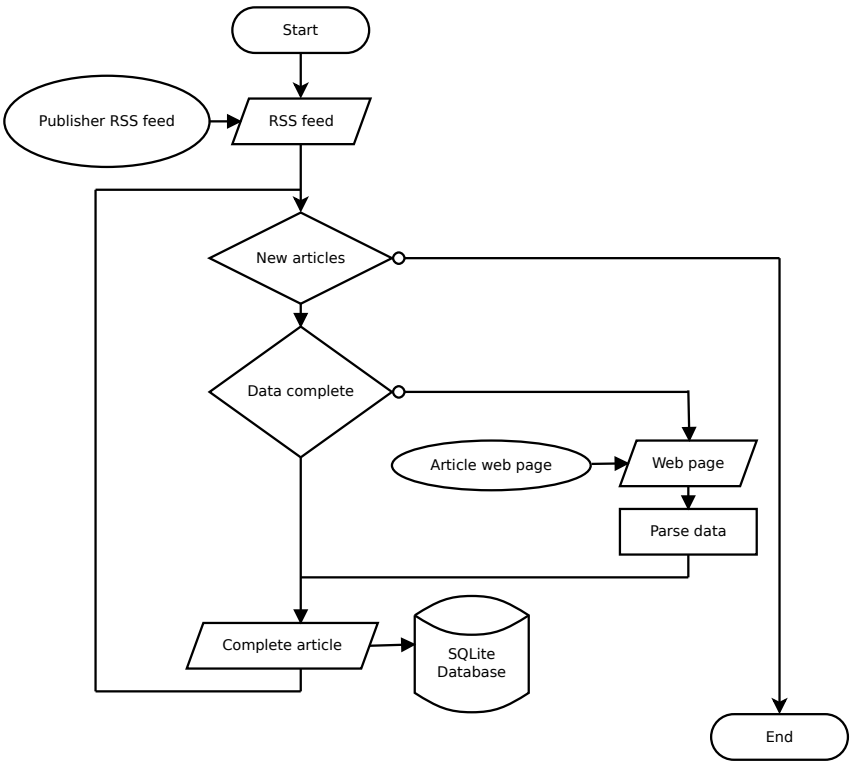


Figure 18: Flowchart describing the fetching of the data.

The per-publisher approach of ChemBrows, instead of a per-journal approach, allows the fast and easy extension of the list of the supported journals. To date (December 1, 2016), the following publishers are supported: ACS, Beilstein-Institut, Elsevier, NAS, NPG, PLOS, RSC, AAAS, Springer, Thieme, Wiley (see the list of abbreviations for acronyms details). A default list of 122 titles is shipped with ChemBrows, but any journal from any of these publishers can be easily added by modifying a configuration file. For an example of a configuration file, see https://github.com/chembrows/ChemBrows/blob/master/journals/acs.ini

4.2 Calculating the hot paperness

ChemBrows uses a score - the so called *hot paperness* - to sort the articles by relevance. The hot paperness of each article is user-specific, as it is calculated on the basis of the articles the

user liked (or not). To obtain these scores and identify relevant articles, ChemBrows first needs to build and train a classifier (figure 19). A classifier is an object which - once trained - will be able to identify "ham" (relevant articles to the user), or "spam" (irrelevant articles to the user). Nowadays, modern webmails providers (*i.e.* Google) use classifiers as email spam filters. While several algorithms exist to predict the category of an input (ham or spam), the procedure to build and use a classifier remains a constant: gather user input →train the classifier →predict the category of new inputs.

In the case of ChemBrows, articles' abstracts and titles are used as input. If the user liked an article, the article is identified as ham. If the article is in the read state but not liked, it is identified as spam. These states are stored in the database, for each article. Then, ChemBrows uses a "bag of words" procedure and builds two bags: one containing the words of the abstracts and titles liked, and one containing the words of the abstracts and titles not liked. Note that these words are free from stop-words (*e.g* words like 'the', 'of', 'a', etc, were removed from the bags). For each word in each bag, the TF-IDF (Term Frequency–Inverse Document Frequency)[35] is calculated. The TF-IDF is the weighted term frequency of a word. The goal of using TF-IDF instead of the raw frequencies of occurrence of a word is to scale down the impact of words that occur very frequently in a given corpus and that are hence less informative than words that occur in a small fraction of the training corpus.

Finally, a LinearSVC (Linear Support Vector Classification) algorithm will receive the two bags of TF-IDFs as an input and yield a trained classifier. This classifier belongs to the class of Support Vector Network classification algorithms[34]. Initially, a Naive Bayes classifier was used, but with a growing input from the user (*i.e.* a growing number of liked/not liked articles), the LinearSVC was found to be more efficient, and provided us with more relevant articles (the relevance improvement was not quantified. Our choice was based on a simple poll of beta testers). It is not the scope of this document to describe how these algorithms work, but additional information and parameters of the algorithms can be found elsewhere[30]. Moreover, the source code of the hot paperness' calculation source code is publicly available: https://github.com/chembrows/ChemBrows/blob/master/predictor.py.

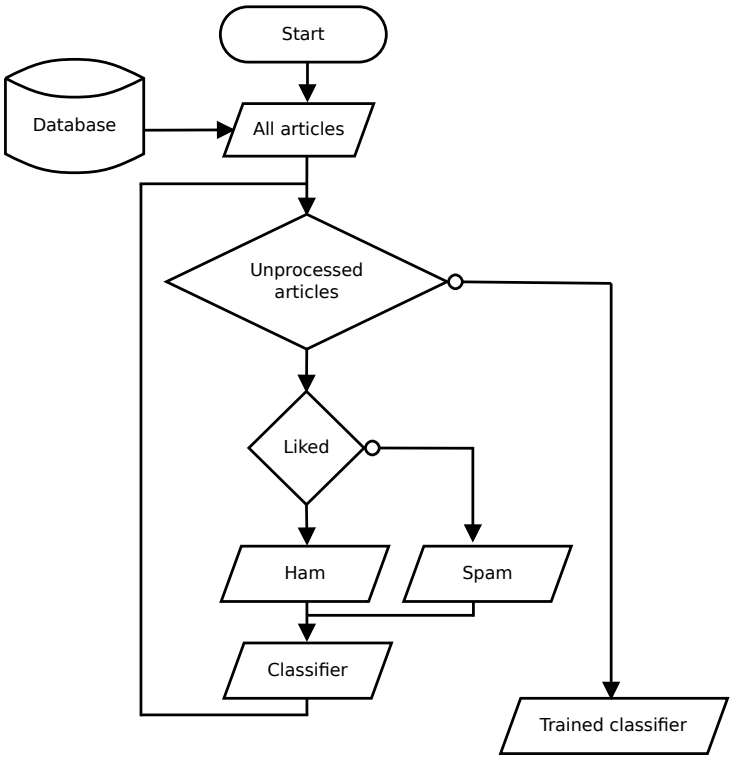


Figure 19: Flowchart describing the training of the classifier.

Once the classifier is trained, each abstract of each article in the database will be analyzed by the classifier to obtain the probability of the abstract to be in the ham category (P(ham), see figure 20). However, while we were testing our software, we noticed that these values are absolute, and not usable without post-processing. Depending on what the beta testers had liked, the most pertinent article could have a P(ham) as low as 1%. We thus decided to scale all the P(ham) values: the most pertinent article (*i.e.* the one with the highest P(ham)) would have a score of 100%, and the others P(ham) would be scaled, based on this maximum. The scaled P(ham) is what we called hot paperness. The hot paperness ranges from 0 to 100, and is visually represented with one to four hot peppers (see figure 17).

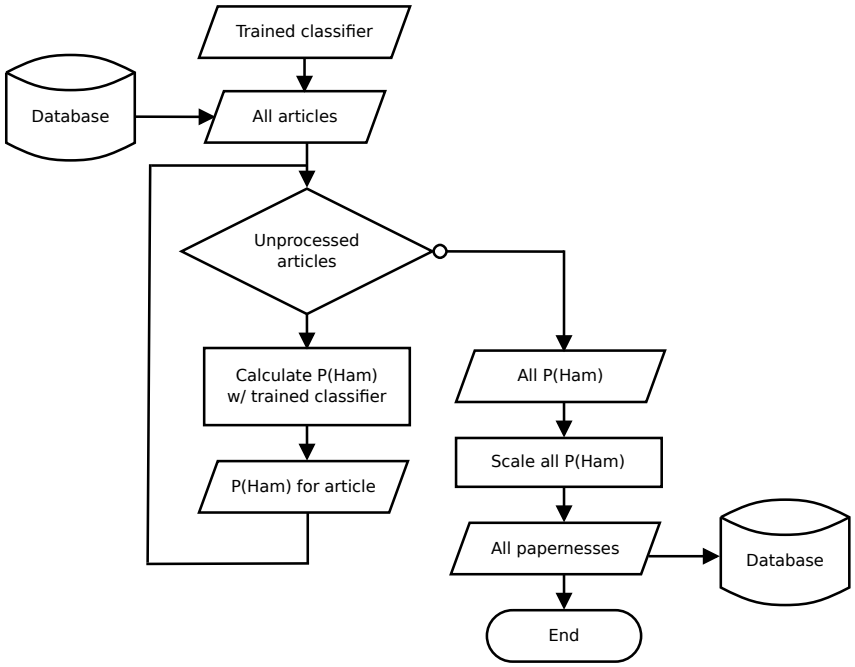


Figure 20: Flowchart describing the calculation of the paperness.

5 Download, Installation and Contributions

The installation files of ChemBrows are available at http://www.chembrows.com. After downloading the program, it may be installed on Windows by double-clicking on the executable file (setup ChemBrows 0.9.6 (32bit).exe). Mac OS X and Linux users should unzip the archive files (tar.bz2), and then execute the package file (ChemBrows.pkg) or launch the executable file (gui), respectively. Once installed, ChemBrows has the capability to update itself during the boot process if a new version is available, thanks to the Esky Python module (see: https://github.com/cloudmatrix/esky). During the development of ChemBrows, we contributed to Esky's code by submitting several pull requests (a pull request is a contribution to a source code repository that uses a version control system, Git in this case: https://git-scm.com. When one submits a pull request to a project, the maintainer of the repository has to "pull" the changes and merge them in order to make the changes effective). The source code of ChemBrows is available at https://www.github.com/chembrows/ChemBrows and is distributed under a GNU General Public License (as specified by the Free Software Foundation, see: http://www.gnu.org/licenses/gpl-3.0.en.html).

6 Conclusion

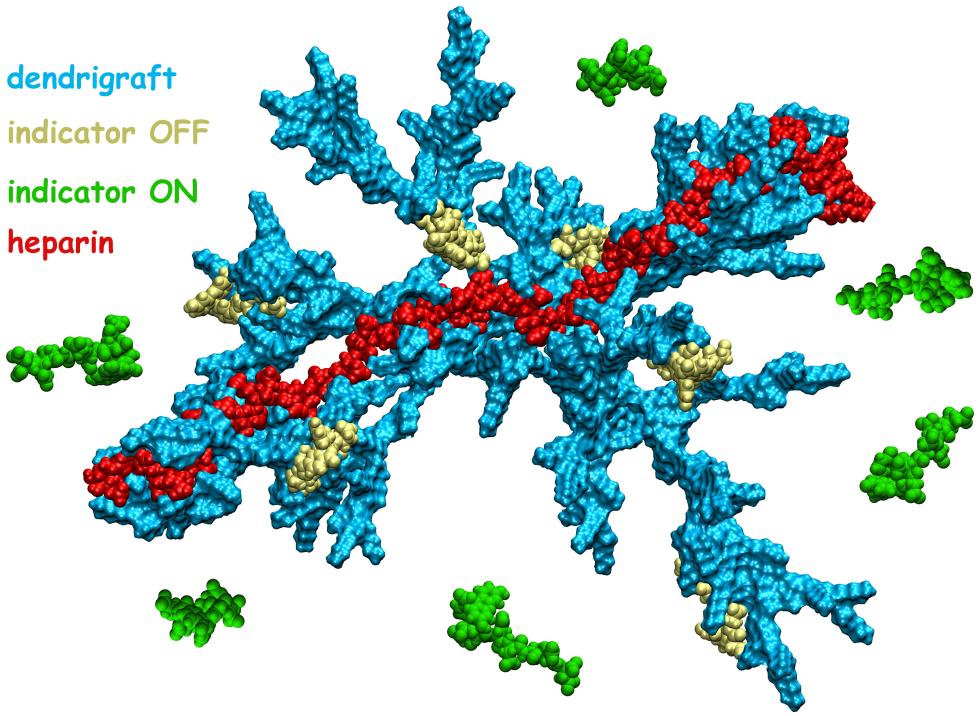
All-in-one RSS reader, keyword-based filter and recommendation engine, ChemBrows offers to the user an unprecedented versatility toward the survey of the literature. We would like to

emphasize that Chembrows is an open-source software that is dedicated to the community of chemists and has noncommercial purpose. Because there are certainly multiple ways to improve upon this software, the source code of ChemBrows is fully available and contributions – including forks toward other disciplines – are enthusiastically welcome.

At the date of December 1, 2016 (211 days after our work was published[36]), more than one hundred users installed ChemBrows, including 9 claiming to be Obi Wan Kenobi. We are currently communicating about ChemBrows on Twitter (@ChemBrows), hoping to reach as many chemists as possible. Numerous improvements are planned, like - as requested by our users - a wizard to add journals that are not included in the original selection of ChemBrows.

Part III

Monitoring clinical levels of heparin in human blood samples with an indicator-displacement assay



Résumé en français

L'héparine est un glycosaminoglycane très largement utilisé comme anticoagulant, principalement au cours d'actes chirurgicaux. Il est important de pouvoir quantifier précisément l'héparinémie des patients pour limiter les risques d'hémorragie. Dans cette partie, nous rapportons qu'un dendrigraft de poly-*L*-Lysine (DGL) est capable de former un complexe multiligand avec un peptide fluorescent, conduisant à la disparition quasi-totale du signal optique qui peut ensuite être restauré en présence d'héparine. Ce système simple permet, pour la première fois, la détection par fluorescence de l'anticoagulant dans le sang humain à des doses cliniques. Nous avons aussi déterminé les paramètres thermodynamiques du phénomène de l'association entre l'indicateur et le récepteur, révélant ainsi son caractère anti-coopératif.

7 Introduction

Heparin is a polysulfated glycosaminoglycan (GAG) that is extensively used as an intravenous anticoagulant, with for example approximately 12 million Americans treated with this molecule each year (figure 21)[37].

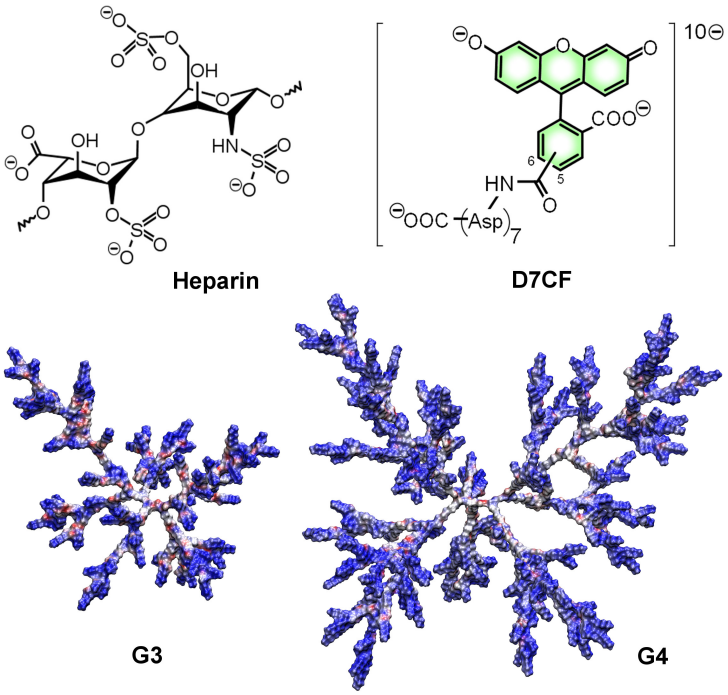


Figure 21: Top left: major repeat unit of heparin. The average molecular weight of most commercial heparin preparations is in the range of 12 to 15 kDa. Top right: bioconjugate between a homopolymer of aspartic acid and 5(6)-carboxyfluorescein, D7CF. Bottom: minimized structures with electrostatic potential surfaces of third- and fourth-generation dendri-graft poly-*L*-lysine (DGL) polymers G3 and G4. For calculation details, see section 16.

It is highly desirable to quantify heparin – administrated at dosing levels ranging from 2–8

USP.mL^{-1} (15–58 µM) during surgery to 0.2–1.2 USP.mL^{-1} (1.5–9 µM) during prophylaxis – in order to prevent thrombophilia or bleeding disorders[38]. In clinical settings, heparin quantification relies mainly on clotting time-based assays[39]. However, these methods are not fully satisfactory due to their inaccuracy and poor implementability at point-of-care[40,41].

To address this issue, the use of fluorescent/colorimetric sensors based on synthetic heparin binders has emerged over the last decade as an interesting alternative for its monitoring[42–51]. Still, only detection and quantification of the anticoagulant in serum was reported in previous studies, imposing therefore additional centrifugation steps from whole blood samples. In order for the sensors to be able to operate in such a competitive and complex medium (*i.e.*, blood), highly efficient and selective binders for heparin are required.

Recently, Smith *et al.* have revealed the ability of cationic dendritic structures to mimic protamine, a heparin neutralizing protein, by binding the polyanionic anticoagulant in a multivalent manner[52–55].

8 Results and discussion

We thus decided to design new fluorescent heparin sensors based on dendrigraft poly-*L*-lysine polymers (DGLs, figure 21)[23]. The advantages of these architectures are numerous: (i) straightforward and green synthesis on multi-gram scale, (ii) strong positive electrostatic potential (at pH 7.8, we made the approximation that only the ε-amine residues of DGLs are protonated, giving a positive net charge of 92 and 280 for G3 and G4, respectively[56]) for interactions with anionic species, (iii) conformational flexibility, allowing the structure to organize its total charge for binding, and (iv) chemical flexibility for further functionalization[57,58] (*e.g.*, immobilization on surfaces).

We envisioned that polycationic DGLs should be able to bind several units of a same anionic fluorescein derivative by electrostatic interactions (although cation-π interactions between the ammonium functions of the polymers and the aromatic core of the dyes cannot be excluded to be part of the binding event[59]), leading eventually to the extinction of the optical signal[60–62]. Indeed, because their emission spectrum (λ_{em} = 535 nm) sufficiently overlaps witht heir absorption spectrum (λ_{ex} = 485 nm), fluorescein derivatives have a high potential for proximity-dependent fluorescence self-quenching that can mainly be attributed to the resonance energy transfer between close fluorescein molecules (the mechanism of self-quenching is still not entirely understood and appears to be a combination of energy transfer to non-fluorescent dimers, dimerization of the dye and collisional quenching interactions between dye monomers[63,64]). Then, removal of the indicator from the host with the highly negatively charged heparin molecule should lead to the restoration of fluorescence, in a typical IDA[17] (figure 22).

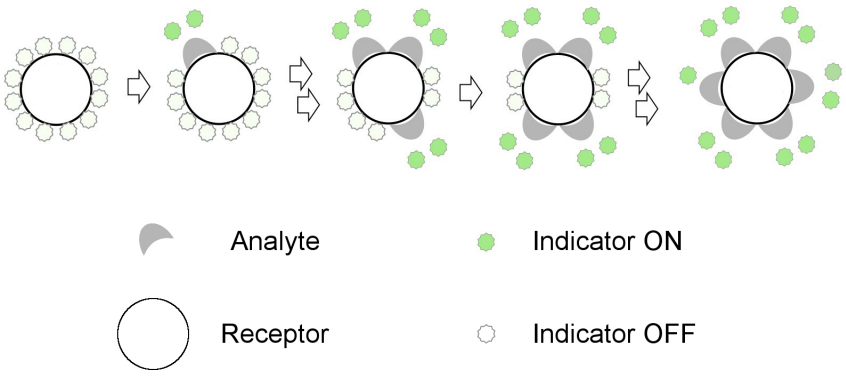


Figure 22: Removal of the indicator (D7CF) from the host (DGL G4) with the analyte (heparin), typical IDA.

