UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES

Institut Charles Sadron, UPR22-CNRS

THÈSE présentée par :

Nicolas CAPIT

soutenue le : **7 décembre 2021**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Chimie Organique et Supramoléculaire

Designing a reversible peptide system for out of equilibrium chemistry exploration

THÈSE dirigée par :

M. GIUSEPPONE Nicolas	Professeur, ICS, Université de Strasbourg
Mme MOULIN Émilie	Directrice de Recherche, ICS, CNRS

RAPPORTEURS :

Mme MARCHI Valérie	Directrice de Recherche, ISCR, CNRS
M. JULLIEN Ludovic	Professeur, Sorbonne Université, ENS Ulm

AUTRES MEMBRES DU JURY :

M. PLASSON Raphaël	Maître de conférences, INRAE, Avignon Université

Résumé en français :

Introduction :

Là ! Vous venez de le refaire. Vous venez de respirer, d'inspirer de l'air (plus ou moins frai) contenant autour de 20% de dioxygène, ce gaz essentiel à notre survie. Pourquoi cela ? Car dans vos cellules il est utilisé pour oxyder de la matière organique, et cette réaction libère l'énergie nécessaire à notre existence. La vie par essence requiert de l'énergie. Tout être vivant utilise une source d'énergie. Ce peut être de l'énergie sous forme de matière organique à l'état réduite dans ses nutriments ou de l'énergie sous forme de lumière dans l'exemple de la photosynthèse. Les êtres vivants, contrairement à nombres de matériaux que l'Homme a sans cesse chercher à rendre stables dans le temps, sont fragiles. Leurs composants ne sont même pas thermodynamiquement stables dans leur environnement. Ce n'est qu'à la mort d'un être vivant que ses constituants évoluent vers l'équilibre. C'est pour se maintenir dans cet état d'hors équilibre qu'un être vivant a besoin d'une source d'énergie. En chimie, on nomme un système capable d'extraire de l'énergie de son milieu et de l'utiliser, un système dissipatif. Si la stabilité a longtemps été l'intérêt principal, les systèmes dissipatifs génèrent de plus en plus d'intérêts.

Les êtres vivants sont en effet capables de nombreuses prouesses. Le mouvement, par exemple, et ce même à l'échelle moléculaire. Les kinésines peuvent se déplacer le long de rails en trainant avec elles un cargo, les ATP-synthases et leurs mouvements rotatoires peuvent transformer de l'énergie d'une forme à une autre ou simplement imprimer ce mouvement à un flagelle et permettre la propulsion d'une bactérie. Dans tous ces cas, il y a dissipation d'énergie et création de travail. Mais il existe d'autres propriétés dissipatives. Les cellules sont capables de remarquables systèmes de communication, tel que la différenciation lors de l'embryogénèse, permise par des systèmes dissipatifs de type réaction-diffusion. Il en est de même pour des composants structuraux. Ainsi les microtubules sont capables de s'assembler dynamiquement en fonction des besoins de la cellule, sous réserve qu'elle dispose de l'énergie nécessaire pour les maintenir. Les concentrations en différentes molécules, notamment les ions, sont aussi maintenus par dissipation d'énergie par de nombreux systèmes de pompes et canaux, assurant ainsi le bon fonctionnement des cellules.

Afin de mettre au point les matériaux de demain possédant des propriétés similaires, il est nécessaire de savoir quels sont les composants nécessaires à la création d'un système dissipatif. Pour cela, il faut donc un système pouvant être mis dans un état d'hors équilibre et qui puisse dissiper l'énergie présente dans son milieu. Cela est l'objectif du présent travail de thèse. Pour ce faire, nous avons pensé à un système capable de polymérisation réversible. Dans un gradient de température, si le système possède des propriétés thermodynamiques adéquates, il y aurait donc différents polymères stables à différentes températures. Les espèces stables dans certaines températures pourraient alors diffuser vers des températures où leur formation ne devraient pas être favorable entraînant des états hors équilibres. Les travaux et résultats seront présentés

comme suit. Tout d'abord, sont abordés un peu plus en détails les phénomènes d'hors équilibre et quelques exemples de systèmes biologiques et de synthèses similaires sont donnés. Puis est abordé la mise au point et l'études des données thermocinétiques de la réaction de polymérisation réversible, à la suite de quoi les résultats d'évolutions du système à différentes températures sont présenté suivis de travaux de modélisations.

Partie 1 : Contextualisation

Il existe trois principes en thermodynamique. Le premier est celui de la conservation d'énergie, le deuxième introduit la notion d'entropie et traite de l'évolution d'un système et le dernier considère que l'entropie d'un cristal parfait est nulle au zéro absolu. En introduisant le principe d'évolution dans celui de conservation d'énergie, il en ressort l'existence d'une grandeur noté *G*, nommée énergie libre de Gibbs qui diminue lors de processus thermodynamique et qui est donc minimale à l'équilibre. Le calcul de ΔG permet donc de savoir si un système va spontanément évoluer vers tel ou tel état. On définit alors l'équilibre comme : « un état stable dans lequel il n'y a ni échange de matière avec le milieu, ni de flux net d'énergie au travers du système ».[1]

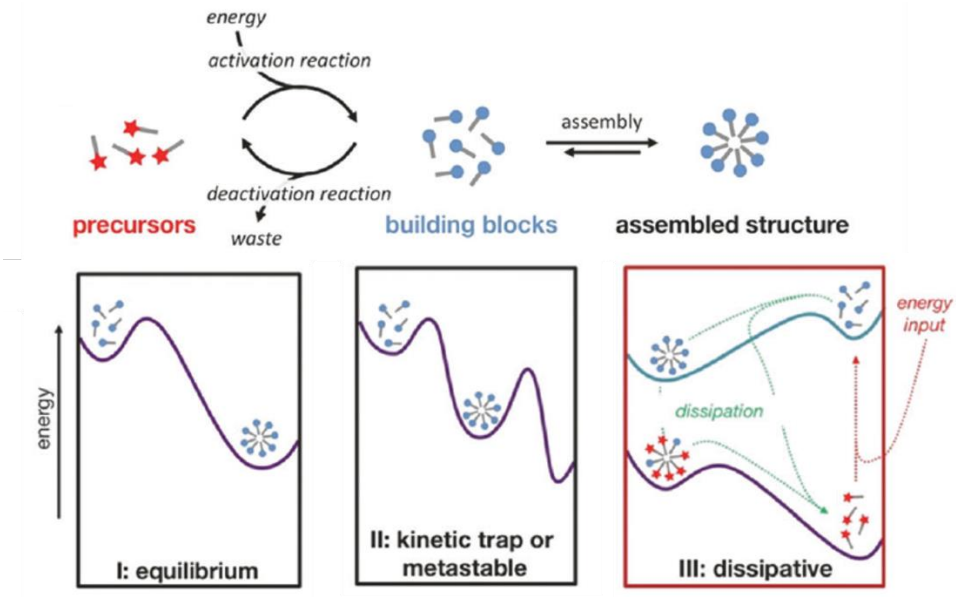


Figure 1 Un système de précurseurs pouvant être activé pour s'auto assembler et en dessous différentes possibilités d'évolutions en fonction du profil d'énergie du système, tiré de [1]

Si l'on considère le système ci-dessus du point de vue de son énergie libre, il y a donc différentes possibilités d'évolution. Tout d'abords, si l'assemblage est plus bas en énergie alors l'évolution vers l'équilibre mènera vers l'assemblage. Parfois, la barrière énergétique menant à l'état le plus stable est cependant trop importante. Le système est alors piégé dans un état qui n'est pas celui d'équilibre, on parle d'état métastable. Si les éléments seuls sont plus stables que l'assemblage alors celui-ci ne se créera pas spontanément. Mais si les éléments peuvent être amenés dans un état énergétique plus important que l'assemblage alors celui se formera mais nécessitera le maintien de cette énergie pour garder les constituants dans cet état énergétique

supérieur. On a alors un système hors équilibre car il y a transfert d'énergie du milieu vers le système qui permet l'assemblage.

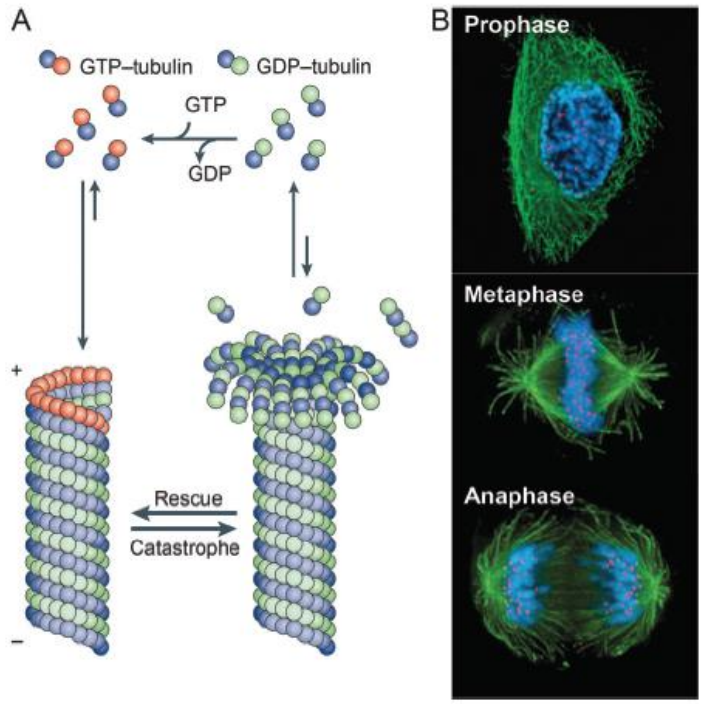


Figure 2 A schéma structurelle d'un microtubule. B Image de microscopie de cellules à différents stades du cycle cellulaire, les microtubules sont marqués en fluorescence verte.[5]

Un exemple d'un tel système dans la nature est celui des microtubules. Les microtubules sont composés de dimères de tubuline qui s'associent entre eux.[2] Les dimères n'ont cependant pas de forte affinité les uns pour les autres et ne s'assemblent pas spontanément. Ils ont en revanche une forte affinité pour une molécule à fort potentiel énergétique, la guanosine triphosphate (GTP). Un dimère de tubuline associé à une GTP possède alors une forte affinité pour d'autres tubuline-GTP et s'auto-assemble pour former le microtubule. Le GTP, instable, finit par être hydrolysé en guanosine diphosphate (GDP), retournant les dimères de tubuline vers un état de faible affinité des uns pour les autres menant au désassemblage du microtubule. On a alors une extrémité par lequel le microtubule s'assemble, riche en GTP, et une extrémité où il y a dissociation à la suite de l'hydrolyse du GTP en GDP. Il en résulte que le microtubule est un assemblage dynamique ce qui permet la réorganisation sans cesse du réseau de microtubule en fonction des besoins de la cellule. Ainsi, le réseau n'est pas du tout le même aux différents stades du cycle cellulaire et la dépolymérisation spontanée du microtubule est un élément clef pour la traction des chromosomes lors de la mitose au niveau des kinétochores.

Des structures dissipatives ont aussi été mises au point en laboratoire. Ainsi des nanoparticules d'or recouvertes avec entre autres des azobenzène sont capable de s'agréger dans certaines conditions. Les azobenzènes sont des molécules capables d'exister en configuration trans ou cis et qui passent d'une configuration à l'autre sous irradiation ultraviolette (UV) ou

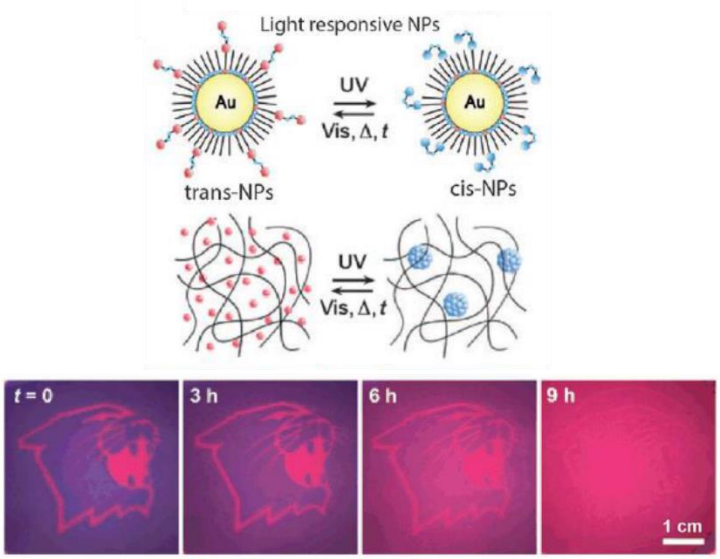


Figure 3 Dessus Fonctionnement schématique de l'isomérisation des azobenzènes recouvrant des nanoparticules d'or et gouvernant leur agrégation en fonction de l'irradiation. Bas Motif obtenu par masquage de lumière UV et sa disparition lors de l'irradiation du gel dans le visible. Figure tirée de [6]

visible. En incorporant ces nanoparticules dans un gel, on crée alors une « photo-encre » réversible. Dans les zones irradiées par des UV, les azobenzènes seront en configuration cis et les nanoparticules s'agrègent, alors que dans les zones sans UV, les particules restent en suspension sans agrégats et si le gel est laissé sous irradiation dans le domaine du visible les agrégats disparaissent au fur et à mesure que la conformation trans redevient majoritaire partout dans le gel.

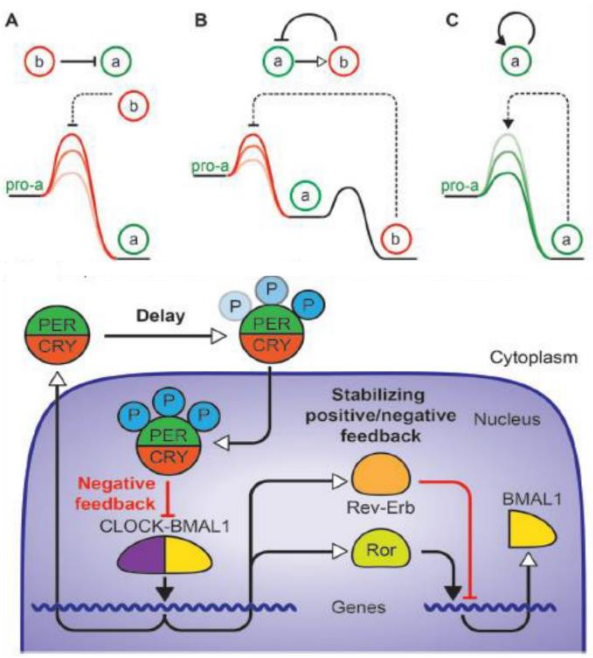


Figure 4 Haut, A b inhibe la formation de a, B rétroaction négative car b inhibe la formation de son précurseur a C boucle de rétroaction positive a amplifie sa formation. Bas, schéma de fonctionnement de CLOCK-BMAL1 intervenant dans le cycle circadien. Figure tirée de [1]

Il existe d'autres possibilité pour avoir des systèmes dissipatifs hors-équilibre. Notamment la mise en place de réseaux de réactions dont les concentrations des différents constituants influent directement sur les réactions de formations d'autres éléments du réseau de réaction. Ainsi comme dans l'exemple A, un élément peut augmenter la barrière énergétique nécessaire à la formation d'une molécule, ce qui conduit à la diminution de la cinétique de la réaction impactée. Lorsque cette intervention se fait en amont de l'élément formé on parle alors de boucle de rétroaction, dans le cas d'une inhibition (augmentation de la barrière énergétique par exemple) on parle de rétro action négative (exemple B), mais il est également possible d'avoir une activation avec par exemple une diminution de la barrière énergétique, on a alors une amplification ou rétroaction positive. Un exemple d'un réseau avec de tels interactions est le cycle circadien. Le cycle circadien, est un oscillateur chimique que nous possédons dont la période d'oscillation est synchrone avec celle d'une journée et permet la régulation de différentes fonctions vitales. En l'occurrence l'association un facteur de transcription CLOCK-BMAL1 à l'ADN dans le noyau des cellules va avoir pour impact la synthèse de la protéine PER-CRY ainsi que celle de BMAL1. BMAL1 peut rentrer dès sa synthèse dans le noyau et ainsi amplifier cet effet (rétroaction positive) alors que PER-CRY doit être phosphorylée pour rentrer dans le noyau et pouvoir inhiber la synthèse de CLOCK-BMAL1, il en résulte des oscillations en concentrations de ces différentes protéines dans le temps d'une période de 24 heures.

Un oscillateur de synthèse connu est le système de Belousov-Zhabotinsky. Ce système repose sur l'oxydation d'acide malonique présent dans le système en présence de bromate, de bromure et de cérium ou de ruthénium. Le bromure réagis en préférence avec les bromates pour créer du brome (R1-3) créant au passage l'espèce $HBrO_2$ (R3) pendant que le brome réagit avec l'acide malonique (R8). Tant qu'il y a du bromure dans le milieu celui-ci réagit avec $HBrO_2$ (inhibition), mais lorsque tout le bromure a réagi, le $HBrO_2$ résiduel réagit avec le bromate et le cérium III dans le milieu et l'oxyde en cérium IV (R5-6) (rétroaction positive) tout en recréant du $HBrO_2$. Cela se poursuit jusqu'à épuisement du cérium III et la mise en place de réactions plus lentes qui vont réduire le cérium et libérer des ions bromures en oxydant l'acide malonique bromé (R10). Il en résulte donc un système où oscillent concentration en ion bromure et cérium IV. Ce système a déjà été couplé à des structures tels que des copolymères à base de ruthénium capable de former des micelles en fonction du degré d'oxydation du celui-ci. La turbidité de la solution oscille alors en même temps que le degré d'oscillation majoritaire du ruthénium. Si le milieu n'est pas agité alors des motifs dus à la compétition entre réaction et diffusion des différentes espèces dans le milieu apparaissent démontrant la possibilité d'avoir des vagues de signaux diffusant au travers d'un matériau ou la possibilité de stocker de l'information en tant que motif de réaction diffusion.

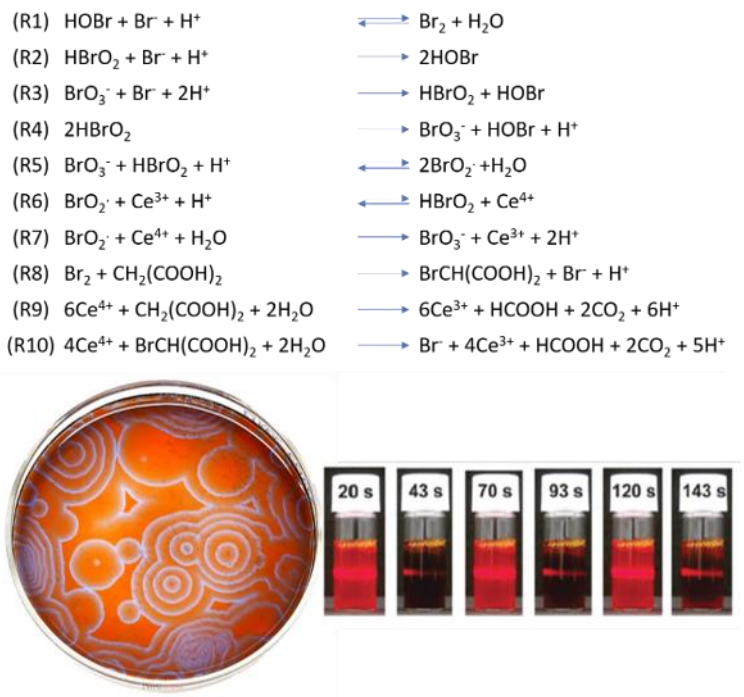


Figure 5 Haut, résumée des réactions intervenant dans le cadre de la réaction de Belousov-Zhabotinsky.[7] Bas gauche, Réaction de Belousov-Zhabotinsky dans une boîte de pétri sur un faible volume laissant apparaître les motifs de réaction diffusion. Bas droite, copolymère formant des miscelles ou non en fonction du degré d'oxydation de Ruthénium réagissant selon une réaction de Belousov-Zhabotinsky, on a donc oscillation de la turbidité de la solution lors de la formation destruction des micelles.[8]

On peut remarquer dans chacun des exemples cités précédemment que chacun des systèmes est construit autour d'une réaction réversible, ainsi que d'une source d'énergie dissipée. Ainsi, nous avons vu l'association réversible de microtubules et nano particules avec comme source d'énergie respective le GTP et le rayonnement UV, et la mise en place d'oscillateur avec comme énergie l'ATP pour la synthèse biologique, et l'oxydation de l'acide malonique pour la réaction de Belouzov-Zhabotinsky.

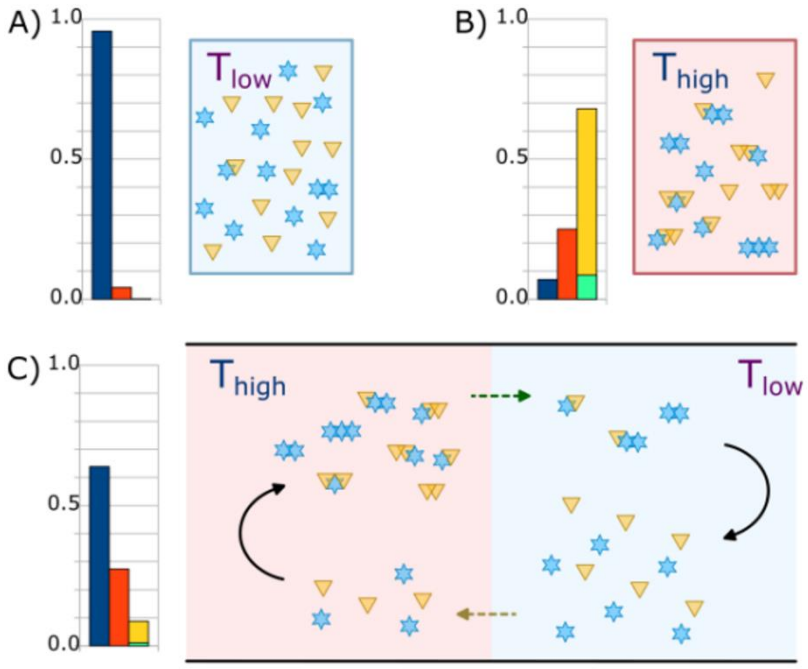


Figure 6 A Composition du système polymérisable à basse température, les couleurs et formes indiquent différents monomères. B Composition du système à haute température, il y a des espèces plus longues absentes à basse température. C système placé dans un gradient de température et boucles de réaction diffusions qui se mettent en place entre la zone chaude et la zone froide poussant le système hors équilibre.[3]

Ce travail de thèse fait partie du projet d'ANR SACERDOTAL. Ce projet vise à établir un système dissipatif grâce à une réaction de polymérisation réversible en utilisant l'énergie disponible dans un gradient de température. Si la réaction de polymérisation permet d'obtenir différentes distributions d'espèce polymérique à différentes températures telles que représentées par les figures A (basse température) et B (hautes température), alors le système dans un gradient de température mènerait à un système hors équilibre si les constantes de diffusion le permettent. En effet, dans la zone chaude du gradient on retrouverait des espèces « chaudes » dans cet exemple des espèces longues (grand degré de polymérisation), ces espèces pourraient alors diffuser vers les zones plus froides du gradient où leur formation ne seraient pas observée si le système était à cette unique température. Il en va de même pour les espèces de la zone froide qui diffuseraient vers la zone chaude, permettant d'obtenir des états hors équilibre par des phénomènes de réaction-diffusion. Ceci permettrait alors la coexistence et l'interaction d'espèces à des températures non permises par un système à l'équilibre, ainsi que l'exploration dynamique d'une grande gamme de séquences de polymères si différents monomères sont utilisés, deux point clefs en vue d'application en chimie combinatoire dynamique. De plus, il serait possible que parmi les espèces formées, certaines puissent catalyser la formation d'autres espèces du milieu et générer ainsi des cycles réactionnels autocatalytiques qui pourraient être en compétition entre eux pour les monomères à la manière de populations d'êtres vivants.

Il faut donc choisir un système polymérisable autour duquel articuler le projet et celui qui a été retenu et le système peptidique. Les protéines possèdent une relation séquence/structure/fonction forte et sont des molécules capables de catalyser des nombreuses réactions à l'instar des enzymes. En créer une large gamme de séquences reviendrait donc à

créer de nombreuses fonction potentiel. De plus un gradient de température permettrait aussi aux peptides d'explorer différentes conformations de repliement pouvant mener ainsi à différentes fonctions. Cependant, la liaison peptidique entre acides aminés est très stable et n'est donc pas réversible. Pour se faire nous avons employé une technologie mise au point dans notre laboratoire, une ligation native dynamique permise par l'incorporation dans la séquence peptidique de N-méthyl-cystéine.

Il convenait donc dans un premier temps de caractériser cette réaction d'échange d'un point de vue thermocinétique en vue de son utilisation au sein d'un gradient de température, puis de mettre au point un système polymérisable réversible et de l'étudier.

Partie 2 : Caractérisation de la réaction d'échange

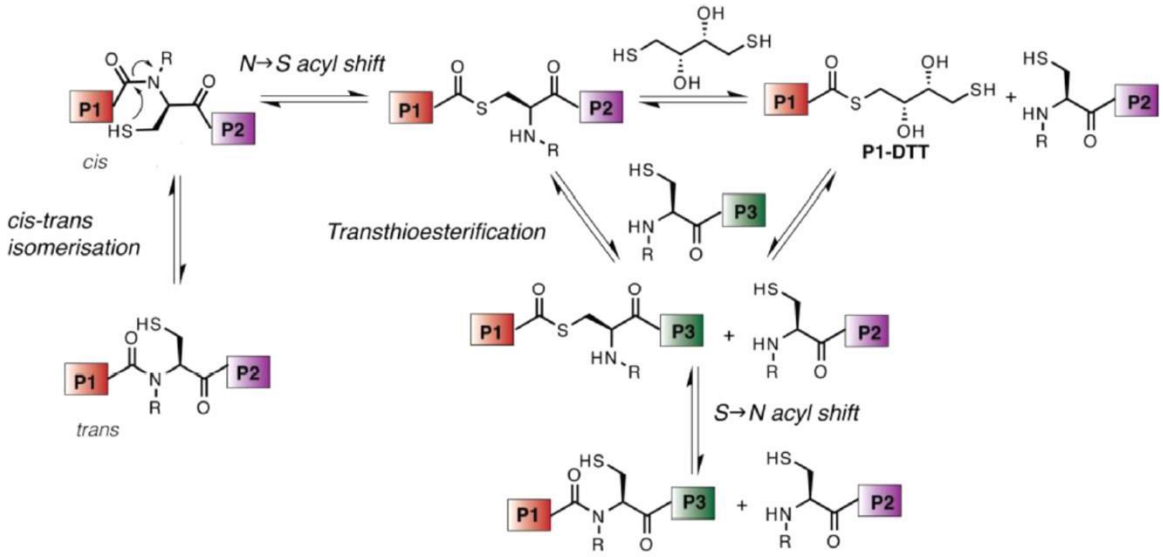


Figure 7 Mécanisme proposé pour la réaction d'échange de fragment peptidique permise par la N-méthyl-cystéine, tiré de [4]

La N-méthyl-cystéine correspond à une cystéine dont un hydrogène de l'azote est substitué par un groupement méthyl. Au sein de la chaîne peptidique, ceci permet de déplacer l'équilibre entre conformère cis et trans tel que décrit dans la figure ci-après. La quasi-totalité du temps la conformation des acides aminé entre eux est de type trans, or déplacer l'équilibre en faveur de la forme cis permet une réorganisation intramoléculaire de type N→S *acyl shift* ce qui permet de passer d'une liaison amide stable à une liaison de type thioester beaucoup plus réactive. C'est cette forme qui va pouvoir réagir avec d'autres groupement nucléophile dans le milieu notamment d'autres thioesters par réaction de transthioesterification. En ayant dans le milieu un autre fragment peptidique possédant une N-méthyl-cystéine en position N terminale, le groupe thiol de ce second fragment peut donc attaquer le thioester former et s'échanger avec le fragment en position C terminal de la séquence initiale. Puis, par un mécanisme inverse à celui du début du mécanisme soit un S→N *acyl shift* on retrouve une liaison de type peptidique et une nouvelle séquence. Cette chimie requiert néanmoins de rester en milieu réduit afin de maintenir les groupes thiol en l'état et empêcher leur oxydation sous forme de disulfures.

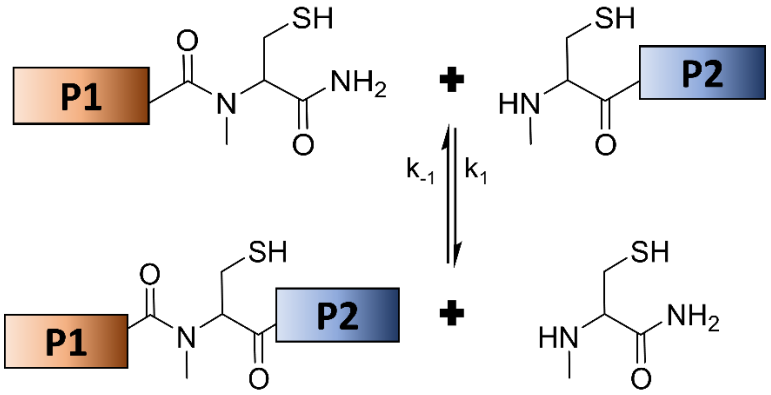


Figure 8 Réaction d'échange de N-méthyl-cystéine modifiée pour s'apparenter à une ligation

Pour passer d'une réaction d'échange à une réaction de ligation/dissociation en vue d'une polymérisation réversible, il suffit de travailler avec des N-méthyl-cystéines terminales. Ainsi, un fragment P1 possédant une N-méthyl-cystéine C terminale alors qu'elle est N terminale pour un fragment P2, mène à la formation du produit d'association P1-Nmetcyst-P2 en libérant une N-méthyl-cystéine qui réagir avec ce produit de ligation pour le dissocier. Afin d'étudier les caractéristiques thermocinétiques de cette réaction, nous avons synthétisé par synthèse sur phase solide assistée par micro-onde les deux peptides suivants :

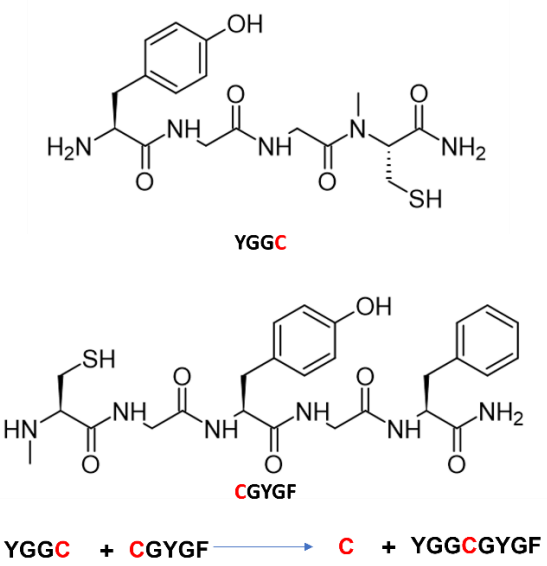


Figure 9 Peptides modèles utilisés pour l'étude thermocinétique de la réaction de ligation/dissociation et ladite réaction.

La tyrosine, permet de suivre la réaction par absorption UV en chromatographie en phase liquide. La phénylalanine sert de boulet pour augmenter le temps d'élution du second peptide et enfin les glycines servent de linker entre les différents acides aminés de façon à maximiser le rendement de la synthèse par le faible encombrement stérique des glycines, ainsi qu'à faciliter au plus l'accès au thioester lors de la réaction de transthioesterification.

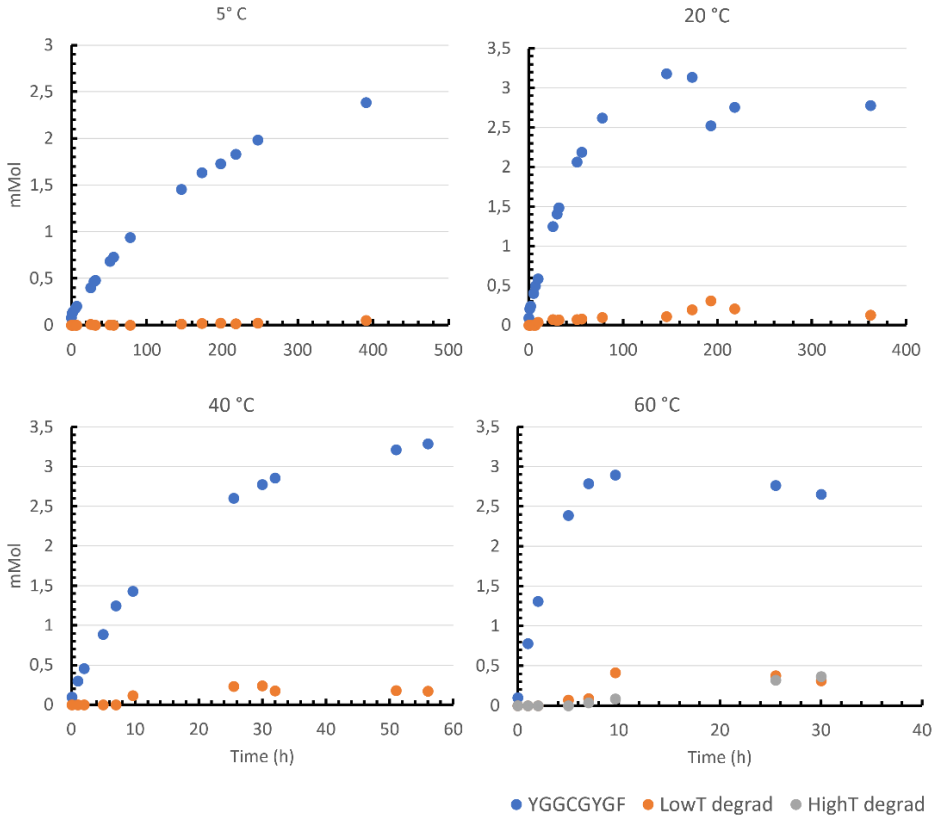


Figure 10 Evolution de la concentration en produit de ligation au cours du temps pour des concentrations initiales en CGYGF de 5 mMol et YGGC de 8,3 mMol à différentes températures dans un milieu contenant 50 mMol de tampon phosphate à un pH de 7, 100 mMol de TCEP et 200 mMol d'ascorbate de sodium.

Les concentrations des différentes espèces sont alors suivies par UPLC (Ultra Performent Liquid Chromatography) couplée à un détecteur UV et un spectromètre de masse afin d'identifier les piques. L'air de chaque pique est ensuite relié à la concentration de chaque espèce par la loi de Beer-Lambert. L'évolution des concentrations en produit de ligation st donnée dans la figure précédente. A partir de ces données expérimentales, nous avons approximé cette réaction au cas modèle de réaction bimoléculaire réversible qui est analytiquement soluble du point de vue cinétique. Nous avons donc réalisé un petit script MATLAB capable d'approximer les courbes obtenues et d'extraire les constantes cinétiques pour chaque température qui sont répertoriée dans le tableau suivant.

Température (°C)	k_1 (L.mol^{-1}.s^{-1})	k_{-1} (L.mol^{-1}.s^{-1})	$t_{1/2}$ (heures)	K (k_1/ k_{-1})
5°C	8.43 10^{-5}	1.58 10^{-4}	117	0.53
20°C	3.67 10^{-4}	4.83 10^{-4}	39.6	0.76
40°C	1.15 10^{-3}	1.07 10^{-3}	13.1	1.08
60°C	4.83 10^{-3}	4.0 10^{-3}	2.8	1.21

Figure 11 Tableau récapitulant les mesures cinétiques de la réaction de ligation/dissociation

A partir des données des constantes d'équilibre obtenues correspondant au rapport des constantes cinétiques d'association et de dissociation, on peut avoir accès aux grandeurs thermodynamiques selon la relation suivante :

$$-RT \ln K = -T\Delta_r S^0 + \Delta_r H^0$$

On peut alors par une régression linéaire avoir accès à l'entropie et l'enthalpie standard pour cette réaction qui sont respectivement de 36 $J.mol^{-1}.K^{-1}$ pour l'entropie et 10 $kJ.mol^{-1}$ pour l'enthalpie de réaction.

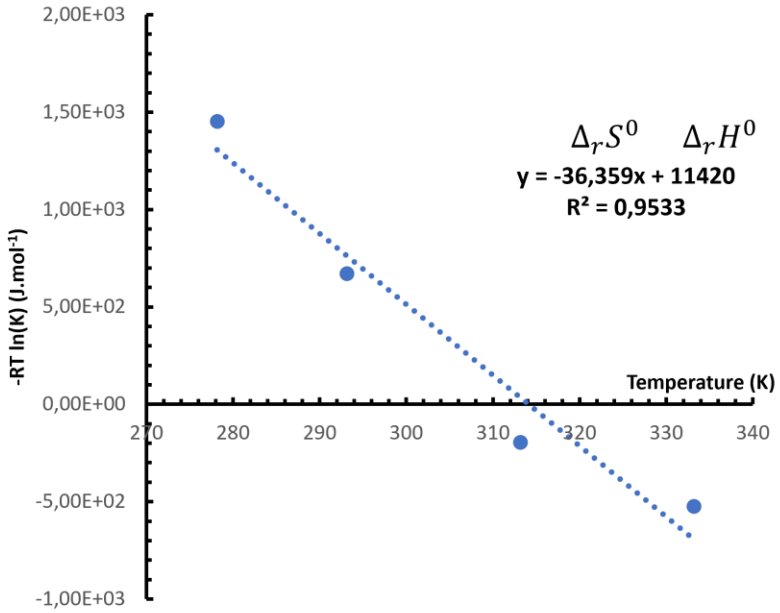


Figure 12 Régression linéaire à partir de la constante d'équilibre donnant accès à l'entropie et l'enthalpie standard de réaction

Partie 3 : Mise au point du système de polymères dynamiques

Comme discuté précédemment, afin d'avoir des réaction d'association/dissociation, il faut des N-méthyl-cystéines en position N et C terminales des séquences peptidiques. Ainsi une séquence débutée et terminée par une N-méthyl-cystéine est capable de réagir avec elle-même. La séquence d'étude choisie, CGYGC est la suivante :

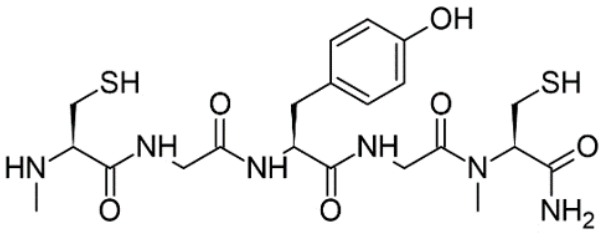


Figure 13 Séquence utilisée pour le système surnommée "living polymer"

Les acides aminés présent y jouent le même rôle que décrit précédemment, et la concentration des différentes espèces peut être suivie par chromatographie en phase liquide couplée à un spectromètre UV. Ce monomère est donc capable de réagir avec lui-même pour former des polymères plus longs mais aussi des espèces cycliques car les extrémités restent réactives ainsi en théorie que toute N-méthyl-cystéine présente au sein du polymère.

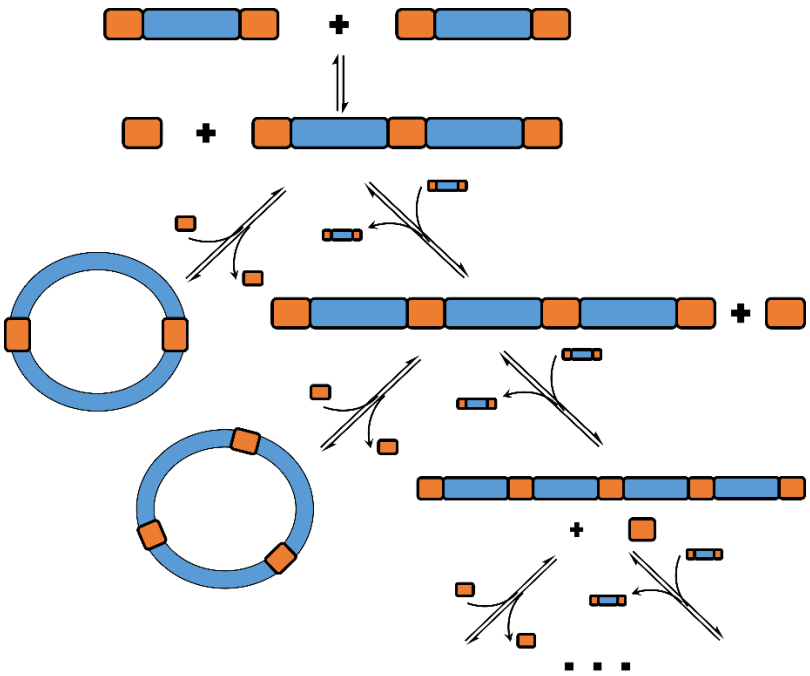


Figure 14 Schéma illustrant les possibilités du système discuté. La partie bleu correspond au motif GYG (glycine,tyrosine,glycine) alors que les morceaux oranges correspondent aux N-méthyl-cystéines

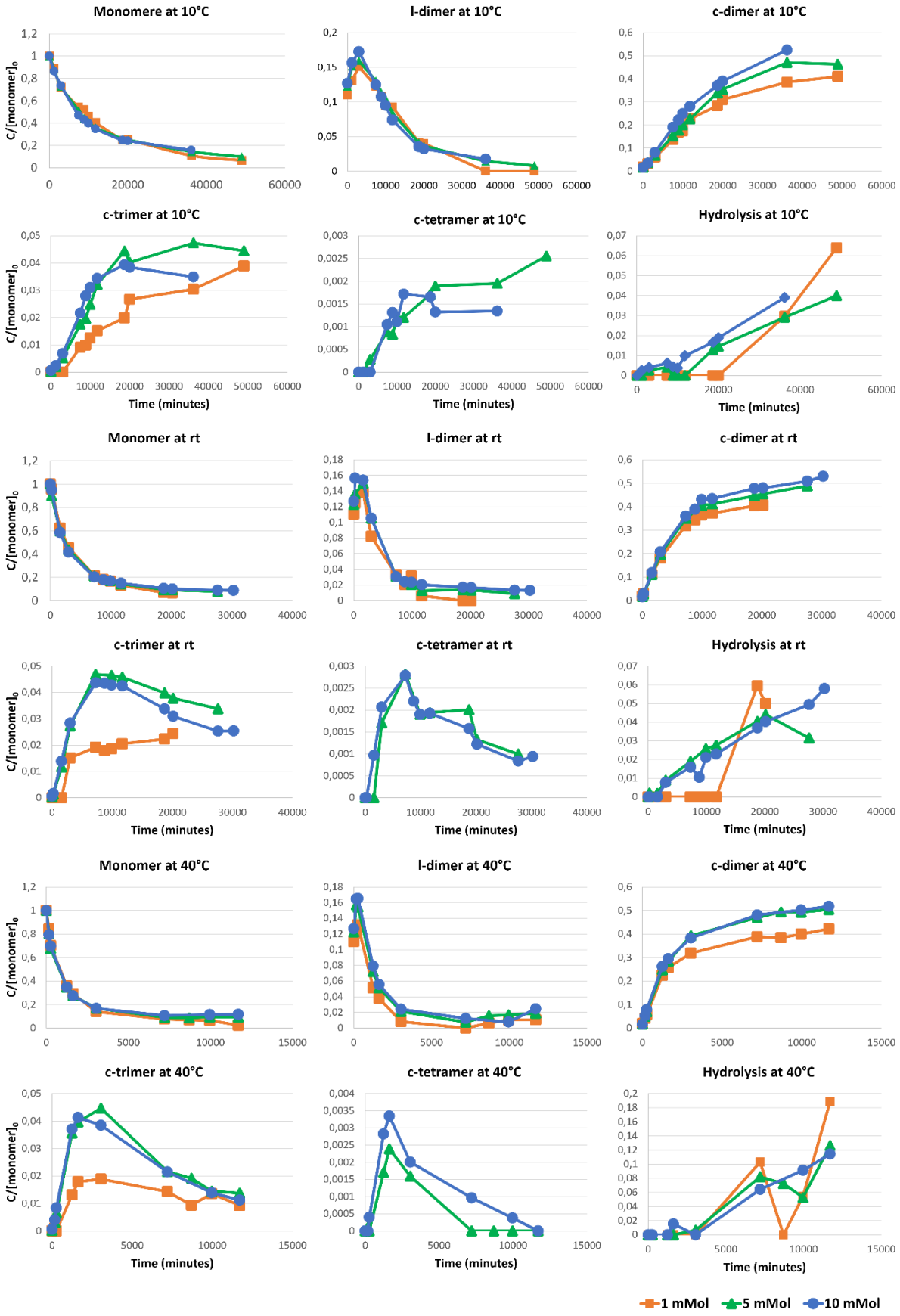


Figure 15 Evolution des concentrations des différentes espèces du système en fonction du temps à différentes températures et pour différentes concentrations initiales en monomère. Le milieu est tamponné à pH 7 par un tampon phosphate, 100 mMol de TCEP et 200 mMol d'ascorbate de sodium sont ajouté

L'évolution d'un tel système a donc été caractérisée pour différentes températures et différentes concentrations initiales en monomère illustré figure 15.

On observe que le système évolue de façon très similaire peu importe la température ou la concentration initiale en monomère. On retrouve ainsi à 20°C le même comportement qu'observé à 40°C mais plus lent. On remarque que le dimère linéaire est ainsi un produit cinétique, rapidement formé au début puis dont la concentration va tendre à fortement diminuer. Il y a cependant quelques différences. Ainsi, on remarque que le trimère cyclique semble être plus stable à 10°C pour 1 mMol, qu'il ne l'est à 20 et 40°C. De plus, le trimère cyclique est observé en de plus grandes proportion à 5 mMol peu importe la température. A l'inverse de plus grandes concentrations initiales semblent mener à une plus grande proportion de dimères cycliques et ce notamment à 20 et 5°C. Le tétramère cyclique n'est quant à lui observé que pour une concentration initiale d'au moins 5 mMol indiquant que les trimère linéaire cyclisent avant de pouvoir former de tétramères.

Vers la fin de l'expérience d'évolution du système, les seules espèces présentes en solution sont quasi-uniquement les espèces cycliques et majoritairement le dimère cyclique. On observe cependant l'apparition de précipité dans les Eppendorf et une diminution de la quantité de matière totale en solution. L'analyse de ces précipités ne change rien quant aux espèces présentes à 5 et 40°C. Cependant le formé à 20°C est composé d'une multitude de polymères non observés en solution comme le montre le chromatogramme analysé suivant :

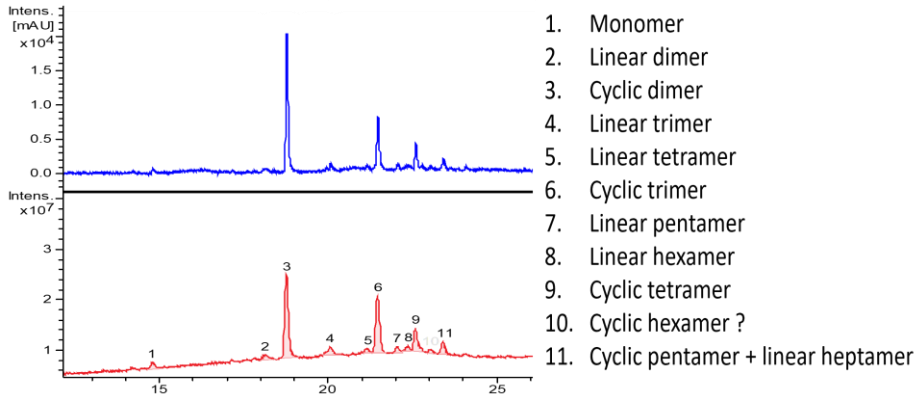


Figure 16 Chromatogramme obtenu pour le précipité formé à 20°C pour une concentration initiale en monomère de 10 mMol. Le graphe du haut correspond au signal UV en fonction du temps d'élution, le graphe du bas au signal de masse en fonction du temps d'élution.

On observe la présence d'espèces linaires allant jusqu'à l'heptamère et des espèces cycliques jusqu'au pentamère. On ne peut dire si ces espèces sont formés en solution et précipitent instantanément ou si elles sont formées à partir d'espèces plus courtes au sein du solide. Il se révèle que le phosphate utilisé comme tampon est lui aussi présent au sein du précipité comme l'a révélé une analyse élémentaire au MEB-EDX. Une autre évolution du système a été menée en utilisant comme tampon du citrate de sodium. Cette fois-ci les Eppendorf n'ont pas été ouvert jusqu'à l'apparition de solide. A 40°C après 20 jours, sont apparus des baguettes cristallines contrairement aux tubes placés à 5 et 20°C. Une analyse du surnageant a montré que les tubes à 40°C ne contenait plus de peptides en solution et la dissolution des baguettes semblait indiquer

qu’elles étaient composées uniquement de dimères cycliques, montrant ainsi une sélection et amplification d’une des espèces du système. A l’inverse les tubes à 20°C montraient la même distribution en espèce qu’avec le tampon précédent et une absence de précipité, laissant supposée que les espèces longues observées précédemment se forment au sein du solide.

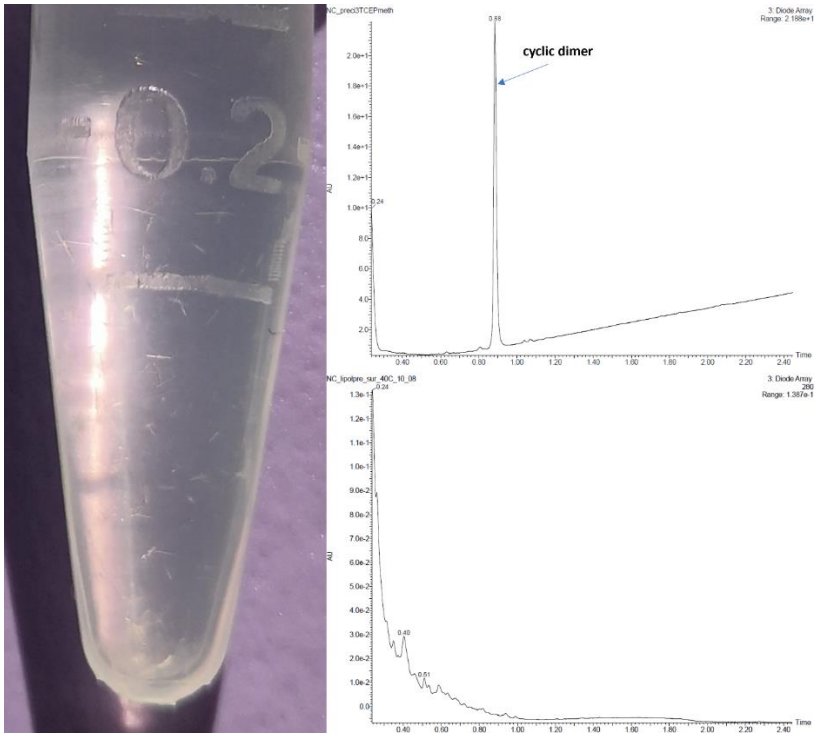


Figure 17 Gauche, photographie d'un Eppendorf dans lequel on apperçoit les baguettes cristallines. Draoite haut, signal UV du chromatogramme de la dillution des baguettes dans du méthanol. Droite bas, signal UV du chromatogramme du surnageant des baguettes.

La structure de ces cristaux a été déterminé par diffraction des rayons X. Ces mesures ont confirmé la composition des cristaux, soit uniquement des dimères cycliques et quelques molécules d’eau. Il semblerait que les cycles s’empilent pour former des sortes de nanotubes dont la cohésion est assurée de deux façons différentes. Par liaison hydrogènes entre le squelette peptidique et le groupe hydroxyle des tyrosines, et par des liaisons hydrogènes entre le squelette peptidique et des molécules d’eau présente dans la structure.

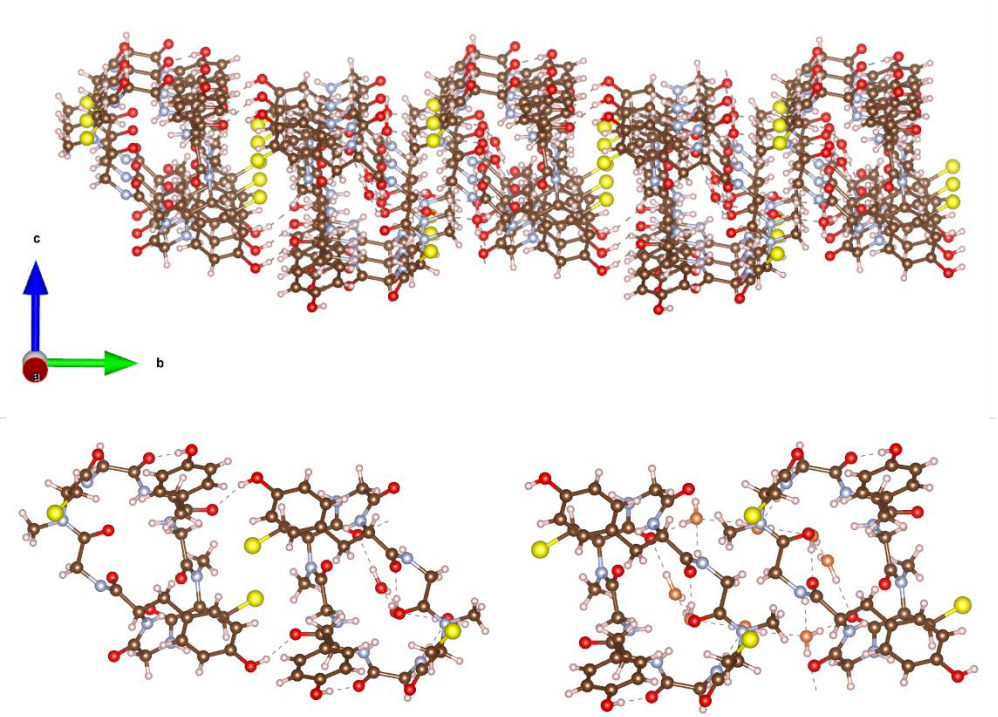


Figure 18 Structure des cristaux obtenue par diffraction au rayons X. Haut, empilement des dimères cycliques sous forme de nanotubes. Bas, les deux types d'interactions entre 2 nanotubes par liaison hydrogène. Les carbones sont en marron, les oxygènes en rouge, les souffres en jaunes, les hydrogènes en blanc, les oxygènes des molécules d'eau sont colorés en orange.

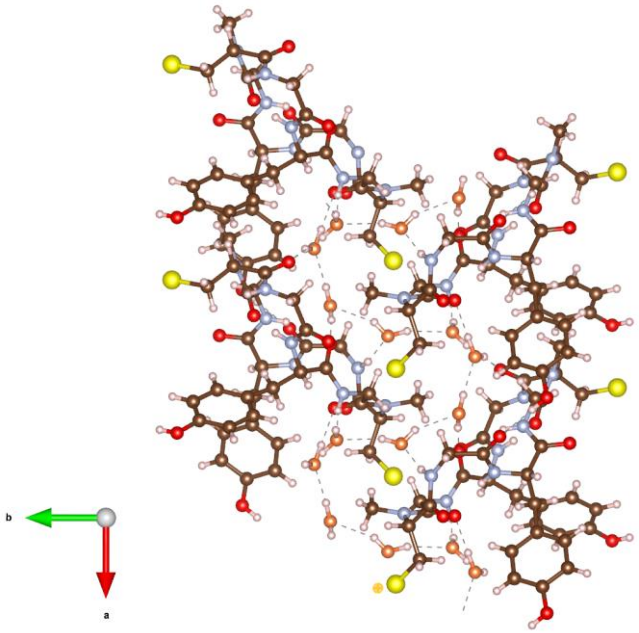


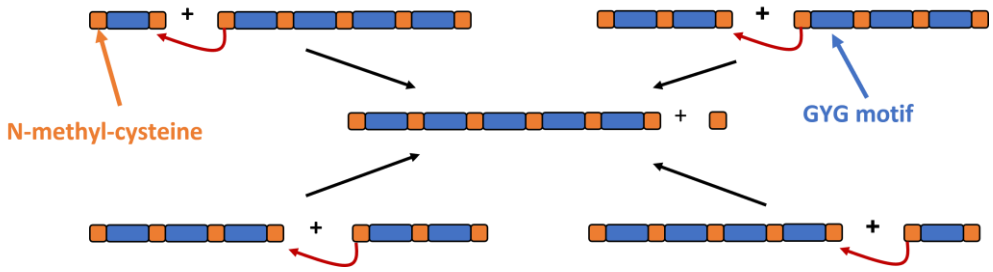
Figure 19 Structure des molécules d'eau au sein de l'empilement de dimères cycliques.

Ces molécules d'eau ne semblent pas intervenir uniquement dans la cohésion latérale des nanotubes mais aussi au sein de l'empilement des cycles en nanotubes comme représentés dans la figure précédente.

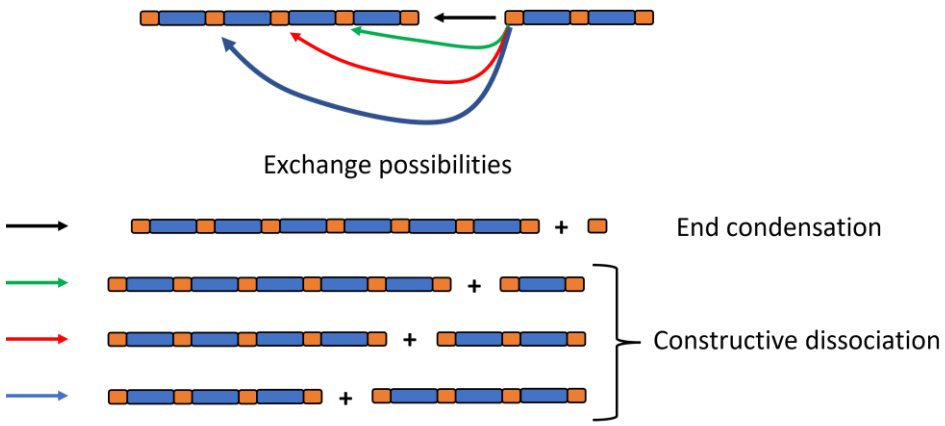
Partie 4 : Etude théorique et modélisation

La modélisation de réseau de réactions chimiques peut être réalisée de deux façons différentes. Elle peut être stochastique et prendre en compte les réactions comme des évènement aléatoire avec certaines probabilités d'occurrence ce qui permet une modélisation plus réaliste des systèmes et une meilleure visibilité d'effet notamment pour de très faibles concentrations. Enfin elle peut être déterministe avec des quantité fixe et moyennée de produits créés et consommés au cours du temps ce qui tend à occulter les effets de faibles concentrations mais requiert bien moins de ressources en termes de temps de calcul. Une première démarche théorique a donc été l'étude de la complexité du système.