8.1 Titration experiments

Upon addition of the simple 5(6)-carboxyfluorescein (CF) to the third-generation dendri-graft poly-*L*-lysine polymer (G3) in aqueous solutions containing 10 mM HEPES and buffered to pH 7.8, only a moderate quenching of the fluorescent signal occurs (figure 23, blue curve, as compared to CF alone, orange curve). We attributed this incomplete quenching to the presence of free CF in solution along the entire titration experiment, as a result of the partial dissociation of the CF–G3 complex in the low- concentration conditions that are necessary for the detection of heparin in clinical levels (*vide infra*).

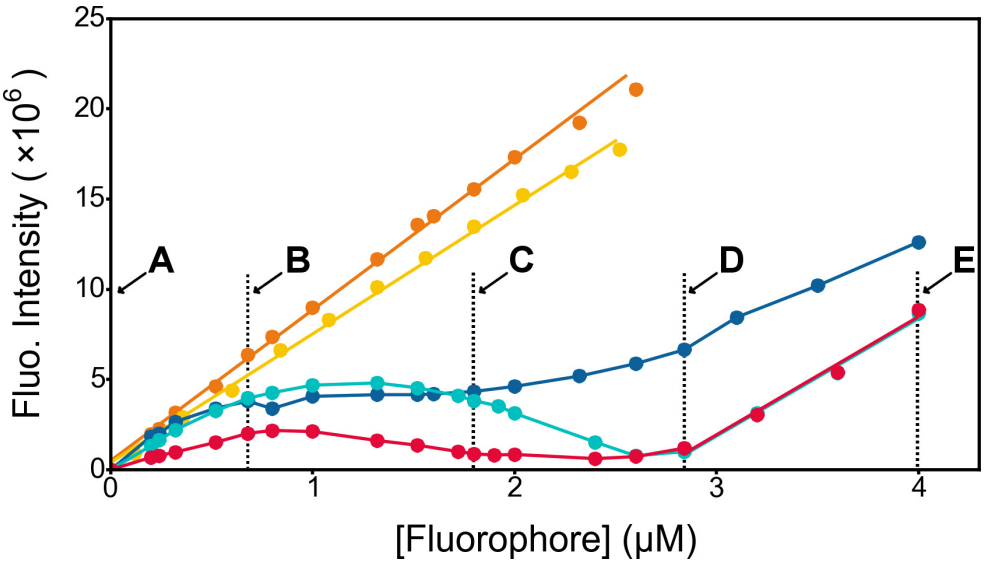


Figure 23: Fluorescence titration of DGLs with CF and D7CF. Orange: CF alone, yellow: D7CF alone, blue: CF in presence of G3, cyan: D7CF in presence of G3, red: D7CF in presence of G4. Conditions: 10 mM HEPES buffer, pH 7.8, [G3] = 163 nM, [G4] = 55 nM, λ_{ex} = 485 nm, λ_{em} = 535 nm. Both DGLs are at same concentration with respect to monomeric lysine residues, [lysine] = 20 mM.

In a sensing context, one would definitively desire to get a true OFF–ON response toward a target analyte. Thus, a new fluorescent reporter was designed, consisting of a bioconjugate between a homopolymer of aspartic acid and 5(6)-carboxyfluorescein (figure 21, D7CF. For the detail of the solid phase synthesis of the D7CF, see section 10.2). We expected that additional negative charges would lead to a much more efficient binding event. Stepwise addition of D7CF to G3 in the previous conditions led to a quite different titration profile (figure 23, cyan curve, as compared to D7CF alone, yellow curve). A first bell-shaped phase (figure 23, A →D) is observed until a fluorescence intensity (FI) minimum (fluorescence quenching yield (1-$FI_{G3/D7CF}/FI_{D7CF}$) x 100 = 95%). Upon further increase of the indicator concentration, a portion of the quenched fluorescence emission was recovered (figure 23, D →E). A similar titration profile was observed when using the fourth-generation dendri-graft poly-*L*-lysine polymer (G4), but with a less intense bell-shaped phase (figure 23, red curve).

8.2 Mechanism

A simple model to explain such binding behaviours is proposed, starting from the same amount of G3 and G4 with respect to the number of lysine residues (figure 24, A). As the degree of polymerization DPn for G3 and G4 are 123 and 365, respectively[23], in samples of identical weight (*i.e.*, at same concentration with respect to monomeric lysine residues), there are about three times more molecules of G3 than G4.about three times more molecules of G3 than G4. Upon addition of DGLs to D7CF, a red-shift of the absorption spectrum of the dye was observed by UV-visible spectroscopy (λ_{max} = 493 - 502 nm), suggesting its stabilization by non covalent interactions (*i.e.*, the reduction of its HOMO–LUMO gap energy) and confirming the formation of a complex between D7CF and DGLs.

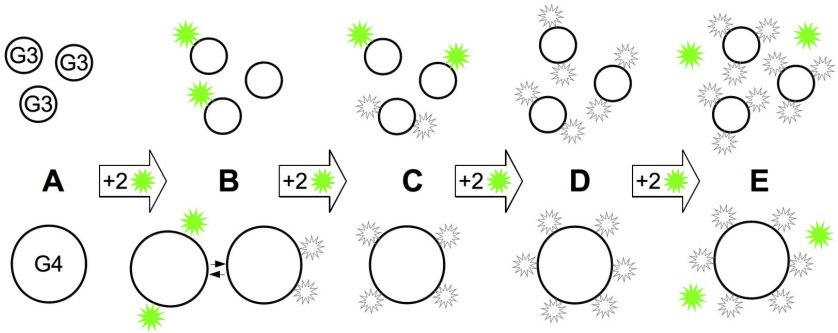


Figure 24: Proposed model for the binding behaviours of D7CF to G3 (top) and G4 (bottom).

From the fluorescence point of view, the addition of D7CF to DGLs leads to either the absence or a partial quenching of the fluorescent signal, depending on the dendrigraft size (figures 23 and 24, B). Upon further dye loading, the fluorescence quenching occurs earlier for the fourth-generation dendrigraft as a result of a higher surface ligand density (figures 23 and 24, C). Finally, both receptors saturation is reached for the same [lysine]/[D7CF] ratio = 7.04 (figures 23

and 24, D), followed by the linear reappearance of the optical signal (figures 23 and 24, E). It should be noted that the regime of this second straight phase is identical to the one of the D7CF alone, meaning that extra dye content – beyond the FI minimum that corresponds to the saturation of the dendrigraft with D7CF – is actually totally free in solution. Signal maximum quenching is reached at far less than 1 equiv. of dendrigraft, which means high binding stoichiometries. In 10mM HEPES buffer at pH 7.8, the maximum number of D7CF ligands per dendrigraft was determined to be 17 and 52 for G3 and G4, respectively. For subsequent heparin sensing experiments, the G4–D7CF complex was preferred since the redistribution of the ligand density on the dendrigraft surface during IDAs perturbs to a lesser extent the fluorescent signal (figures 23 and 24, D →C →B →A).

8.3 Thermodynamic study

Quantitative analysis of the interactions between G4 and D7CF was performed by analytical centrifugation at different G4/D7CF ratios. Experiments were realized using Vivacon®500 concentrators (10 kDa membrane cut-off) to selectively retain receptor–fluorescent probe complexes but not free D7CF; the concentration of which being determined by fluorescence spectroscopy (see section 10.3 for the detailed procedure).

A binding isotherm representing the number of bound ligands n per G4 was plotted as a function of free ligand concentration $[L]$ (figure 25A). n was determined from equation (3):

$$n = \frac{[L]_{tot} - [L]}{[S]_{tot}} \tag{3}$$

where $[L]_{tot}$ is the total concentration of ligand, $[L]$ is the free ligand concentration, and $[S]_{tot}$ the receptor concentration. Then, n was plotted over $[L]$ to graphically determine n_{max}, the maximum number of ligands per molecule of receptor. The isotherm displays a rapid increase of n with $[L]$ suggesting a rather high ligand–substrate association constant. Surprisingly, the corresponding Scatchard plot ($n/[L]$ was plotted versus n) appeared to be non-linear (figure 25B), which is indicative of either: (i) multiple non-equivalent independent binding sites, or (ii) multiple equivalent non-independent binding sites.

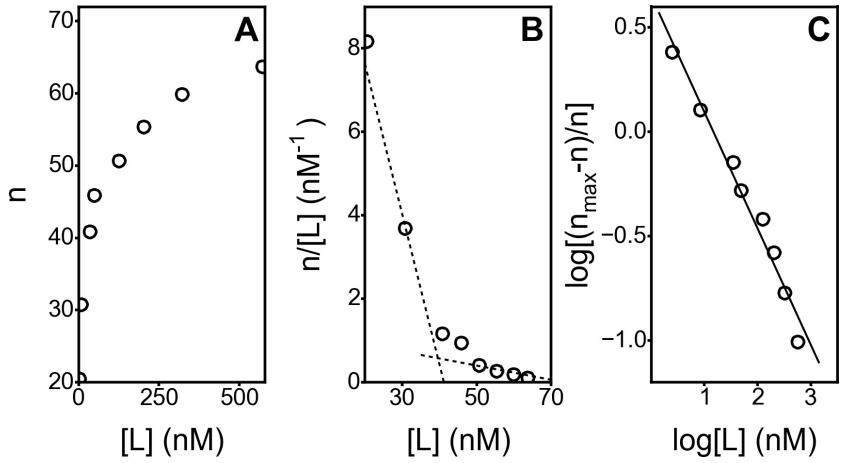


Figure 25: Binding isotherm (A) of the D7CF ligand with the DGL G4 polymer and the corresponding Scatchard (B) and Hill (C) plots. n is the mean number of ligands per polymer, n_{max} is the maximum number of ligands per polymer, and $[L]$ is the free ligand concentration. Dotted lines: Scatchard plot fitted with a two binding sites model[65,66]. Conditions: 10 mM HEPES buffer, pH 7.8, [G4] = 195 nM.

To test both hypotheses, the binding isotherm can be directly fitted according to the corresponding models.

Two non equivalent independent binding sites

$$n = \frac{n_1 k_1 [L]}{1 + k_1 [L]} + \frac{n_2 k_2 [L]}{1 + k_2 [L]} \tag{4}$$

Equation (4) was used to fit the experimental points from the isotherm experiments. n_1 is the number of ligands for the first binding site, and k_1 the association constant for this site. Fitting with Origin the results reported in table 1 (the confidence intervals are the standard deviation of the fitted parameters).

Site number	n	k (M^{-1})
1	41.2 ± 2.9	$3.56 \times 10^8 \pm 7.9 \times 10^7$
2	38.6 ± 5.0	$3.02 \times 10^6 \pm 1.5 \times 10^6$

Table 1: Fitted parameters for the ”Two non equivalent independent binding sites” model.

As it is well known that the first portion of the Scatchard plot can be fitted with the linear equation (5)[67], the second portion cannot be fitted with the parameters previously obtained. A simple linear regression was applied to this portion, to give a slope of -0.0174 and an intercept of 1.25.

$$y = -k_1 n + n_1 k_1 \tag{5}$$

Multiple equivalent non-indepedent binding sites

Whereas a two sets of binding sites model turned out to be in poor agreement with the experimental data (figure 25B, dotted lines), a linear Hill plot of the experimental data was obtained (figure 25C) using equation (8), derived from the original Hill equation (equation (6))[66] and equation (7), where K_D is the dissociation constant of the complex, z the Hill coefficient, and θ the fraction of occupied sites on the receptor. n as been previously defined as the number of occupied sites on the receptor.

$$\theta = \frac{1}{(\frac{K_D}{[L]})^z + 1} \tag{6}$$

$$\theta = \frac{n}{n_{max}} \tag{7}$$

$$log\,\frac{n_{max} - n}{n} = z\,log\,K_D - z\,log([L]) \tag{8}$$

The slope of the straight line (*i.e.*, Hill coefficient n_H = 0.56) suggests negative cooperativity of D7CF binding to G4. By negative cooperativity we mean here the alteration in the affinity of a binding site, when other sites are occupied by a ligand. To put it simply, the more D7CF ligands are bound to G4, the less the incoming dyes have available positives charges for further binding. From the Hill plot, a global dissociation constant K_D = 14.8 nM was also determined between G4 and D7CF in 10mM HEPES buffer at pH 7.8.

8.4 Proof of concept

In presence of our G4–D7CF sensor, increased levels of heparin in buffered water correlated linearly with FI response (figure 26A. Before addition of the GAGs, the curve offset on the y-axis is related to the intrinsic fluorescence emission of the G4–D7CF complex).

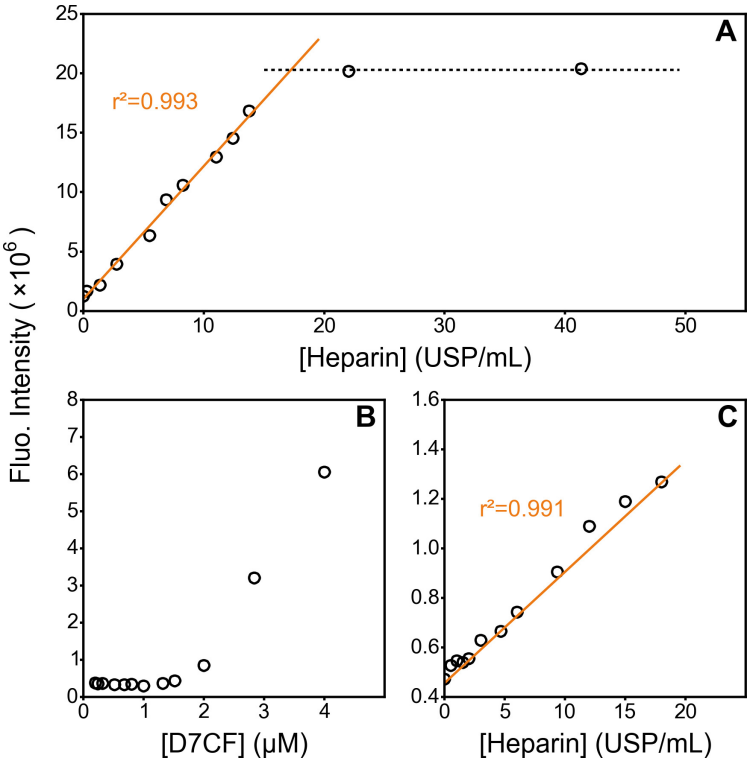


Figure 26: (A) Fluorescence titration of heparin with a G4–D7CF complex in 10mM HEPES buffer, pH 7.8. [G4] = 55 nM, [D7CF] = 2.84 µM. (B) Fluorescence titration of G4 with D7CF in 10mM HEPES buffer, pH 7.8 and in presence of 5% human blood. [G4] = 55 nM. (C) Fluorescence titration of heparin with a G4–D7CF complex from heparinized human blood samples. Conditions: 5% human blood, 10 mM HEPES buffer, pH 7.8, [G4] = 55 nM, [D7CF] = 1.52 µM, λ_{ex} =485 nm, λ_{em} = 535 nm. Heparin concentrations given in abscissa are the actual heparin levels in blood samples.

The following plateau indicates the complete displacement of the indicator by heparin from the dendrigraft surface (FI = 20.3 x 10^6 corresponding to [D7CF] = 2.8 µM in solution, see: figure 23, yellow calibration curve). These observations validate the proof-of-principle of our IDA for heparin monitoring. Before testing the assay from heparinized human blood samples, a titration of G4 with D7CF in 10 mM HEPES buffer, pH 7.8 and in presence of 5% human blood was realized to determine the optimal receptor/indicator ratio (figure 26B).

Two differences from previous titrations in pure buffer can be spotted: (i) the lower amount amount of D7CF that is necessary to reach dendrigraft saturation ([lysine]/[D7CF] = 13) and (ii) the absence of the bell-shaped phase. Such differences could be explained by the partial complexation of G4 with blood components (*e.g.*, proteins), resulting in both less available dendrigraft surface (= faster saturation) and tighter indicators confinement (= more efficient quenching). In a typical sensing experiment, a sample of human blood containing heparin of a known concentration was added to a buffered aqueous solution of G4 and D7CF (final blood/buffer ratio = 0.05). The fluorescent emission at 535 nm (excitation at 485 nm) was used to detect and quantify heparin within the range of clinically relevant concentrations (0–18 USP.mL^{-1}, figure

26C). Limits of detection (LOD) and quantification (LOQ) of heparin in human blood were determined as 0.12 USP.mL^{-1} and 0.30 USP.mL^{-1}, respectively. Because blood has a strong absorption band with a maximum centred around 410 nm, fluorescence intensities are significantly lower in comparison with measurement in pure buffer.

8.5 Control experiments

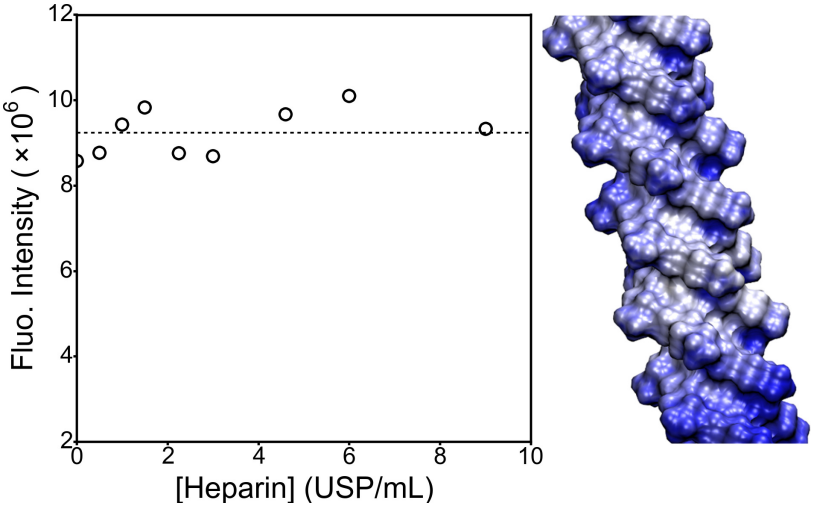


Figure 27: Left: fluorescence titration of heparin with a PLL400–D7CF complex. Conditions: 10 mM HEPES buffer, pH 7.8 in presence of 5% serum, [PLL400]= 50 nM ([lysine] = 20 mM), [D7CF] = 1.52 µM, λ_{ex} = 485 nm, λ_{em} = 535 nm. Right: electrostatic potential surface of a poly-*L*-lysine fragment in a α-helix conformation (the structure was generated using Spartan'10 software[68], without energy minimization.

Finally, a heparin sensing assay with α-poly-*L*-lysine (PLL_{400})– as a linear equivalent ofDGLs – in presence of serum was unsuccessful (figure 27). The D7CF molecules are fully displaced from the polymer even in absence of heparin. This definitely highlights the importance of the receptor topology for a strong association event between the binding partners when working in complex media. Additional control experiments in serum with other GAGs were also performed (figure 28). Whereas hyaluronic acid (one negative charge by dimeric unit) was not able to displace the indicator, it turned out the chondroitin sulfate A (two negative charges by dimeric unit) was almost as efficient as heparin in this role; suggesting that electrostatic interactions are the predominant interactions between GAGs and DGLs. It must be noted that the mean concentration of endogenous total GAGs in serum is about 15nM[69], which is about two orders of magnitude below heparin therapeutic concentrations, thus excluding possible interferences during the monitoring of the exogenous anticoagulant (in addition, other biologically relevant anions at physiological concentrations such as Cl^-, HPO_4^{2-} and ATP^{4-} were not able to displace the indicator from the dendrigraft surface (data not shown)).