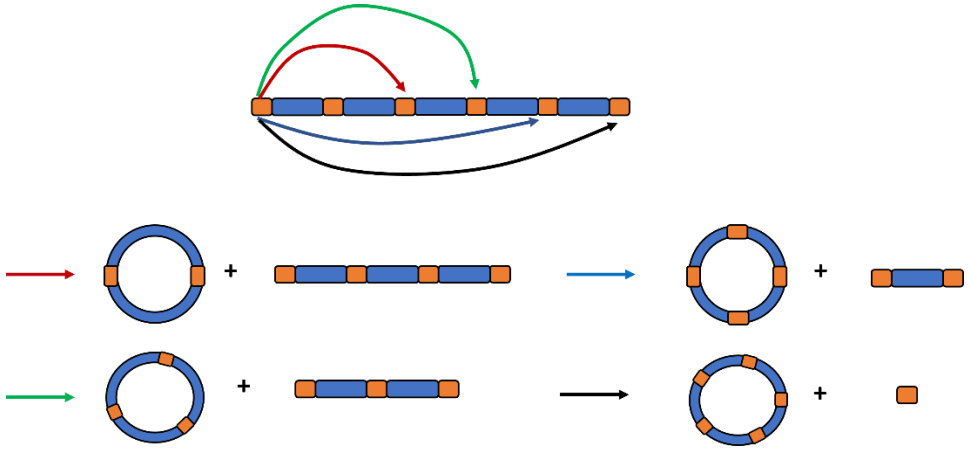
Pour se faire nous avons compté le nombre de réaction possibles pour former par exemple une espèce de longueur N dans un système où l'espèce la plus longue observée serait de taille M. Ainsi le premier type de réaction est celui d'élongation où des polymères fusionnent par leur extrémité.



Aux réactions d'élongations viennent ensuite s'ajouter les réactions des polymères sur les N-méthyl-cystéines présentes au sein même des séquences d'autres polymères générant d'autres espèces.



Enfin aux réactions entre espèces linéaires, viennent s'ajouter les réactions de cyclisation tel qu'illustrée ci-dessous qui ne se font pas nécessairement entre extrémité mais entre N-méthyl-cystéine N-terminale et une autre N-méthyl-cystéine de la séquence autre que la deuxième car le monomère cyclique n'est pas observé.



Le nombres de réactions dans un système menant à la formation d'un N-mère est donc le suivant :

$$\frac{M^2}{2}+\frac{M}{2}-N^2+N+MN-3-\delta_{N\leq\frac{M}{2}}+\delta_{M=N}$$

Avec δ valant 1 si la condition en indice est remplie et 0 sinon. Le nombre de réactions permettant de former une N-mère cyclique quant à lui est de :

$$M-N+1$$

On a donc une complexité en M^3 pour notre système lorsque l'on somme le nombre de réaction sur la totalité des M espèces présentes d'où l'intérêt de partir vers un système déterministe simple peu gourmand en ressources.

Pour ce faire nous avons utiliser le modèle d'Euler appliquer aux systèmes d'équations différentielles obtenues pour notre système. Ce type de modèle est illustré avec l'exemple ci-dessous :

$$A+B\rightarrow C \quad k_1$$

$$[A](t+\Delta\mathrm{t})=[A](t)+d[A]\approx[A](t)+\Delta A=[A](t)-k_1[A](t)[B](\mathrm{t})\Delta t$$

$$[B](t+\Delta t)=[B](t)-k_1[A](t)[B](\mathrm{t})\Delta t$$

$$[C](t+\Delta t)=[C](t)+k_1[A](t)[B](\mathrm{t})\Delta t$$

La quantité d'espèce formé au cours de chaque pas de temps dépends alors des concentrations du pas de temps précédent ainsi que des constantes cinétiques de chaque réaction. Un script MATLAB a donc été écrit, qui à chaque pas de temps actualise les concentrations des différentes espèces selon les différentes réactions exposées précédemment. La figure suivante donne un exemple d'évolution du système donnée par notre modèle. On y retrouve l'aspect produit cinétique du dimère linéaire dont la concentration augmente au début puis diminue au fur et à mesure que la cyclisation prend de l'ampleur. Il reste à mettre en place un algorithme

d’inférence qui à travers ce modèle et les données expérimentales permettrait d’extraire les constantes cinétiques notamment de cyclisation qui ont été approximée à la main pour le tracée de l’évolution donnée en exemple.

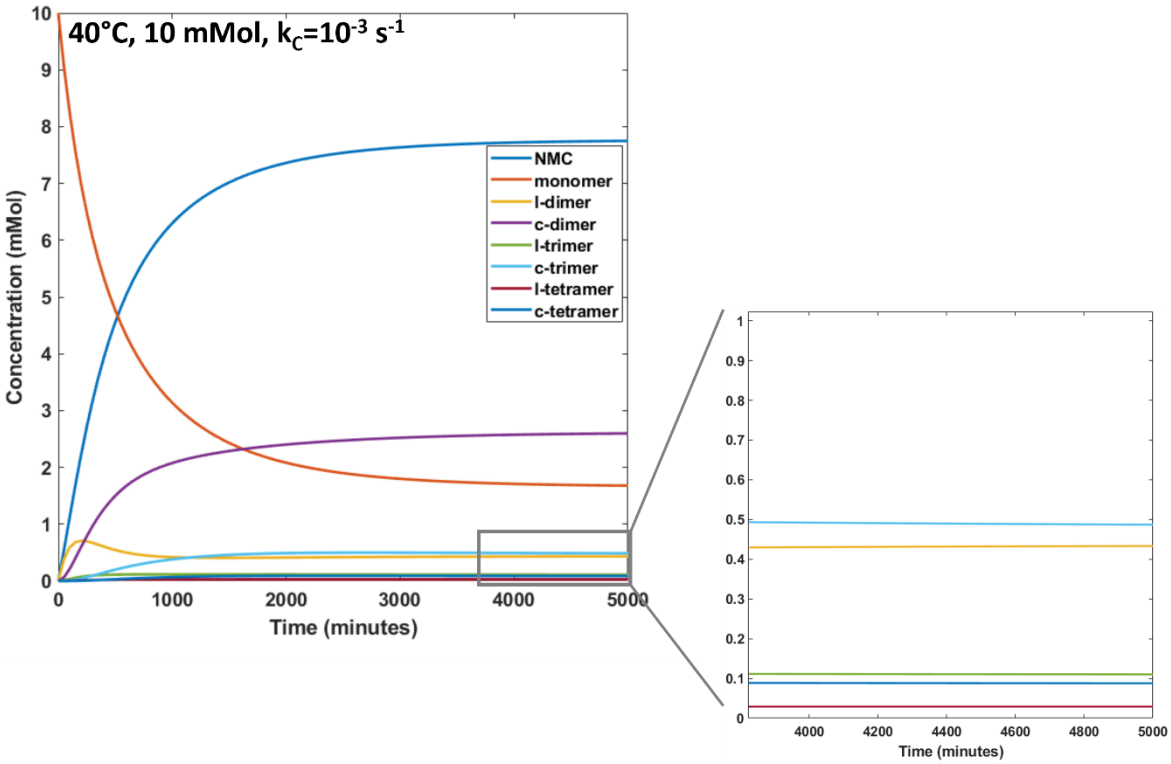


Figure 20 Exemple d'évolution du système généré par notre implémentation du modèle d'Euler. Ici les constantes cinétiques ont été choisies égales à celles obtenue dans le chapitre 2 pour 40°C avec une concentration initiale en monomère de 10 mMol. Les constantes de vitesse pour les réactions de cyclisation ont été fixées à $10^{-3}s^{-1}$.

Conclusion :

Au cours de ces travaux, nous avons caractérisé d'un point de vue thermocinétique la réaction d'échange de fragments peptidiques permise par la N-méthyl-cystéine afin de permettre à nos collaborateurs du projet ANR de dimensionner l'expérience en gradient de température. La réaction fonctionne relativement bien de 5 à 60°C et ce sur une gamme de temps allant de la journée au mois en fonction de la température.

A partir de cette chimie nous avons mis au point un système de polymères dynamiques capable d'explorer différentes tailles mais aussi structure avec l'apparition de formes cycliques. Même si ce système ne semble pas a priori présenter différentes distributions d'espèces en solution fonction de la température, on observe différent comportement au niveau de la phase solide qui se crée. Avec un tampon phosphate à 20°C on observe la formation de structures allant jusqu'à l'heptamère linéaire et le pentamère cyclique alors qu'avec un tampon à base de citrate d'ammonium, on observe l'apparition au bout de quelques semaines de cristaux résultant de la sélection de l'unique espèce dimère cyclique dont la structure a été caractérisée.

Une approche théorique en vue de modéliser le système a été entamée. La complexité et le nombre de réaction a calculé a été évaluée et une approche déterministe a été choisie. Un modèle simple basée sur la méthode d'Euler semble pouvoir reproduire les comportements observés dans l'expérience.

Les pistes de recherche restantes dans le projet sont la mise au point d'un algorithme d'inférence cinétique à partir du modèle afin d'extraire des données expérimentales les constantes cinétiques des différentes réactions notamment de cyclisation, puis le système devrait être prêt pour une expérience en gradient de température. Ainsi l'observation de cristaux de dimère cyclique dans une région du gradient de température plus basse constituerait une propriété émergente car non observée dans le système à l'équilibre thermique à de telles températures. Cependant, le consensus actuel au sein du projet est de générer le gradient de température par irradiation différentielle de la solution qui contiendrait des nanoparticules d'or par un rayonnement laser. Il reste à voir si le système reste dynamique en présence d'or car les thiols ont une forte affinité pour ce dernier.

Enfin ce système est facilement adaptable à une exploration de séquences. Il suffit d'avoir différents monomères possédants chacun une séquence différente entre les deux N-méthyl-cystéines. Si les espèces plus longues ne sont observées dans les solides que parce qu'elles sont peu solubles, incorporer une lysine ou un autre acide aminé hydrophile pourrait permettre une meilleure solubilité.

Bibliographie:

1. Merindol, R. & Walther, A. Materials learning from life: Concepts for active, adaptive and autonomous molecular systems. *Chemical Society Reviews* **46,** 5588–5619 (2017).
2. Hess, H. & Ross, J. L. Non-equilibrium assembly of microtubules: from molecules to autonomous chemical robots. *Chemical Society Reviews* **46,** (2017).
3. *AAPG2019 SACERDOTAL Survival of reactive systems for driving property-directed molecular evolution*.
4. Ruff, Y., Garavini, V. & Giuseppone, N. Reversible Native Chemical Ligation: A Facile Access to Dynamic Covalent Peptides. *Journal of the American Chemical Society* **136,** (2014).
5. Giuseppone, N. & Walther, A. *Out-of-Equilibrium (Supra)molecular Systems and Materials*. (Wiley, 2021). doi:10.1002/9783527821990
6. Klajn, R., Wesson, P. J., Bishop, K. J. M. & Grzybowski, B. A. Writing self-erasing images using metastable nanoparticle "inks." *Angewandte Chemie - International Edition* **48,** 7035–7039 (2009).
7. Field, R. J., Koros, E. & Noyes, R. M. Oscillations in chemical systems. II. Thorough analysis of temporal oscillation in the bromate-cerium-malonic acid system. *Journal of the American Chemical Society* **94,** (1972).
8. Ueki, T., Shibayama, M. & Yoshida, R. Self-oscillating micelles. *Chemical Communications* **49,** (2013).

Table of contents

“All we have to decide is what to do with the time that is given us”

- Gandalf, The Fellowship of the Ring, J. R. R. Tolkien

Acknowledgements/Remerciements

Je tiens tout d'abord à remercier mon directeur de thèse Pr. Nicolas Giuseppone pour m'avoir donné l'opportunité de mener à bien ce projet. Il a su me faire confiance et me donner une seconde chance et m'a permis d'avoir la liberté dont j'avais besoin au cours de moments difficiles. Merci beaucoup. Nos échanges ont toujours été enrichissants et ont permis d'aiguiller ce travail dans de bonnes direction. Merci aussi à ma co-directrice Émilie Moulin pour son aide et pour les bonnes discussions au cours des moments de détentes.
Merci à Andréas Vargas, pour tes suggestions toujours pertinentes et pour les analyses de STEM, et à Pierre Lutz avec qui je ne manquais pas de discuter quand il passait, toujours d'une grande gentillesse.

J'ai eu l'occasion de partager mon laboratoire avec de nombreuses personnes au cours de ce travail. Tout d'abord Marie-Céline, qui est toujours de bonne humeur et avec le sourire qui me réconforte dans le fait que je ne suis pas le seul que les machines ont l'air de bouder. Odile, qui en plus de nos échanges quotidiens m'assistait dans mes (més)aventures avec l'HPLC. Maintenant, on est prêt à tout. Joachim avec qui on parlait de tout, et n'importe quoi, de la chimie à la géopolitique en passant par Starcraft 2. Je te souhaite beaucoup de réussite et j'espère qu'un jour je pourrai égaler ton niveau en russe. Yali, my last peptide companion. We did not speak that often but when we did, we always had good laughs. I wish you a great future in Shanghai. Enfin Philippe, nous n'avons pas tant eu l'occasion d'échanger avec les différents évènements de ces deux dernières années mais je sais que tu sauras mener une thèse brillante.

Raphaël, mon ami « parisien » avec qui je regrette de ne pas avoir passé plus de temps, j'aurai dû trouver plus de prétextes d'utiliser le lyophilisateur ! Merci d'avoir toujours été là pour m'aider, tant d'un point de vue scientifique, qu'à garder le moral et à retrouver le sourire, sans compter le thé que tu m'as fourni. Pierrick, mon petit stagiaire devenu grand ! Bon courage pour la suite, je ne doute pas que ta motivation à toute épreuve saura t'emmener loin. Je remercie également Chris Rete, pour avoir en quelques mois essayer de me transmettre le plus de connaissances possibles sur les peptides, la N-méthyl-cystéine et la synthèse alors que je découvrais un laboratoire de recherche de chimie organique pour la première fois. Un grand merci à tous les autres membres de l'équipe SAMS que j'ai pu côtoyer au cours de ces trois ans : Vasyl, Christian, Lara, Soichi, Chuan, Wengzhi, Xiaoqin, Jean-Rémy, Damien, Alexis, Frédéric, Maria, Alessandro, Dania, Sergio, Thiebault, Mélodie, Flavio.
I had the opportunity to surround myself with an amazing group of friends during my stay at ICS. Fedir, thank you very much for all your support and advice when things were rough, thanks for the great talks and wonderful memories. I hope we will find a way to continue our music

composition sessions. Anastasiia, Imen, Vaibhav, our weekend lunches were always among the highlights of the week, thanks for sticking with me through thick and thin, it was a blessing to all have each other for support.

Yann, merci de m'avoir supporté pendant deux ans, surtout qu'en plus il y avait le confinement. Tu as été comme un frère pour moi au cours de ces deux dernières années. Merci pour tous ces instants de complicité et de rigolade. Romain, Thibaud, Amaury, Axel, Quentin, les amis je ne sais trop où j'en serai si je ne vous avais pas. Même si c'était surtout via discord, nos games et nos discussions faisaient toujours chaud au cœur. Ma petite Lucie, merci d'avoir toujours été là pour moi quand j'en avais besoin.

Merci à ma famille, et à son soutien à toutes épreuves quelques soient mes décisions. Malgré la distance et les circonstances, je sais que vous êtes toujours à mes côtés.

Je tiens aussi à remercier François Guyot, qui m'a aidé à ne pas perdre le nord et à rester productif pendant cette situation particulière qu'à été le confinement. Merci à Jean-Marc Strub pour les analyses HPLC du précipité et pour m'avoir appris beaucoup de choses concernant l'exploitation des spectres de masse. Merci à Cedric Leuvrey qui m'a permis d'image et d'analyser par microscopie électroniques des échantillons. Merci à Alexandra Elbakyan pour son engagement pour une recherche scientifique sans barrières.

Je remercie Valérie Marchi d'avoir accepté d'examiner ce travail.

Enfin merci à Ludovic Jullien qui m'a permis de participer à ce projet et sans qui cette thèse n'aurait pas eu lieu. Merci pour les conseils avisés dans les situations de doutes et merci d'avoir accepté d'examiner ce travail.

Abbreviations and units

ADP	Adenosine diphosphate
ANR	Agence Nationnale de la Recherche
Asc	Ascorbic acid
ATP	Adenosine triphosphate
Boc	tert-Butoxycarbonyl
BZ	Belousov-Zhabotinsky
°C	Celsisus degree
Da	Dalton
DCL	Dynamic Combinatorial Library
DIEA	N,N-di-isopropyl-N-ethylamine
DKP	Diketopiperazine
DMBA	di-methoxy-benzoic acid
DTT	Dithiothreitol
DMF	N,N-dimethylformamide
dyNCL	Dynamic native chemical ligation
E. coli	Escherichia coli
Fmoc	fluorenylmethoxycarbonyl
g	Grammes
G	Gibbs free energy
GDP	Guanosine diphosphate
GSH	Glutathione
GTP	Guanosine triphosphate
GSSG	di-glutathion (oxidized form)
H	Enthalpy
HBTU	(2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate

HFIP	1,1,1,3,3,3-hexafluoropropan-2-ol
HOBT	Hydroxybenzotriazole
HPLC	High performance liquid chromatography
h	hours
J	Joules
kJ	kilo Joules
K	Kelvin
L	liters
min	minutes
mMol	milli Molar (millimoles per liter)
Mol	Molar (moles per liter)
MS	Mass spectrometry
µm	micrometers
NMC	N-methyl-cysteine
p	pressure
PBS	Phosphate buffer saline
Q	Heat
R	Gas constant
RNA	Ribonucleic acid
rt	room temperature (20°C)
s	seconds
S	entropy
SEM	Scanning electronic microscopy
SPPS	Solid phase peptide synthesis
T	Temperature
TEM	Transfer electronic microscopy
TFA	trifluoro acetic acid
trt	trityl
U	Internal energy
UPLC	Ultra-Performance Liquid Chromatography

UV	Ultraviolet light
V	Volume
W	Work

All amino acids used for peptide synthesis in this work are L – α - amino acids

Abbreviation	Amino acid
C	N-methyl-cysteine
G	Glycine
Y	Tyrosine
F	Phenylalanine
Fmoc-ser-(trt)-OH	N-Fmoc-S-trityl-serine

Introduction

There! You did it again! You just breathed. Inhaled some (more or less) fresh air, containing around 20% of di-oxygen, this crucial gas for your survival. Why is it so important? In your cells, it is used to oxidize organic matter and through this oxidation the energy that you rely on to survive is created. Life in its essence requires energy to be sustained, and in all its forms exploits an energetic source from which it extracts what it needs for its survival. Traditionally, many of the materials that we make are made to last: sturdy concrete for our houses, resistant coatings, the plastic of your screen, on the other hand living organisms are quite fragile. Their components are for the most part not meant to be the most stable under their living conditions. They need to be repaired, replaced, and all these tasks require energy. Only when a living organism dies do its component evolve towards thermodynamic equilibrium with its environment. This ability of life to take energy from its surroundings and to use it differently makes it what is called a dissipative system. Just like life they are not stable and survive through the energy they dissipate, and if chemistry has mainly focused on making stable components, they have only begun to be of interest lately.

Living organisms are capable of many feats; motion is one of them, and even on the molecular level. Kinesins can walk along rails carrying cargo with them, ATP-synthases can rotate and transform energy from one type to another or simply propel flagella which may propel the organism in return. These are examples of energy dissipation through the creation of work. Cells display remarkable sensing and communication properties. An embryo does switch from a single cell to multiple ones which through differentiation will yield different organs and structures. The reaction-diffusion processes behind this are allowed by the energy dissipated by the system. The creation of different biological processes synchronized with external ones, such as our circadian clock requires out of equilibrium components to react in sequential and timely manners. Structures can arise in micro-organisms, such as the tracks for the previously mentioned kinesins, microtubules. These structures not being at equilibrium allows them to be transitory and adapt to the environment. Notably their creation and disappearance are key in the mechanisms behind the crawling movement of cells. What is more fascinating is how all these properties intertwine with one another. If the materials of the future are to exhibit such properties, then out-of-equilibrium chemistry must be further investigated.

To perform such chemistry, the system must be able to be brought out of equilibrium through an energy input. In biology, the central energy input comes from electron transfer reactions. Even in plants, the light harvested is used so that electrons may transit through the embedded proteins of in the thylakoid membranes. But this energy is then stored in a multitude of ways and converted between them. ATP coined the “molecular unit of energy currency”, stores it through its high energy tri-phosphate link and releases it upon hydrolysis. Concentration gradients across membranes are also key to energy related processes as the favorable reflux of compounds can be used as an energy source, notably in the formation of ATP.

In order to study dissipative systems, we proposed to recreate one. In a temperature gradient, energy in the form of heat is dissipated as the energy is transferred from the hotter region to the cooler ones. By introducing possible chemical reactions in a temperature gradient, the reaction

between species and their diffusion towards other zones of the gradient represents an energy outlet and may lead to the creation of a chemical dissipative system. If such a system is to be sustainable, the reactions must be reversible, else once all the system has reacted the dissipation goes back to being a temperature gradient one. We thus investigated the creation a dynamic polymerizable system which could be used in such gradient experiments. Given the possibility for polymerization, the system could give rise to emergent behaviors, to this end we chose amino acids as building blocks, for there is a well-established link between sequence, structure and function for amino acids maximizing the possibility for "life-like" properties to emerge.
The objective then becomes creating a dynamic polymer system from amino acids which could behave differently according to temperature to dissipate the energy from the gradient. Through the energy input, reversible ligation and different states explored throughout the gradient, this system could also serve in dynamic combinatorial chemistry as the evolution of the monomers would differ depending on the location in the gradient and multiple structures could arise. Specifically, if different monomers are present the energy would allow for amino acids an exploration both in terms sequences, and folding patterns throughout the gradient, generating candidates for amplification mechanisms through interaction with desired target molecules.

In this manuscript, I will describe our steps towards the creation of such dynamic amino acid-based polymer system. In a first chapter, I will address the groundwork for such project, going through out-of-equilibrium processes in biology and what chemists have manage to recreate. The choice of the amino acid chemistry and what it entails in terms of ligation chemistry will also be addressed. Finally, this work is part of the temperature gradient experiment project, and thus preliminary modelling work was done to anticipate the desired chemistry for the system. This preliminary modelling work will be presented as it set the goal for the thermodynamic parameters for the ligation reaction. The next part will focus on the discussed dynamic ligation reaction, the N-methyl-cysteine dynamic native chemical ligation. This chapter deals with the kinetics and thermodynamic study as well as the refinement of experimental conditions. Once the exchange reaction characterized, we will turn towards the "living polymer" system we design exploiting the dynamic ligation reaction. The diversity of compounds generated the influence of temperature as well as structures observed will be addressed. Finally, the last chapter deals with the modelling and theory behind the living polymer, with the characterization of the complexity of the system as well as the development of algorithmic tools such as a finite element model for the system and some protein property forecast tools. The manuscript will then be concluded by a short summary of the results obtained and our perspectives for the project followed by the experimental section.

Chapter 1. Groundwork

1. Thermodynamic equilibrium

There are three laws in thermodynamics, and a main statement. This statement also called zeroth law is that if two systems are at thermal equilibrium with a third system then they are at thermal equilibrium with each other [1]. The first law is the conservation of energy. The total energy of an isolated system cannot change. If energy comes to pass into or out of the system such as work or heat, then the internal energy adjusts to account for the loss or gain of energy. The second law of thermodynamics states that the total entropy of an isolated system always increases over a thermodynamic process. And finally, the third law links entropy to disorder stating that at absolute zero, the entropy of a perfect crystal is null [2]. All possible processes abide by these laws. If one takes a closer look at the second law, it relates the heat exchanged by the system to its environment by $dS = \frac{dQ}{T} + dS_i$ with T the temperature and $dS_i \geq 0$ the entropy generated by the irreversible processes. For isolated systems there is no heat exchange, so we indeed have $dS = dS_i \geq 0$. This means that as spontaneous processes happen the entropy must increase until it reaches its maximum at which point equilibrium is reached and the system can no longer spontaneously evolve as this would imply yet another increase in entropy.

If we look back at the first law dealing with energy conservation and flow through the system, we obtain that $dU = dQ - dW$ meaning that the change of internal energy of the system dU corresponds to the energy gained by heat transfer dQ minus the energy lost through work dW. However, from the second law we know that $dS \leq \frac{dQ}{T}$ which in turn means that $dS \leq \frac{dU+dW}{T}$ or $dU + dW - TdS \leq 0$. Now if we consider a system where the only work is volumetric at constant pressure and temperature this can be rewritten as $d(U + pV - TS) \leq 0$ with p the pressure and V denoting the volume of the system. From here we see appear the definition of the enthalpy H, the sum of the internal energy of a system and the product of the volume and pressure $U + pV$. We can therefore conclude that for a transformation taking place at constant pressure and volume $d(H - TS) \leq 0$ also written as $dG \leq 0$ with G Gibb's free energy. Thus, during chemical processes at constant temperature and pressure Gibb's free energy diminished until, as denoted by second law an equilibrium is reached and its minimum value is reached. Gibb's free energy is the value that is mostly looked at especially in biology when it comes to chemical reactions. Another way of putting it would be that a process leading to a negative variation of free energy $\Delta G \leq 0$ is thermodynamically favorable and could spontaneously happen whereas one with a positive one could not.

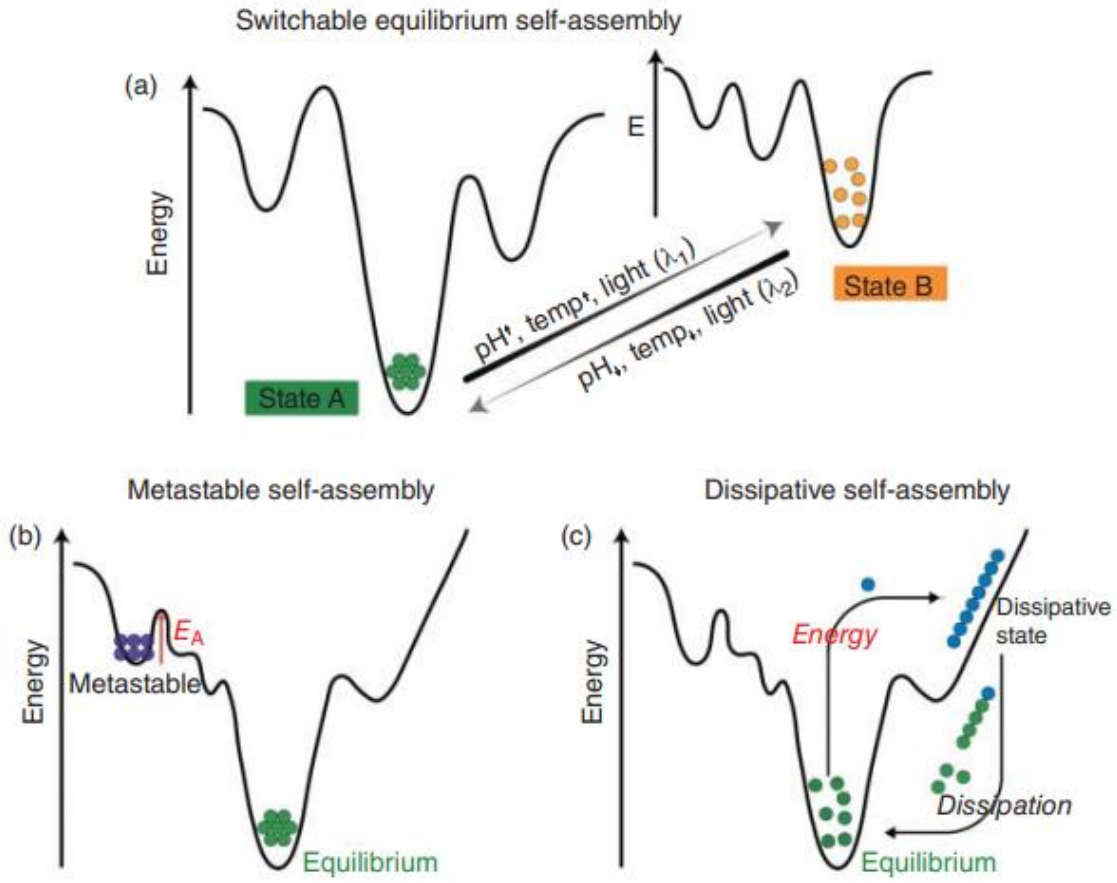


Figure 1-1 Differences in terms of energy landscapes between (a) classical switchable self-assembly, (b) kinetically controlled metastable self-assembly, and (c) dissipative self-assembly. Figure taken from ref.[3]

Thermodynamic equilibrium is then defined as: "a stable state in which no matter is exchanged with the environment (else the system would not be isolated, and the previous statements wouldn't hold) and there is also no net flux of energy through the system" [4]. The free energy from the system G can be linked to the equilibrium concentrations of species at equilibrium through $\Delta_r G^0 = -RT\ lnK$.

Now, even in equilibrium one can observe some complex, life-like behaviors. Self-assembly for example is the result of "the non-covalent interaction of two or more molecular subunits to form an aggregate whose novel structure and properties are determined by the nature and positioning of the individual components"[5]. If we consider equilibrium chemistry, the self-assembled state corresponds the minimum in the energy landscape as illustrated in figure 1-1 (a) and (b) with the unassembled states being higher in energy. As the system evolves the parts assemble as this minimizes the energy off the free system of the system until, once equilibrium is reached, there is a distribution of the part population between the assembled and the unassembled state based on their energy levels and the internal energy of the system[6]. The parts join and leave the assembly at the same rate resulting in now net flow of matter or energy in the system, equilibrium. Self-assembly is ubiquitously found in biology, whether in protein assembly, cytoskeleton structures[7]. Here we demonstrate how it intervenes in the formation of cell's phospholipid bilayers.

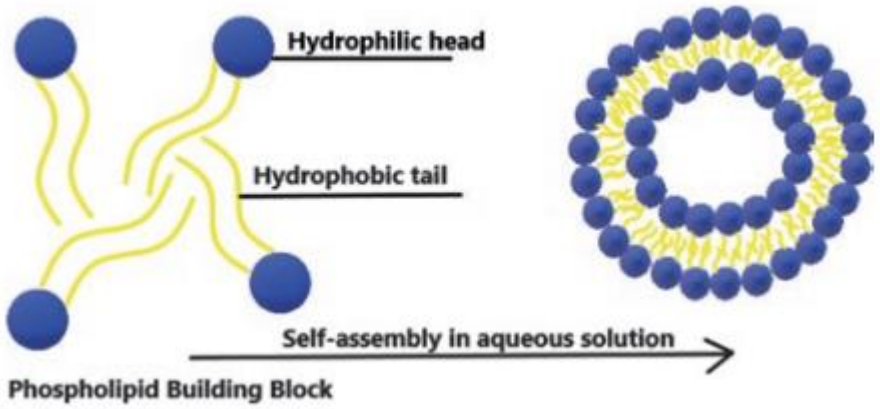


Figure 1-2 The formation of liposomes membranes by the self-assembly of phospholipids in aqueous solution[7]

The core component of cell membranes is a bilayer composed of phospholipids (as shown on figure 1). These molecules have a hydrophilic head and two hydrophobic tails and through hydrophobic interaction can self-assemble in water. Water molecules interact between each other notably through hydrogen bonding[8]. The hydrophobic tails of phospholipid do not interact strongly with water through hydrogen bonding or Debye interactions with water. This would lead to a reorientation of the hydrogen bonds to form a "cage" around the molecules. The presence of the hydrophobic tails would mean a reduction in mobility for the water molecules around it which translates to a lower entropy for the system[9]. The more favorable configuration is one where the tails come together forming a hydrophobic phase with the heads interacting in a favorable manner with water at the interface. The water molecules no longer in contact with the hydrophobic tails no longer adopt restricted configurations resulting in a counterintuitive increase in entropy. Even though phospholipid bilayers are thermodynamically favorable, biological membranes are more complex systems and are not at equilibrium as this would result in the death of the cell[10].

In the laboratory chemists have however developed self-assembly systems which simply evolve towards equilibrium. Figure 1-3 is a summary of some recent examples of such systems. Each of those examples makes use of different interactions and cleaver design of assembling parts. Mirkin's group[11] (as illustrated in (a)) makes use of DNA base-pair recognition through hydrogen bonding between complementary bases to link nanoparticles. The nanoparticle attached strand is composed of three parts. i) The linking part to the gold nanoparticle, non-binding and incorporating a thiol to attach to the gold surface and a recognition sequence that can bind to a linker present in the solution. ii) The terminating sequence of the linkers is a "sticky end" and allows for binding between them leading to assembly of the nanoparticles. iii) Finally, a spacing sequence which can bind to the linker between the recognition sequence and the sticky end gives control on the spacing between nanoparticles during self-assembly. This leads to the formation of solids exhibiting different crystallization lattices. Another use of DNA, showed in (b), reported by the group of Liedl[12] was the formation of helices as gold nanoparticles once again covered with strands of DNA assembled around DNA helix-bundles in specific locations due to sequence recognition between the strands. On both examples, DNA is used to create the required attractive interactions between the components that self-assemble, but the use of biological molecules is not necessary for self-assembly to occur, and repulsive interactions also lead to interesting results. As illustrated in (c) Akcora and coworkers grafted

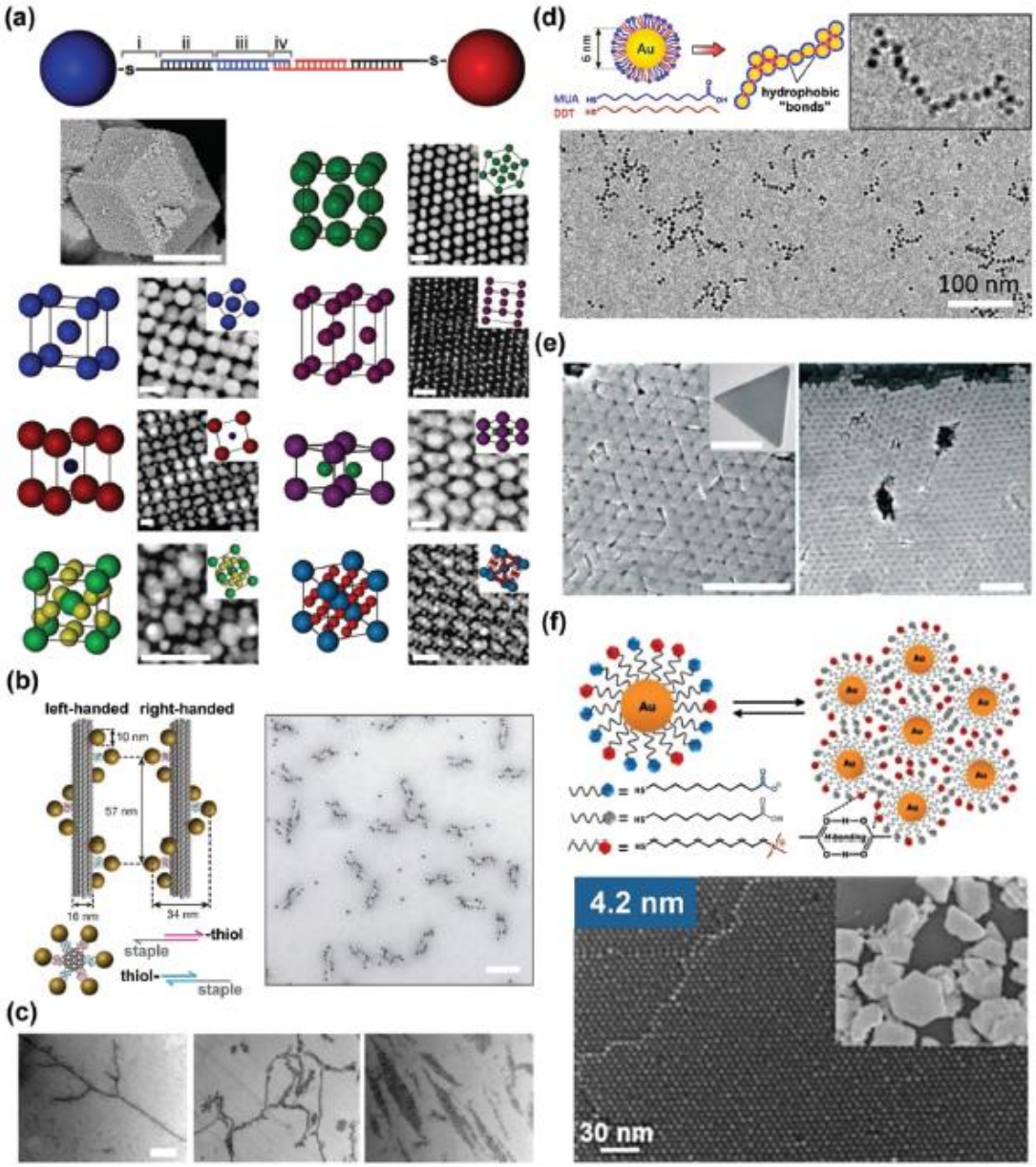


Figure 1-3 Equilibrium self-assembly, ESA, based on nanoparticles. (a) By controlling particle sizes and the nature of capping DNA ligands, nanoparticles can be crystallized into various types of superlattices. The DNA strands that drive the assembly comprise (i) an alkylthiol moiety and a nonbinding region, (ii) a recognition sequence that binds to a DNA linker, (iii) a spacer sequence of programmable length to control interparticle distance, and (iv) a ''sticky end'' sequence that drives self-assembly via DNA hybridization. The SEM image in the top-left is of a bcc microcrystal made of 20 nm gold nanoparticles (scale bar = 1 µm). Other schemes and TEM images illustrate ordering in nanoparticle superlattices isostructural with bcc, CsCl, Cr_3Si (left column, top to bottom), fcc, hcp, AlB_2 and Cs_6C_{60} (right column, top to bottom). Scale bars in TEM images = 50 nm. Reproduced from ref [11] *(b) Left- and right-handed nanohelices formed by nine gold nanoparticles attached to the surface of DNA origami 24-helix bundles. Each attachment site consists of three 15-nucleotide-long single-stranded extensions of staple oligonucleotides. Nanoparticles are capped with thiol-modified DNA strands complementary to staple extensions. The scale bar in the TEM image = 100 nm. Adapted from ref* [12]*. (c) Spherical silica nanoparticles grafted with polystyrene brushes and mixed with a polystyrene matrix form a range of anisotropic structures depending on the grafting density, molecular mass of the polymers, and annealing time. Scale bar = 0.5 µm. Adapted from ref*[13]*. (d) Gold nanoparticles functionalized with mixed monolayers of DTT (hydrophobic) and MUA (hydrophilic) ligands form chain-like structures via hydrophobic ''patches''.*[14] *(e) Positively charged nanotriangular prisms (see inset, scale bar = 100 nm) assemble into close-packed, mono- and multilayer arrays when electrostatic repulsions are offset by van der Waals attractions. Scale bars in SEM images = 1 µm.*[15] *(f) Scheme of nanoparticles decorated with mixed SAMs comprising ligands with terminal carboxylic (MUA) or quaternary amine (TMA) groups. Although the particles have net positive charges, their crystallization is driven by hydrogen bonding between MUAs. SEM images of 3D crystals resulting from the self-assembly of 4.2 nm, AuNPs at pH ~ 4.*[16] *Figure adapted from ref.*[4]

silica-nanoparticles with polystyrene combs. The inorganic phase of nanoparticles tends to be immiscible with an organic one. The silica nanoparticles prefer interacting between each other, but by grafting the polystyrene those interactions are shielded, as when the particles come too close to one another they distort the polymer brush, which is not entropically favored. Although even if the nanoparticles are spherical and grafted homogenously, their incorporation in polystyrene leads to their agglomeration of the nanoparticles in anisotropic structures as predicted by their modeling work[13]. Including different interactions in the chemical system creates a competition between them which controls the properties of the assembly. By coating gold nanoparticles with both hydrophilic (Mercapto-undecanoic acid) and hydrophobic (Dodecanethiol), they observe assemblies, the hydrophobic ligands can assemble in hydrophobic patches around them enabling nanoparticle assembly by hydrophobic interaction as illustrated in (d), the hydrophilic ligand stabilizing the free surface of the assembly to the solution[14]. Even electrostatic repulsive interactions can be used, and they can be compensated by Van der Waals interactions by nanoparticle cores as reported by [15] and illustrated in (e). Or through hydrogen bonding between other ligands present around the nanoparticle (f)[16].

However, unlike these examples suggest. Not all systems will spontaneously evolve to minimize free energy. If one considers the energy landscape as illustrated in figure 1-1 there is always an energy requirement to change states. For systems spontaneously evolving towards equilibrium, this energy barrier must be low enough for the energy of the system, such as thermal energy, to overcome it. The system can then evolve towards the energy minimum. Some systems however can be trapped in a kinetic trap or metastable state. Their current state is not the most stable, however the system lacks the required energy to by-pass the energy barrier. One of the most common examples of such systems is carbon. The most stable form of carbon under atmospheric pressure and room temperature is graphite[17], however diamonds do not turn into graphite over time because the activation energy required for it is way above the energy of the system at room temperature.

A system which exploits the metastability of its component while also not including nanoparticles nor DNA has been proposed by Aida[18], (figure 1-4) displaying the potential for organic molecules to create subassemblies. In this system at 25°C, the monomer **M** adopts a caged-like structure, favored by intramolecular hydrogen bonding, and does not form any assemblies. When the temperature is increased, monomers can unfold and act as an initiator for the polymerization of fibers through inter-molecular hydrogen bonding. The same can be achieved with an activator where the hydrogen bonding amines have been methylated favoring the unfolded state and initiating the polymerization. Although energy must be brought to the system to bring it to a higher energy state for the structures to appear, the structures themselves are not out-of-equilibrium as will be discussed later, if the heating is stopped, the structures are stable and do not disappear.

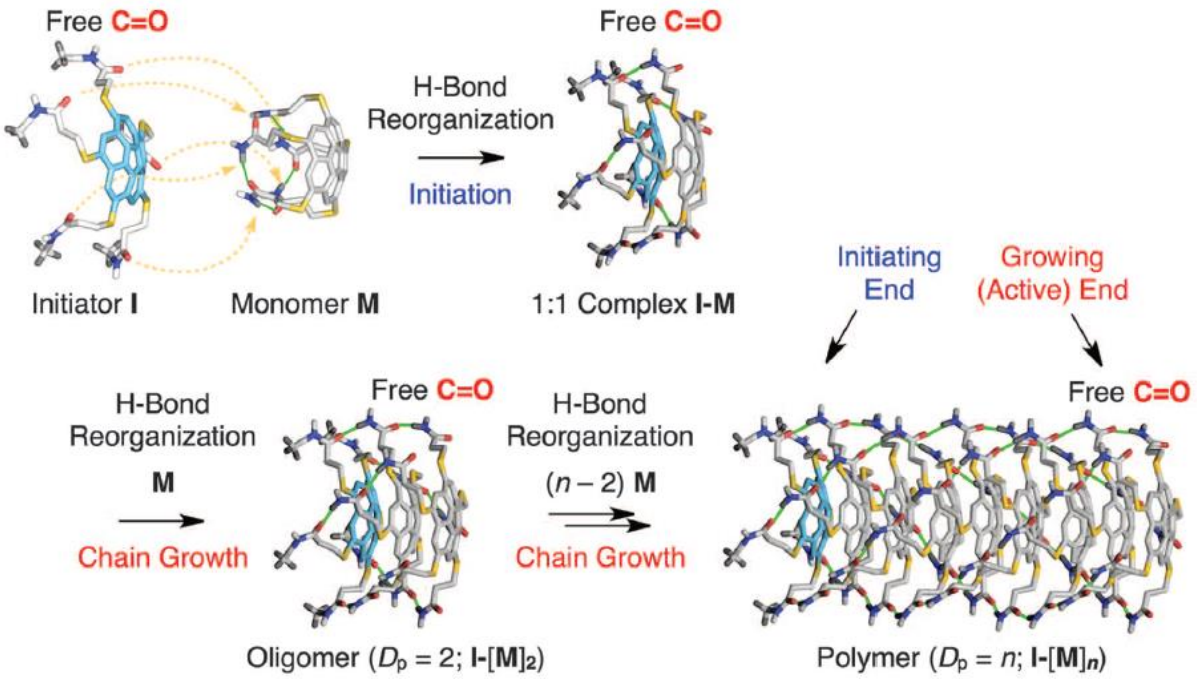


Figure 1-4 Schematic representation of the chain growth supramolecular polymerization of monomer M initiated with I, where the chain growth is accompanied by the H-bond reorganization of M. The growing polymer carries an initiator unit at one end (the initiating end), whereas the other end [the growing (active) end] adopts a structure analogous to I, with free amide C=O groups. This structural feature prevents bimolecular coupling of the propagating ends.

2. Out of equilibrium, dissipative systems

a. Reaching out of equilibrium

If we go back to the definition of thermodynamic equilibrium: "a stable state at which no matter is exchanged with the environment and there is also no net flux of energy through the system" we can see two ways to bring the system to an out of equilibrium state. If matter is removed or gained by the system, and if there is a systemic flux of energy. In both cases, the environment must interact with the system to bring it out of equilibrium. The second law of thermodynamics showed that the entropy of an isolated system must increase and using the first law we demonstrated that this translated for isobaric isochoric transformations to a decrease in free energy of a system. However, if a system interacts with its environment these principles now apply to a bigger system: the initial system and its environment. This means that the free energy of a system may increase, so long as it is accompanied by a greater decrease in the environment resulting in an overall decrease. This allows a system to be brought out of equilibrium without violating the thermodynamic principles. As illustrated in figure 1-1, the energy inputs allow a system to reach high energy states from which it will relax back towards equilibrium, dissipating the received energy along the way, through irreversible processes like thermal dissipation. Those corresponds to the dS_i in the second law, resulting in the increase of entropy. As soon as the energy input stops, these irreversible steps dissipate all the surplus energy as the system evolves towards equilibrium, or in the case of life, death. Thus, entropy production is intrinsically associated with out of equilibrium processes[4].

We can then breakdown two important factors to obtain an out of equilibrium chemical reaction network.

First an external energy input which will allow the out of equilibrium processes. There are many ways energy can be brought to a system. Just to give a few examples: light excitation of molecules like photo switches, molecules able to change structure through electro-magnetic irradiation[19 20 21 22], the use of magnetic fields to transfer energy to magnetic systems [23 24 25], chemical energy where a molecule is used as fuel to keep the system out of equilibrium [26 27] and even flow fields.[28]

Second, the system must possess at least two irreversible reactions which control the activation and deactivation of the chemical processes[6] (this will be touched upon later). If we have a precursor at the minimum in the energy landscape (figure 1-1), it must be brought to a higher energy state by an activation reaction in which it can partake in assemblies or reactions. In parallel there must also be a way for the high energy state to be brought down to the initial precursor state.

b. The potential for out of equilibrium chemistry

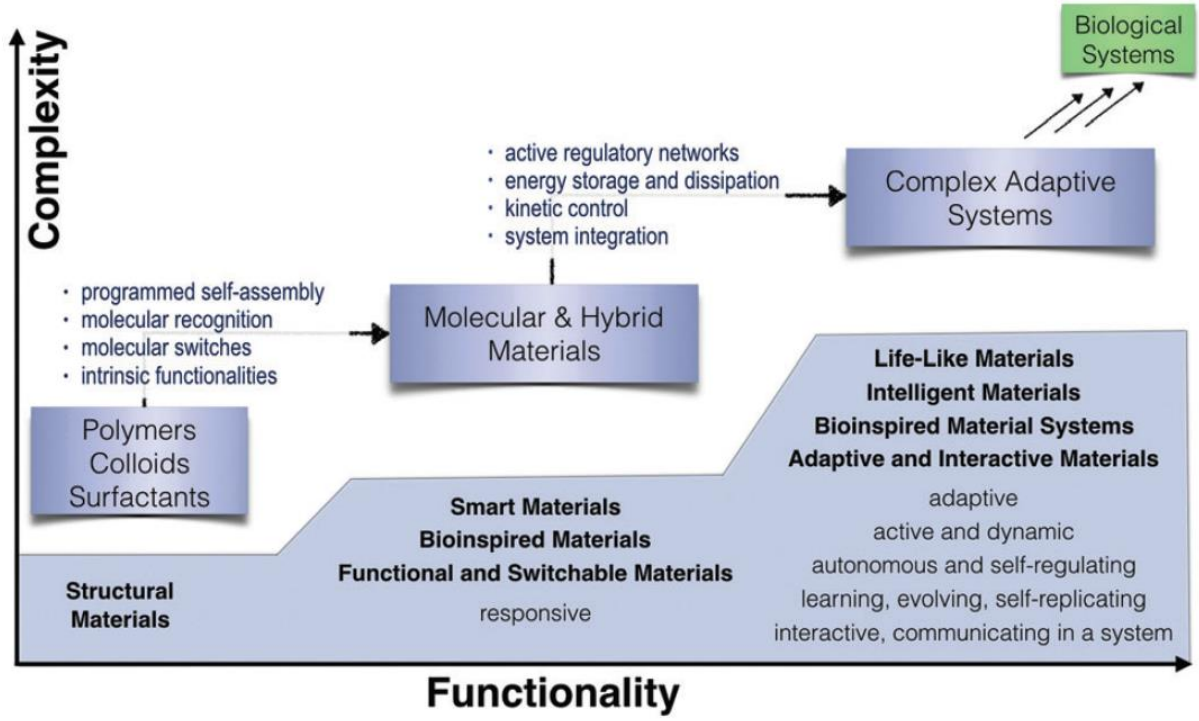


Figure 1-5 Correlation between complexity and functionality in molecular[29]

For the longest of times, material sciences have focused on making stable materials, whether at equilibrium or metastable. Over the past decades, the study of biological structures and research on the building blocks, materials were made smarter. They can self-assemble, have intrinsic functionalities such as conductivity and respond to external triggers. In a sense they have been made more responsive, but such responses still rely on external inputs[29].

Living organisms through the flux of energy available to them display impressive features. They can adapt to their media, replicate, amplify signals, and evolve. If we are to create intelligent materials, able to adapt by themselves, display systemic communication, heal, replicate, and even evolve, then there is still work to be done on the understanding of the out of equilibrium processes which rule biology[6].

In ref. [29] Mérindol identifies four main categories of function that rely on out of equilibrium processes:

- Temporal control
- Information processing
- motion and mechanical work
- autonomous structures

For each of those categories I will provide an example of a biological process and a how recent advances in chemistry try to reproduce them in abiotic systems.

3. Properties linked to out of equilibrium processes

a) Temporal control

There are many forms of periodic cycles among living organisms, an example that everyone is familiar with is the day and night cycle. These cycles are permitted by oscillating concentrations of signals of varying periods based on the cycle. To have such oscillations there are certain required behaviors, feedbacks. There two types of feedback, negative and positive. Negative feedback takes place when the product of a process makes that process less likely. For example, of figure 1-6 A, the signal **b** increases the energy barrier of the process creating **a** from its precursor pro-**a**. To have proper oscillations the feedback process must be delayed in time, else the system can reach a steady state if **a** directly reduces its own formation speed. B gives an example of delayed feedback; pro-**a** is processed into **a** which is then transformed into **b** which has an influence on the formation of **a** from pro-**a** (and therefore on its own formation as well). This means that as the process produces more and more **a** (and afterwards **b**), it does so slower and slower until the barrier is too high and no more a is produced. This process is enough to obtain oscillations (if **b** can be degraded to allow for a new cycle to start) but biology often incorporates a positive feedback loop where the product of a process increases the speed of that process (as seen on C). Positive feedback loops allow the kickstart of each new cycle. If we look at these systems considering our previous discussion on equilibrium, we can see that the systems remain out of equilibrium for as long as there is something to keep the oscillations going.

Going back to the circadian clock (depicted in F) which has a period of 24 hours, and its oscillations or coupled to other temporal cycles in the organism. A protein complex CLOCK-BMAL1 will bind to DNA and lead to the transcription of its inhibitors the PER and CRY proteins. However once synthesized in the cytoplasm PER and CRY first must undergo phosphorylation before being able to enter the nucleus and inhibit the CLOCK-BMAL1 complex which reduces their transcription thus their concentration and the cycle starts anew. In parallel CLOCK-BMAL1's interaction with DNA will also have a positive and negative feedbacks on the transcription of BMAL1 which makes for more robust oscillations[30–32].

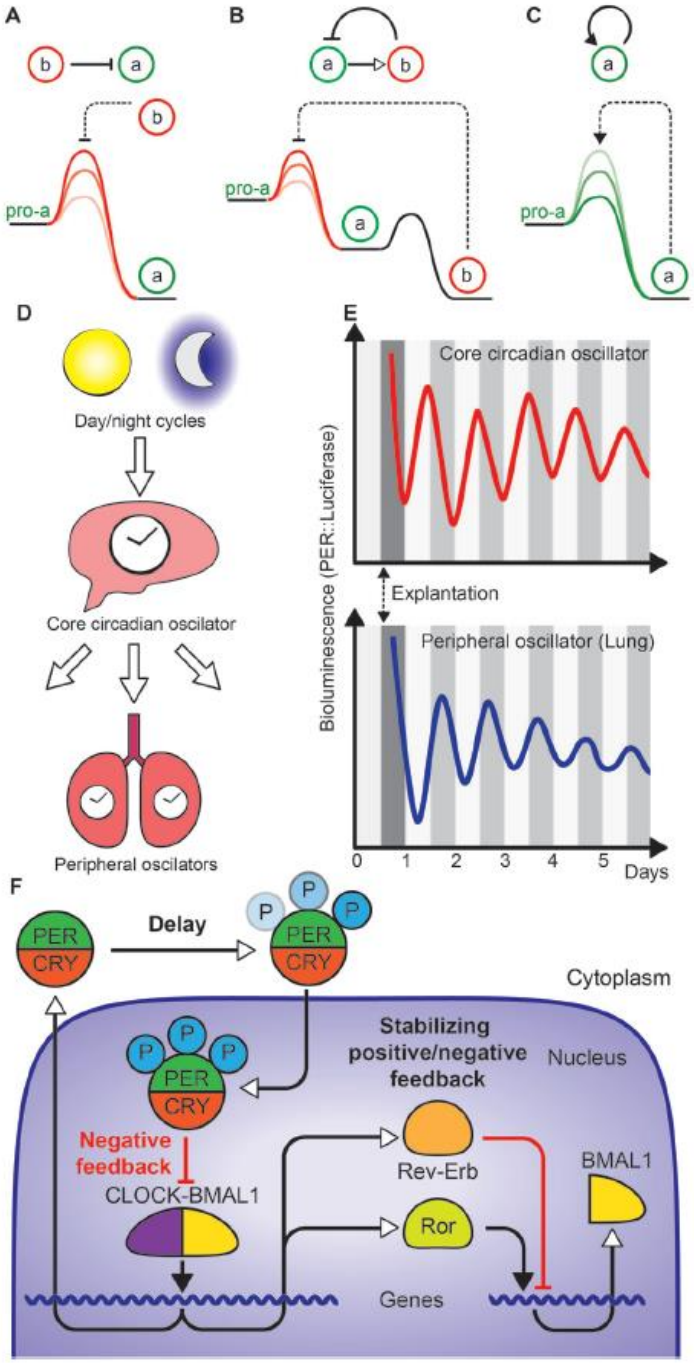


Figure 1-6 Biological regulatory networks for autonomous temporal control. (A) Schematic representation of the down-regulation of ''a'' production by ''b'', symbolized by an arrow with a flat head (top), and an exemplarily corresponding reaction profile, where the amount of signal ''b'' is positively correlated with the activation barrier of formation of ''a'' (bottom). (B) Schematic representation and reaction rate profile of a delayed negative feedback, where the delay induced by the transformation of ''a'' into ''b'' is represented by an arrow with empty head. (C) Autocatalytic positive feedback for the formation of ''a'', symbolized by an arrow with a filled head (top), and the corresponding reaction profile, where ''a'' decreases its own activation barrier of formation (bottom). (D) Schematic representation of the hierarchical circadian network, in which the core clock synchronizes the metabolic clock in the lungs while being regulated by the day/night cycles. (E) Real time visualization of PER expression using PER::luciferase fusion proteins in the supercharismatic nucleus (core oscillators) and in lungs (peripheral oscillators) in mice. Tissues were explanted at day 0.5, white and dark lines show day/night cycles. (F) Schematic representation of the feedback loops controlling circadian oscillations. Figure taken from ref. [29]

(R1) $HOBr + Br^- + H^+ \rightleftharpoons Br_2 + H_2O$

(R2) $HBrO_2 + Br^- + H^+ \rightarrow 2HOBr$

(R3) $BrO_3^- + Br^- + 2H^+ \rightarrow HBrO_2 + HOBr$

(R4) $2HBrO_2 \rightarrow BrO_3^- + HOBr + H^+$

(R5) $BrO_3^- + HBrO_2 + H^+ \rightleftharpoons 2BrO_2^{\cdot} + H_2O$

(R6) $BrO_2^{\cdot} + Ce^{3+} + H^+ \rightleftharpoons HBrO_2 + Ce^{4+}$

(R7) $BrO_2^{\cdot} + Ce^{4+} + H_2O \rightarrow BrO_3^- + Ce^{3+} + 2H^+$

(R8) $Br_2 + CH_2(COOH)_2 \rightarrow BrCH(COOH)_2 + Br^- + H^+$

(R9) $6Ce^{4+} + CH_2(COOH)_2 + 2H_2O \rightarrow 6Ce^{3+} + HCOOH + 2CO_2 + 6H^+$

(R10) $4Ce^{4+} + BrCH(COOH)_2 + 2H_2O \rightarrow Br^- + 4Ce^{3+} + HCOOH + 2CO_2 + 5H^+$

Figure 1-7 Reactions taking place during the Belousov-Zhabotinsky oscillating reaction, the ionic species are likely to be forming complexes and not be in free form. Adapted from ref.[33]

In the laboratory, the most famous oscillating reaction is the Belousov-Zhabotinsky (BZ). In a solution of sulfuric acid, potassium bromate, cerium sulfate and malonic the concentration of bromide ions and cerium (IV) oscillate. According to reactions 5 and 6 the more $HBrO_2$ in solution the more bromate ions react with it to form the $BrO_2^{\cdot}$ radical which in turn oxidizes cerium III to cerium IV creating $HBrO_2$ in the process. This is the positive feedback loop, the autocatalysis of $HBrO_2$ in the system. However, as the concentration of cerium IV increases, the speed of reaction 10 increases which liberates bromide ions back in the system which will react with the free $HBrO_2$ acting as a negative feedback loop. This reaction has been extensively used to explore out of equilibrium oscillating systems due to its tunability through many factors[34–36], but the harsh conditions required make it incompatible with a lot of chemical systems especially biocompatible ones.