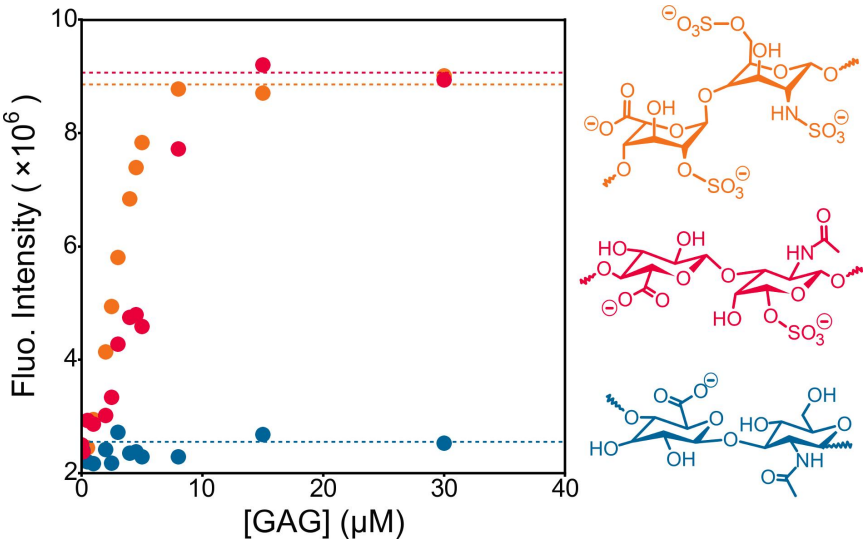


Figure 28: Fluorescence titration of heparin (orange), chondroitin sulfate A (red), and hyaluronic acid (blue) with a G4–D7CF complex.20,24 Conditions: 10 mM HEPES buffer, pH 7.8 in presence of 5% serum, [G4] = 55 nM ([lysine] = 20 mM), [D7CF] = 1.52 µM, λ_{ex} = 485 nm, λ_{em} = 535 nm. The reported GAG concentrations are based on the amount of disaccharide. The molecular weights of the GAG repeat units are assumed as that of the sodiated analogues. The curve offset on the y-axis is related to the intrinsic fluorescence emission of the G4–D7CF complex.

9 Conclusion

In conclusion, we reported a simple indicator-displacement assay, involving a dendrigraft and a new designed dye, which allows for the first time the turn-ON fluorescent detection and quantification of heparin in human blood and at actual clinical levels. Qualitative and quantitative analyses of the involved supra- molecular interactions were proposed. Also, this study revealed the ability of dendrigrafts to bind very efficiently some GAGs. By intervening in GAG-mediated biological processes, such accessible and non-immunogenic[70] materials may find new biomedical applications. We are currently working in this direction.

10 Experimental

10.1 General

All commercial chemicals were used as received without further purification. DGLs were provided by the COLCOM company (G4: batch #1207-04, G3: batch #1204-03). Poly-L-lysine hydrochloride (PLL400 , Mw=66 kDa) was purchased from Alamanda, and GAGs from Sigma-Aldrich: Heparin sodium salt from porcine intestinal mucosa (207 USP/mg, ref: H3399-50KU), Chondroitin sulfate A from sodium bovine trachea (ref: C9819-5G), Hyaluronic acid sodium

salt from streptococcus equi (ref: 94137-10MG). Sheep serum was purchased from Sigma-Aldrich (ref: S2263-100ML), while human blood samples – collected in sodium citrate tubes – were provided by the Etablissement Français du sang (EFS). As serum and blood are viscous liquids, low-binding tips for automatic pipettes were used. All solutions were prepared in 10 mM HEPES buffer, pH 7.8. The buffer was prepared using MilliQ deionized water (18 MΩ). HPLC purifications were performed using an Armen SPOT Prep system, on a Reveleris flash cartridge (reversed phase C18, 40 µM particles, 40 g) operated at 9 $mL.min^{-1}$. HPLC analyses were performed on a Waters equipment including a Waters 996 photodiode array detector and a Waters 2690 separation module. The analytical column (Thermo Scientific 3 µm C18 particles, 50 x 2.1 mm) was flooded at 0.2 $mL.min^{-1}$. For the two previous HPLC setups, a gradient from 90:10 H_2O (0.1 % TFA)/MeCN (0.1 % TFA) to 10:90 H_2O (0.1 % TFA)/MeCN (0.1 % TFA) over 30 minutes was operated, with UV/Vis monitoring at 215 and 437 nm. ESI-MS spectra were recorded on a Synapt G2 mass spectrometer. Fluorescence measurements were performed on a Berthold Tristar LB 941 microplate reader (λ_{ex} = 485 nm and λ_{em} = 535 nm). All experiments were performed at least in duplicate.

10.2 Synthesis of the D7CF

Assembly of the linear protected peptide $Fmoc-Asp(OtBu)-(Asp(OtBu))_6-Chlorotrityl^{TM}$ resin was carried out manually with a traditional Fmoc/tBu strategy in a glass reactor with a glass frit. 250 mg of Chlorotrytil^{TM} resin (1.3 mmol/g loading) were made to swell in 5 mL of DCM. Fmoc-Asp(OtBu)-OH (74 mg, 0.18 mmol) was added, followed by DIEA (0.1 mL, 0.6 mmol). The suspension obtained was stirred for 90 min at room temperature. The supernatant was removed by suction and then the resin was capped with DCM/MeOH/DIEA (17:2:1, 5 mL, 1 h), then DCM/MeOH/DIEA (17:2:1, 5 mL, 10 min). Fmoc deprotection reactions were performed with DMF/piperidine (8:2, 3x3 mL, 5 min each) and the resin was washed with DMF (5x3 mL, 1 min each). A suspension of Fmoc-Asp(OtBu)-OH (123 mg, 0.3 mg), PyBop (156 mg, 0.3 mg) and DIEA (0.1 mL, 0.6 mmol) in 5 mL of DMF was stirred for 45 min. Each coupling reaction was performed in duplicate in order to avoid the risk of deletion. The final introduction of the 5(6)-carboxyfluorescein (CF) label was performed as follow, reducing lighting exposure as much as possible[71]: 50 mg of $H_2N-Asp(OtBu)-(Asp(OtBu))_6-Chlorotrityl^{TM}$ were made to swell in 5 mL of DMF. CF (113 mg, 0.3 mmol), DIC (0.05 mL, 0.3 mmol) and HOBt (46 mg, 0.3 mmol) were added and the suspension was stirred for 1 h. The fully deprotected and labeled peptide was recovered directly upon acid cleavage with TFA/TIPS/H_2O (95:2.5:2.5, 4 mL, 1 h), followed by TFA/TIPS/H_2O (95:2.5:2.5, 2x3 mL, 1 min each). The combined filtrates were concentrated under reduced pressure. Crude D7CF was obtained by precipitation with Et_2O.
Analyses and purification of the synthesized D7CF were performed by HPLC column, as previously mentioned (see section 10.1). Retention time = 11.06 minutes. MS: m/z = 1180.28 $\rightarrow [M-H]^-$. Yield = 49%.

10.3 Methods for titration

Typical isotherm experiment

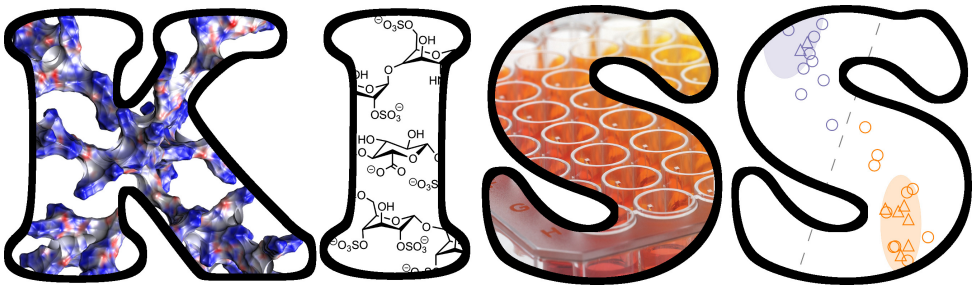
Eppendorf tubes were filled with 100 µL of a stock solution of G4 (final concentration: 195 nM) and variable volumes of a D7CF solution (final concentrations: from 1 to 14 µM). The volume of each sample was completed to 2 mL with buffer. The samples were shaken for 2 hours at room temperature. Then, 500 µL of the previous solutions were centrifuged at 3500 g (Sigma 302-K, 8000 rpm) for 1 min over an inert membrane (Sartorius Vivacon®500, Molecular Weight Cut-Off: 10 kDa). The process allowed 20 µL of solutions to cross the membranes, which were transferred into 96 wells plates. The volume of each well was completed to 150 µL with buffer, and the fluorescence intensity was measured. The same procedure (samples containing only D7CF) was repeated for calibration.

Typical Heparin titration in human blood

Eppendorf tubes were filled with 100 µL of pure human blood. Variable volumes of a stock solution of heparin in 10 mL HEPES buffer (pH 7.8) were added and the final volumes were adjusted to 200 µL with buffer. Heparinized blood samples were allowed to stir for 1 h at room temperature before use. In a typical titration experiment, microplate wells were filled with 95 µL of buffer, 20 µL of a G4 solution in buffer (final concentration in the well: 55 nM), 20 µL of D7CF solution in buffer (final concentration in the well: 1.52 µM), and finally 15 µL of the previous heparinized blood samples. The fluorescence intensity was then followed for 2 hours, with measurements every 2 min.

Part IV

A KISS (keep it simple, sensor) array for glycosaminoglycans



Résumé en français

Nous démontrons ici que le système simple décrit dans la partie III peut évoluer très facilement vers une barrette de senseurs, capable de différencier cinq analogues de l'héparine. Composé d'un dendrigraft de poly-*L*-Lysine de troisième génération et d'un court peptide fluorescent, ce nouveau système est la première barrette de senseurs composé d'un unique récepteur et d'un unique indicateur. En fonction de la quantité d'indicateur sur le récepteur, des glycosaminoglycanes chargés négativement (GAG) induisent une variation du signal fluorescent selon s'ils déplacent ou compactent les indicateurs à la surface du récepteur. Cette stratégie unique permet non seulement l'identification aveugle de GAGs purs avec un niveau de précision de 100 %, mais aussi la différenciation de mixtures.

11 Intoduction

Glycosaminoglycans (GAGs) are an important class of exogenous biopolymers that display numerous therapeutic activities[72]. Because GAGs are structurally analogous (figure 29), orthogonal and lengthy analytical techniques – involving NMR spectroscopy, HPLC, and capillary electrophoresis – are necessary to identify them[73,74].

The ability to routinely discriminate and assess the purity of these anionic polysaccharides is mandatory in order to prevent public health disasters. Indeed, the use of contaminated heparin in patients undergoing dialysis resulted in serious acute hyper-sensitivity reactions (including some resulting in death) during the late 2000s[21]. Contaminants in adulterated lots included chondroitin sulfate A, dermatan sulfate (chondroitin sulfate B), and hyaluronic acid. To address this issue, some research groups developed pattern-based recognition assays for various GAGs by using arrays of nanoparticles, macrocycles, or liposomes[19,20,75,76]. Despite the success of those works, extensive synthetic steps were systematically involved to obtain the multiple receptors and/or indicators that were necessary to construct the arrays; therefore impeding their use for cheap routine screening of GAGs. Herein a much simpler sensor array for GAGs, from a single receptor and a single indicator, is presented.

We recently reported that a cationic dendrigraft poly-*L*-lysine polymer (DGL) was able to form a multi-ligand complex with a conjugate between an anionic homopolymer of aspartic acid and 5(6)-carboxyfluorescein; leading to the extinction of the optical signal as a result of its aggregation on the surface of the receptor[77]. Fluorescence could be restored upon the introduction of heparin in an indicator-displacement process[17]. Although this sensing ensemble allowed us to detect and quantify – for the first time – heparin in human blood samples at clinically relevant levels, it was not possible to discriminate between GAGs such as heparin and chondroitin sulfate. In this previous study, the fourth-generation dendrigraft poly-*L*-lysine polymer and the seven-membered homopolymer of aspartic acid and 5(6)-carboxyfluorescein bioconjugate were necessary to work in presence of complex biological fluids in order to ensure sufficient

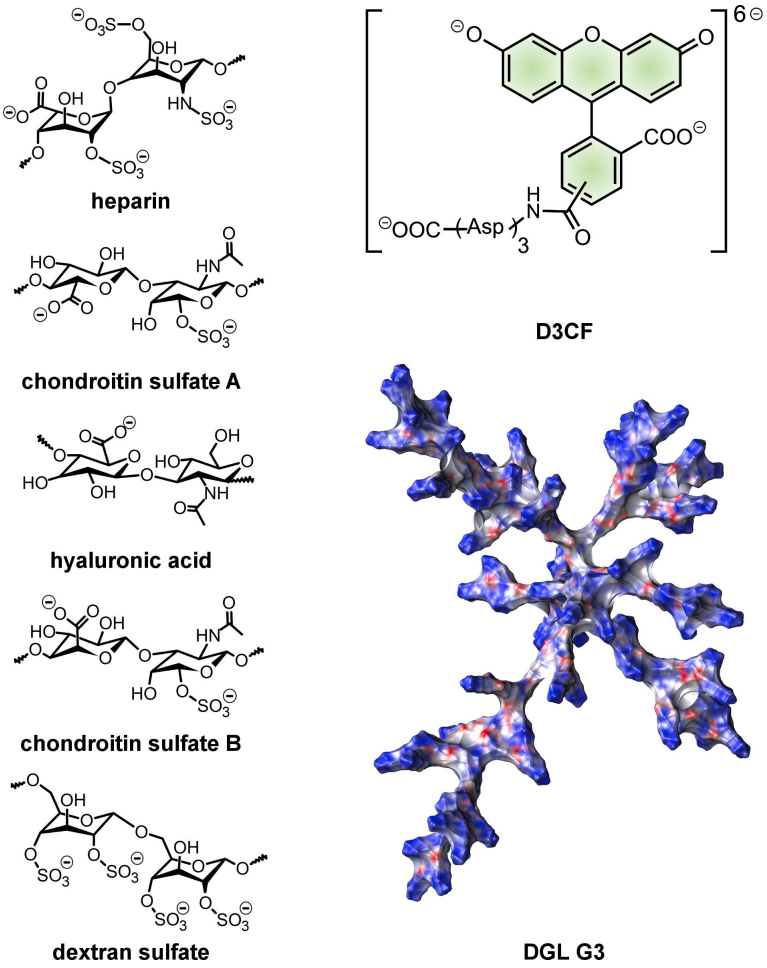


Figure 29: Left, from top to bottom: major repeat unit of selected GAGs. Top right: bioconjugate between a homopolymer of aspartic acid and 5(6)-carboxyfluorescein, D3CF. Bottom right: minimized structure with electrostatic potential surface of the third-generation dendrigraft poly-*L*-lysine polymer DGL G3.

binding between the partners. For the present work, since the experiments were all performed in a much less competitive medium (10 mM HEPES buffer, pH 7.8), the more accessible DGL G3 and shorter D3CF (figure 29) were chosen in order to keep these assays for GAGs as simple as possible (given that the simple carboxyfluorescein was not able to bind efficiently G3 in HEPES buffer (10 mM, pH 7.8)[77], we selected an intermediate peptidic chain length for the indicator).

12 Results and discussion

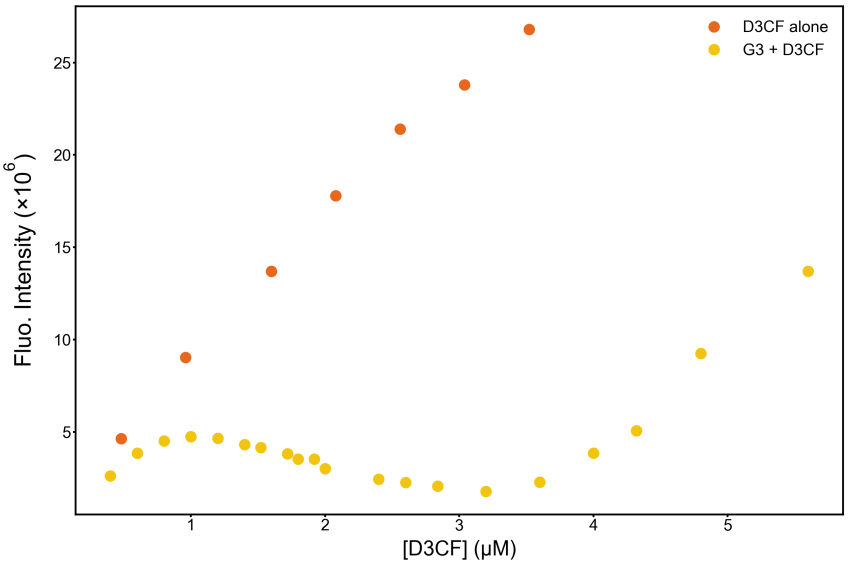


Figure 30: Fluorescence titration of the G3 with D3CF. Conditions: 10 mM HEPES buffer, pH = 7.8, [G3]=163 nM ([lysine] = 20 mM), λ_{ex} = 485 nm, λ_{em} = 535 nm.

We envisioned that, using an array that was based on the variable loading of the fluorescent indicator on the receptor, each GAG could generate a specific fluorescent response pattern. Indeed, because they display specific electrostatic surfaces, the complexes can be considered as "differential" receptors for GAGs (figure 31)[78]. The loadings were calculated from the titration curve of DGL G3 with D3CF (figure 30). For the rationalization of the titration curve's profile, see section 8.

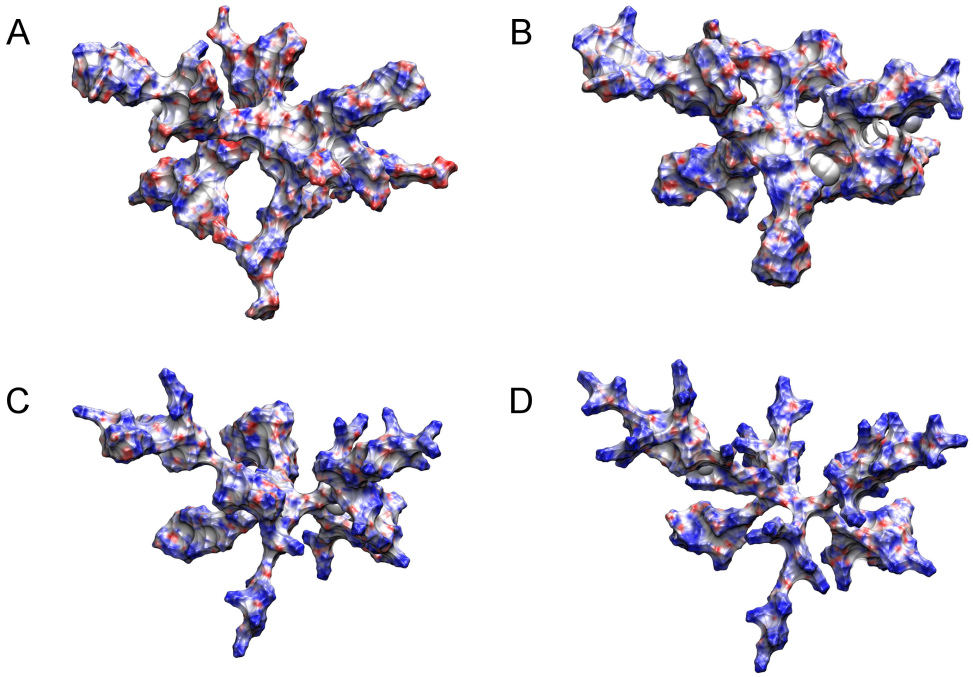


Figure 31: Minimized structures of G3-D3CF complexes at various loadings, with electrostatic potential surfaces. Loadings: 100% (A), 75% (B), 50% (C), 25% (D). For calculation details, see section 16.

Upon addition of dextran sulfate to G3-D3CF complexes in water containing 10 mM HEPES and buffered to pH 7.8, distinct optical behaviours were detected depending on the loading of the indicator on the receptor (figure 32). Mass concentrations for GAGs were used in this study in order to: (i) be able to blind test the polysaccharides, and (ii) get rid of any approximation approximation about their structure that may lead to rough molar concentration calculations.