Other types of systems have been developed by combining synthetic material as well as biological resulting in hybrid systems. Some enzymes have very attractive properties when it comes to their activity depending on conditions which they modify which makes for easy feedback loops. Urease for example is more active at pH 7. In presence of urea at pH 3 the activity is low but as the urea is hydrolyzed to create ammonia which increases the pH, the reaction becomes faster and faster leading to a positive feedback loop (figure 1-8-A)[37]. Another example would be trypsin, a protease (an enzyme capable of cleaving peptide sequences from other proteins). The precursor to trypsin is trypsinogen which differs from trypsin by an additional peptide sequence which renders trypsinogen inactive. Trypsin can cleave this peptide sequence increasing its concentration in a positive feedback loop. If trypsin can also cleave and activate a molecule that can after a delay act as its inhibitor as shown in C and D of figure 1-8 then the negative feedback allows for oscillations[38]. This process as opposed to the BZ system requires an open reactor microfluidic system to evacuate the inhibitor and allow a new cycle to start.

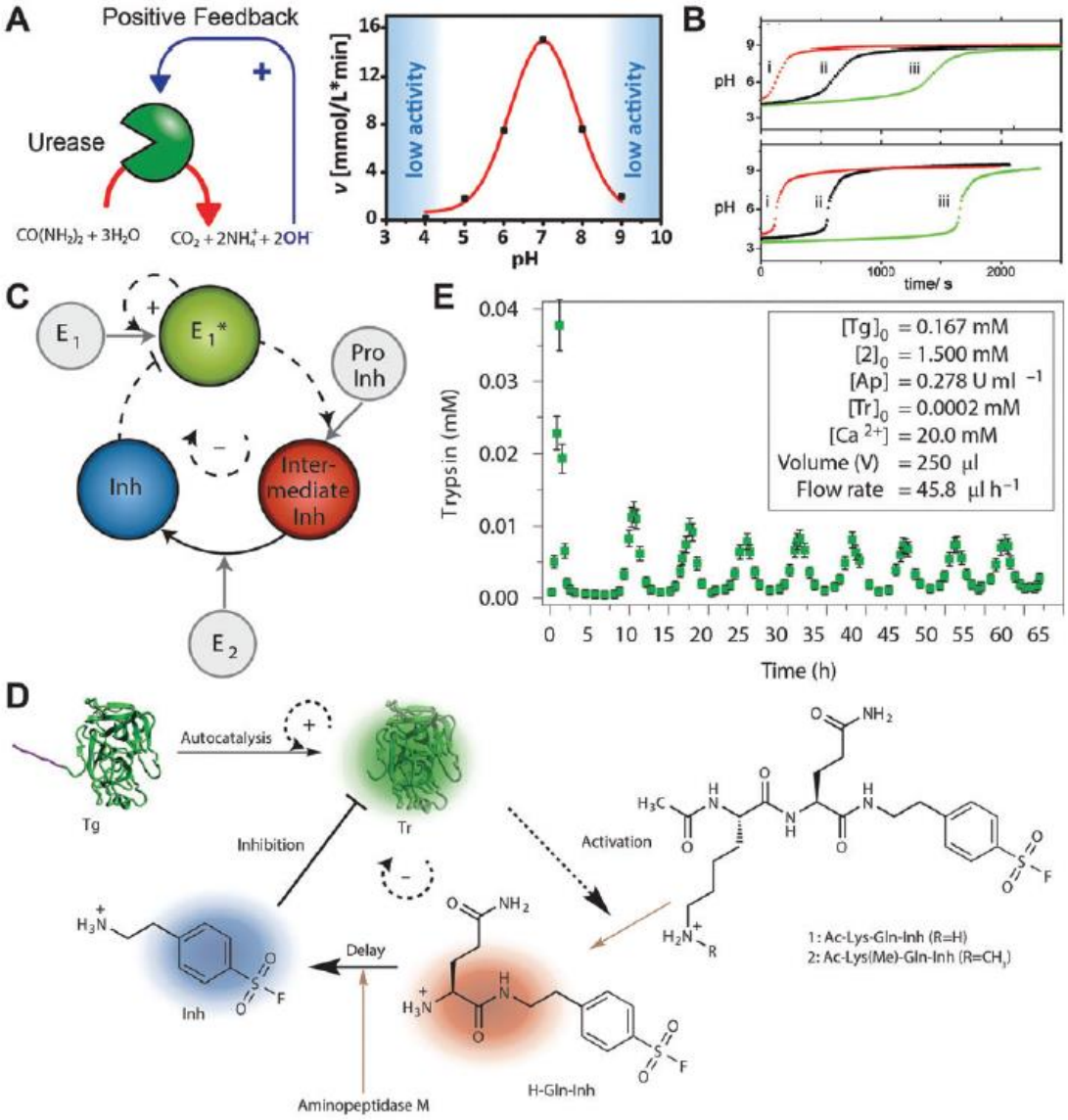


Figure 1-8 Systems for temporal regulation. (A) Schematic representation of positive feedback for urea/urease system under acidic conditions together with the bell-shaped activity curve of urease.[37] *(B) pH profile of a system containing urea (5 mM), urease (1.4 U/mL) and increasing concentrations of (a) acetic acid (i = 0.2, ii = 0.58, iii = 0.93 mM) or (b) sulphuric acid (i = 0.05, ii = 0.11, iii = 0.18 mM).*[39] *(C) Schematic representation of the feedback mechanism in the trypsin oscillator and (D) actual molecular network. (E) Trypsin oscillations obtained with the conditions shown in the inset.* [38] *Figure taken from* [29]

b) Autonomous structures

The cytoskeleton is the structure responsible for many cellular processes. It comes into play for propulsion, motion of the cell, movement of organelles, chromosomes during mitosis, vesicle excretion etc. It is composed of three types of structures: microtubules, actin filaments and intermediate filaments[40 41]. Microtubules shown on figure 1-9 are the largest of these structures. Proteins known as tubulin α and β assemble as dimers and can bind GTP (guanosine triphosphate) a highly energetic molecule[42]. When a tubulin dimer binds GTP, it adopts a "straight" configuration and gains a strong affinity for other GTP-dimers resulting in assemblies named protofilaments which in turn assemble as microtubules. A microtubule cross section contains 13 protofilaments. Tubulin dimers however have a GTPase activity, meaning that the catalyze the hydrolysis of GTP into GDP (guanosine diphosphate) and free phosphate. The tubulin-GDP dimer remains stable inside the microtubule, maintained in a straight compact conformation by its interactions with other tubulin dimers. However, at the extremity of the microtubule when the GDP-dimer is not interacting with a dimer behind it, it adopts a twisted conformation and depolymerizes from the microtubule. This makes microtubules a dynamic polarized structure with two different extremities. One marked by a "+" where it polymerizes and has a high concentration of GTP-dimers known as the "GTP cap" which prevents depolymerization on this end, and the other one marked "-" from which it depolymerizes[42].

If GTP runs low in the system, the microtubule can lose its GTP cap and starts depolymerizing rapidly from both extremities, this event is known as a catastrophe. The GTP cap can be restored by addition of GTP in the system resulting in a rescue of the microtubule.

Microtubules can easily be seen as an out-of-equilibrium system with a chemical fuel GTP as its energy source. If one takes a solution of buffer, GTP, tubulin, microtubules, GDP and phosphate then microtubules will appear and assemble until GTP runs out of the system and results in catastrophes for all microtubules which then disappear. In animal cells there are structures that stabilize extremities of microtubules and enables the organization of intracellular traffic, the centrosomes. During the interphase the centrosome is duplicated, and a copy migrates to each pole of the cell during mitosis organizing the microtubule network between them. The microtubules then attach to the duplicated chromosomes and allow the drag a copy of each to the two poles of the cells. [43]

Actin filaments are similar to microtubules in the sense that they assemble from ATP-actin complexes with ATP (adenosine triphosphate) another highly energetic molecule. At the positive end ATP-actin polymerizes, then the ATP gets hydrolyzed in ADP and at the negative end the filament disassembles. The fact that actin and microtubules rely on a different type of chemical fuel allows for their existence and independent control. During mitosis, once the chromosome copies have migrated to different poles actin rings will form around the membrane on the cell and contract separating the cell in two[41]. Mitosis occurrences result of oscillating temporal cycles discussed previously further emphasizing the interactions and communications between these out of equilibrium processes in biology[44].

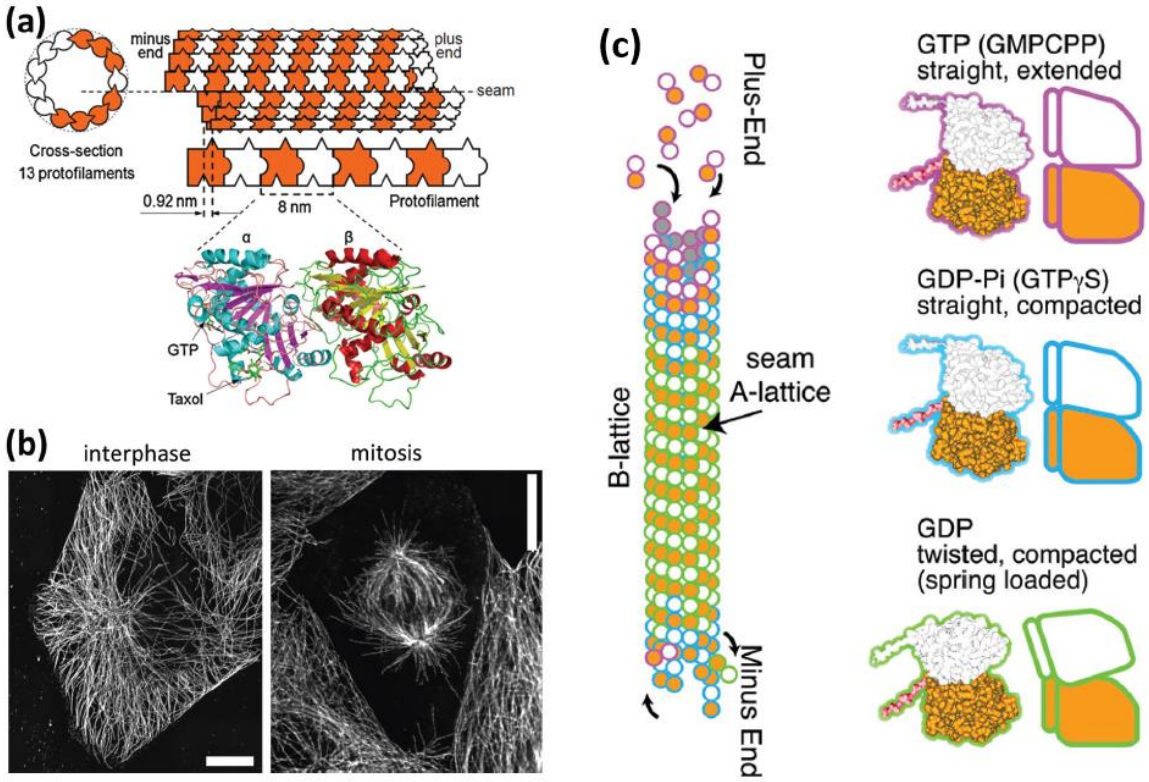


Figure 1-9 (a) Schematic representation of amicrotubule with 13 protofilaments.Protofilaments are made from tubulin dimers which bind to each other in ahead-to-tail fashion. Protofilaments bind to each other with an offset which results in a seam in the microtubule structure. The ribbon diagram is derived from electron crystallography[45] *and shows the GTP and Taxol binding sites. (b) During interphase, the microtubule cytoskeleton is radially organized, with microtubules emerging from a microtubule-organizing center, centrosomes, located in the vicinity of the nucleus (visible as the microtubule free central region). During mitosis, two spindles of oriented microtubules form, then the microtubules attach to the chromosomes initially located along the centerline of the dividing cell and separate them by exerting forces generated from depolymerization. Scale bars: 10 mm. (c) Microtubule lattice with GTP cap at plus and minus ends, and GDP tubulin in the body. The lattice is a B-lattice, with alpha–alpha lateral interactions. (b) Tubulin conformations GTP, straight and extended conformation, GDP-Pi straight and compacted conformation, and GDP in the compacted and twisted conformation. Figure adapted from ref.* [42]

These examples show interesting characteristics for out of equilibrium structures, they require energy (or stimuli as denoted on figure 1-10) for the parts to assemble and when the stimuli stop, the structures might be meta-stable and last for some time but will eventually vanish as the system relaxes towards equilibrium. The oscillating systems discussed previously can be used to create such systems as examples will show us. However, here, the structural component is the active molecule that harnesses the energy from the system (tubulin catalyzes the hydrolysis of GTP) and has direct influence on the structure, whereas for temporal oscillations the system had to go through another species for the negative feedback to take place, making the system more responsive than active. This can lead to delayed responses and non-homogenous behaviors especially at the macroscale which will be discussed later on.

In part 2) we introduced photo-switches, molecules capable through absorption of radiations to modify their structures, thereby harnessing the light's energy to create higher energy states and dissipating energy in the process. These structure changes can sometimes be associated with pH modifications, an example is given in figure 1-10 B, a nanoparticle system which agglomerates at high pH is associated with Merocyanine, this photo-switch upon irradiation converts to spiropyran and releases a proton thus lowering the pH and solubilizing the nanoparticle agglomerates. Once the light stimulus stops the spiropyran reverts to merocyanine, recapturing a proton and the system reverts to agglomerated particles. When this system was

incorporated in a hydrogel (figure 1-10 D), blue light masking was able to create patterns. In the regions of the gel where the light is masked, the photo-switch is in the merocyanine form leading to higher pH and the nanoparticles agglomerate. The agglomerates make darker regions in the hydrogel making the masked motives visible. With these methods the stimuli indirectly impact the assembly, the photo-switches change the conditions leading to the self-assembly and the structures are not maintained out-of-equilibrium, the overall system is through the photo-switches. [46]

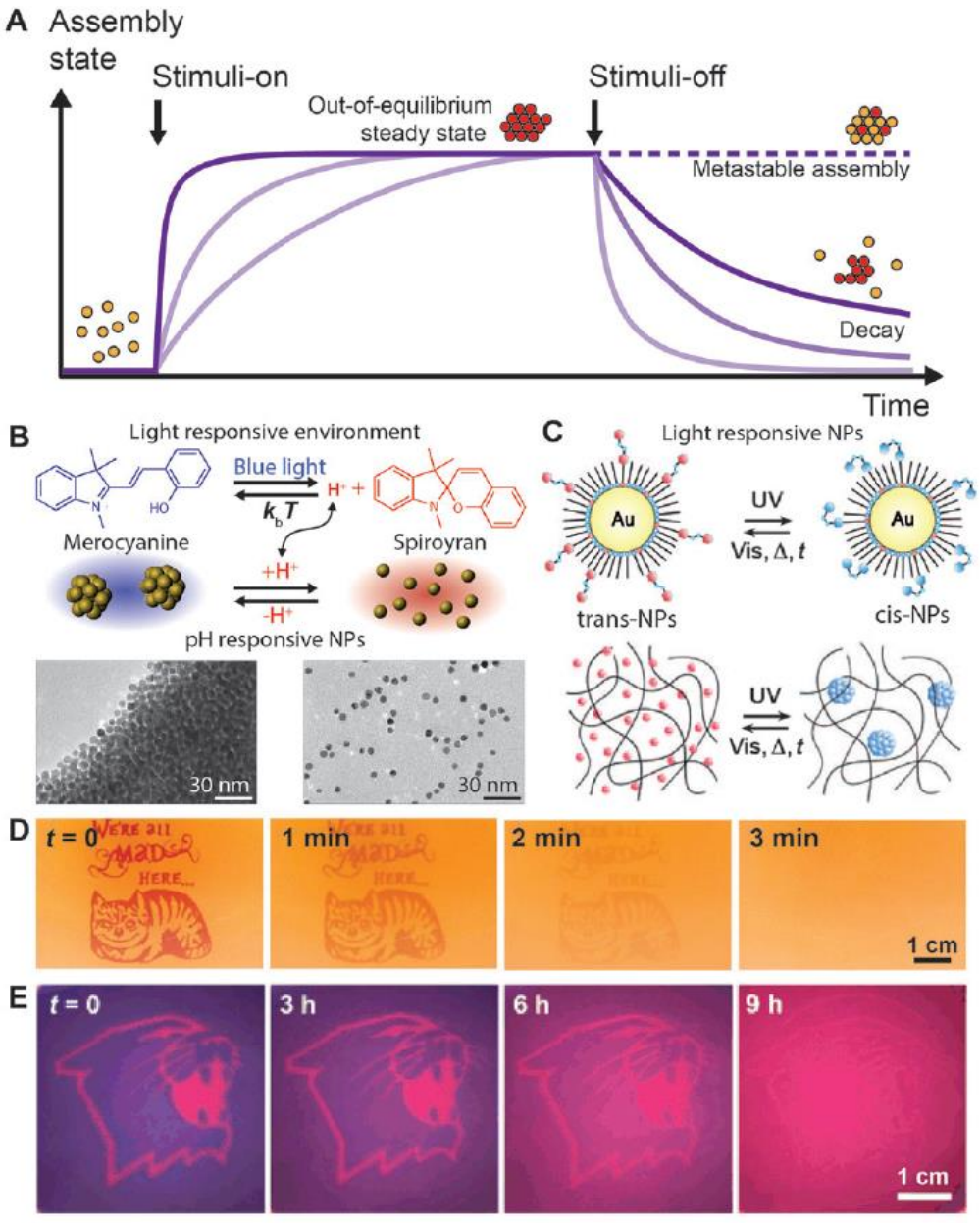


Figure 1-10 Autonomously decaying structures. (A) Temporal behavior of stimuli induced assembly regarding the speed of formation and the structural stability.[29] *Active environments: (B) schematic representation of blue-lightinduced aggregation of pH-responsive nanoparticles associated with the photoinduced merocyanine to spiropyran transformation, which releases protons. (D) Self-erasing image due to the spontaneous reaggregation of blue light dispersed pH responsive NPs.* [46] *Active structures: (C) schematic representation of UV-induced aggregation of azobenzene-decorated Au-NPs. (E) Self-erasing images due to the spontaneous decay of the azobenzene-decorated Au-NP aggregates.* [22]

Azobenzenes are another type of photoswitch able to change from *trans* to a *cis* configuration upon light irradiation, affecting molecular properties such as dipolar moments. Gold nanoparticles decorated with azobenzenes have been demonstrated to self-assemble upon UV irradiation is illustrated in figure 1-10 C & D. When the azobenzenes adopt a *cis* configuration, the nanoparticles interact together through Debye interactions resulting in agglomeration. When the UV light is turned off the azobenzene revert to *trans* and the agglomerates disperse. This technic was also used with masking to create motives in hydrogels. In this example however, the active component (the photoswitch) directly influences the assembly of the structure.[22]

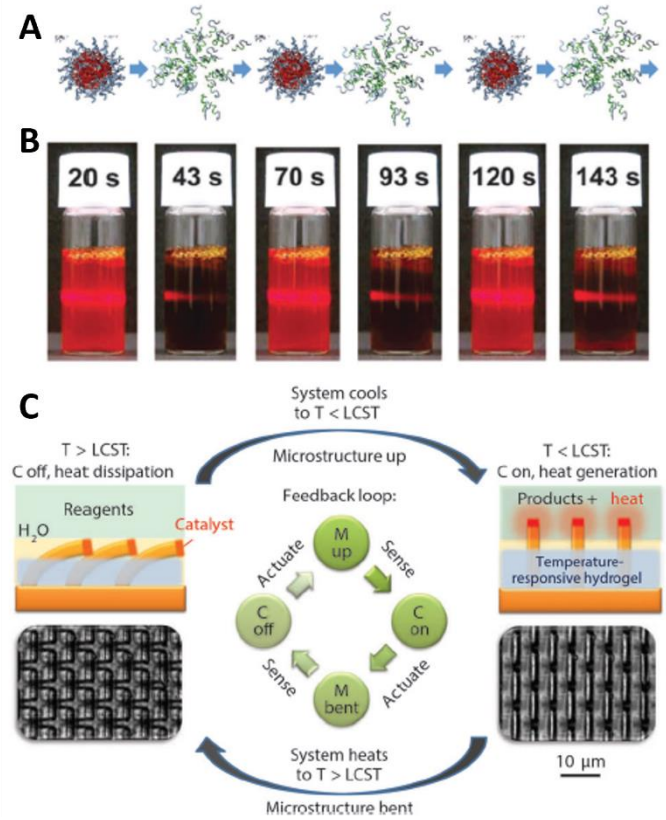


Figure 1-11 (A) Scheme of the assembly and disassembly of structural elements at the core of an oscillating structure. (B) Variation of turbidity associated with micelle formation.[47] *(C) Mechano- chemical oscillations of rigid posts embedded in a temperature responsive hydrogel. The catalyst at the top of the posts contacts the reagents, leading to heat-up and subsequent bending of the posts into the reagent free layer until the temperature drops to reinitiate a cycle.*[48]

It is also possible to combine the previously discussed oscillating processes with the creation of structures as life does with the example of microtubules in the mitosis cycle. The BZ reaction allows for redox cycles as the concentrations of cerium IV and bromide oscillate alternatively. Poly(N-isopropyl acrylamide) hydrogels (PNIPAM) with covalently bound ruthenium tris-(2,2'-bipyridine) catalysts have been developed where the ruthenium is used instead of the cerium in the BZ reactions oscillating between Ru^{2+} and Ru^{3+} . This change in charge modifies the backbone of the polymer gel leading to the alternating shrinking and expansion of the gel. [34] Similar systems have been developed such as a block copolymer which can periodically assemble into micelles, modifying the turbidity of the solution as shown in figure 1-11 B.[47] There are however other possible oscillating synthetic systems, one of which is illustrated in figure 1-11 C. In this system a heat responsive hydrogel is designed to have a catalyst sticking out of the gel. In a microfluidic experiment, the reactants for the catalyzed reaction flow over the hydrogel and react with the presence of the catalyst. The exothermic reaction leads to an

increase in temperature which in turn induces the collapse of the hydrogel, separating the catalyst from the reagent layer. As the heat dissipates, the hydrogel swells again and the catalyst becomes in contact with the reactant again and a new cycle begins.[48]

c) Motion and mechanical work

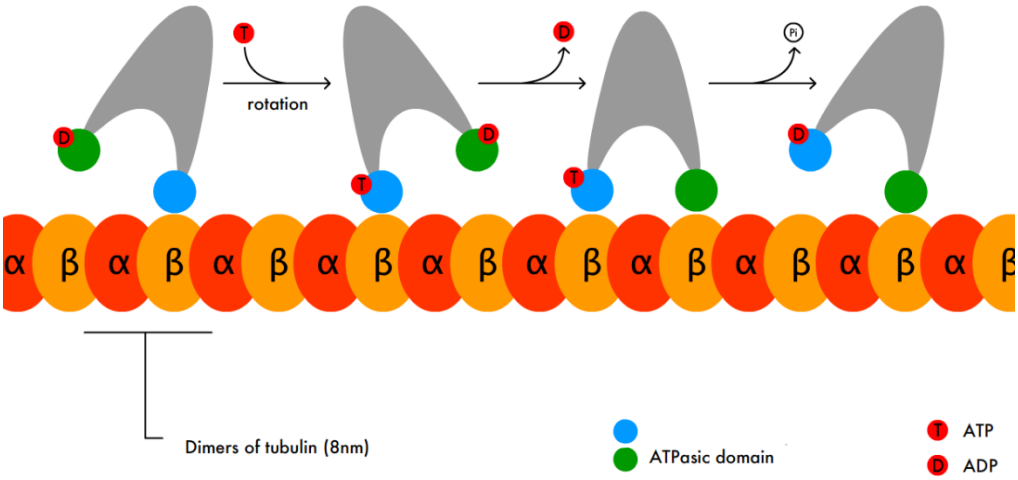


Figure 1-12 Biological motion. Schematic representation of the coordinated walk of kinesin motors on microtubules fueled by ATP hydrolysis. Taken from ref.[49]

Another way energy can be dissipated is through motion. Energy can be converted into mechanical work by molecular motors which fulfill many roles in living organisms. The microtubules presented before notably serve as tracks for vesicular transport inside the cell. The vesicles are attached to the microtubules via motion proteins that allow transport across the microtubule. Kinesins will move towards the "+" extremity[50] of the microtubule whereas dyneins will move towards the "-" extremity[51]. An example of such protein is shown in figure 1-12. The protein is composed of two subunits, a and b, which will alternate catalyzing the hydrolysis of ATP and use the resulting energy to move a dimer of tubulin forward. These molecules are dubbed molecular walkers.

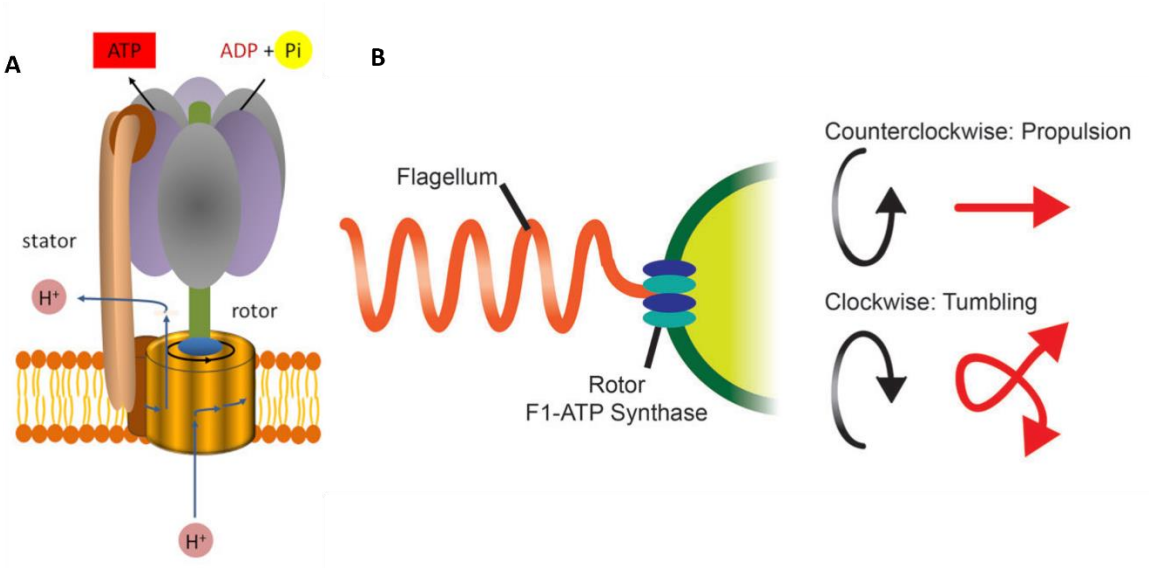


Figure 1-13 (A) Representation of the ATP-synthase complex, (B) the flagellar propulsion of bacteria. Adapted from refs.[50,52]

The ATP-synthase (figure 1-13 A) the protein complex responsible for the synthesis of ATP, an omnipresent source of chemical energy in biology[29], is able to transform energy, available in the form of chemical energy, in a rotational motion similar as to how motors transform fuel in a rotational motion. An exergonic flux of protons resulting from a pH gradient across a membrane gives the required energy for the rotor part of the molecule to rotate[53,54], this rotation leads to modifications on the spherical outcropping of the ATP synthase, allowing it to bind ADP and phosphate, catalyze the ATP formation and then release it alternatively on three different sites.[51] This protein's capacity to transform energetically favorable proton fluxes in rotational movement is also exploited by *E. coli* where a left-handed chiral filament is affixed to the rotor part[55]. This leads to a forward propulsion when the rotor turns counterclockwise and a tumbling motion when the rotor rotates clockwise as illustrated on figure 1-13-B.

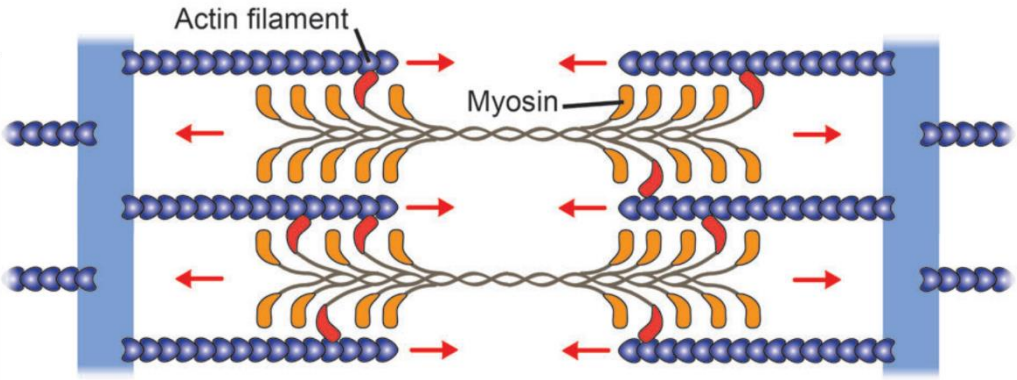


Figure 1-14 The organization of myosin motors and actin filaments in the muscle sarcomere. Adapted from ref.[50]

Biology is also capable of amplifying these molecular motions to the macroscopic scale as can be demonstrated by your muscles. In muscle cells, sarcomeres, Actin filaments are interlaced with myosin subunits as represented on figure 1-14. When calcium concentration increases in the cell, the myosin associated with ADP will bind to the actin filament and bend, bringing actin filaments subunits (i.e., the vertical blue bars in the scheme) closer and releasing the ADP. ATP can then bind to the bent myosin causing it to unbind from the actin filament. The myosin will then catalyze the hydrolysis of ATP to recover its initial configuration and the myosin-ADP

complex will once again bind to the actin filament. This cycle, similar to a rowing motion, transforms the chemical energy from ATP-hydrolysis in molecular mechanical work. Since these actin/myosin units are in series, the synchronous contractions of all the units allow the tissue to contract macroscopically. [5056,57]

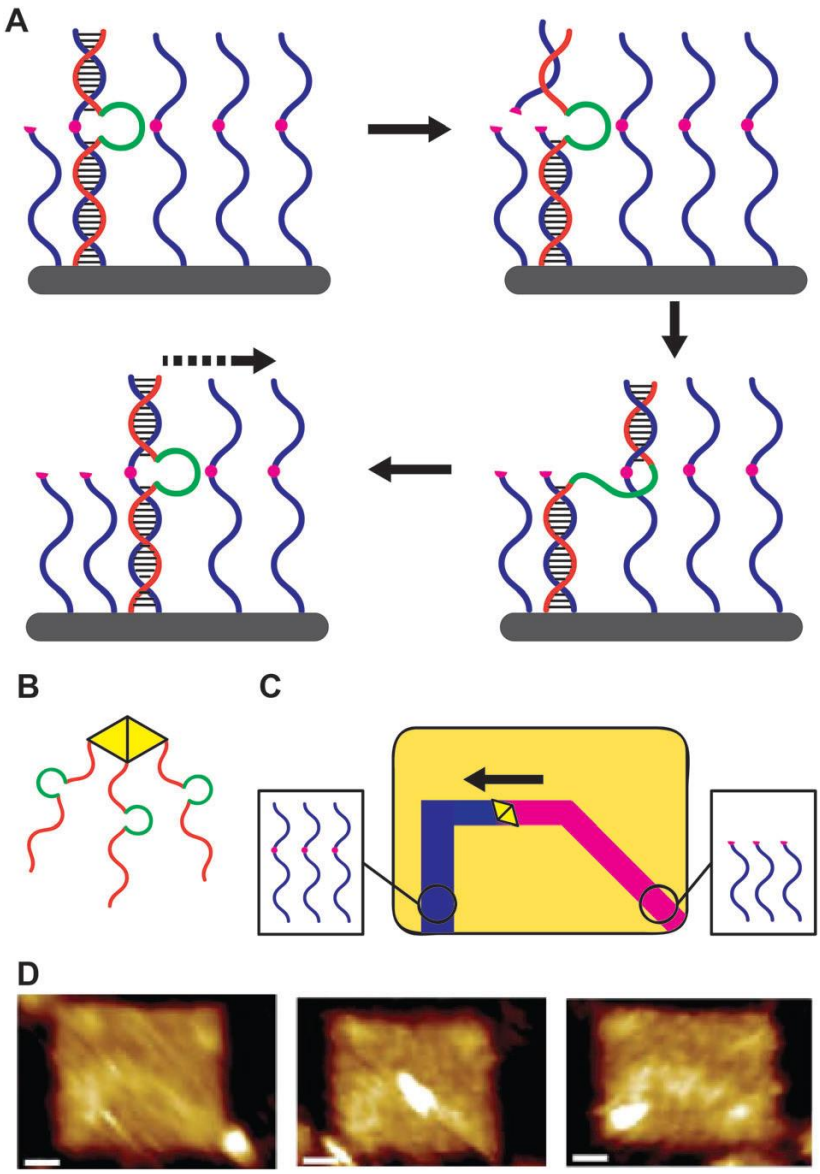


Figure 1-15 DNAzyme molecular walker. (A) Walker mechanism, where the DNA–aptamer sequence (green) that forms, in presence of magnesium, a DNAzyme capable of hydrolyzing the ribose base (pink dot) of the DNAtrack. Subsequent release of the cleaved sequence leads to processive rebinding to the next strand and movement.[58] *(B) Multipedal DNAzyme spider and (C) its progress on a self-assembled prescriptive track of ribosecontaining strands on a DNA tile. Blue tracks are the intact ones, pink are the consumed ones, and orange territory cannot be explored as it does not present track strands on its surface. (D) AFM images of the spider at the start, in the middle and at the end of the track (from left to right).*[59]

Similar types of motions have been achieved by synthetic molecules. For example, molecular walkers based on the hybridization of DNA and the presence of a metastable chemical fuel (here an RNA base) as illustrated on figure 1-15- A have been developed. The molecular walker (in red) consists of a single strand DNA which possesses a catalytic loop (in green). When the walker attaches through hybridization to a strand affixed to the support (in blue), the catalytic loop in presence of magnesium hydrolyzes the RNA base present in the support affixed strands. The walker then hybridizes with a supported strand next to it while the now free portion of blue strand detaches. The rest of the walker detaches from the remaining of the first blue strand, positioning the catalytic loop in front of the RNA base, allowing the process to start again. Since the tracks are one use only, as they act as the fuel and get consumed in the process, the movement is unidirectional as the walker only moves towards intact strands.[58] Such DNA walkers have been grouped together forming three legged "spiders", exhibiting a cooperative behavior as the spider moves along the track as evidenced in figures 1-15 B to D.[59]

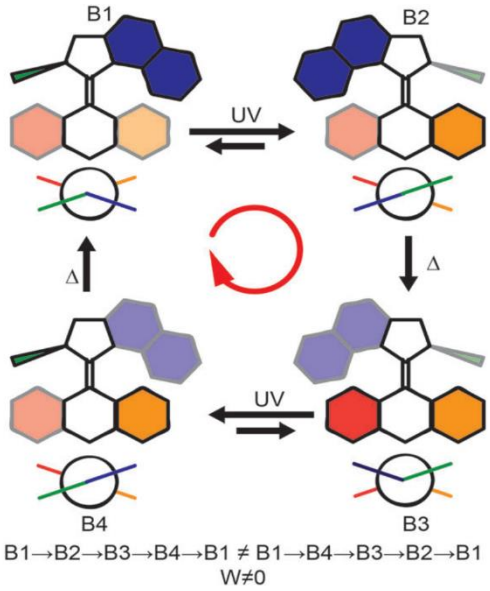


Figure 1-16 Schematic representation of a non-reciprocal isomerization cycle capable of generating work: a molecular motor.[60]

Synthetic molecules capable of transforming energy in rotational motion has also been developed. An example of such motors is given in figure 1-16. This motor uses energy in the form of UV light to perform a rotation which can be decomposed in four steps. The first step is a photoisomerization where the absorption of UV leads to a change of conformation of the double bond, resulting in almost half a turn of the motor. The structure then relaxes by performing a thermo-helix inversion restoring a conformation similar to the starting one but with the top part inverted. Then, the motor can undergo a second photo-isomerization which followed by another thermo-helix inversion leads to the complete rotation of the motor. The key element is the chirality of the steric hindering groups which enable a unidirectional rotation of the motor.[60]

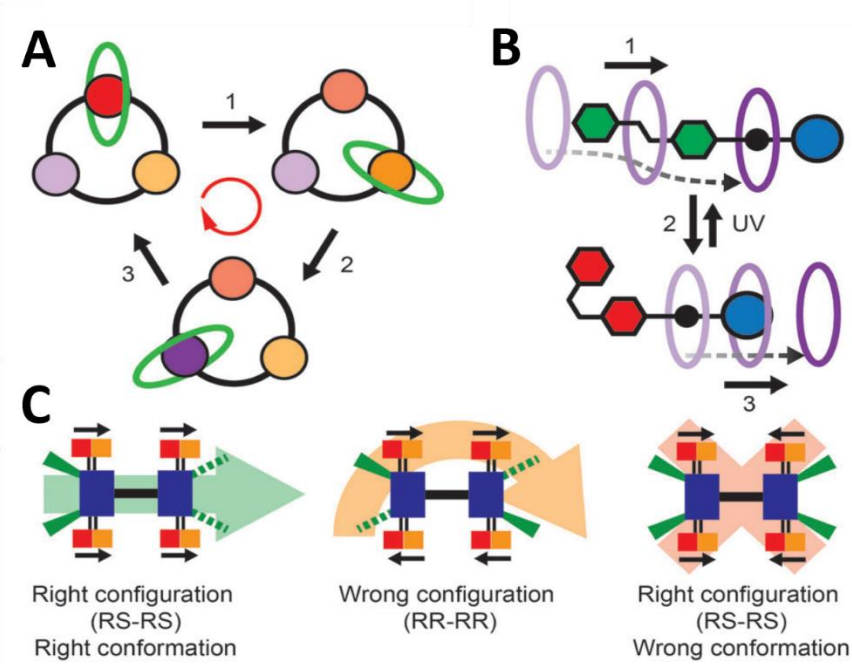


Figure 1-17 (A) A catenane rotating unidirectional along a macrocycle.[61] *(B) Light powered uniaxial translation of a linear molecule through a macrocycle.*[62] *(C) Molecular cars with various configuration and conformation showing either linear motion (left), random motion (center) or no motion (right) upon electrical excitation.*[63]

There are other examples of work generation such as catenanes rotating unidirectionnaly along a macro cycle (figure 1-17-A) this time the energy comes from a chemical fuel which modifies the binding site allowing the catenane to move to the next station. However, even though work is generated, it required a lot of external input in the form of alternating chemical fuel resulting in slow rotations[61]. Another light-based system has been developed (1-17-B) with a linear molecule threaded through macro-cycles which is able to move through them unidirectionally with the conformation changes of a double bond through UV irradiation once again.[62]

Molecular motors such as the ones detailed in 1-16 have also been incorporated in larger structures such as molecular cars (1-17-C), where four motors were linked together. Upon UV irradiation, the "cars" which possessed the correct configuration and conformation when deposited on a copper plate were able to move forward with 0.6 nm steps, as the four motors worked together in a coordinated fashion.[63]

Such motors were also incorporated in polymer gels, acting as reticulating points for the gel as the motor was functionalized to have 4 anchor points as can be seen in figure 1-18. This meant that upon UV irradiation the motor would entangle the polymer network of the gel, leading to its contraction[64], therefore in a similar way to the myosin subunits in sarcomeres, amplifying the molecular motion to a macroscale effect. These gels were also functionalized with a modulator unit able to perform a reversible ring opening during visible light irradiation which allowed the polymer network to unentangle and recover the volume that was lost by the work of the motor[65].

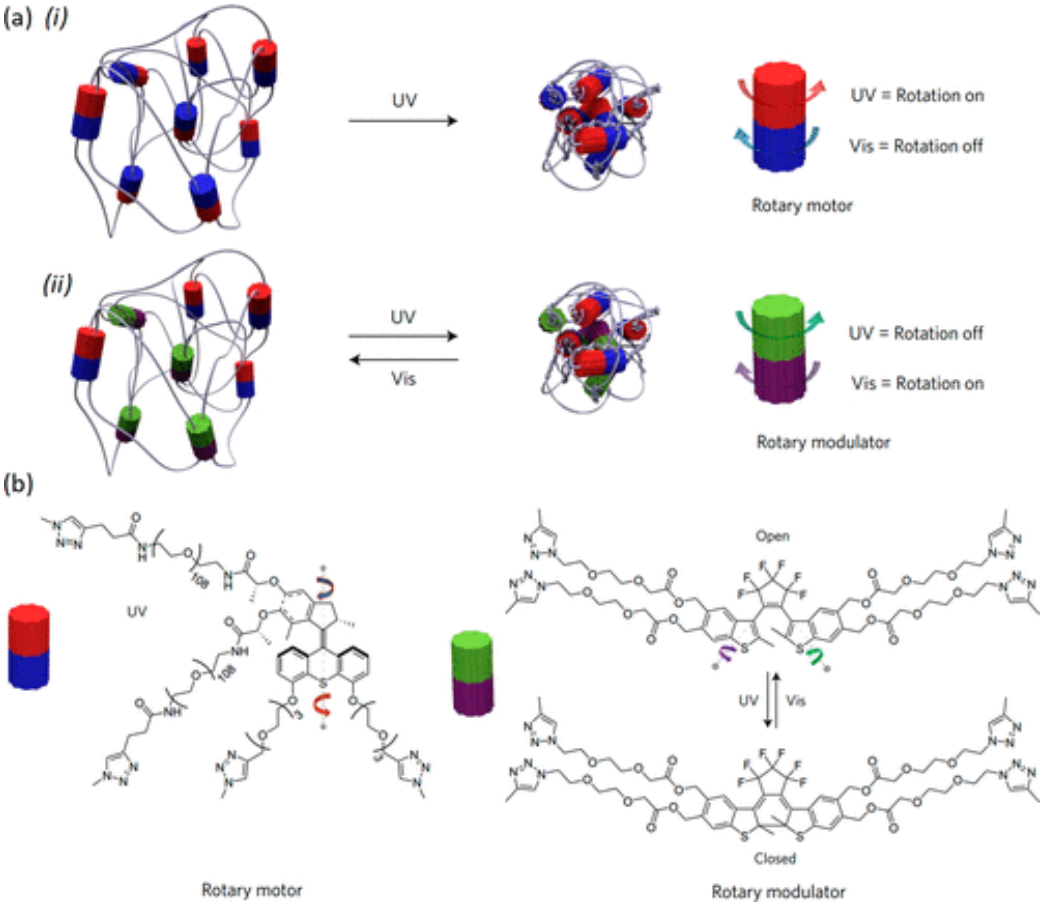


Figure 1-18 (a) Functioning principle and chemical design of an integrated motor/modulator system. (i) Schematic representation of a reticulated polymer–motor gel that, under UV light irradiation, produces a continuous unidirectional out-of-equilibrium rotation of the motor part (red and blue cylinders) with subsequent twisting of the polymer chains, which leads to a macroscopic contraction of the entire network. A twist is defined as a new entanglement within a pair of polymer chains. (ii) Schematic representation of a polymer–motor–modulator gel that proceeds as in (i) when exposed to ultraviolet light but that untwists the polymer chains when exposed to visible light by activating the modulator part (green and purple cylinders). Visible-light activation of the modulator allows its rotation in the opposite direction of the elastic torque initially produced by the motor and thus leads to a full re-expansion of the network at thermodynamic equilibrium. (b) Chemical structures that compose the polymer–motor–modulator system shown in (a(ii)). Figure taken from [65]

When it comes to creating work at the molecular scale, it seems that light is a great energy source as it doesn't generate waste in the system. However, it requires external light input as opposed to chemical fuel which could be present in the system from the start. Chemical systems however face a problem when it comes to scaling up the size of the system, as when the dimensions increase above the diffusion length of the either the fuel, or a potential signaling molecule if trying to incorporate temporal control on the system which would be the next step towards creating intelligent material. Another important property for such materials is information processing. A property that once again relies heavily on out of equilibrium processes in biology.

d) Information Processing

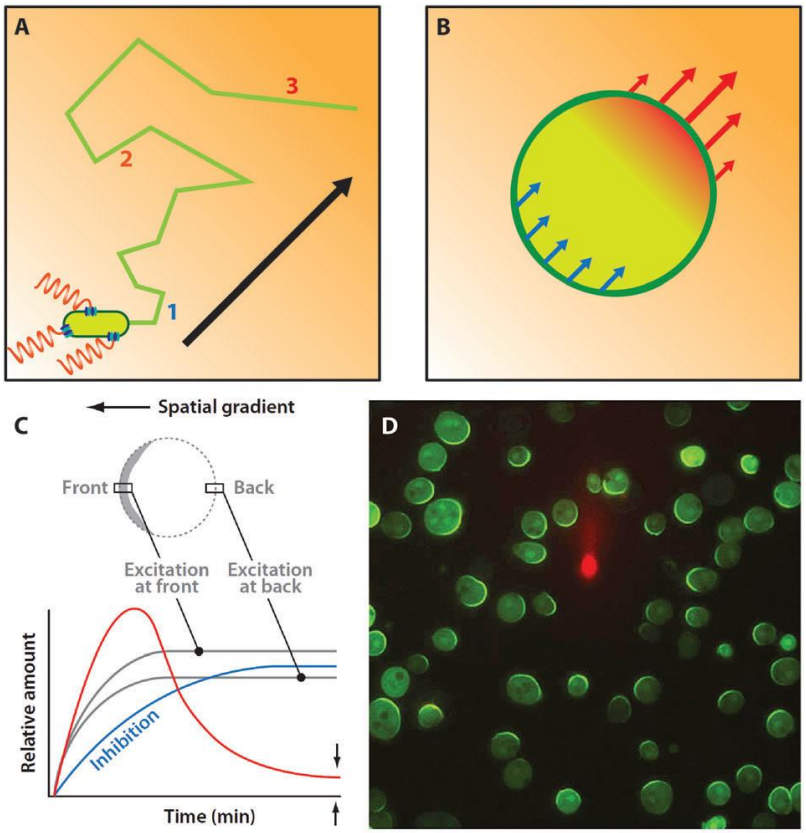


Figure 1-19 Examples of biological adaptation on different levels. (A) Chemotaxis of flagellar bacteria. As the chemo-attractant increases (orange background) the frequency of tumbling decreases and allows larger linear moves into the direction of the chemo-attractant. (B) Chemotaxis of eukaryotes. A logic network transfers the gradient of chemo-attractant into a thresholded signal gradient and regulates actin polymerization at the origin of motion. (C) Response of the Locally Exited Globally Inhibited (LEGI) network to the gradient of chemo-attractant. Grey lines: local excitation. Blue line: inhibition feedback to allow for threshold sensing between front and back. Red line: net signal response at the gradient front (excitation minus inhibition). (D) Response of immobilized Dictyostelium discoideum in a gradient of cyclic AMP provided by a micropipette labeled in red in the center. The signaling proteins of the cells are fused with a green fluorescent protein. The green crescent fluorescence oriented to the center are representative of the gradient sensing of the cells. A) & B) adapted from ref.[29] *C) from ref.*[66] *and D from ref.*[67]

The ability to sense the environment and act accordingly is key for the survival of life even in "simple" organisms. Being able to detect nutrient sources and move towards them while avoiding toxic or dangerous environment plays directly in Darwinian laws of evolution. Gradients are common out of equilibrium scenarios. With diffusion or conduction processes striving to homogenize the media, there is a constant either energy or matter flux through the system, resulting in an out of equilibrium state. It is those gradients of attracting conditions such as high nutrient concentrations called chemoattractants, or of toxic environments, that microorganisms have adapted to sense in a process called chemotaxis. *E. coli* for example is able modify its swimming behavior (explained in c)) by sensing different chemical gradients such as pH, nutrient, and oxygen levels. The sensing is done through several trans-membrane histidine-aspartate phosphorelays, which through phosphorylation exert a negative feedback loop on the ATP-synthases correlated to the flagella.[68,69] This enables the bacteria to switch from linear motion to tumbling. When the chemo-attractants concentrations are low, the bacteria's movement is chaotic with a lot of tumbling, however as the concentration increases the movement becomes more and more linear allowing the bacteria to move faster as the

concentrations increase. *E. coli* is 2 µm in diameter but larger organisms can sense the gradient. An example for it is the eukaryote *Dictyostelium* which is around 10 µm in diameter. Chemo-attractants, such as cyclic adenosine monophosphate, are sensed through a locally excitable and globally inhibited (LEGI) network. This process is composed of two different feedback loops. A fast positive feedback loop with a short lifetime activator, and a slower negative feedback loop with a long lifetime inhibitor. The long-lasting inhibitor can diffuse across the entire cell cancelling out the signal of background concentration, only at the front of the gradient is the positive feedback loop strong enough to overcome the inhibition. This allows the cell to promote actin polymerization and pseudopodia formation (i.e. initiate movement) in the direction of the gradient.[66,67,70,71]

Secretion of signaling molecules in select spots and the resulting gradient also allows cells of a micro-organism to communicate. When combining the gradients created with feedback loops discussed in part a), many different patterns can arise. This can be related to the reaction-diffusion model hypothesized by Turing[72], with the different feedback loop caused by the signal molecules playing the role of the reactions. The resulting Turing patterns have been observed on many animals (figure 1-20).[73] Gradients and feedback loops play an important role during embryogenesis when the cells start specializing in what will become the different organs of the embryo in a process called morphogenesis.[74]

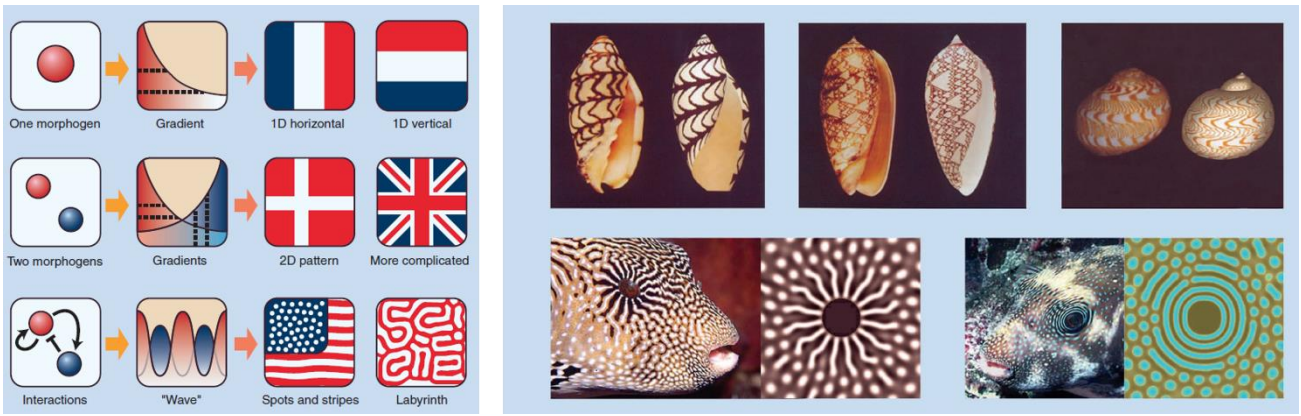


Figure 1-20 Schematic drawing showing the difference between the morphogen gradient model and Turing model. (A) A morphogen molecule produced at one end of an embryo forms a gradient by diffusion. Cells "know" their position from the concentration of the molecule. The gradient is totally dependent on the prepattern of the morphogen source (boundary condition). (B) Adding a second morphogen produces a relatively complex pattern; but with no interactions between the morphogens, the system is not self-regulating. (C) With addition of the interactions between the morphogens, the system becomes self-regulating and can form a variety of patterns independent of the prepattern. [Artwork by S. Miyazawa]

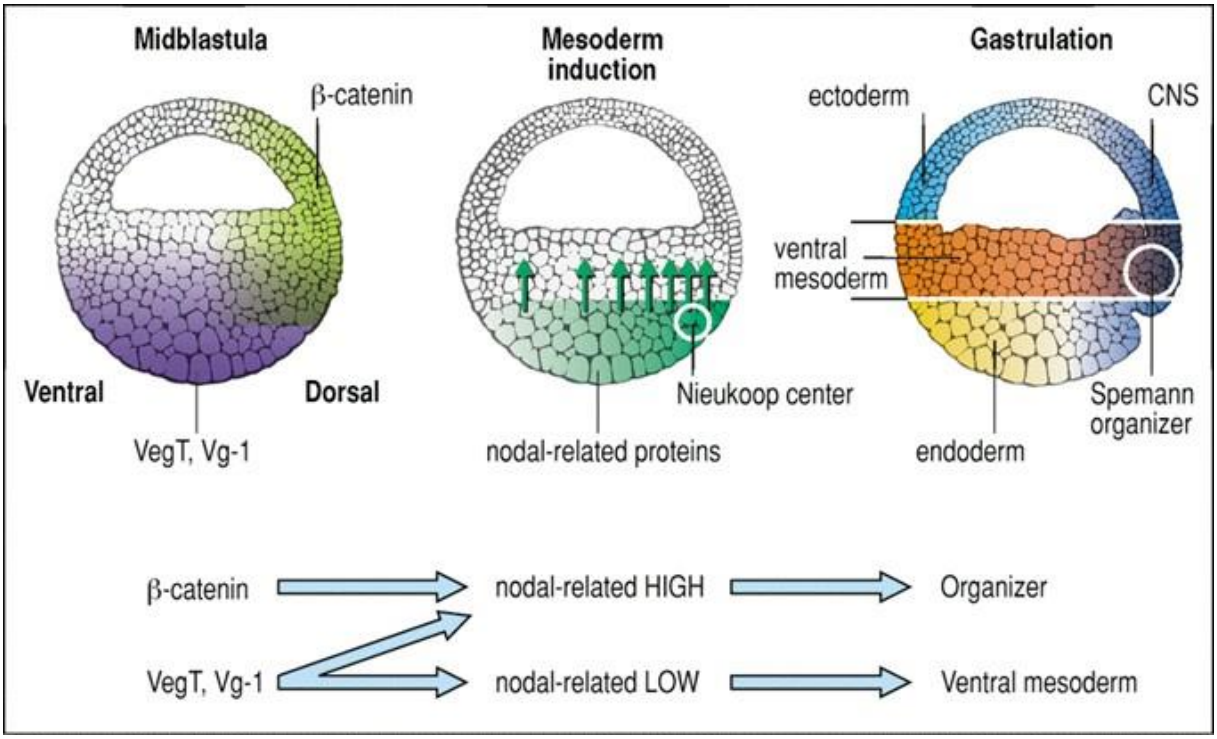


Figure 1-21 Model for the induction of the mesoderm in Xenopus[75]

A good example is the induction of the mesoderm (layer of cells which will differentiate to become most of the inner organs except for the central nervous system) for frogs as illustrated in figure 1-21. The embryo at the midblastula stage has a concentration pattern is certain transcription factors inherited from the initial post fecundation single cell state. The lower (posterior) part of the embryo presents higher concentrations of VegT and Vg-1 transcription factors, this will lead to the secretion of mesoderm inducing molecules (Xrn). However, the dorsal part is richer in β-catenin (inherited pattern), the dorsal cells exposed to both high concentrations of Xrn and β-catenin differentiate in the Spemann organizer instead of ventral mesoderm. The Spemann organizer will in turn through the secretion of transcription factors and their resulting gradient induce the differentiation of different parts of the embryo[74]. This shows the ability of cells to use logic networks, as the information Xrn AND β-catenin leads to a different output than just either of them.