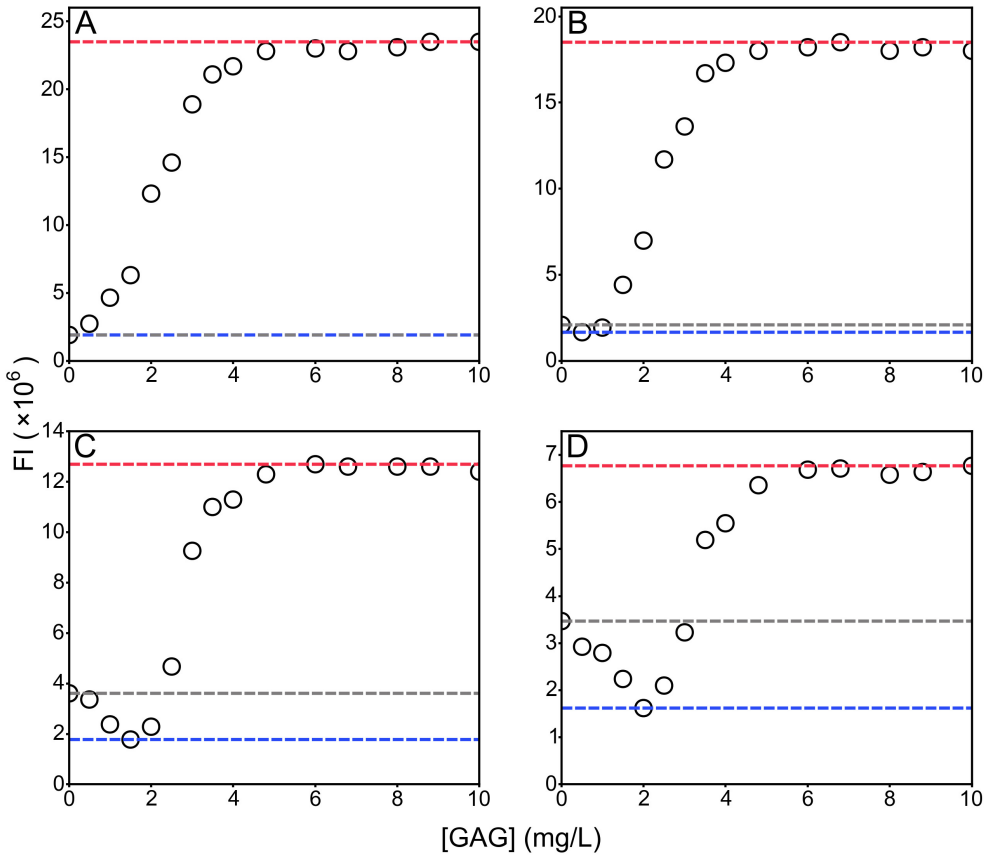


Figure 32: Fluorescence titrations of the G3-D3CF complexes with dextran sulfate. Loadings: 100% (A), 75% (B), 50% (C), 25% (D). Dashed lines: grey = FI_0, blue = FI_{min}, red = FI_{max}. Conditions: 10 mM HEPES buffer, pH = 7.8, [G3] = 163 nM ([lysine] = 20 mM), λ_{ex} = 485 nm, λ_{em} = 535 nm.

At 100% loading, increased levels of dextran sulfate correlated positively with the fluorescence intensity (FI) response (figure 32A). At 75% loading, a decrease of the FI occurred upon addition of dextran sulfate until a minimum value FI_{min} (figure 32B), followed by the restoration of the optical signal. Similar titration curves could be observed at lower loadings, although reaching FI_{min} required increasing amounts of dextran sulfate (figure 32C and 32D). For all these titration curves, it could be noted that (i) the initial fluorescence FI_0 is in agreement with the titration curve of G3 with D3CF (see figure 30), and (ii) a final plateau with a specific maximum fluorescence value FI_{max} is reached at a GAG concentration of 4.8 $mg.L^{-1}$.

12.1 Mechanism: the compaction-displacement indicator assay

A simple model to explain such unexpected behaviors, which depend on the loading of the indicator on the receptor, is proposed (figure 33).

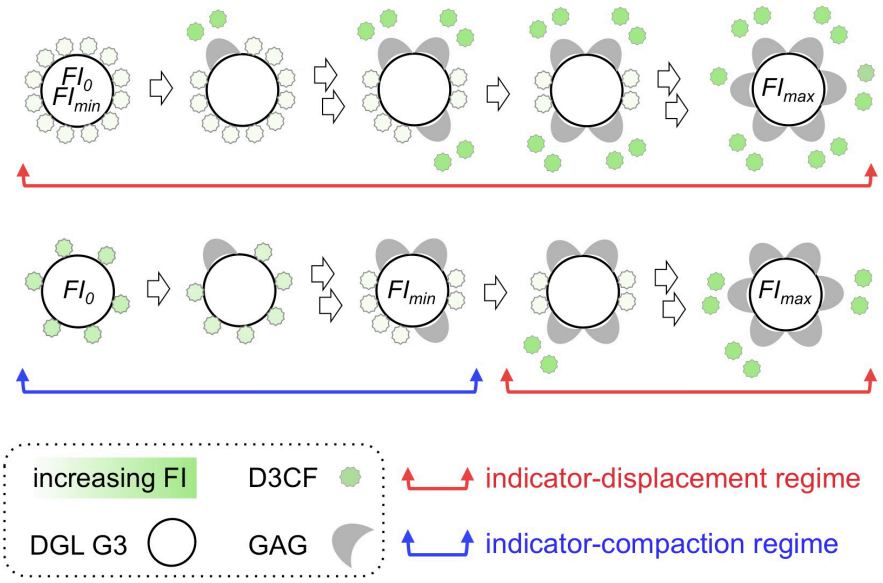


Figure 33: Proposed model for the various binding behaviours of GAGs to G3-D3CF complexes.

Starting from a fully loaded G3 with D3CF, FI_0 is equal to FI_{min} since the distance between the indicators on the receptor's surface is at its smallest, resulting in maximum fluorescence self-quenching (figure 33, top). The subsequent stepwise addition of GAG leads to the continuous displacement of the indictor from thedendrigraft's surface and to the increase of FI, until a plateau that indicates that all the fluorophore is actually free in solution according to the calibration curve of D3CF (see figure 30). In contrast, a lower loading gives a less compacted starting arrangement of the indicators on the dendrigraft's surface, yielding a higher FI_0 (figure 33, bottom).

Additional GAG molecules can therefore bind to the receptor without displacing but with compacting the fluorophores on the receptor's surface. This results into the diminution of the optical signal until FI_{min} that is correlated with maximum compaction. Upon this point, additional GAG molecules restitute the indicator-displacement regime. In this model, (i) the lower the loading is, the higher the GAG concentration necessary to reach full compaction is, and (ii) the same amount of GAG is needed to fully displace the indicator whatever the loading is.

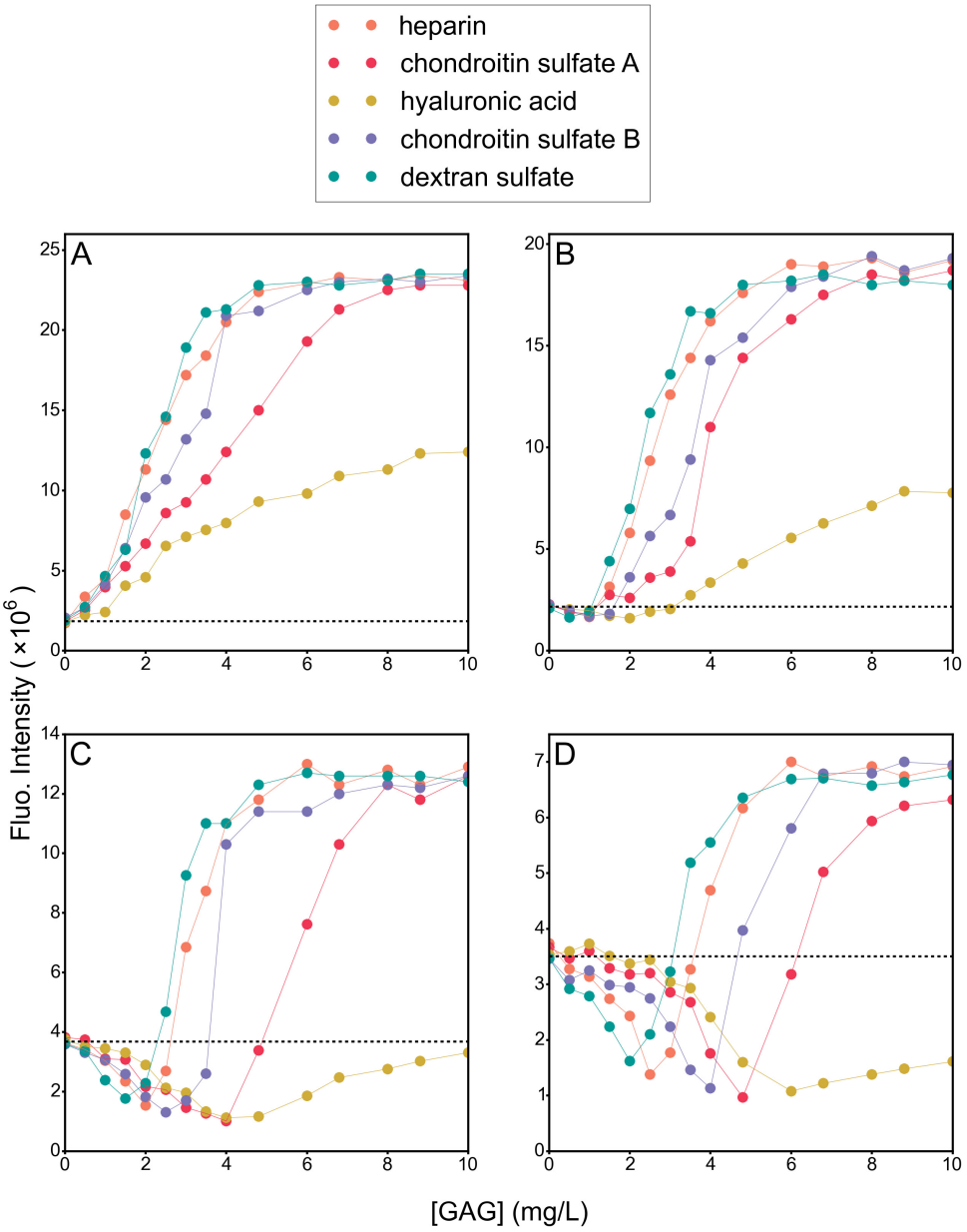


Figure 34: Fluorescence titrations of the G3-D3CF complexes with GAGs. Loadings: 100% (A), 75% (B), 50% (C), 25% (D). Conditions: 10 mM HEPES buffer, pH = 7.8, [G3]=163 nM ([lysine] = 20 mM), λ_{ex} = 485 nm, λ_{em} = 535 nm.

Such a model could be used to virtually explain the similar behaviors that were observed upon the addition of other GAGs, each one displaying specific indicator-displacement or indicator-compaction regimes (see figure 34). As a consequence, differential optical responses were observed when the selected GAGs were submitted at a concentration of 3.5 mg.L^{-1} to our sensor array (figure 35).

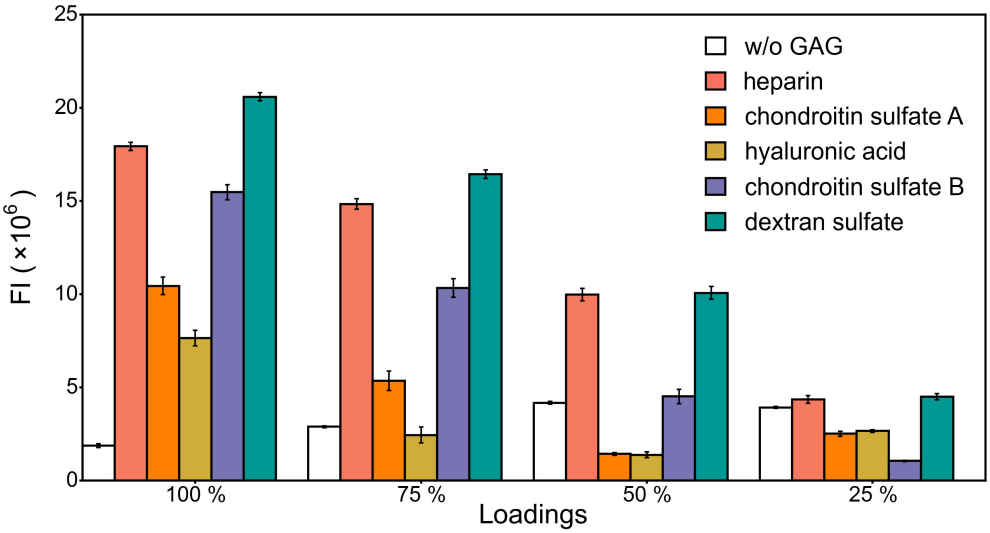


Figure 35: Fluorescence intensities from the addition of GAGs to G3-D3CF complexes. Conditions: 10 mM HEPES buffer, pH = 7.8, [G3] = 163 nM ([lysine] = 20 mM), [GAG] = 3.5 mg.mL^{-1}, λ_{ex} = 485 nm, λ_{em} = 535 nm.

12.2 Linear Discriminant Analysis: concept

In this section, we will describe the Principal Component Analysis (PCA) and the Linear Discriminant Analysis (LDA) concepts, and as this document is mainly intended for chemists, we will avoid any unecessary mathematics.

One of the main purposes of PCA and LDA is dimensionality reduction. To put it simply, if a data set is defined in a space with k dimensions, PCA and LDA will allow to represent it in a subspace of n dimensions, where $n \leq k$ and with minimal loss of information. For example, let's assume one wants to represent a set of points defined in a two dimensions space (figure 36) in a one dimension space.

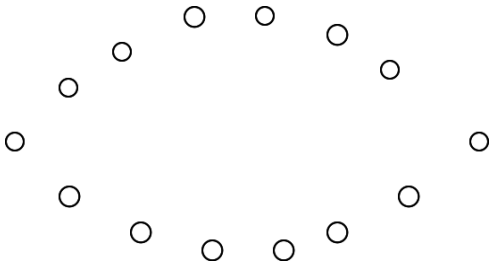


Figure 36: A set of data points defined in a two dimensions space.

To do so, one would have to find the principal component of the data, *i.e* the straight line where the projections of the points would be more spread out (figure 37).

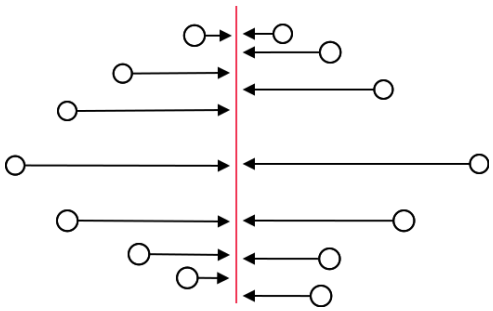


Figure 37: A set of data points projected on a vertical line.

As the data is not very spread out here, the variance will be low. This vertical line is probably not the principal component. A better choice (the best, actually) would be a horizontal line (figure 38). One can not obtain a higher variance with this set of points, and the horizontal line is therefore the principal component. For a 2D data set, this principal component could be used as a new x-axis, and after the projection of each point onto this axis, it leads to the best 1D representation of the original 2D data set. This would be the perfect example of dimensionality reduction.

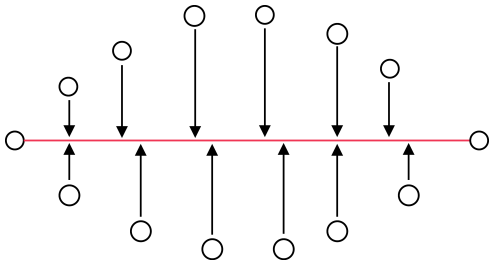


Figure 38: A set of data points projected on a horizontal line.

Using a mathematical vocabulary, the horizontal and vertical lines are called eigen vectors, and indicate a direction. For a data set of k original dimensions, there are k possible eigen vectors. Each eigen vector has its own eigen value (a number describing how spread out the data is in this direction). The eigen vector with the highest eigen value is the main component (in this case, the horizontal line). The second eigen vector has to be orthogonal to the first (and a potential third one would have to be orthogonal to the second one, and so on): it is the vertical line depicted in figure 37. When performing dimensionality reduction, it is possible to calculate the *percentage of captured variance ratio* for the new axis, defined for each eigen vector as in equation (9). This percentage indicates ”how much of the total information” was captured by the dimensionality reduction. Figure 41 shows an example of a 4D →2D reduction. The two main eigen vectors captured 99.5 % of the information available from the four original dimensions. This means that the original problem - impossible to represent and hard to analyse - was greatly simplified, and almost no information was lost during the procedure.

$$ratio = \frac{eigenValue_i \times 100}{\sum_{i=1}^{k} eigenValue_i} \tag{9}$$

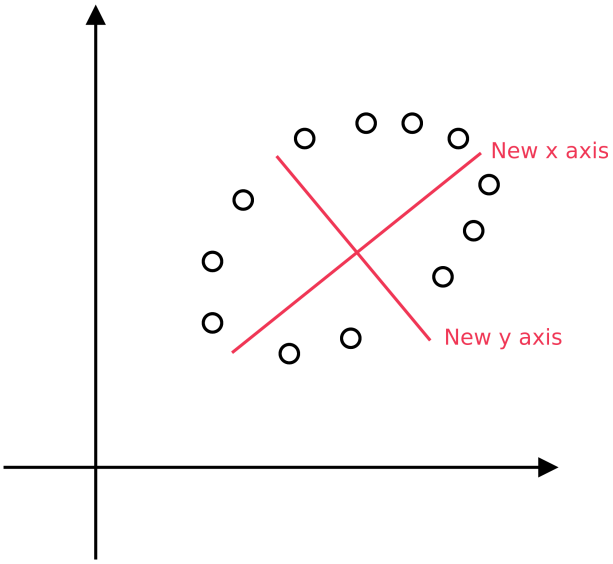


Figure 39: The new x and y axes.

The eigen vectors provide another way to frame the data in, but the original data were not modified. Also note that the new axes are combinations of the "old" ones, and are entirely new dimensions (*i.e* the new x-axis contains a part of the old x-axis and a part of the old y-axis).

LDA is similar, but while PCA is qualified of *unsupervised* (no supplementary input is injected to the procedure), LDA is *supervised*, and needs an extra information: the class labeling. In our case, there are four experimental dimensions: the fluorescence intensity value for each of the four G3-D3CF complexes. To create a PCA score plot, a set of points $(FI_{100\%}, FI_{75\%}, FI_{50\%}, FI_{25\%})$ is required, but LDA also requires the class label of each point (*i.e* the nature of the analyte: heparin, dextran sulfate, etc). Then, while the PCA procedure finds the eigen vectors maximizing the variance between all the points, the LDA procedure will find the eigen vectors maximizing the variance between the classes (see figure 40). Note that while the eigen vectors provided by PCA are all orthogonal to each other, this is not necessarily the case for the ones provided by LDA[79].
Here it becomes obvious that - while similar - LDA and PCA are not used for the same purposes. As no class labeling is involved in PCA, no bias is introduced in the representation of the data. Therefore, PCA is a tool to spot patterns in the data, while LDA is generally used for data classification (see section 12.3 for an concrete example of classification).