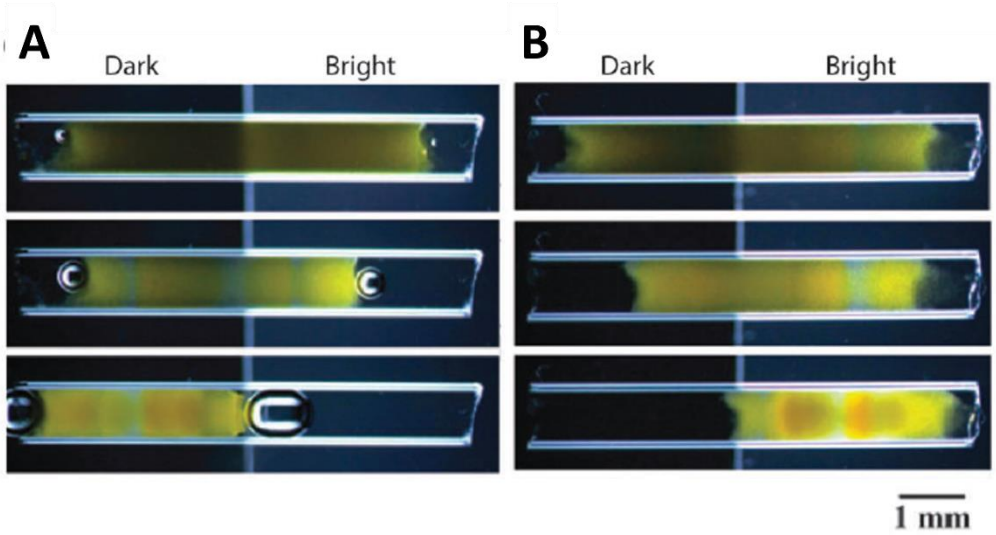


Figure 1-22 (A and B) Phototactic motion of a self-oscillating BZ gel under anisotropic illumination in a glass capillary top at t=0min down at t4150 min. (A) The gel is photophopic under low illumination (dark area 33 mW cm^{-2}, bright area 106 mW.cm^{-2}) and (B) phototrophic under high illumination (dark area 205 mW cm^{-2}, bright area 904 mW cm^{-2}).[76]

Chemotaxis is not uncommon; catalytic nano and micromotors show apparent diffusion coefficients correlated with their chemical fuel gradients.[77] However, chemotaxis was also observed in the case of enzyme with their corresponding substrate in micro-fluidic devices.[78] Non-fuel gradient related chemotaxis has been possible when working with catalytic motors fueled by hydrogen peroxide oxidation and the resulting oxygen formation[79–81]. Since this reaction is pH dependent, pH gradients permit directionality. This is also observed with motions based on peptide rearrangements where the pH once again influences the reorganization. Light, and the resulting phototrophic and photophobic behaviors, can also be used to influence systems even if it's not the origin of its non-equilibrium. As an example, autonomously self-oscillating hydrogels based on the BZ reaction with ruthenium instead of cerium, displayed different behaviors based on illumination intensity. Under low illumination the gel became photophobic whereas at high intensity it became phototrophic as shown on figure 1-22.[76]

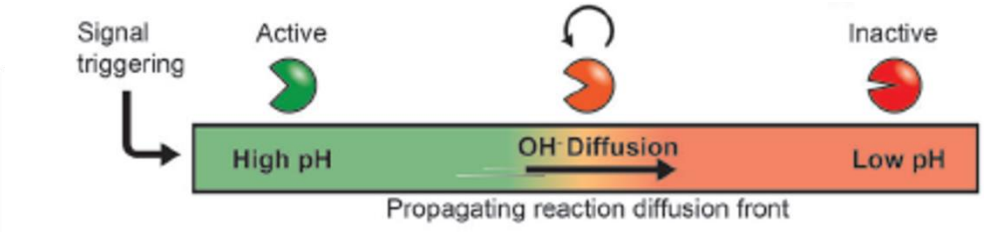


Figure 1-23 Enzymatic propagation of a pH reaction-diffusion front with positive feedback. Adapted from ref.[29]

Reaction diffusion systems, where diffusion is slower than the reactive process, allow the communication between different parts of potential smart materials, as the signal molecule has time to diffuse from one part to another. These types of long-range communication can be created with positive feedback loops such as the urea/urease discussed previously.[37] In a gel at low pH where the enzyme is inactive, the addition of a base would activate the enzyme locally.

The ammonia produced locally can then diffuse and an activation wave courses through the material.[82] To have a more spatially restrained response, it is necessary to have a negative feedback loop or an inhibitor.

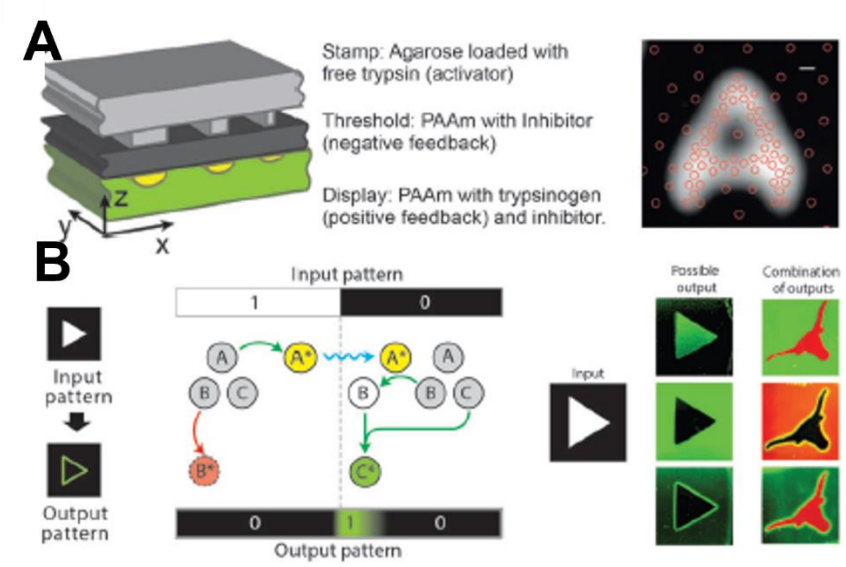


Figure 1-24 A) Spatial sensing device for trypsin. The stamp (light grey top) contains trypsin, the network activator. In the threshold layer (dark grey middle) the inhibitor erases low trypsin concentrations. The display layer (green down) contains the full autocatalytic reaction network and generates a localized fluorescent signal in the presence of trypsin. (right) Actual example of output (bright area) and stamp contact point (red circle), fluorescence appears only when the density of contact points overcome the threshold level.[83] *(B) (left) Network of strand displacement reactions leading to edge recognition. UV simultaneously activates A and deactivates B; the fluorescent probe C is only activated near the edge where A* and B can react by diffusion. (right) Examples of outputs with one and two display reaction networks.*[84]

In figure 1-24 A, the trypsin/inhibitor system discussed previously is used to recover a spatial input. An agarose stamp loaded with trypsin is applied on a gel layer loaded with the trypsin inhibitor underneath with lies a gel loaded with trypsinogen and lower concentrations of inhibitor. This setup allows for a response of the positive feedback loop of trypsin only in the places where there is enough diffusion to overcome the inhibition, i.e., directly underneath the stamped regions and nowhere else.[84,85] It is possible to have more complex systems yielding more information as illustrated in B. This system is based on DNA light interaction with 3 DNA strands. Input light irradiation, strand A gets activated while strand B gets deactivated. The activated A strand can react with B and C to generate the activated C strand, however since light deactivates B, this can only happen at the edges of illuminated areas with the diffusion of activated A strands. This allows not only to sense the illuminated areas but also gives the information on the edges of the illumination pattern.[83]

Diffusion is the mechanism behind the communication within the system, its tuning is key for the rise of smart materials. There are different ways this can be achieved. Complexation agents can be added to the system to slow down the diffusion speed such as in the system showed in figure 1-25 A. Here, a low mobility complexation agent (poly (sodium acrylate)) reduces the diffusion speed of protons that act as the positive feedback signal in a thiourea-iodate-acrylate system and doesn't interact with the thiourea which acts as the inhibitor, therefore finetuning the relative diffusion speed between them. This led to the first rational implementation of Turing patterns in unstirred systems.[86]

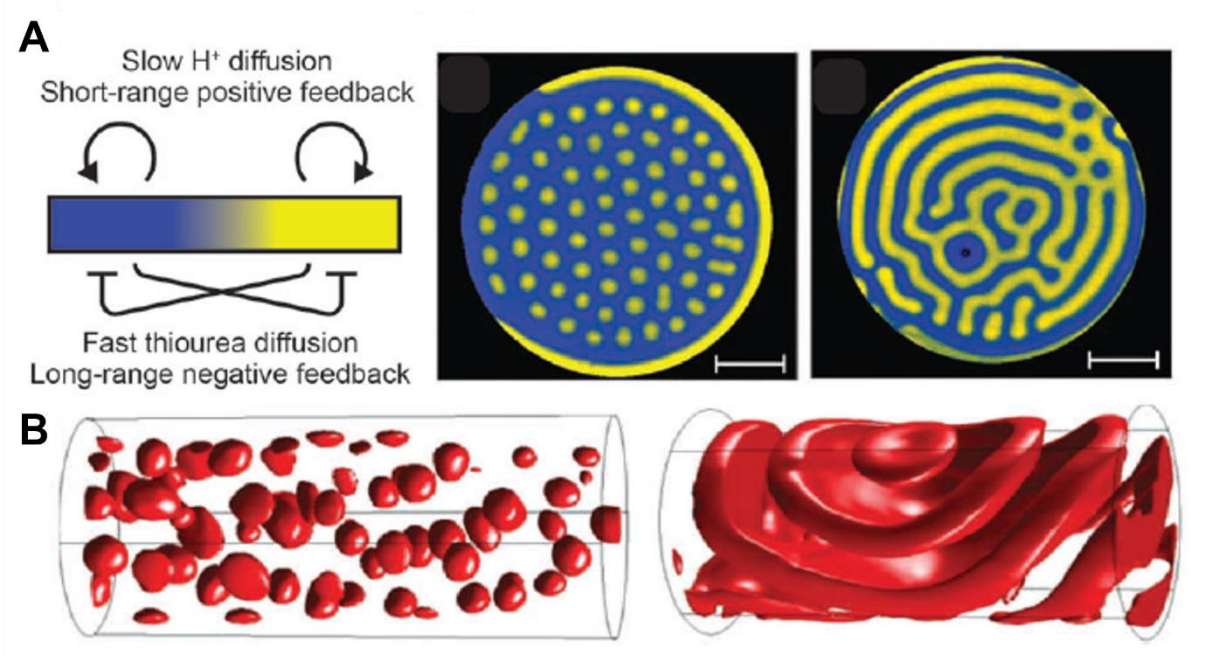


Figure 1-25 Schematic representation of the short-range positive feedback and long-range negative feedback leading to the appearance of 2D Turing patterns in one-side-fed unstirred reactor with thiourea–iodate–sulfite reaction (scale bar 4 mm).[87] *(F) Topographically reconstructed concentration fields for 3D Turing patterns obtained in capillary with a diameter of 0.6mm.*[86]

Another possibility is to modify the media in which the process takes place. When the BZ reaction was run in a water-in-oil microemulsion, the different species separated with the charged ones remaining in the aqueous microphase whereas neutral could diffuse rapidly through the oil, therefore modifying the diffusion speeds according to the water-to-oil volume ratio. This allowed to create a photosensitive device which could memorize illumination patterns for several hours as the 3D chemical patterns observed could be linked back to different illumination profiles (figure 1-25 B).[86]

Understanding and tuning of out-of-equilibrium processes is the key to unlocking the behaviors found in biology in our future materials. There is still much to investigate when it comes to linking all those behaviors together. This PhD work was based on an ANR project (SACERDOTAL) to perform out-of-equilibrium experiments with reaction-diffusion at its core to keep the system out-of-equilibrium.

4. The project at the origin of this PhD work

This PhD is part of the ANR project oriented towards the Survival of reactive systems for driving property-directed molecular evolution (SACERDOTAL). The goal of this project is to create Darwinian Chemistry out-of-equilibrium in an abiotic system. The system revolves around a reversible polymerization/depolymerization reaction, allowing for the creation of different species from monomers.

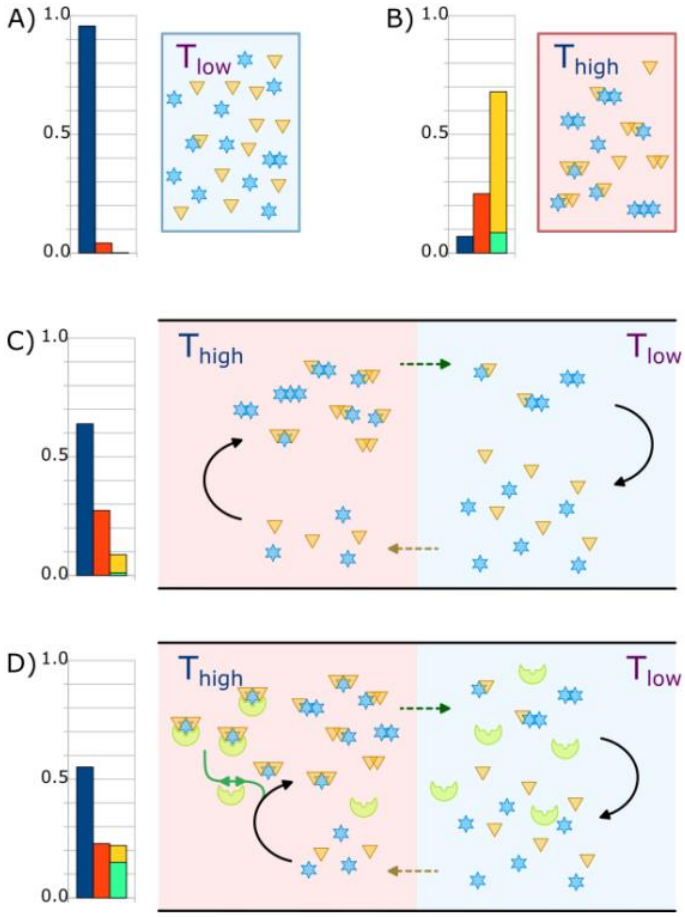


Figure 1-26 Scheme representing the species present in a reversible polymerizable system whose ligation reaction is influenced by temperature. A) Shows the composition of the system at cold temperatures, which is composed primarily of the different monomers introduced (star and triangles). B) Shows the composition of the system at high temperature, where the monomers have almost disappeared, and oligomers have formed. C) Integration of the polymerizable system in a temperature gradient, allowing the monomers from the cooler regions to diffuse to the hotter ones and the oligomers to experience cooler temperatures before their degradation. D) Introduction of a target molecule in green which amplifies the specie with which it interacts the most.

The monomer system at different temperatures should yield different species distributions as shown on figure 1-26. If the ligation reaction is endothermic (the enthalpy of the system increases over the process $\Delta H > 0$) then at low temperatures (A), the energy would not allow the different monomers (represented by the triangles and stars) to ligate and form long polymeric species, resulting in a specie distribution heavily in favor of monomers. Whereas at higher temperatures (B), the energy is high enough for elongation to take place, modifying the distribution in favor of longer species. By applying a temperature gradient to such system, reaction-diffusion processes would arise, bringing the system out of equilibrium. In the hot

region, the accumulation of high-length species would result in their diffusion towards cooler regions whereas in cooler regions the monomers would diffuse towards the hotter regions where they are consumed. Furthermore, the reversibility of the reaction would ensure that the monomers are regenerated in the colder regions. If the gradient size and diffusion coefficients are finely tuned relatively to the kinetics of the ligation reaction, high temperature species can diffuse to lower temperatures where the energy should not be able to sustain them, bringing the system out-of-equilibrium (C). Different properties can arise from such systems as discussed previously. Different species may interact together, either stabilizing or destabilizing each other (essentially acting as feedback loops discussed previously), thus modifying the distribution of species compared to equilibrium at a given temperature.

This experiment has many similarities with dynamic combinatorial libraries experiments (DCLs). In these experiments a system generates a myriad of compounds, one is selected, and amplified based on a specific property. This process could also be achieved with the SACERDOTAL project (D). A target could be introduced in the system which would stabilize a specific specie therefore increasing its concentration. However, if the target has a low diffusion constant or is immobilized in the system, then the interaction with the reaction-diffusion cycles can furthermore locally enrich the system in the selected specie.

In a last step, the heating of the system, or monomer concentration could be periodically modified, therefore selecting the species in a "survival assay" based on their resistance to hydrolysis for example.

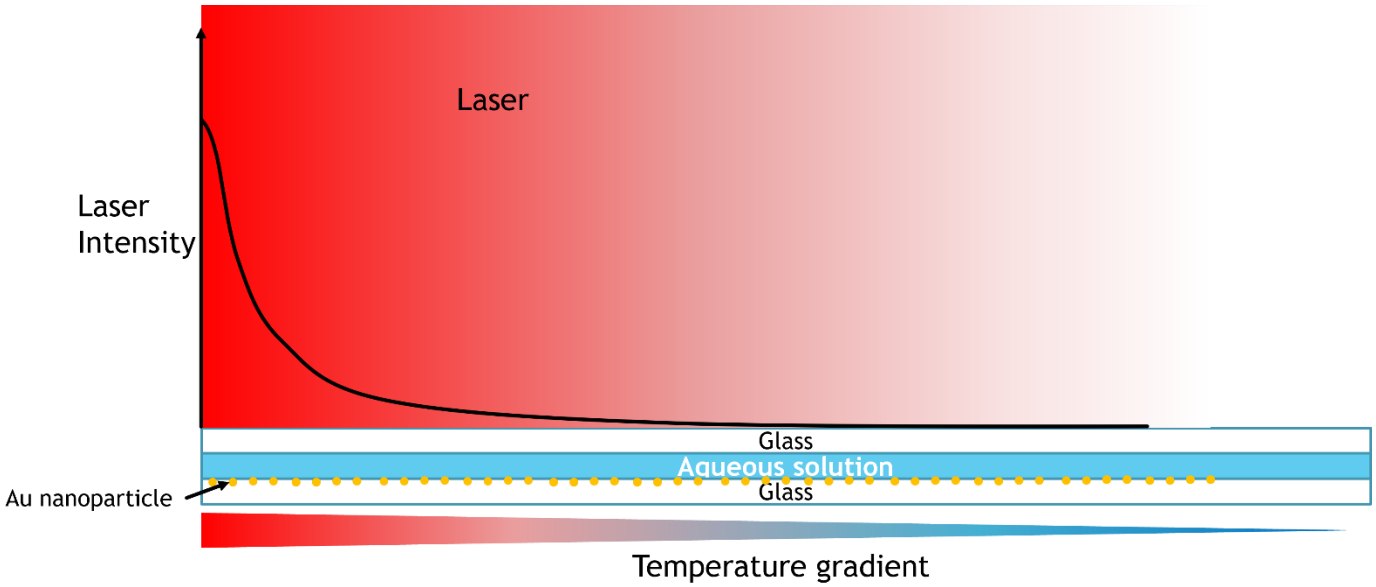


Figure 1-27 Scheme representing the experimental setup for a temperature gradient experiment. The reactive system is in the aqueous solution between two glass sheets and is heated depending on its location through differential laser intensity, transformed into heat by gold nano particle plasmonic effects.

To have reaction-diffusion systems based on thermic gradients, the system must not allow convection processes. Convection arises in fluids as more efficient ways to dissipate the energy of the system, however as opposed to diffusion where the species move slowly through

Brownian motion, convection leads to bulk movement of the solution. In these convection cells, hot fluid reaches the colder regions to cool off and colder fluids flows towards hotter regions. In order to maintain the temperature gradient across the fluid, heat must only be dissipated through conduction. The ability of a fluid to develop convection processes is denoted by the Rayleigh number in fluid mechanics.

$$Ra = \frac{g\beta}{\nu\alpha}(T_s - T_\infty)L_c^3$$

The Rayleigh number (Ra) is an empirical, dimensionless number, which depends on the acceleration due to gravity g, the thermal expansion coefficient of fluid β, the viscosity of the fluid ν, the difference in temperature between the surface T_s and further in the fluid T_∞, and of the characteristic length of the system L_c. Empirically, convective processes arise when the Rayleigh number is above 1700 for a fluid. To avoid such processes taking place, our system would be comprised of a thin layer of the monomer solution sandwiched between two glass plates. This ensures small characteristic sizes and low enough Rayleigh numbers based on the extrema of the temperature gradient.

The creation of the gradient in a precise manner with (resolutions at the micrometer scale), would be done by a method developed by our partner Guillaume Baffou, working at the Fresnel institute in Marseilles. When illuminating gold nanoparticles with a laser, through plasmonic effects, they dissipate the energy in thermal processes, heating their surrounding media based on the intensity of light they receive (figure 1-28). The tuning of the illumination profile can thus give rise to different shapes lengths of temperature gradients.[88] Furthermore, this technique has been used to reach temperatures well above 100°C in aqueous system without ebullition taking place[89] (therefore mixing the system, disrupting diffusive processes) by exploiting the lack of rugosity at the surface of glass lamellae and thus the high energy required by gaseous water to nucleate bubbles at their surface. This enables bigger temperature differences across the gradient to facilitate the appearance of out-of-equilibrium processes.

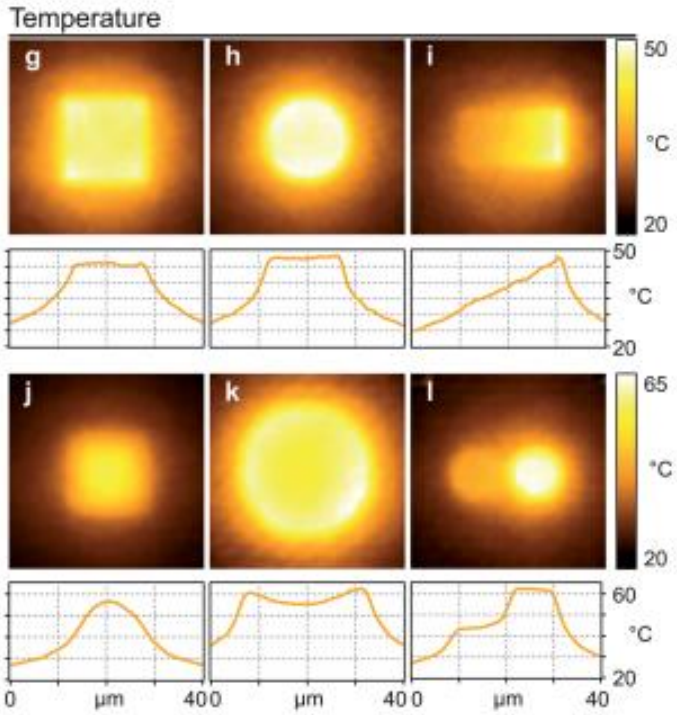


Figure 1-28 Thermal measurements on various gold NP patterns. (a–f) TIQSI measurement of the heat source density delivered by gold nanoparticle assemblies. (g–l) TIQSI measurement of the associated temperature distributions along with horizontal crosscuts. Taken from ref.[88]

5. The Amino acid toolkit

When it came to choosing what kind of polymer system was to be used for this project, several aspects were taken into consideration. We wanted a system that could be soluble in aqueous solution for easier experimental conditions as well as a system with high likelihood of "life-like" emerging behavior. Proteins as shown in the previous discussions are at the heart of out-of-equilibrium processes in biology. The biological motors are protein complexes, enzymes are for the most part proteins, and proteins interact in very interesting ways with temperature.

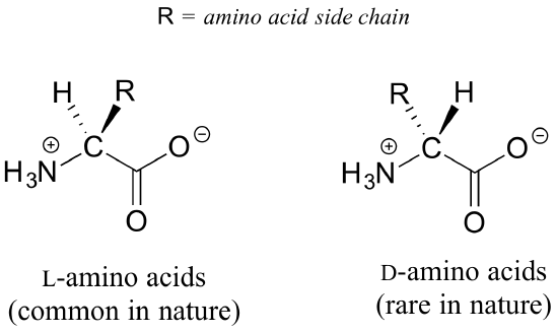


Figure 1-29 Representation of L and D amino acids.

Proteins are amino acid polymers. An amino acid is composed of a backbone common to all amino acids, an amine linked to a carbon called the alpha carbon, which is linked to a hydrogen, a side chain, and a carboxylic acid function. The chirality of the alpha carbon defines two classes of amino acids, L amino acids, which are almost exclusively used by biology, and D amino acids which are widely found but rarely used in biology[90]. Amino acids are polymerized by the condensation between the amine and the carboxylic acid of another amino acid, leading to amide (peptide) bonds. There are 22 proteinogenic amino acids (amino acids naturally found in proteins), which all differ by their sidechain.

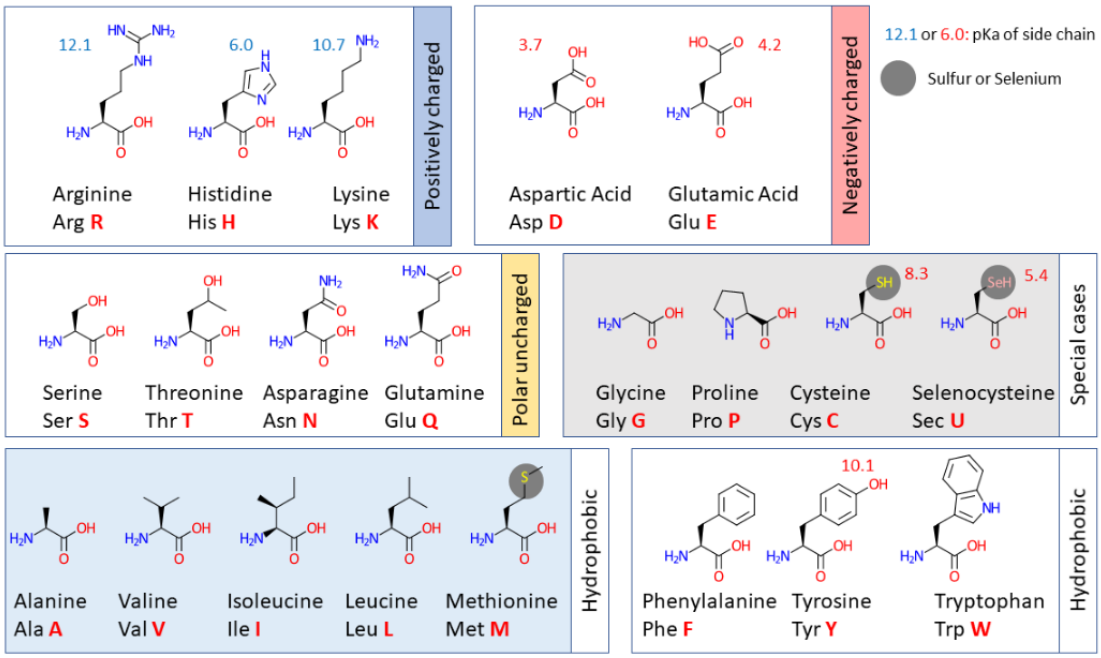


Figure 1-30 The 21 most common proteinogenic amino acids. (Pyrolysine is the 22nd but is extremely rare with occurrences only in some archaea strands).

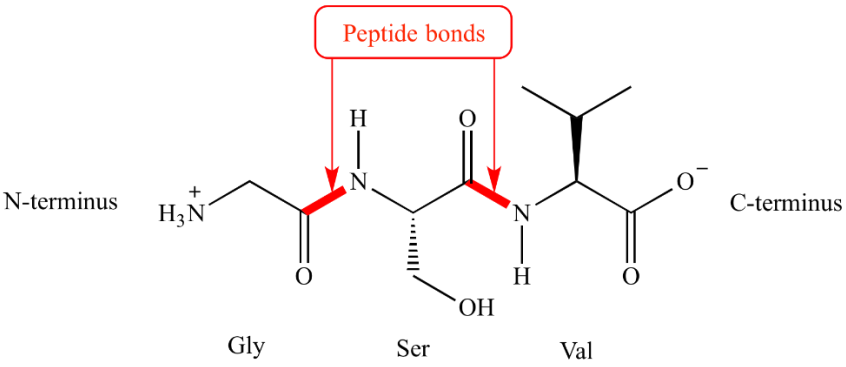


Figure 1-31 Scheme representing the peptide bond in a protein sequence

The sequence of amino acids, also called the primary structure of proteins, can adopt many different structures. This is referred to as protein folding. The backbone of the polymer can interact with itself via hydrogen bonding between the hydrogen of the nitrogen in the amide bond and the oxygen of a different amide bond in the polymer to create structures.