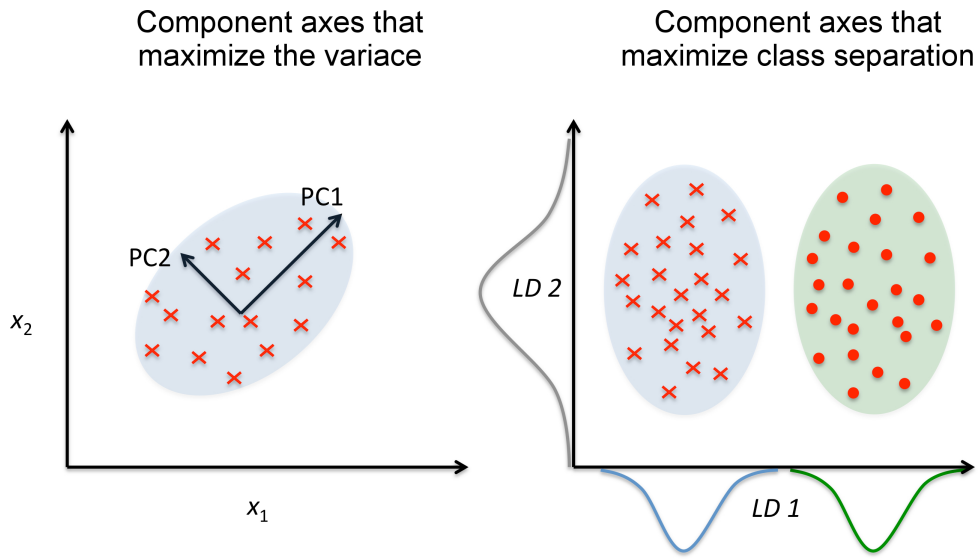


Figure 40: Differences between PCA (left) and LDA (right). Image generously provided by Sebastian Raschka[80], and adapted.

The mathematics and algorithms used to determine the eigen vectors are out of the scope of this document, and will not be described here, but more information can be found elsewhere[81]. For this work, LDA was performed with the LDA class of the scikit-learn[30] Python module, through a Python script which automated the determination of the eigen vectors. However, the LDA class of the scikit-learn module originaly lacked the *captured variance ratio* feature described above, and was not returning the eigen values. It was therefore impossible to know how much of the total variance was captured by each eigen vector. As scikit-learn's source code is open and available on GitHub (https://github.com/scikit-learn/scikit-learn), we modified it to add the feature, and sent our modifications to the developers of the module (two pull requests were accepted and merged: #5216 and #6027). We believe in the use of free and open-source softwares, and hope this contribution will be helpful to others.

12.3 Identification of GAGs with LDA

Differentiation was demonstrated and validated by linear discriminant analysis (LDA), which is a supervised procedure that allows the classification of data[81] (see section 12.2). In a first place, known samples of GAGs (figure 41, triangles) were used as a training input to the procedure, giving a captured variance ratio of 93.4% for the first linear discriminant and 6.1% on the second one. Decision boundaries (figure 41, dashed lines) were then generated by using Bayes' rule[82]. These boundaries delimit areas of the reduced space that are exclusively attributable to a single GAG. Finally, unknown samples (figure 41, circles) were tested against the training set, achieving 100% of identification with 50 correct cases out of 50, and demonstrating therefore the robustness of this assay.

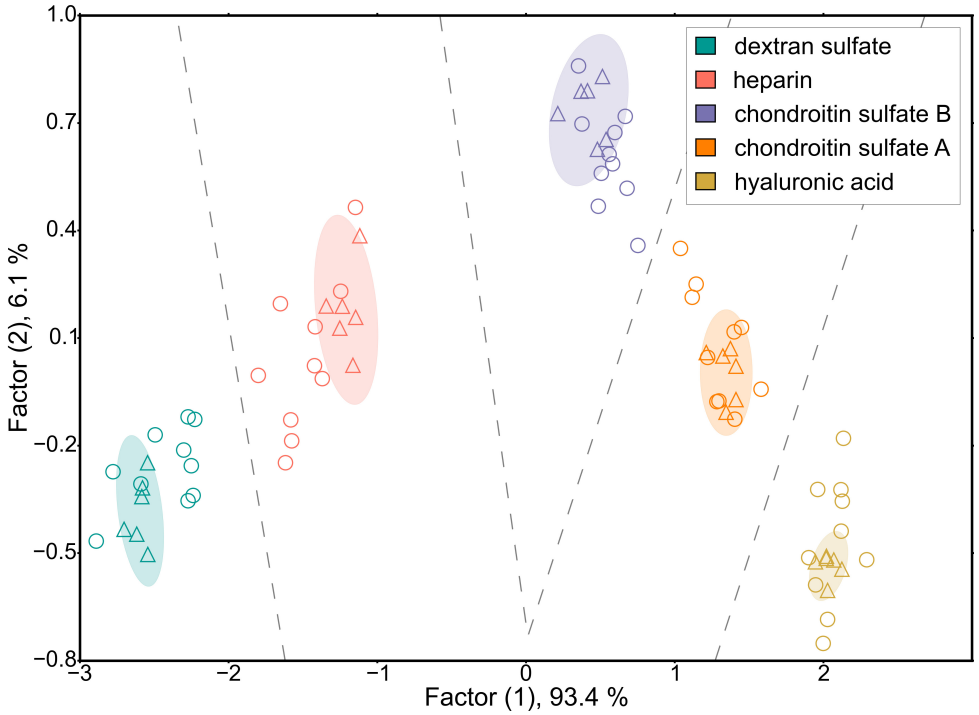


Figure 41: LDA canonical score plot for the response of the sensor array to GAGs. Dashed lines: decision boundaries, triangles: training set, circles: blind tests, ellipses: confidence limits at 99% for the training set.

In order to further demonstrate the applicability of our sensor array, we tested its ability to assess the purity of samples of heparin contaminated with 10%, 20%, 30%, and 50% of chondroitin sulfate B, which is the most prevalent impurity in pharmaceutical heparin preparations[83]. From the resulting score plot it was clear that all mixtures were separated with good resolution (figure 42).

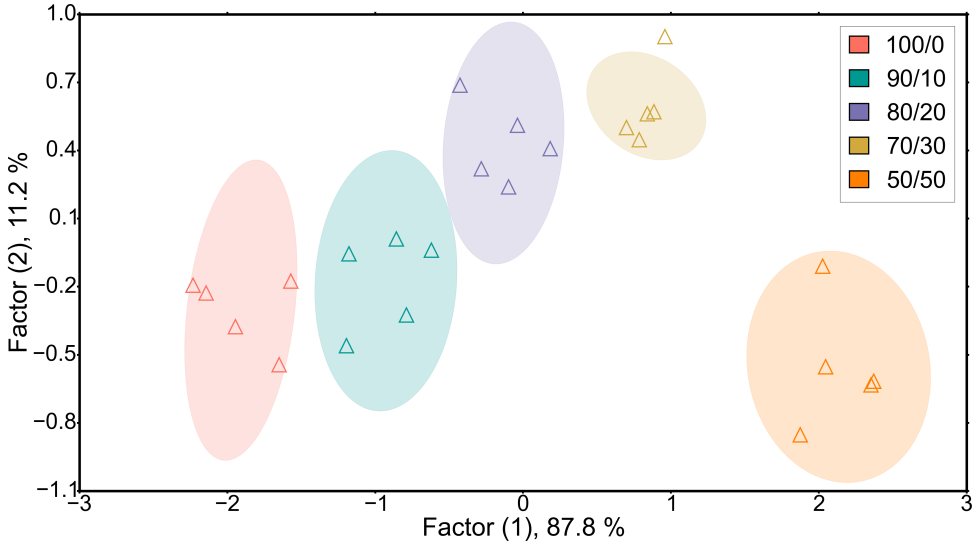


Figure 42: LDA canonical score plot for the response of the sensor array to heparin/chondroitin sulfate B mixtures. Ellipses: confidence limits at 99%.

13 Conclusion

In conclusion, we reported a simple sensor array that allows not only the unambiguous identification of structurally similar pure glycosaminoglycans, but also the differentiation of mixtures of glycosaminoglycans. The simplicity of this array relies on the use of only one commercial receptor and one highly accessible fluorescent indicator. Furthermore, a new sensing approach is introduced in this study that one could describe as a compaction/ displacement-indicator assay. We are confident that such an original approach will find future applications in the supramolecular analytical field.

14 Experimental

14.1 General

All commercial chemicals were used as received without further purification. DGL G3 was provided by the Colcom company (batch #1204-03). GAGs were purchased from Sigma-Aldrich: Heparin sodium salt from porcine intestinal mucosa (207 USP/mg, ref: H3399-50KU), Chondroitin sulfate A sodium salt from bovine trachea (ref: C9819-5G), Hyaluronic acid sodium salt from streptococcus equi (ref: 94137-10MG), Chondroitin sulfate B sodium salt from porcine intestinal mucosa (ref: C3788-25MG), Dextran sulfate sodium salt from Leuconostoc spp. (ref: D6924-1G). All solutions were prepared in 10 mM HEPES buffer, pH 7.8. The buffer was prepared using MilliQ deionized water (18 MΩ). HPLC purifications were performed using an Armen SPOT Prep system, on a Reveleris flash cartridge (reversed phase C18, 40 µM particles, 40 g) operated at 9 mL.min^{-1}. For the purification procedure, a gradient from 90:10 H_2O (0.1 % TFA)/MeCN (0.1 % TFA) to 10:90 H_2O (0.1 % TFA)/MeCN (0.1 % TFA) over 30 minutes was operated, with UV/Vis monitoring at 215 and 437 nm. LC/MS analyses were performed on an UPLC Acquity H-Class equipment (Waters), with a Kinetex column (Phenomenex, reversed phase C18, 100 x 2.1 mm, 2.6 µm particles). Chromatograms were recorded by a TUV Photodiode detector at 210 nm, and mass spectra were recorded by a Synapt G2-S mass spectrometer (Waters, ESI, positive ions introduction). For the analysis procedure, the gradient shown in table 2 was operated at 0.5 mL.min^{-1}.

Time (min)	% A (H_2O + 0.1 % Formic acid)	%B (MeCN + 0.1 % Formic acid)
0	90	10
12	20	80
13	0	100

Table 2: Gradient used for the HPLC analysis

Fluorescence measurements were performed on a Berthold Tristar LB 941 microplate reader (λ_{ex} = 485 nm and λ_{em} = 535 nm), at 20°C. All fluorescence experiments were performed at least in triplicate.

14.2 Synthesis of the D3CF

Assembly of the linear protected peptide Fmoc – Asp(OtBu) – $(Asp(OtBu))_2$ – Chlorotrityl™ resin was carried out manually with a traditional Fmoc/tBu strategy in a glass reactor with a glass frit. 50 mg of Chlorotrytil resin (1.3 mmol/g loading) were made to swell in 5 mL of DCM. Fmoc-Asp(OtBu)-OH (15 mg, 0.036 mmol) was added, followed by DIEA (0.02 mL, 0.12 mmol). The suspension obtained was stirred for 2 h at room temperature. The supernatant was removed by suction and then the resin was capped with DCM/MeOH/DIEA (17:2:1, 5 mL, overnight), then DCM/MeOH/DIEA (17:2:1, 5 mL, 2h). Fmoc deprotection reactions were performed with a mixture DMF/piperidine/formic acid 75:20:5 (1 x 2 mL x 1 min; 3 x 2 mL x 3 min; 1 x 2 mL x 15 min). Addition of formic acid to the deprotection mixture is intented to prevent the formation of aspartimide side products[84]. The resin was washed with DMF (5x3 mL, 1 min each). A suspension of Fmoc-Asp(OtBu)-OH (25 mg, 0.06 mmol), PyBop (31 mg, 0.06 mmol) and DIEA (0.02 mL, 0.12 mmol) in 2 mL of DMF was stirred for 2 h. Each coupling reaction was performed in duplicate in order to avoid the risk of deletion. The final introduction of the 5(6)-carboxyfluorescein (CF) label was performed as follow, reducing lighting exposure as much as possible[71]: all the previously prepared quantity of H_2N – Asp(OtBu) – $(Asp(OtBu))_2$ – Chlorotrityl™ was made to swell in 5 mL of DMF. CF (113 mg, 0.3 mmol), DIC (0.05 mL, 0.3 mmol) and HOBt (46 mg, 0.3 mmol) were added and the suspension was stirred for 1 h. The fully deprotected and labeled peptide was recovered directly upon acid cleavage with TFA/TIPS/H_2O (95:2.5:2.5, 4 mL, 1 h), followed by TFA/TIPS/H_2O (95:2.5:2.5, 2x3 mL, 1 min each). The combined filtrates were concentrated under reduced pressure. Crude D3CF was obtained by precipitation with Et_2O. Analyses and purification of the synthesized D3CF were performed by LC/MS, as previously mentioned (see section 14.1). Retention time = 4.22 and 4.26 min (two main peaks were observed since 5(6)-carboxyfluorescein is a mixture of two regio-isomers. m/z:722.15 →[M + H^+]. Yield = 86 %.

14.3 Analytical methods

Typical fluorescent sensing experiment

In a typical sensing experiment (for 100% loading), microplate wells were filled with 55 µL of buffer, 10 µL of a G3 solution in buffer (final concentration in the well: 163 nM), and 20 µL of D3CF solution in buffer (final concentration in the well: 3.2 µM). The mixture was allowed to equilibrate for 15 min. Finally, 35 µL of a 15 $mg.L^{-1}$ GAG solution were added (final concentration in the well: 3.5 $mg.L^{-1}$). The fluorescence intensity was then followed for 45 min, with

measurements every 2 min. The fluorescence intensities finally used in the treatments were a mean of the last 3 points of a measurement.

Part V

In Silico Three-Dimensional Structures of Poly-*L*-Lysine Dendrigrafts

Résumé en français

Nous présentons dans cette partie une méthodologie originale et simple pour la construction de structures tridimensionelles de dendrigrafts de poly-*L*-Lysine (DGLs). Nous étudions ensuite les caractéristiques structurelles de ces polymères par des simulations de dynamique moléculaire durant plusieurs microsecondes. Cette méthodologie repose sur l'encodage des caractéristiques expérimentales des DGLs (*i.e.* degré de polymérisation, rapports de branchement, charges) en chaînes alphanumériques, qui seront ensuite interprétées par le programme de mécanique moléculaire Amber. Ce travail ouvre des perspectives vers l'exploration *in silico* des propriétés des dendrigrafts et des polymères hyperbranchés.

15 Introduction

Dendrigrafts of Poly-L-Lysine (DGLs) recently complemented the family of polycationic dendritic macromolecules that include the prominent PolyAMidoAMine (PAMAM) and PolyEthylenImine (PEI) polymers (figure 43)[23]. Because their syntheses in aqueous conditions are thermodynamically controlled by precipitation and kinetically controlled by steric hindrance, DGLs share features with both PAMAM dendrimers (*i.e.* repeated iterative and geometric growth), and hyperbranched PEI (*i.e.* random distribution and narrow molecular weight distribution). In contrast to PAMAM and PEI, DGLs are biodegradable[85], exhibit low cellular toxicities[86], and turned out to be non-immunogenic[70]. As a consequence, DGLs have recently gained a huge interest in the biomedical field during the last quinquennium, with numerous applications in drug and gene delivery[85–97], biomaterials and tissue engineering[98–103], bio-imaging[96,104–108], and biosensing[77,109,110].

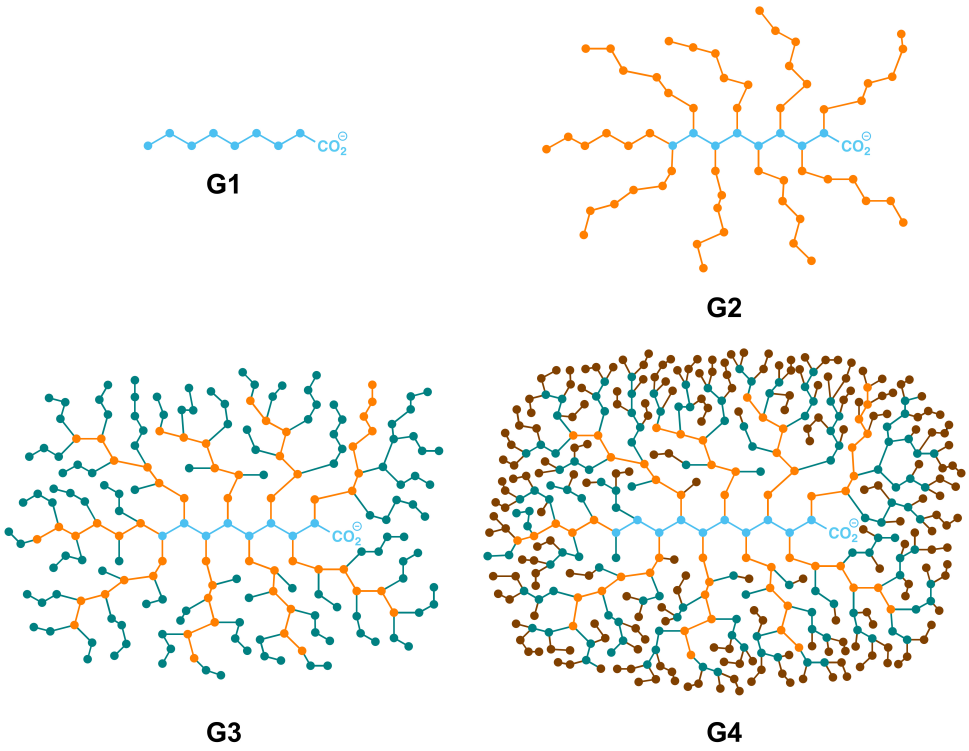


Figure 43: Schematic representation of first- to fourth-generation DGLs G1-G4 (each dot represents a L-lysine residue, pending free amino groups are not represented)

Despite this growing use of DGLs, a deeper understanding of the molecular level properties of these macromolecules is mandatory. In this context, we decided to explore the possibility to construct DGLs *in silico* from scratch (due to the polymeric nature of these macromolecules, there is no X-ray structure available), and subsequently perform Molecular Dynamics (MD) simulations. As DGLs are made of lysine residues, one could assimilate them to proteins, and try to use softwares such as AMBER to produce primary and tertiary structures. However, some of the residues are linked to each other through isopeptidic bonds (*i.e.* amide linkages that involve the nitrogen atom of the lateral chain of lysine residues). These unconventional type of bond makes the DGLs impossible to build by traditional softwares without heavily tweaking them. Also, a manual approach toward an initial set of three-dimensional (3D) structures for DGLs - with respect to known experimental data - suffers from several flaws: i) it is highly error-prone since the fourth-generation G4 consists of 365 residues, ii) it is likely to be biased toward an homogeneous distribution of the residues and unlikely to be representative of a polymerization event. iii) it is highly time-consuming since it would take weeks to obtain a small number of potentially incorrect structures. For these reasons, a new and automated method has been designed for the construction of 3D structures of DGLs.

Our first approach (naive) is detailed in section 16. We used it for the published works in parts III and IV. Even though the structures it provided were not accurate or suitable for molecular dynamics, they were the first to correctly represent the DGLs' topology. When we later

met Prof. Gérald Monard (Université de Nancy, http://gmonard.wikidot.com), our research evolved toward a more systematic and realistic modeling of the DGLs. This work is still in progress, and the preliminary results are discussed in section 17.

16 Naive approach

Previously reported characterization data[23], such as the mean branching ratio (BR) and the mean degree of polymerization (Dp) were used as a starting point to develop a homemade Python script. The branching ratio was defined as the number of ε-branched Lys residues to the total number of residues (Dp). From these parameters, it is possible to calculate the number of peptidic and isopeptidic bonds, as well as the number of free α- and free ε-amine functions in DGLs (figure 3).

Generation	Dp	BR	Free α amines	Free ε amines	Isopept. bonds	Pept. bonds
G1	8	0	1	8	0	7
G2	48	12.3	7	42	6	41
G3	123	24.0	31	93	30	92
G4	365	24.7	85	281	84	280
G5	963	24.4	236	728	235	727

Table 3: Properties of the five generations of DGLs

The purpose of the Python script was to ”mimic” the synthesis of the DGLs through the manipulation of SMILE string chains. The starting point was the first generation polymer G1 (a linear peptide of 8 residues):

[O-]C(=O)[C@@H](CCCCN)NC(=O)[C@@H](CCCCN)NC(=O)[C@@H(CCCCN)NC(=O)[C@@H](CCCCN)NC(=O)[C@@H](CCCCN)NC(=O)[C@@H](CCCCN)NC(=O)[C@@H](CCCCN)NC(=O)[C@@H](CCCCN)N

to which some markers were added:

[O-]C(=O)[C@@H](CCCCN[E0])NC(=O)[C@@H](CCCCN[E1])NC(=O)[C@@H](CCCCN[E2])NC(=O)[C@@H](CCCCN[E3])NC(=O)[C@@H](CCCCN[E4])NC(=O)[C@@H](CCCCN[E5])NC(=O)[C@

@H](CCCCN[E6])NC(=O)[C@@H](CCCCN[E7])N[X7]

, where [En] and [Xn] are ε and α positions that are available for further polymerization. Then, the sequential approach of the algorithm randomly replaced these markers with Lys building blocks, with respect to the BR and Dp parameters. The use of two different markers allows the discrimination of the protonation state of α- and ε- amines. At pH 7.8, we made the hypothesis that only the ε-amine residues of DGLs were protonated[56].
The SMILE string chains were then converted to 3D structures using the Discovery Studio software, and the equilibrium geometries were minimized using Spartan'10 at the MMFF level of theory in vacuum[68]. Finally, VMD[111] (version 1.9.1) and the APBS[112] plugin (version 1.3, default settings) allowed us to generate electrostatic potential maps (figures 21, 29 and 31).

17 Accurate Modeling

17.1 Protonation state

Determining the protonation state of each generation of DGL is mandatory in order to perform accurate modeling experiments. However, when a molecule bears several acidic functions, the protonation state of each of them influences the intrinsic acidity of the other ones, resulting in a shift of the functions' pKa. This phenomenon is called the *polyelectrolyte effect* (for a full description of this phenomenon, see the review of Koper and Borkovec[113]). Moreover, these pKa are also influenced by several other parameters, like the solvent exposition of the acidic functions, or their capacity to form hydrogen bonds. Therefore, while the pKa of the α-amine of an isolated lysine residue is 9.16 (see figure 44), the literature reports an average value of 7.7 for proteins' N-termini, with a low value of 6.8[56,114].
We previously assumed that only the ε-amines of DGLs were protonated at physiological pH. Herein, we present new experimental data that conforts this previous hypothesis.

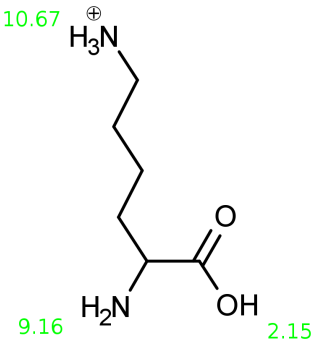


Figure 44: The lysine amino acid at physiological pH 7.4; *green*: pKa of the acidic and basic functions

We performed the pH-metric titrations of native DGLs (counterion: TFA) in deionized water. The samples were first acidified with 0.5 mL of 1M nitric acid, and then titrated with 0.5M sodium hydroxide with a 702 SM Titrino automatic titrator. A typical titration curve for such an experiment is shown in figure 45. One would expect the pH to be stable below 1 mL of added NaOH since it was the required volume to neutralize the nitric acid used to acidify the sample. However the titration curve displayed a "lack" of nitric acid (0.18 mL, first peak of the derivative). This gap could be explained by the fact that not all the amine functions were protonated in the commercial powders (for the synthetic route to DGLs, see Collet *et al.*[23]). Therefore, a part of the nitric acid added in the first place has been consumed for the protonation of these neutral amines. If some amino groups were not in their protonated state, they obviously do not carry TFA counterions, which has consequences on the calculated molecular weights of the DGLs (see *vide infra*).

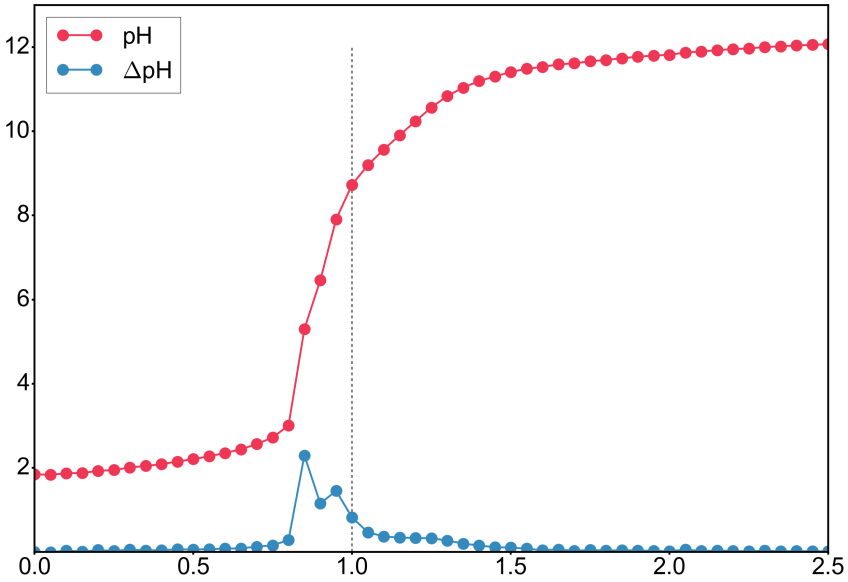


Figure 45: pH-metric titration of a DGL G2 in water, acidified with 0.5 mL of 1M nitric acid. pH is plotted versus the added volume of NaOH.

The second peak of the derivative gave us the quantity of the α-amino groups in the samples. Due to the polyelectrolyte effect, the third peak of the derivative - corresponding to the pKa of the ε-amines - was barely visible. The deprotonation of these functions takes place on a wide range of pH, making it impossible to determine the number of ε-amines. Therefore, this number was calculated from the ratio $\frac{\alpha}{\epsilon}$ (see table 3). We then wrote the electroneutrality equation to calculate the protonation fraction Θ (equation 10), with C_a the concentration of added acid, C_b the concentration of added base, C_{TFA} the missing concentration of TFA, C_{H^+} and C_{OH^-} the concentration of free H^+ and free OH^-, and C_p the total concentration of amino groups.

$$\Theta = \frac{C_a - C_b - C_{TFA} - C_{H^+} + C_{OH^-}}{C_p} \tag{10}$$

$$\Theta = \frac{C_a - C_b - C_{TFA} - 10^{-pH} + \frac{10^{-pK_w}}{10^{-pH}}}{C_p} \tag{11}$$

Equation 10 could also be expressed as a function of the pH, leading to equation 11. A typical output of such calculation is shown on figure 46. From such curves, the protonation degree of DGLs G2, G3 and G4 were extrapolated at pH 7.4 (blue point in figure 46), and are reported in table 4. These results show that the percentages of non-protonated amine functions correspond to the percentage of α-amine functions in the DGLs. This suggested that only the ε-amines were protonated at a physiological pH. Table 5 reports the net charges of the DGLs (based on our hypothesis), and their respective molecular weights after a correction accounting for the missing TFA counterions.

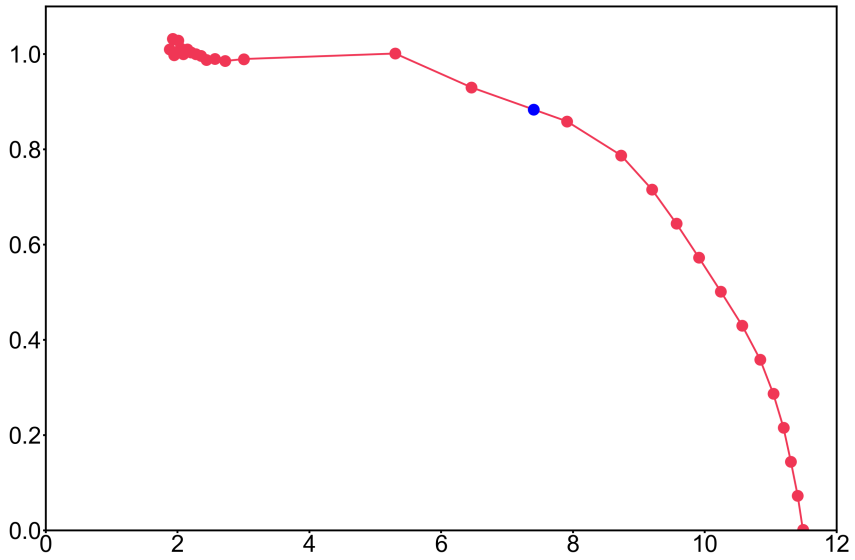


Figure 46: Protonation fraction vs pH, for DGL G2. The blue point is the exptrapolated Θ value at pH 7.4

Generation	Θ (%)	$\frac{\epsilon}{\epsilon+\alpha}$ (%)
G2	88	86
G3	82	75
G4	82	78

Table 4: Percentages of protonation compared to percentages of ε-amines.

Generation	Dp	MW ($g.mol^{-1}$)	Net charge
G2	48	10081	+41
G3	123	24481	+92
G4	365	70945	+280

Table 5: Number of residues, molecular weights after correction and net charges of DGL G2-G4

17.2 Encoding the structures

As a first step, the macromolecules G1-G4 have been deconvoluted into six discrete building blocks which encompass all the possible arrangements (*i.e.* peptidic or isopeptidic bonds, charged or neutral nitrogen atoms) of the *L*-lysine residues within the dendrigrafts (figure 47). These building blocks will be later used as starting points to create custom Amber "residues" (see section 17.3).

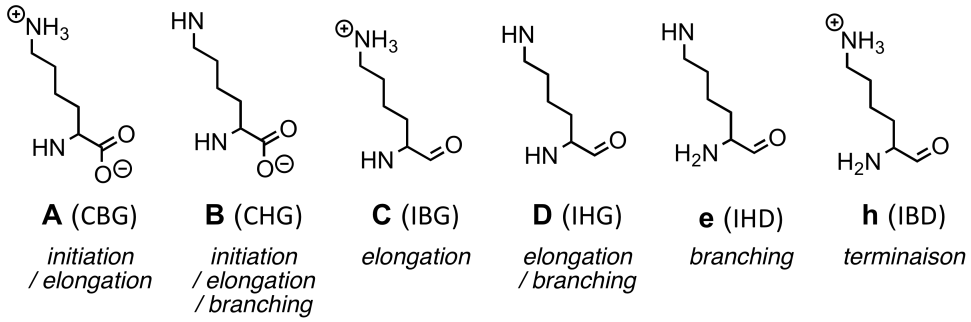


Figure 47: All possible *L*-Lysine buildings blocks involved in the construction of DGLs G1-G4.

Each building block was subsequently labeled with a one-letter code that unambiguously identifies it. As a consequence, G1 could be written as the following string: ACCCCCCh (from C-terminal to N-terminal). For the higher generation G2-G4, numbers have been introduced

within the strings, which addressed the locations of the residues where branched chains were growing from (starting from the C-terminal amino acid). For instance, a possible representation of a G2 polymer using these molecular descriptors would be an alphanumeric string of 48 letters, and 6 discrete numbers (figure 48, top). These alphanumeric strings encode not only the position of every *L*-Lysine residue in the polymers, but also their connections and protonation states (figure 48, bottom). A home-made Python script allowed us to generate a set of eight random – but respecting the initial experimental parameters (*i.e.* Dp, BR, and net charge) – sequences for each generation of DGL G2 to G4. To later stay consistent in our analyses, each structure was grown from the corresponding structure in the previous generation (*e.g.* $G3_{2/8}$ was grown from $G2_{2/8}$)

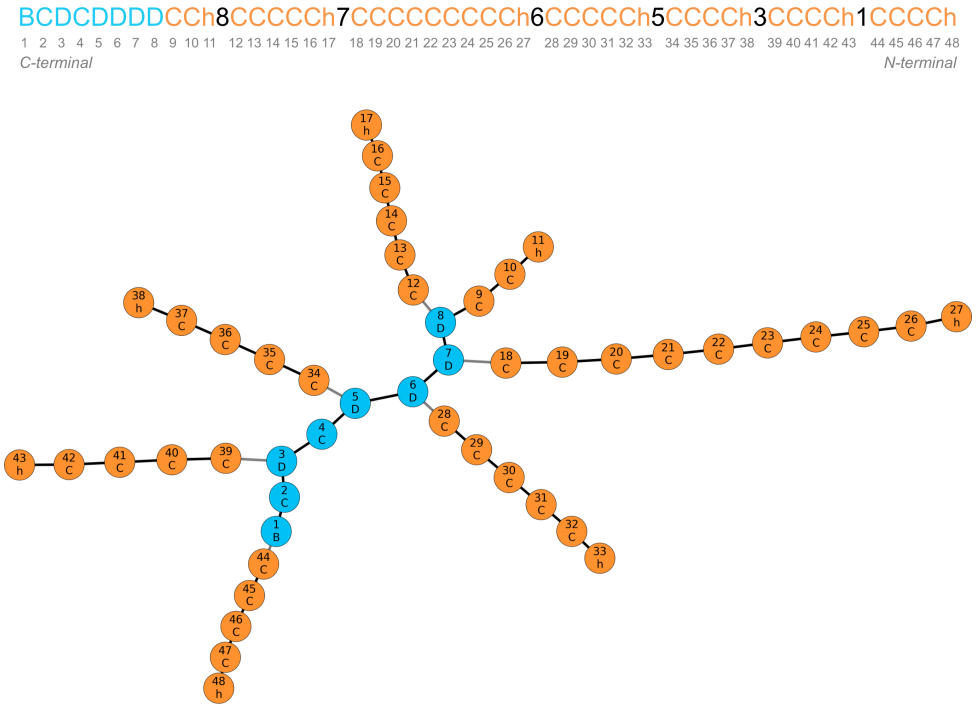


Figure 48: 2D schematic representation of a particular structure of a DGL G2.

17.3 Reading and building the structures

In a first place, molecular mechanic (MM) parameters (partial charges, angles, dihedrals) of the six building blocks were determined in order to create custom residues. To do so, the antechamber module of the AmberTools15 Package was used[115] (base used for the charge derivation: Hartree-Fock 6/31-G*). To respect Amber's naming convention, we also established a one to three letters equivalence (figure 47). The alphanumeric-coded polymers could then be read by a homemade Nucleic Acid Builder (NAB) program, developped specially to build the DGLs' branched structures. Each one-letter coded residue was read by the program, translated to its three-letters equivalent, and added (at the right position !) to the growing dendrigraft. After each addition, a short conjugated gradient minimization was performed (cut-off: 15 Å,

100 iterations maximum) to ensure that no bond or atom would overlap. A powerfull desktop computer can build a fourth-generation dendrigraft within an hour of calculations. Finally, the 3D structures were saved for subsequent simulations (see section 17.4).

17.4 Molecular Dynamics

The results presented in this section are preliminary.

We chose to run the MD calculations using a Generalized Born implicit solvent (representing water). By doing so, eight trajectories - lasting ten microseconds each ! - for eight different structures of the DGL G2 were obtained (approximately 7 hours of calculations per trajectory, on the nodes of the ROMEO HPC center: https://romeo.univ-reims.fr). Such trajectory durations are not obtainable with an explicit solvent and for several possible structures of the DGL, therefore impeding a reliable conclusion about the potential folding of these polymers. Also, the use of a cluster of Graphics Processing Units (GPU) instead of the traditionnal Central Processing Units (CPU), along with a few optimizations like hydrogen mass repartitioning[116], greatly reduced the calculation times. In the near future, we should be able to obtain trajectories lasting at least one microsecond for several structures of the higher-generation DGLs.
In a first place, the two-dimensional Root Mean Squared Deviation (RMSD, measurement of how similar a structure's internal atomic coordinates, by comparison with some reference molecule coordinates) was plotted (figure 49). This figure is a 2D matrix of frames during the 10 µs trajectory. Each frame represents the molecule's structure at the time of the frame, and the matrix compares each frame to the others (the matrix is therefore symmetrical). The highest the RMSD is, the more different the two compared structures are (consequently, the diagonal of the matrix appears as a black line). For a folded protein, one would expect clear areas with low RMSD values (see the graphical abstract of Roe and Cheatham III[117] for an example of such a 2D-RMSD), demonstrating that the protein adopts a small number of very particular conformations. We did not observe this behaviour, for any of the eight trajectories simulated, leading us to think the DGL G2 does not adopt folded conformations.

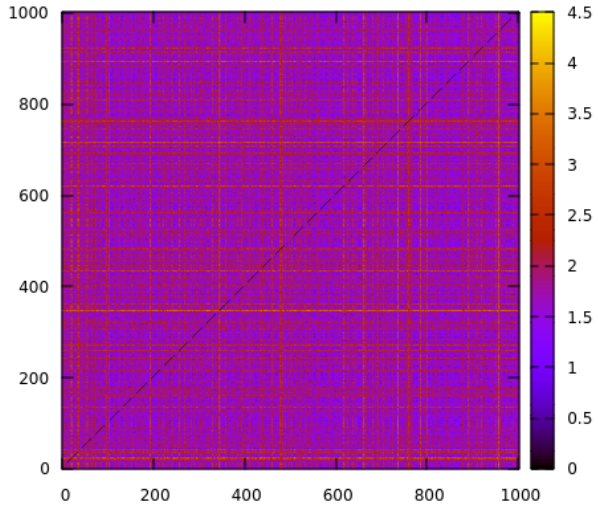


Figure 49: 2D RMSD of the core (residues 1 to 8) of one possible structure of the DGL G2.

We pushed the investigations further and used the cpptraj's tool *secstruct* on our trajectories (figure 50). This tool calculated the secondary structural propensities for the residues in a molecule, using the rules defined by Kabsch and Sander [118]. None of the residue seemed to adopt a durable secondary structure, except when an IBD residue follows an IBG residue. In this particular case, the two residues adopted a "Bend" conformation (a region where the curvature is superior to 70°, see Kabsch and Sander [118] for more details). We believe this pattern is an artefact due to the unconventional branched nature of the DGLs, and we are currently trying to explain the phenomenom and adapt our analysis.

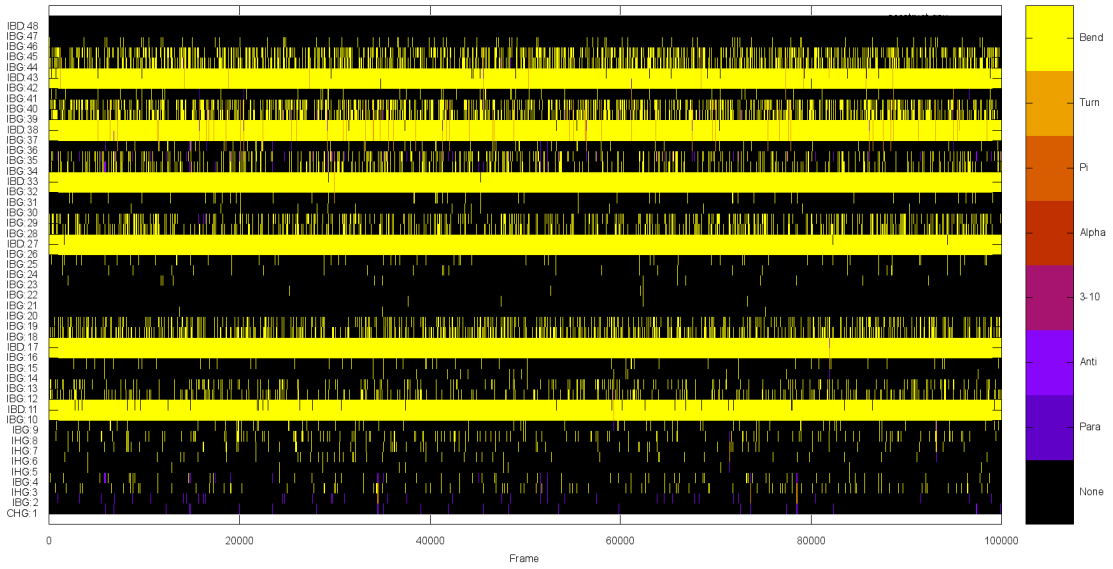


Figure 50: Secondary structure for each residue vs frames of the simulation

We finally decided to investigate the radius of gyration and the localization of the charges for the DGL G2. The radius could be easily obtained with the *radgyr* cpptraj's command and was determined to be 21.7 Å, which correlated reasonnably well with the experimental hydrodynamic radius previously reported[119]. Figure 51 shows the probability (from 0 to 1) to find the nitrogen atom of an ammonium function *versus* the center of mass of the molecule. This density of probability turned out to display a quasi-normal distribution, centered at approximately 22 Å. It is therefore reasonnable to assume that the charges of a DGL G2 are preferentially localized at the periphery of the molecule.

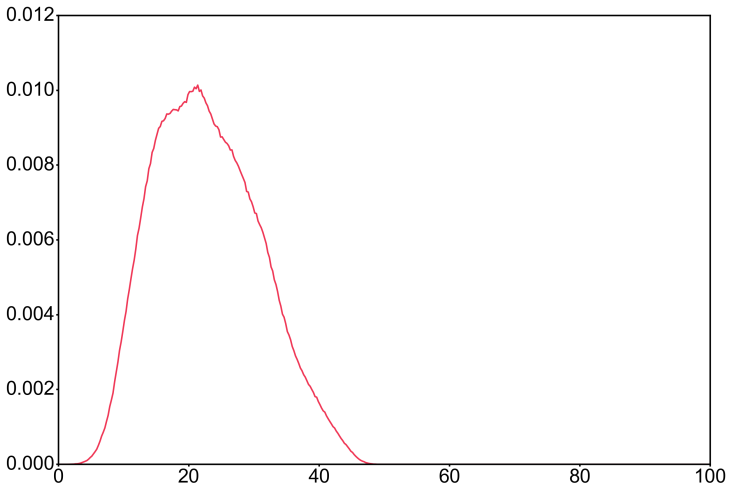


Figure 51: Probability of presence of an amonium function vs the distance (Å) to the center of mass of the molecule

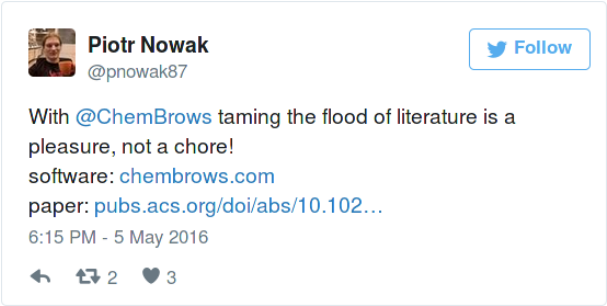
18 Conclusion

In conclusion, Molecular Dynamics simulations on the second generation Poly-*L*-Lysine Dendrigrafts unrevealed the extreme flexibility displayed by this macromolecule at physiological pH, showing a myriad of conformational states accessible to it without folded states. This conformational plasticity (*i.e.* this ability to adaptably display positive charges to make ion pairs) may explain why DGLs are such efficient binders for anionic ligands even in the most competitive media[77]. More broadly, this work provides to the community a general approach toward the construction of dendrigrafts, but also hyperbranched polymers *in silico*. We are confident that Molecular Dynamics will become a pervasive tool to explore the structural features of large dendrigrafts and hyperbranched polymers at the molecular level. We plan to extend our work to the third- and fourth-generation of the Poly-*L*-Lysine Dendrigrafts, in order to provide more detailled information about their structures. Further analyses will for example include a description of the global shape of these molecules, as well as the three-dimensionnal density of the charges.

Part VI

General conclusion

In order to respect our motto, we tried to Keep thIs document as Short and Simple as possible. To conclude, we would like to highlight the relevance of our work for the community. First of all, ChemBrows could greatly facilitate the day-to-day research work by providing the latest articles in a field[36]. To quote one of our user:



To date, more than one hundred people use our software, and we already received several messages of encouragement, giving us the intencive to persue the development of ChemBrows. Secondly, by using a commercial dendrigraft as a receptor and a short fluorescent peptide as a probe, we were able – for the first time – to detect and to quantify the widely used anticoagulant heparin in the most complex biological fluid (blood). We also investigated the thermodynamic parameters of the multi-ligand complex, which unrevealed the anti-cooperative binding of the probe to the receptor. This work, published in 2015, is already cited four times (at the date of December 1, 2016)[77].

We later adapted our sensing ensemble to create the first KISS array: a sensor array composed of a unique receptor, and a unique reporter[109]. While simple, our array was able to unambiguously discriminate five similar glycosaminoglycans. The involved mechanism, so-called compaction-displacement indicator assay, is new in the supramolecular analytical chemistry field. While developing the discrimination procedure, we improved the open-source code of a tool dedicated to machine learning, and we made our changes public.

Finally, we started a collaboration with Prof. Gérald Monard to tackle the missing description of the Poly-*L*-Lysine Dendrigrafts at the atomic level. To the best of our knowledge, we are the first to build such polymers *in silico*. While growing randomly, our three-dimensional structures respect known experimental parameters, and are fully suitable for molecular mechanics. Recent molecular dynamics simulations on the second-generation dendrigraft unrevealed the extreme flexibility of the macromolecule at physiological pH. We are currently working toward the *in silico* exploration of higher generations of Poly-*L*-Lysine Dendrigrafts.

References

[1] J.-M. Lehn, *Supramolecular Chemistry*, Wiley-VCH, 1995.

[2] M. E. Csete and J. Doyle, Reverse Engineering of Biological Complexity, *Science*, 2002, **295**, 1664, DOI: 10.1126/science.1069981.

[3] D. N. Bolon and S. L. Mayo, Enzyme-like proteins by computational design, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, 14274, DOI: 10.1073/pnas.251555398.

[4] L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Rothlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, C. F. Barbas, D. Hilvert, K. N. Houk, B. L. Stoddard and D. Baker, De Novo Compututational Design of Retro-Aldol Enzymes, *Science*, 2008, **319**, 1387, DOI: 10.1126/science.1152692.

[5] D. A. Moffet, L. K. Certain, A. J. Smithith, A. J. Kessel, K. A. Beckwith and M. H. Hecht, Peroxidase Activity in Heme Proteins Derived from a Designed Combinatorial Library, *J. Am. Chem. Soc.*, 2000, **122**, 7612, DOI: 10.1021/ja001198q.

[6] J. L. A. Jonathan W. Steed, *Supramolecular Chemistry, Second Edition*.

[7] L. Vial, R. F. Ludlow, J. Leclaire, R. Pérez-Fernández and S. Otto, Controlling the biological effects of spermine using a synthetic receptor, *J. Am. Chem. Soc.*, 2006, **128**, 10253, DOI: 10.1021/ja062536b.

[8] P. T. Corbett, J. Leclaire, L. Vial, K. R. West, J.-L. Wietor, J. K. M. Sanders and S. Otto, Dynamic combinatorial chemistry, *Chemical reviews*, 2006, **106**, 3652, DOI: 10.1021/cr020452p.

[9] R. L. Koder, J. L. R. Anderson, L. A. Solomon, K. S. Reddy, C. C. Moser and P. L. Dutton, Design and engineering of an O(2) transport protein., *Nature*, 2009, **458**, 305, DOI: 10.1038/nature07841.

[10] X. Zhang, Z. Zhang, X. Xu, Y. Li, Y. Li, Y. Jian and Z. Gu, Bioinspired therapeutic dendrimers as efficient peptide drugs based on supramolecular interactions for tumor inhibition, *Angew. Chem. Int. Ed.*, 2015, **54**, 4289, DOI: 10.1002/anie.201500683.

[11] E. J. Howe, B. O. Okesola and D. K. Smith, Self-assembled sorbitol-derived supramolecular hydrogels for the controlled encapsulation and release of active pharmaceutical ingredients, *Chem. Commun.*, 2015, **51**, 7451, DOI: 10.1039/C5CC01868D.

[12] L. C. Clark and C. Lyons, Electrode systems for continuous monitoring in cardiovascular surgery, *Ann. NY Acad. Sci.*, 1962, **102**, 29, DOI: 10.1111/j.1749-6632.1962.tb13623.x.

[13] J. Wang, Glucose biosensors: 40 Years of advances and challenges, *Electroanalysis*, 983.

[14] *Gold Book*, International Union of Pure and Applied Chemistry (IUPAC), DOI: 10.1351/goldbook.b00663.

[15] E. V. Anslyn, Supramolecular Analytical Chemistry, *J. Org. Chem.*, 2007, **72**, 687, DOI: 10.1021/jo0617971.

[16] L. You, D. Zha and E. V. Anslyn, Recent Advances in Supramolecular Analytical Chemistry Using Optical Sensing, *Chem. Rev.*, 2015, 7840, DOI: 10.1021/cr5005524.

[17] B. T. Nguyen and E. V. Anslyn, Indicator–displacement assays, *Coord. Chem. Rev.*, 2006, **250**, 3118, DOI: 10.1016/j.ccr.2006.04.009.

[18] Z. Zhong and E. V. Anslyn, A colorimetric sensing ensemble for heparin, *J. Am. Chem. Soc.*, 2002, **124**, 9014, DOI: 10.1021/ja020505k.

[19] R. B. C. Jagt, R. F. Gómez-Biagi and M. Nitz, Pattern-based recognition of heparin contaminants by an array of self-assembling fluorescent receptors, *Angew. Chem. Int. Ed.*, 2009, **48**, 1995, DOI: 10.1002/anie.200805238.

[20] S. G. Elci, D. F. Moyano, S. Rana, G. Y. Tonga, R. L. Phillips, U. H. F. Bunz and V. M. Rotello, Recognition of glycosaminoglycan chemical patterns using an unbiased sensor array, *Chem. Sci.*, 2013, **4**, 2076, DOI: 10.1039/c3sc22279a.

[21] D. B. Blossom, A. J. Kallen, P. R. Patel, A. Elward, L. Robinson, G. Gao, R. Langer, K. M. Perkins, J. L. Jaeger, K. M. Kurkjian, M. Jones, S. F. Schillie, N. Shehab, D. Ketterer, G. Venkataraman, T. K. Kishimoto, Z. Shriver, A. W. McMahon, K. F. Austen, S. Kozlowski, A. Srinivasan, G. Turabelidze, C. V. Gould, M. J. Arduino and R. Sasisekharan, Outbreak of adverse reactions associated with contaminated heparin., *N. Engl. J. Med.*, 2008, **359**, 2674, DOI: 10.1056/NEJMoa0806450.

[22] A. P. Umali, S. E. LeBoeuf, R. W. Newberry, S. Kim, L. Tran, W. A. Rome, T. Tian, D. Taing, J. Hong, M. Kwan, H. Heymann and E. V. Anslyn, Discrimination of flavonoids and red wine varietals by arrays of differential peptidic sensors, *Chem. Sci.*, 2011, **2**, 439, DOI: 10.1039/c0sc00487a.

[23] H. Collet, E. Souaid, H. Cottet, A. Deratani, L. Boiteau, G. Dessalces, J.-C. Rossi, A. Commeyras and R. Pascal, An expeditious multigram-scale synthesis of lysine dendrigraft (DGL) polymers by aqueous N-carboxyanhydride polycondensation., *Chem. Eur. J.*, 2010, **16**, 2309, DOI: 10.1002/chem.200901734.

[24] A. Commeyras, H. Collet, E. Souaid, H. Cottet, B. Romestan and O. Trambouze, *Method of preparing grafted polylysine dendrimers*, 2006, http://www.google.com/patents/WO2006114528A1?cl=en.

[25] A.-M. Caminade, D. Yan and D. K. Smith, Dendrimers and hyperbranched polymers, *Chem. Soc. Rev.*, 2015, **44**, 3870, DOI: 10.1039/C5CS90049B.

[26] *Dendrimers and Other Dendritic Polymers*, ed. J. M. J. Fréchet and D. A. Tomalia, Wiley-Blackwell, 2001, DOI: 10.1002/0470845821.

[27] D. Tomalia, *Dendrimers: Building Blocks for Nanoscale Synthesis*, 2004, vol. 37.

[28] S. J. Teertstra and M. Gauthier, Dendrigraft polymers: Macromolecular engineering on a mesoscopic scale, *Prog. Polym. Sci.*, 2004, **29**, 277, DOI: 10.1016/j.progpolymsci.2004.01.001.

[29] S. V. der Walt, S. C. Colbert and G. Varoquaux, The NumPy Array: A Structure for Efficient Numerical Computation, *Comput. Sci. Eng.*, 2011, **13**, 22, DOI: 10.1109/mcse.2011.37.

[30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and É. Duchesnay, Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.*, 2011, **12**, 2825.

[31] J.-F. Lutz, Slow science, *Nature Chem.*, 2012, **4**, 588, DOI: 10.1038/nchem.1415.

[32] L. E. Pence and H. E. Pence, Accessing and Managing Scientific Literature: Using RSS in the Classroom, *J. Chem. Educ.*, 2008, **85**, 1449, DOI: 10.1021/ed085p1449.

[33] E. Gibney, How to tame the flood of literature, *Nature*, 2014, **513**, 129, DOI: 10.1038/513129a.

[34] C. Cortes and V. Vapnik, Support-Vector Networks, *Mach. Learn.*, 1995, **20**, 273, DOI: 10.1023/A:1022627411411.

[35] C. Manning, P. Raghavan and H. Schuetze, *Introduction to Information Retrieval*, Cambridge University Press, Cambridge, England, 2008, vol. 1, DOI: 10.1017/CBO9781107415324.004.

[36] J.-P. Francoia and L. Vial, ChemBrows: An Open-Source Application Software To Keep Up to Date with the Current Literature, *J. Chem. Educ.*, 2016, DOI: 10.1021/acs.jchemed.6b00024.

[37] I. Capila and R. J. Linhardt, Heparin-Protein Interactions, *Angew. Chem. Int. Ed.*, 2002, **41**, 390.

[38] J. S. Ginsberg, Management of Venous Thromboembolism, *N. Engl. J. Med.*, 1996, **335**, 1816, DOI: 10.1056/NEJM199612123352407.

[39] D. M. A. Funk, Coagulation assays and anticoagulant monitoring., *Hematology Am. Soc. Hematol. Educ. Program*, 2012, **2012**, 460, DOI: 10.1182/asheducation-2012.1.460.

[40] M. K. Rommers, N. Van der Lely, T. C. G. Egberts and P. M. L. A. van den Bemt, Anti-Xa activity after subcutaneous administration of dalteparin in ICU patients with and without subcutaneous oedema: a pilot study., *Crit. Care*, 2006, **10**, R93, DOI: 10.1186/cc4952.

[41] S. Kitchen, Problems in laboratory monitoring of heparin dosage., *Br. J. Haematol.*, 2000, **111**, 397.

[42] S. M. Bromfield, E. Wilde and D. K. Smith, Heparin sensing and binding - taking supramolecular chemistry towards clinical applications., *Chem. Soc. Rev.*, 2013, **42**, 9184, DOI: 10.1039/c3cs60278h.

[43] H. Liu, P. Song, R. Wei, K. Li and A. Tong, A facile, sensitive and selective fluorescent probe for heparin based on aggregation-induced emission., *Talanta*, 2014, **118**, 348, DOI: 10.1016/j.talanta.2013.09.055.

[44] Z. Liu, Q. Ma, X. Wang, Z. Lin, H. Zhang, L. Liu and X. Su, A novel fluorescent nanosensor for detection of heparin and heparinase based on CuInS2 quantum dots., *Biosensors & bioelectronics*, 2014, **54**, 617, DOI: 10.1016/j.bios.2013.11.050.

[45] J. Zhao, Y. Yi, N. Mi, B. Yin, M. Wei, Q. Chen, H. Li, Y. Zhang and S. Yao, Gold nanoparticle coupled with fluorophore for ultrasensitive detection of protamine and heparin., *Talanta*, 2013, **116**, 951, DOI: 10.1016/j.talanta.2013.08.010.

[46] S. M. Bromfield, A. Barnard, P. Posocco, M. Fermeglia, S. Pricl and D. K. Smith, Mallard blue: a high-affinity selective heparin sensor that operates in highly competitive media., *J. Am. Chem. Soc.*, 2013, **135**, 2911, DOI: 10.1021/ja311734d.

[47] S. Wang and Y.-T. Chang, Discovery of heparin chemosensors through diversity oriented fluorescence library approach., *Chem. Commun.*, 2008, 1173, DOI: 10.1039/b717058k.

[48] L. Vial and P. Dumy, Nanomolar heparin detection with an artificial enzyme., *ChemBioChem*, 2008, **9**, 2950, DOI: 10.1002/cbic.200800420.

[49] J. C. Sauceda, R. M. Duke and M. Nitz, Designing fluorescent sensors of heparin., *ChemBioChem*, 2007, **8**, 391, DOI: 10.1002/cbic.200600361.

[50] W. Sun, H. Bandmann and T. Schrader, A fluorescent polymeric heparin sensor., *Chem. Eur. J.*, 2007, **13**, 7701, DOI: 10.1002/chem.200700677.

[51] A. T. Wright, Z. Zhong and E. V. Anslyn, A functional assay for heparin in serum using a designed synthetic receptor., *Angew. Chem. Int. Ed.*, 2005, **44**, 5679, DOI: 10.1002/anie.200501437.

[52] S. M. Bromfield, P. Posocco, M. Fermeglia, J. Tolosa, A. Herreros-López, S. Pricl, J. Rodríguez-López and D. K. Smith, Shape-persistent and adaptive multivalency: rigid transgeden (TGD) and flexible PAMAM dendrimers for heparin binding., *Chem. Eur. J.*, 2014, **20**, 9666, DOI: 10.1002/chem.201402237.

[53] S. M. Bromfield, P. Posocco, C. W. Chan, M. Calderon, S. E. Guimond, J. E. Turnbull, S. Pricl and D. K. Smith, Nanoscale self-assembled multivalent (SAMul) heparin binders in highly competitive, biologically relevant, aqueous media, *Chem. Sci.*, 2014, **5**, 1484, DOI: 10.1039/c4sc00298a.

[54] S. M. Bromfield, P. Posocco, M. Fermeglia, S. Pricl, J. Rodríguez-López and D. K. Smith, A simple new competition assay for heparin binding in serum applied to multivalent PAMAM dendrimers., *Chem. Commun.*, 2013, **49**, 4830, DOI: 10.1039/c3cc41251b.

[55] A. C. Rodrigo, A. Barnard, J. Cooper and D. K. Smith, Self-assembling ligands for multivalent nanoscale heparin binding., *Angew. Chem. Int. Ed.*, 2011, **50**, 4675, DOI: 10.1002/anie.201100019.

[56] R. L. Thurlkill, R. L. Thurlkill, G. R. Grimsley, G. R. Grimsley, J. M. Scholtz, J. M. Scholtz, C. N. Pace and C. N. Pace, pK values of the ionizable groups of proteins, *Protein Sci*, 2006, **15**, 1214, DOI: 10.1110/ps.051840806.

[57] B. Maret, T. Regnier, J.-C. Rossi, L. Garrelly, L. Vial and R. Pascal, Reduction with tris(2-carboxyethyl)phosphine (TCEP) enables the use of an S-sulphonate protecting group for thiol-mediated bioconjugation, *RSC Adv.*, 2014, **4**, 7725, DOI: 10.1039/c3ra47407k.

[58] T. Liu, F. Oukacine, H. Collet, A. Commeyras, L. Vial and H. Cottet, author, DOI: 10.1016/j.chroma.2012.11.074.

[59] D. A. Dougherty and J. C. Ma, The Cation−π Interaction, *Chem. Rev.*, 1997, **97**, 130, DOI: 10.1021/cr9603744.

[60] M. Bonizzoni, S. R. Long, C. Rainwater and E. V. Anslyn, PAMAM dendrimer-induced aggregation of 5(6)-carboxyfluorescein., *J. Org. Chem*, 2012, **77**, 1258, DOI: 10.1021/jo201360u.

[61] J. C. Rainwater and E. V. Anslyn, Amino-terminated PAMAM dendrimers electrostatically uptake numerous anionic indicators., *Chem. Commun.*, 2010, **46**, 2904, DOI: 10.1039/b925229k.

[62] S. Jockusch, N. J. Turro and D. A. Tomalia, Aggregation of Methylene Blue Adsorbed on Starburst Dendrimers, *Macromolecules*, 1995, **28**, 7416, DOI: 10.1021/ma00126a020.

[63] B. Walter, *Ann. Phys*, 1888, **34**, 316.

[64] R. F. Chen and J. R. Knutson, Mechanism of fluorescence concentration quenching of carboxyfluorescein in liposomes: energy transfer to nonfluorescent dimers., *Anal. Biochem.*, 1988, **172**, 61.

[65] A. K. Bordbar, A. A. Saboury and A. A. Moosavi-Movahedi, The Shapes of Scatchard Plots for Systems with Two Sets of Binding Sites, *Biochem. Educ.*, 1996, **24**, 172, DOI: 10.1016/0307-4412(95)00122-0.

[66] J. N. Weiss, The Hill equation revisited: uses and misuses, *FASEB*, 1997, **11**, 835.

[67] I. M. Klotzt and D. L. Hunstonl, author, DOI: 10.1021/bi00792a013.

[68] Y. Shao, L. F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. a. DiStasio, R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Y. Lin, T. Van Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C.-P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. Woodcock, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer, J. Kong, A. I. Krylov, P. M. W. Gill and M. Head-Gordon, Advances in methods and algorithms in a modern quantum chemistry program package., *Phys. Chem. Chem. Phys.*, 2006, **8**, 3172, DOI: 10.1039/b517914a.

[69] F. N. Lamari, A. D. Theocharis, A. P. Asimakopoulou, C. J. Malavaki and N. K. Karamanos, Metabolism and biochemical/physiological roles of chondroitin sulfates: analysis of endogenous and supplemental chondroitin sulfates in blood circulation., *Biomed. Chromatogr.*, 2006, **20**, 539, DOI: 10.1002/bmc.669.

[70] B. Romestand, J.-l. Rolland, A. Commeyras, I. Desvignes, R. Pascal and O. Vandenabeele-trambouze, Dendrigraft Poly-L-lysine: A Non-Immunogenic Synthetic Carrier for Antibody Production, *Biomacromolecules*, 2010, **11**, 1169, DOI: 10.1021/bm9012056.

[71] P. J. A. Weber, J. E. Bader, G. Folkers and A. G. Beck-sickinger, A fast and inexpensive method for N-terminal fluorescein-labeling of peptides, *Bioorg. Med. Chem. Lett.*, 1998, **8**, 597, DOI: 10.1016/S0960-894X(98)00084-5.

[72] N. Volpi, Therapeutic Applications of Glycosaminoglycans, *Curr. Med. Chem.*, 2006, **13**, 1799, DOI: 10.2174/092986706777452470.

[73] S. Beni, J. F. K. Limtiaco and C. K. Larive, Analysis and characterization of heparin impurities, *Anal. Bioanal. Chem.*, 2011, **399**, 527, DOI: 10.1007/s00216-010-4121-x.

[74] M. Guerrini, Z. Zhang, Z. Shriver, A. Naggi, S. Masuko, R. Langer, B. Casu, R. J. Linhardt, G. Torri and R. Sasisekharan, Orthogonal analytical approaches to detect potential contaminants in heparin., *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 16956, DOI: 10.1073/pnas.0906861106.

[75] M. Twomey, T. Vokatá, M. R. Kumar and J. H. Moon, Differential interactions of conjugated polymer nanoparticles with glycosaminoglycans in synthetic urine, *Chem. Commun.*, 2015, **51**, 4065, DOI: 10.1039/C5CC00110B.

[76] E. K. Nyren-Erickson, M. K. Haldar, Y. Gu, S. Y. Qian, D. L. Friesner and S. Mallik, Fluorescent liposomes for differential interactions with glycosaminoglycans, *Anal. Chem.*, 2011, **83**, 5989, DOI: 10.1021/ac2009993.

[77] J.-P. Francoia, R. Pascal and L. Vial, Monitoring clinical levels of heparin in human blood samples with an indicator-displacement assay, *Chem. Commun.*, 2015, **51**, 1953, DOI: 10.1039/C4CC08563A.

[78] A. T. Wright and E. V. Anslyn, Differential receptor arrays and assays for solution-based molecular recognition., *Chem. Soc. Rev.*, 2006, **35**, 14, DOI: 10.1039/b505518k.

[79] J. R. Beveridge, *The Geometry of LDA and PCA Classifiers Illustrated with 3D Examples*, 2001.

[80] S. Raschka, *Python Machine Learning*, Packt Publishing, 2015.

[81] S. Stewart, M. A. Ivy and E. V. Anslyn, The use of principal component analysis and discriminant analysis in differential sensing routines., *Chem. Soc. Rev.*, 2014, **43**, 70, DOI: 10.1039/c3cs60183h.

[82] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning Data*, Springer, New York, 2009, DOI: 10.1007/978-0-387-84858-7.

[83] G. A. Neville, F. Mori, K. R. Holme and A. S. Perlin, Monitoring the purity of pharmaceutical heparin preparations by high-field 1H-nuclear magnetic resonance spectroscopy., *J. Pharm. Sci.*, 1989, **78**, 101.

[84] T. Michels and D. Rudolf, Acid-Mediated Prevention of Aspartimide Formation in Solid Phase Peptide Synthesis, *Org. Lett.*, 2012, **14**, 5218, DOI: 10.1021/ol3007925.

[85] Y. Kodama, T. Nakamura, T. Kurosaki, K. Egashira, T. Mine, H. Nakagawa, T. Muro, T. Kitahara, N. Higuchi and H. Sasaki, Biodegradable nanoparticles composed of dendrigraft poly-l-lysine for gene delivery, *Eur. J. Pharm. Biopharm.*, 2014, **87**, 472, DOI: 10.1016/j.ejpb.2014.04.013.

[86] M. Tang, H. Dong, Y. Li and T. Ren, Harnessing the PEG-cleavable strategy to balance cytotoxicity, intracellular release and the therapeutic effect of dendrigraft poly-L-lysine for cancer gene therapy, *J. Mater. Chem. B*, 2016, **4**, 1284, DOI: 10.1039/C5TB02224J.

[87] H. Yao, K. Wang, Y. Wang, S. Wang, J. Li, J. Lou, L. Ye, X. Yan, W. Lu and R. Huang, Enhanced blood–brain balancerrier penetration and glioma therapy mediated by a new peptide modified gene delivery system, *Biomaterials*, 2015, **37**, 345, DOI: 10.1016/j.biomaterials.2010014.10.034.

[88] L. Han, Y. Guo, H. Ma, X. He, Y. Kuang, N. Zhang, E. Lim, W. Zhou and C. Jiang, Acid active receptor-specific peptide ligand for in vivo tumor-targeted delivery, *Small*, 2013, **9**, 3647, DOI: 10.1002/smll.201300279.

[89] S. An, Y. Kuang, T. Shen, J. Li, H. Ma, Y. Guo, X. He and C. Jiang, Brain-targeting delivery for RNAi neuroprotection against cerebral ischemia reperfusion injury, *Biomaterials*, 2013, **3454**, 8949, DOI: 10.1016/j.biomaterials.2013.07.060.

[90] Y. Liu, Y. Guo, S. An, Y. Kuang, X. He, H. Ma, J. Li, J. Lv, N. Zhang and C. Jiang, Targeting Caspase-3 as Dual Therapeutic Benefits by RNAi Facilitating Brain-Targeted Nanoparticles in a Rat Model of Parkinson's Disease, *PLoS ONE*, 2013, **8**, e62905, DOI: 10.1371/journal.pone.0062905.

[91] J. Hofman, M. Buncek, R. Haluza, L. Streinz, M. Ledvina and P. Cigler, In Vitro Transfection Mediated by Dendrigraft Poly(L-lysines): The Effect of Structure and Molecule Size, *Macromol. Biosci.*, 2012, **13**, 167, DOI: 10.1002/mabi.201200303.

[92] Y. Liu, J. Li, K. Shao, R. Huang, L. Ye, J. Lou and C. Jiang, A leptin derived 30-amino-acid peptide modified pegylated poly-l-lysine dendrigraft for brain targeted gene delivery, *Biomaterials*, 2010, **31**, 5246, DOI: 10.1016/j.biomaterials.2010.03.011.

[93] R. Huang, S. Liu, K. Shao, L. Han, W. Ke, Y. Liu, J. Li, S. Huang and C. Jiang, Evaluation and mechanism studies of PEGylated dendrigraft poly-L-lysines as novel gene delivery vectors, *Nanotechnology*, 2010, **21**, 265101, DOI: 10.1088/0957-4484/21/26/265101.

[94] G. Hu, H. Zhang, L. Zhang, S. Ruan, Q. He and H. Gao, Integrin-mediated active tumor targeting and tumor microenvironment response dendrimer-gelatin nanoparticles for drug delivery and tumor treatment, *Int. J. Pharm.*, 2015, **496**, 1057, DOI: 10.1016/j.ijpharm.2015.11.025.

[95] S. Huang, K. Shao, Y. Liu, Y. Kuang, J. Li, S. An, Y. Guo, H. Ma and C. Jiang, Tumor-Targeting and Microenvironment-Responsive Smart Nanoparticles for Combination Therapy of Antiangiogenesis and Apoptosis, *ACS Nano*, 2013, **7**, 2860, DOI: 10.1021/nn400548g.

[96] S. Liu, Y. Guo, R. Huang, J. Li, S. Huang, Y. Kuang, L. Han and C. Jiang, Gene and doxorubicin co-delivery system for targeting therapy of glioma, *Biomaterials*, 2012, **33**, 4907, DOI: 10.1016/j.biomaterials.2012.03.031.

[97] Z. Sideratou, N. Sterioti, D. Tsiourvas, L.-A. Tziveleka, A. Thanassoulas, G. Nounesis and C. M. Paleos, Arginine end-functionalized poly(l-lysine) dendrigrafts for the stabilization and controlled release of insulin, *J. Colloid Interface Sci.*, 2010, **351**, 433, DOI: 10.1016/j.jcis.2010.07.072.

[98] C. Lorion, C. Faye, B. Maret, T. Trimaille, T. Régnier, P. Sommer and R. Debret, Biosynthetic support based on dendritic poly(L-lysine) improves human skin fibroblasts attachment, *J. Biomater. Sci. Polym. Ed.*, 2013, **25**, 136, DOI: 10.1080/09205063.2013.843966.

[99] L. Keller, S. Eap, J. Schiavi, O. Huck, L. Jacomine, C. Gauthier, V. Sebastian, P. Schwinté, N. Jessel and F. Fioretti, A living thick nanofibrous implant bifunctionalized with active growth factor and stem cells for bone regeneration, *Int. J. Nanomedicine*, 2015, 1061, DOI: 10.2147/ijn.s72670.

[100] S. Eap, T. Bécavin, L. Keller, T. Kökten, F. Fioretti, J.-L. Weickert, E. Deveaux, N. Benkirane-Jessel and S. Kuchler-Bopp, Nanofibers Implant Functionalized by Neural Growth Factor as a Strategy to Innervate a Bioengineered Tooth, *Adv. Healthc. Mater*, 2014, **3**, 386, DOI: 10.1002/adhm.201300281.

[101] A. Cadiere, B. Couturaud, J. Boismard, P. L. Cann, A. Gérard, A. Mas, C. Faye, L. Garrelly and B. Roig, Assessment of poly- l -lysine dendrigrafts for virus concentration in water: use of MS2 bacteriophage as proof of concept, *J. Appl. Microbiol.*, 2013, **115**, 290, DOI: 10.1111/jam.12209.

[102] C. Mendoza-Palomares, A. Ferrand, S. Facca, F. Fioretti, G. Ladam, S. Kuchler-Bopp, T. Regnier, D. Mainard and N. Benkirane-Jessel, Smart Hybrid Materials Equipped by Nanoreservoirs of Therapeutics, *ACS Nano*, 2012, **6**, 483, DOI: 10.1021/nn203817t.

[103] F. Fioretti, C. Mendoza-Palomares, M. Helms, D. A. Alam, L. Richert, Y. Arntz, S. Rinckenbach, F. Garnier, Y. HaiĚkel, S. C. Gangloff and N. Benkirane-Jessel, Nanostructured Assemblies for Dental Application, *ACS Nano*, 2010, **4**, 3277, DOI: 10.1021/nn100713m.

[104] L. Han, M. Liu, D. Ye, N. Zhang, E. Lim, J. Lu and C. Jiang, Tumor cell membrane-targeting pH-dependent electron donor-acceptor fluorescence systems with low background signals, *Biomaterials*, 2014, **35**, 2952, DOI: 10.1016/j.biomaterials.2013.12.020.

[105] J. Li, S. Huang, K. Shao, Y. Liu, S. An, Y. Kuang, Y. Guo, H. Ma, X. Wang and C. Jiang, A choline derivate-modified nanoprobe for glioma diagnosis using MRI., *Sci. Rep.*, 2013, **3**, 1, DOI: 10.1038/srep01623.

[106] R. Huang, L. Han, J. Li, S. Liu, K. Shao, Y. Kuang, X. Hu, X. Wang, H. Lei and C. Jiang, Chlorotoxin-modified macromolecular contrast agent for MRI tumor diagnosis, *Biomaterials*, 2011, **32**, 5177, DOI: 10.1016/j.biomaterials.2011.03.075.

[107] T. A. Theodossiou, Z. Sideratou, D. Tsiourvas and C. M. Paleos, A novel mitotropic oligolysine nanocarrier: Targeted delivery of covalently bound D-Luciferin to cell mitochondria, *Mitochondrion*, 2011, **11**, 982, DOI: 10.1016/j.mito.2011.08.004.

[108] C. Kojima, B. Turkbey, M. Ogawa, M. Bernardo, C. A. Regino, L. H. Bryant, P. L. Choyke, K. Kono and H. Kobayashi, Dendrimer-based MRI contrast agents: the effects of PEGylation on relaxivity and pharmacokinetics, *Nanomedicine*, 2011, **7**, 1001, DOI: 10.1016/j.nano.2011.03.007.

[109] J.-P. Francoia and L. Vial, A KISS (keep it simple, sensor) array for glycosaminoglycans, *Chem. Commun.*, 2015, **51**, 17544, DOI: 10.1039/C5CC07628E.

[110] B. Couturaud, A. M. Bondia, C. Faye, L. Garrelly, A. Mas and J. J. Robin, Grafting of poly-l-lysine dendrigrafts onto polypropylene surface using plasma activation for ATP immobilization – Nanomaterial for potential applications in biotechnology, *Journal of Colloid and Interface Science*, 2013, **408**, 242, DOI: 10.1016/j.jcis.2013.06.065.

[111] W. Humphrey, A. Dalke and K. Schulten, VMD: visual molecular dynamics., *J. Mol. Graph.*, 1996, **14**, 33, DOI: 10.1016/0263-7855(96)00018-5.

[112] N. A. Baker, D. Sept, S. Joseph, M. J. Holst and J. A. McCammon, Electrostatics of nanosystems: application to microtubules and the ribosome., *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 10037, DOI: 10.1073/pnas.181342398.

[113] G. J. M. Koper and M. Borkovec, Proton binding by linear, branched, and hyperbranched polyelectrolytes, *Polymer*, 2010, **51**, 5649, DOI: 10.1016/j.polymer.2010.08.067.

[114] C. N. Pace, G. R. Grimsley and J. M. Scholtz, Protein ionizable groups: pK values and their contribution to protein stability and solubility, *J. Biol. Chem.*, 2009, **284**, 13285, DOI: 10.1074/jbc.R800080200.

[115] D. A. Case, R. M. Betz, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao and P. A. Kollman, AMBER 16, University of California, San Francisco, 2016.

[116] C. W. Hopkins, S. L. Grand, R. C. Walker and A. E. Roitberg, Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning, *J. Chem. Theory Comput.*, 2015, **11**, 1864, DOI: 10.1021/ct5010406.

[117] D. R. Roe and T. E. Cheatham III, PTRAJ and CPPTRAJ: software for processing and analysis of molecular synamics trajectory data, *J. Chem. Theory Comput.*, 2013, **9**, 3084, DOI: 10.1021/ct400341p.

[118] W. Kabsch and C. Sander, Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers*, 1983, **22**, 2577, DOI: 10.1002/bip.360221211.

[119] M. Martin, A. Papillaud and A. Commeyras, Determination of Dendrigraft Poly-L-Lysine Diffusion Coefficients by Taylor Dispersion Analysis, *Biomacromolecules*, 2007, **8**, 3235, DOI: 10.1021/bm070268j.

Summary

In this thesis, we report that a ”tree-like” dendrigraft of poly-*L*-Lysine (DGL) is able to form a multi-ligand complex with a fluorescently labelled peptide, leading to the almost complete extinction of the optical signal that can be restored upon the introduction of heparin. This simple system allows, for the first time, the turn-ON fluorescent sensing of the anticoagulant in human blood at clinically relevant levels.

Then, we demonstrate that this sensing ensemble can evolve toward a sensor array. Depending on the loading of the indicator on the receptor, negatively charged glycosaminoglycans (GAGs) induce a positive or negative variation of the fluorescent signal as they displace the indicators from the receptor or they compact the indicators on the receptor's surface, respectively. This unique strategy allows not only the blind identification of pure GAGs with a level of accuracy of 100%, but also the differentiation of mixtures.

We also report an original and simple methodology for the construction of three-dimensional structures of DGLs, and the subsequent investigation of their structural features using molecular dynamics simulations. This methodology relies on the encoding of the polymers' experimental characterizations (*i.e.* degrees of polymerization, branching ratios, charges) into alphanumeric strings that are ”readable” by the Amber simulation package. This work opens avenues toward the *in silico* exploration of dendrigrafts and hyperbranched polymers.

Finally, we developed ChemBrows, an in-house software that will significantly help scientists/teachers/ students to tame the flood of publications. Working as an enhanced RSS reader that integrates keyword-based filters and a machine-learning-based recommendation engine, ChemBrows is available on multiple platforms as a free and open-source software at www.chembrows.com.

Keywords : Dendrigrafts, Sensors, Glycosaminoglycans, Fluorescence, Literature survey, Python

Résumé

Dans cette thèse, nous rapportons qu'un dendrigraft de poly-*L*-Lysine (DGL) est capable de former un complexe multi-ligand avec un peptide fluorescent, conduisant à la disparition quasi-totale du signal optique, qui peut ensuite être restauré en présence d'héparine. Ce système simple permet, pour la première fois, la détection de l'anticoagulant dans le sang humain à des doses cliniques.

Ensuite, nous démontrons que ce système simple peut évoluer vers une barrette de senseurs. En fonction de la quantité d'indicateur sur le récepteur, des glycosaminoglycanes chargés négativement (GAG) induisent une variation du signal fluorescent selon s'ils déplacent ou compactent les indicateurs à la surface du récepteur. Cette stratégie unique permet non seulement l'identification aveugle de GAGs purs avec un niveau de précision de 100 %, mais aussi la différenciation de mixtures.

Nous présentons également une méthodologie originale et simple pour la construction de structures tridimensionelles des DGLs. Nous étudions ensuite les caractéristiques structurelles de ces polymères par dynamique moléculaire. Cette méthodologie repose sur l'encodage des caractéristiques expérimentales des DGLs (*i.e.* degré de polymérisation, rapports de branchement, charges) en chaînes alphanumériques, qui seront ensuite interprétées par le programme de mécanique moléculaire Amber. Ce travail ouvre des perspectives pour l'exploration *in silico* des propriétés des dendrigrafts et des polymères hyperbranchés.

Finalement, nous avons développé ChemBrows, un logiciel qui assiste les scientifiques/enseignants/étudiants dans leur veille bibliographique. Fonctionnant comme un lecteur de flux RSS amélioré qui intègre des filtres par mots-clés et un moteur de recommandation basé sur l'apprentissage machine, ChemBrows est un logiciel libre et open-source: www.chembrows.com.

Mots clés : Dendrigrafts, Senseurs, Glycosaminoglycanes, Fluorescence, Veille bibliographique, Python