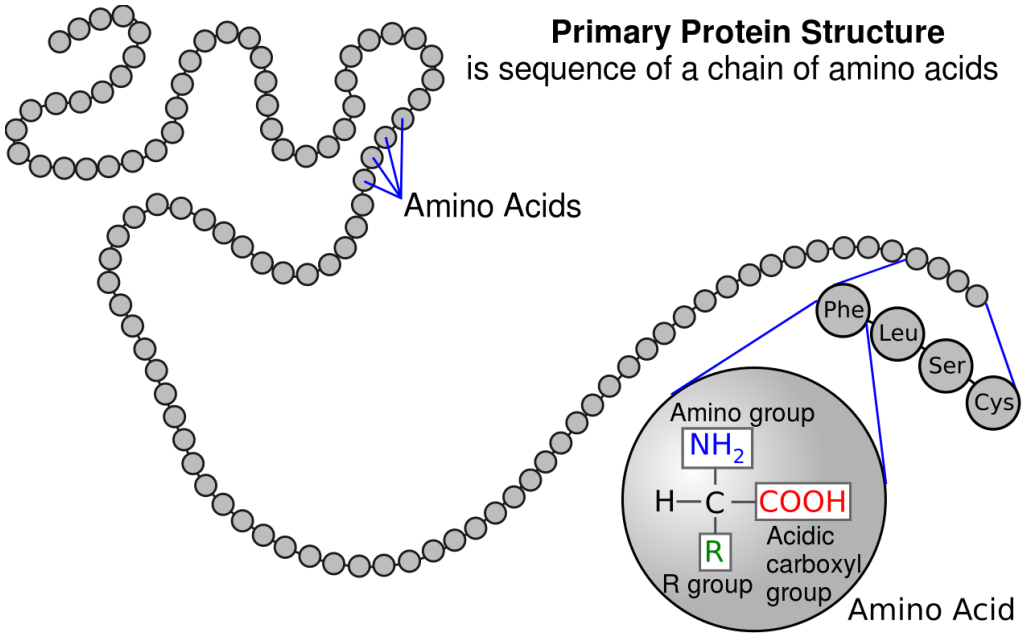


Figure 1-32 Representation of the primary structure of a protein. Taken from ref. [91]

In the α-helix, the hydrogens of the peptide bonds create hydrogen bonds with the oxygen of the third amino acid down the sequence[92], the displacement between the amine interacting with carboxylic group results in a helical structure with 3.6 amino acids per turn. Hydrogen bonds can also appear between the backbone of different parts of the sequence. β-sheets are an example of this. Two segments of a protein form hydrogen bonds and align each other, depending on the alignment the β-sheet can be either Parallel or antiparallel.[51] These different folding domains then interact with each other resulting in a three-dimensional structure called the tertiary structure.[51]

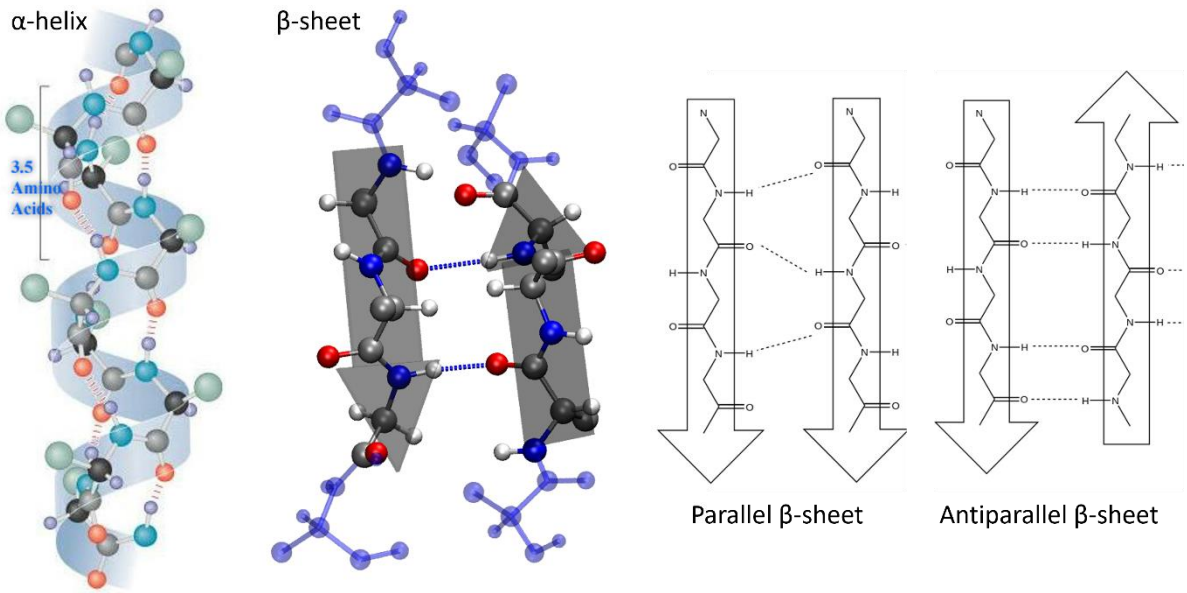


Figure 1-33 (Left) Representation of an α-helix motif and the hydrogen bonds between the backbone of its constituting amino acids. (center and right) Representation of a β-sheet structure (parallel and antiparallel) and of the hydrogen bonds between the sheets. Adapted from ref.[51,93]

It is this tertiary structure that gives a protein its properties. Proteins that destined to the membrane possess structures with hydrophobic amino acids that will, through hydrophobic interactions, embed the protein in the membrane whereas free globular proteins will fold with hydrophilic amino acids on the exterior to keep them soluble. As an example, enzymes are proteins that catalyze certain reactions. They do so by binding to the reactants (called substrate for the enzyme) decreasing the activation energy required for the transition states leading to the products. Enzymes are mostly substrate specific, meaning that they only catalyze a specific reaction for a specific compound. To achieve this, the amino-acid sequence is folded so that the enzyme can specifically recognize and bind the substrate and catalyze the reaction. The modification of sometimes only one amino acid in the sequence of a protein may lead to the complete inactivation of an enzyme or drastically modify the structural properties of the protein as is the case with many genetic diseases.

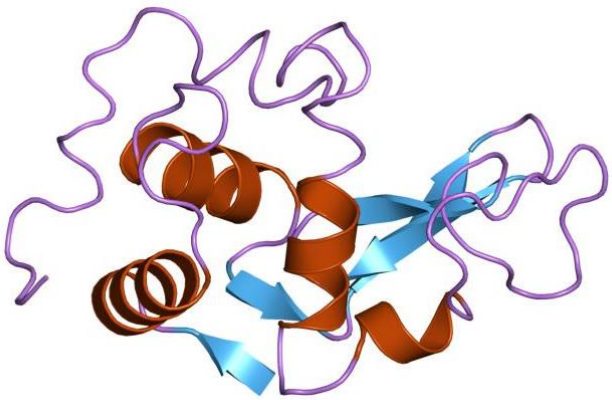


Figure 1-34 The tertiary structure of a protein, with folded domains interacting with one another.

This relation between sequence, structure and thus function, makes generating random amino acid sequences very promising to discover properties. Furthermore, a single protein can adopt different stable structures with sometimes only one of them being active. In biology, specific protein complexes such as chaperones can help proteins fold in their active tertiary structure, but proteins can later misfold as well. Prions are misfolded proteins that have lost their intended activity but whose misfold also make them: resistant to degradation by the organism, able to misfold correctly folded proteins of the same type. They are responsible for different diseases such as mad cow disease and its human counterpart Creutzfeldt–Jakob disease.[94] Since the interactions behind protein folds are mostly weak interactions (hydrogen bonding, electrostatic interactions, hydrophobic interactions), they can be overcome by temperature. This means that in the hotter regions of the temperature gradient in the experimental set up, amino acid sequences will unfold and their diffusion towards cooler regions (where the available energy might not have allowed their formation in the first place) where they can interact between each other or possibly fold, allowing the system to explore the folding space on top of the sequence space for possible properties.

If amino acids are to be used, such system, might not produce very long sequences. Short sequences of amino acids are called peptides. Complex folds are therefore not to be expected, but alpha helix domains do not require long sequences to arise, and other helicoidal shapes are possible. β-sheets on the other hand are not likely to happen within a sequence but different sequences might interact with each other in a similar pattern.[95]

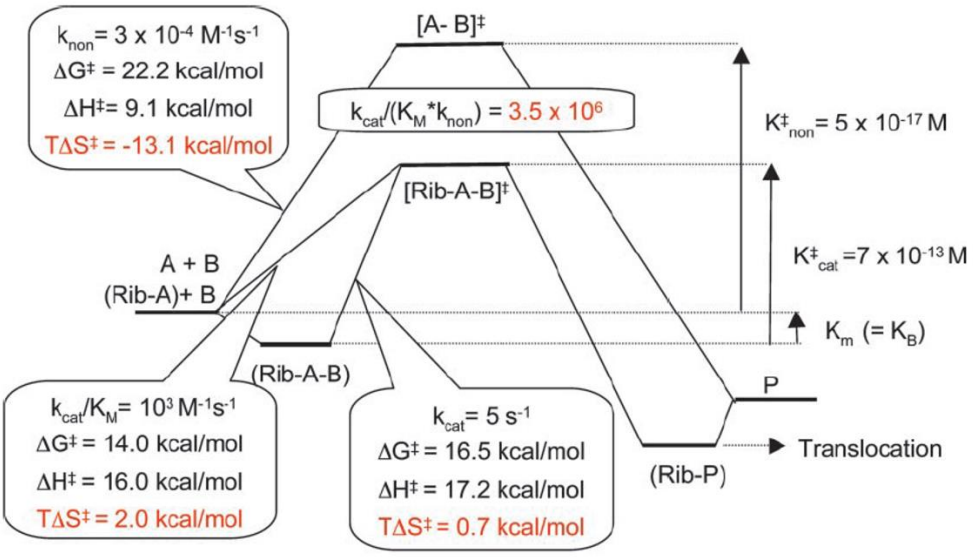


Figure 1-35 Activation parameters at 25°C for the second-order uncatalyzed (knon) and ribosome catalyzed peptide bond formation (kcat/KM), calculated from the concentration of total amine at pH 7.5. The broken line shows the first-order reaction in the ribosomal active site (kcat). K‡ values represent the equilibrium constant between the ground state and transition state calculated from the differences in free energy. Taken from ref.[96]

The major drawback of working with an amino acid polymer system is the ligation and hydrolysis reaction. The activation energy required to perform abiotic peptide ligation can be overcome with thermal energy, however the uncatalyzed process is quite slow with kinetic constants around $10^{-4}\ Mol^{-1}.s^{-1}$. In biology the ligation process, takes place in a ribosome,

a protein/RNA complex which will catalyze the peptide bond formation as well as couple it to the hydrolysis of guanosine triphosphate (GTP) to acquire the energy to overcome the activation barrier. This results in much faster kinetic constants around $10^3 Mol^{-1}.s^{-1}$ with each is seven orders of magnitude faster than the abiotic reaction.[96] Similarly, the hydrolysis of the peptide bond in water at neutral pH and 25°C is between 350 and 700 years[97], which makes it nonreversible at feasible time scales. Therefore, if amino acids are to be the components of our polymer system, a new ligation strategy must be investigated.

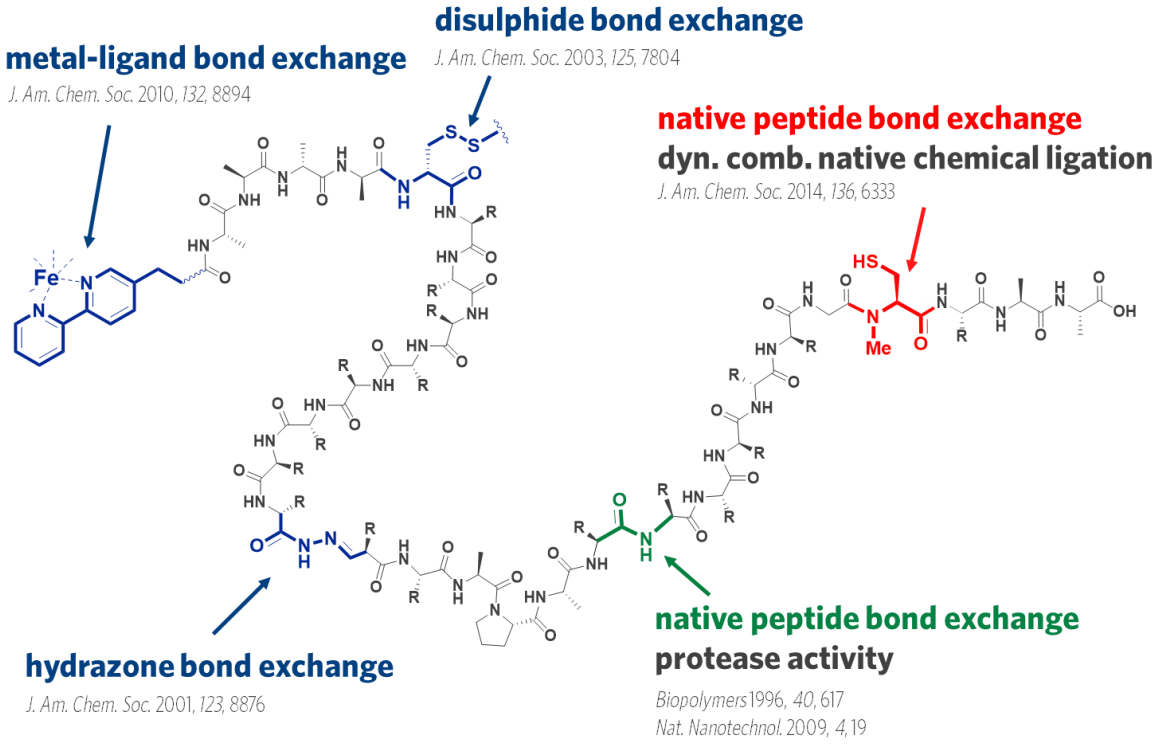


Figure 1-36 Examples of exchange reactions compatible with the peptide side-chain functional groups. Taken from ref.[98]

Finding reversible chemical ligation reactions is at the core of DCLs discussed previously. This emerging field has been using several strategies for the construction of dynamic peptide libraries, which do not interfere with amino acid side chains. However, most of these methods do not give reversibility at the amino acid level. Instead, it allows a dynamic system of peptide fragments to ligate and dissociate with one another.

A first possibility is to use noncovalent interactions. Since the energy associated with such interaction is low, the fragments can assemble and disassemble. As an example, in metal-ligand bond exchange, peptide fragments possess complexing agents and metal complexation acts as joints between peptide fragments. [99,100]

Hydrazone bond exchange has also been reported for the creation of DCLs. In this system, hydrazide terminated peptides react with protected aldehyde beginning fragments under acidic conditions to create hydrazones. These hydrazones can then undergo fragment exchange between each other. This strategy is robust but requires different modifications on both ends of the peptide fragments.[101]

Various enzymes can catalyze transamidation processes, thus exchanging fragments between different peptides. Notably, non-specific proteases such as thermolysin have been used for the

generation of peptide libraries.[102] Another enzyme, Sortase A can also catalyze the formation and hydrolysis of LPXTG depsipeptide patterns (a depsipetptide is a sequence in which amino acids and hydroxy acids alternate) with X any amino acid.[103] However, as described previously, the enzyme might face folding problems in the hotter areas of the gradient and if the folds are irreversible the system might be short lived.

The side chain of amino acids can also be used to create ligation between peptide fragments. The side chain of a cysteine consists of a sulfur group. The oxidation of this sulfur leads to disulfide bonds which are key structural elements for protein structures but that can also be reduced back to the sulfur state by reducing agents therefore creating reversible ligation between cysteine containing fragments.[104] The Otto groups has developed, a system which can be considered out-of-equilibrium by using this type of chemistry. This system is based on the subunits shown in figure 1-37. It consists of a peptide sequence with a switchable amino acid linked by a peptide bond to a dithiol-benzaldehyde. The two thiols act as the dynamic links between subunits through disulfide bond exchange. This allows the system to assemble in rings of different diameter depending on the number of subunits in each ring.[105]

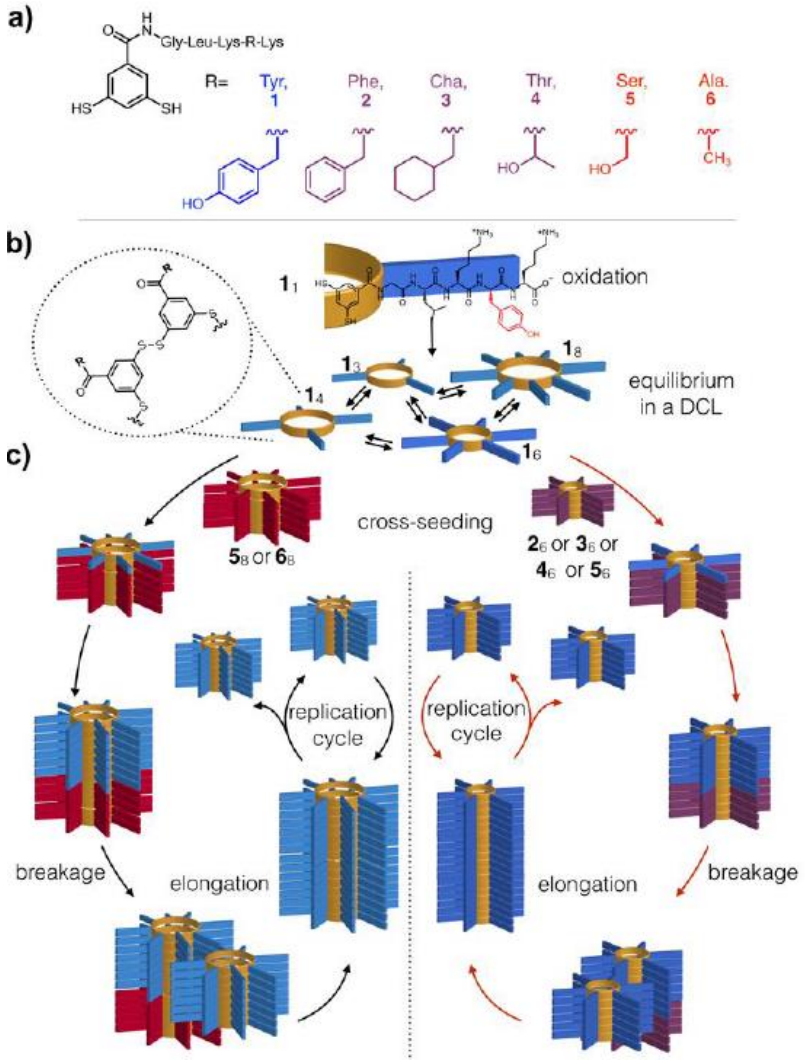


Figure 1-37 (a) Structures of the dithiol building blocks. (b) Oxidation of the tyrosine containing building block leads to formation of a dynamic combinatorial library of differently sized macrocyclic disulfides. (c) Selective formation of 1_6 or 1_8 upon cross-seeding with other pre-formed replicators made from building blocks 2-6 and schematic representation of a tentative mechanism through which these pre-existing replicators direct the formation of the new replicator 1_6 and 1_8. Adapted from ref. [105]

The formed macrocycles can then stack through β-sheet formation between the peptide sequences. This allows the system to create fibers. The system behaves differently depending on the mechanical and oxidative stress applied to the system[104,105] and displays some templating properties, notably from aliphatic amines which speed up the formation of hexamers for phenylalanine systems.[106] This example shows that peptide sequences when brought out of equilibrium here through mechanical agitation, may allow different behaviors to emerge.

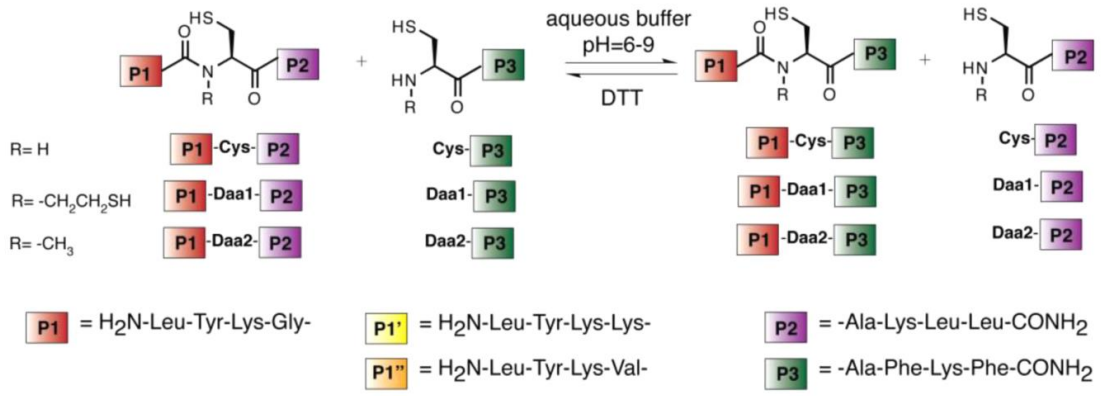
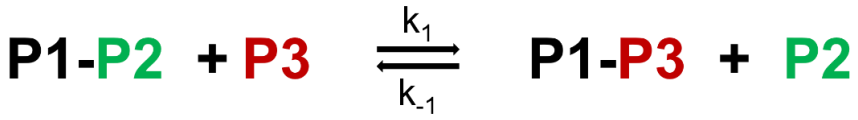


Figure 1-38 Reversible peptide bond formation using dynamic native chemical ligation (dynNCL). Dynamic NCL based on the use of N-(2-thioethyl)-cysteine (SEtCys) and N-methyl-cysteine.

The last strategy, which we used in this project, is dynamic native chemical ligation mediated by alkyl-cysteines which was originally developed in our laboratory, and which will be discussed in detail in Chapter 2.[107] In these systems, N-alkylated cysteines allow the peptide bond to undergo a N-S acyl shift and switch to a thioester bond. Peptide fragments from different thioesters can be exchanged through trans-thioesterification giving rise to a dynamic system without the requirement of a catalyst.

6. Preliminary work

Before the beginning of this PhD project, our modelling partner Raphaël Plasson, working at the INRA (Institut National de la Recherche Agronome) in Avignon, proceeded to model a very simple system to help visualize potential effects what the chemistry should be for such a system to be possible. The model corresponds to a capillary filled with the reactive mixture placed in a temperature gradient. A finite elements method[1] is then used to evolve the system based on the following interactions:



$$\Delta H_{P1P2\text{-}P1P3} \neq 0$$

$$K = \frac{k_1}{k_{-1}}$$

P1 4AA P2 2AA et P3 10AA

$$D_{P1P2} = 0{,}92\ D_{P2}\ ,\ D_{P3} = 0{,}75\ D_{P2},\ D_{P1P3} = 0{,}64\ D_{P2}$$

Figure 1-39 Simple exchange system used for the preliminary modelling work

In this system, a peptide composed of two fragments, P1 and P2, respectively 4 and 2 amino acids long, could exchange its P2 fragment in a reversible manner with another peptide fragment present in solution, P3. This reaction is characterized by it forwards (k_1) and backwards (k_{-1}) kinetic constants, its equilibrium constant (K) and a non-null reaction enthalpy (ΔH) given that the ligation process must be influenced by temperature.

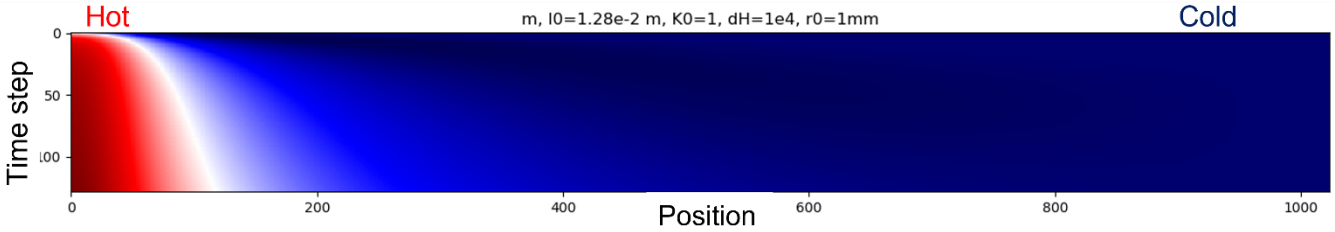


Figure 1-40 Heatmap of the total concentration in the modelled system. Red areas show above average, blue below.

Starting from equimolar concentrations of peptide fragments, when letting the system evolve with a gaussian gradient between 200°C (achieved in aqueous media by the gold nanoparticle heating system)[89] and room temperature, a concentration gradient appears as shown by the heatmap (figure 1-29). In the hot region, the more slow-diffusive specie P1-P3 is formed, depleting the P1-P2 pool of the region allowing for its diffusion towards the hotspot and generating P2 fragments which will diffuse towards the cooler regions. This results in a higher total concentration of species near the hotspot. The maximum of total concentration was shown

[1] More information on the modelling of reaction networks is given in chapter4.

to increase with the reaction enthalpy and peaks with gradients lengths of millimeters. The concentration effects for this system are of 0,5% for a reaction enthalpy of 1 kJ/mol for the ligation reaction (i.e., at the maximum the total concentration is 0,5% higher than it would be at equilibrium) and increase up to 3% for a reaction enthalpy of 10 kJ/mol.

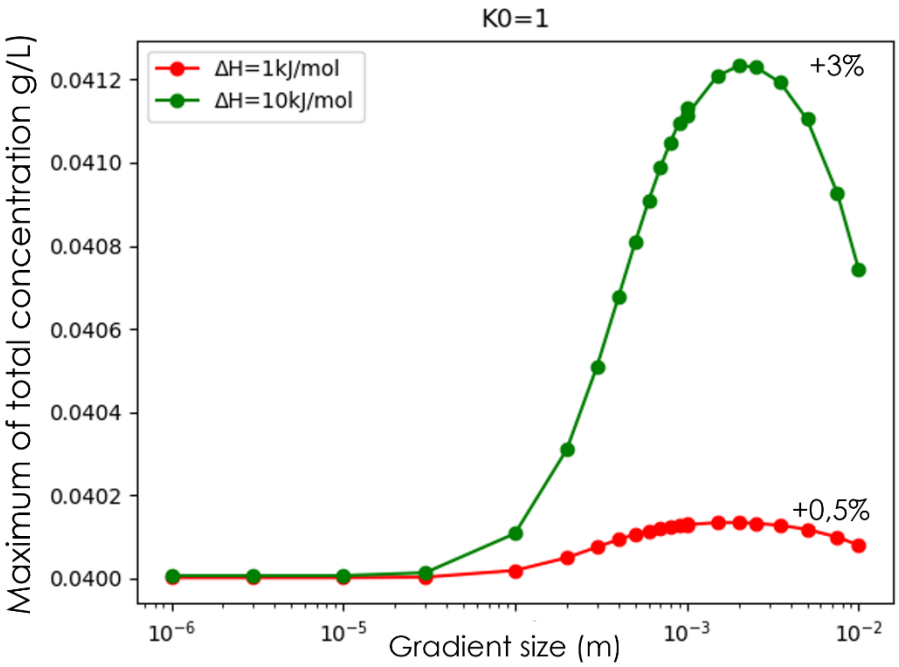


Figure 1-41 Evolution of the maximum local concentration in the system with gradient size and reaction enthalpy values.

The entropy of the system was also modelled (figure 1-42). Initially, it drops as the heat from the gradient starts being dissipated by the system, but then the entropy increases again as the reaction-diffusion processes emerge and the system is brought out-of-equilibrium. The heatmap of the evolution of entropy with time shows a transition between a one maxima regime at the beginning, to two maxima. When the reaction diffusion is not set up, the hotspot corresponds to the location receiving more energy in the form of heat. When a two maxima regime appears, the second one is further down the gradient. This can be explained the apparition of dissipative systems such as reaction-diffusion cycles able to reduce the entropy in between the two high values and transferring it to a maximum further down the gradient.

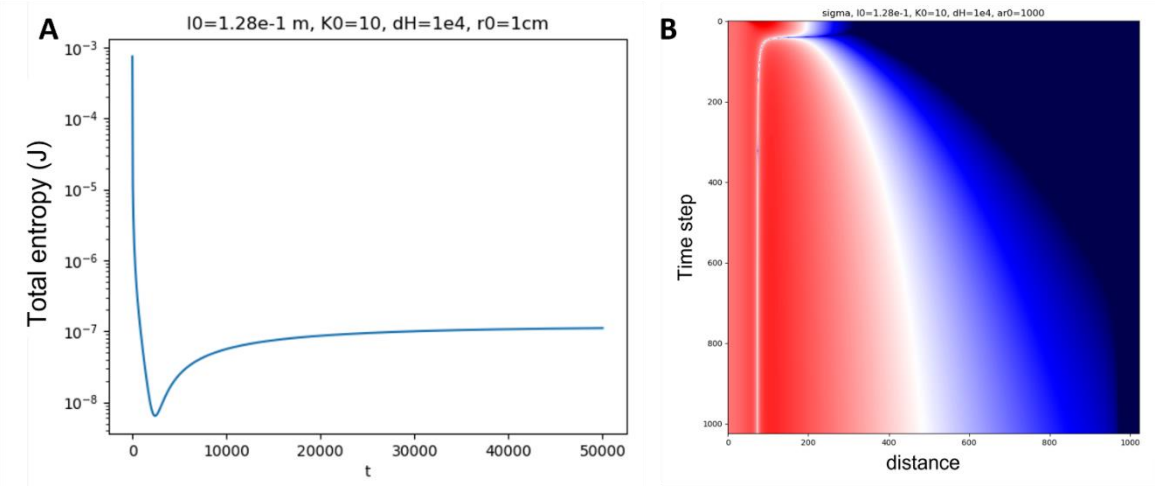


Figure 1-42 (Left) Evolution of the total entropy of the system with time in the modelled system. (Right) Heatmap of the local entropy in the simple modelled system.

7. Objectives of this PhD project

In this introductory part, we established that systems are maintained out of equilibrium by constant input of energy in the system. Biology is in a permanent out of equilibrium state gives rise to many properties and chemical behaviors in several domains, namely: temporal control, information processing, motion and mechanical work, and autonomous structures. Researchers have been able to reproduce many of these behaviors with synthetic abiotic systems however, the new generation of smart materials and systems will probably come from the incorporation of all these features together and this remains to be a challenge.

This work wants to take advantage of the reaction diffusion cycles that can be generated through a temperature gradient. To achieve this, a monomer system whose ligation and dissociation give rise to different species distribution at different temperatures must be designed. The objective is to use the temperature gradient to allow high temperature created polymers to diffuse to lower temperatures, where the energy cannot sustain their creation, thus bringing the system out of equilibrium. We selected amino acids as monomers for such system. Amino acid sequences heavily influence the structure and properties of their polymers and even amino acid oligomers, peptides can exhibit folding and structural modifications linked to their sequence.

The peptide bond however is a quite stable chemical bond in non-catalyzed media. To have a dynamic system we settled on the dynamic native chemical ligation permitted by the incorporation of alkyl-cysteines in the peptide sequences. These enable reversible peptide fragment exchanges.

Preliminary napkin modelling of the system, led to the conclusion that a simple exchange reaction can already give rise to out of equilibrium effects such as concentration gradients across the temperature gradient and variations in entropy across the system with multiple maxima. This work also led to the conclusion that to maximize out of equilibrium effects, the system should have high reaction enthalpy. The out of equilibrium effects were observed with diffusion constants for small oligopeptides but the diffusion constants mostly impact the size of the gradient.

8. References:

1. E.A. Guggenheim. *An Advanced Treatment for Chemists and Physicists*. (1985).
2. Kittel, C. & Kroemer, H. *Thermal Physic ·*. (1980).
3. Giuseppone, N. & Walther, A. *Out-of-Equilibrium (Supra)molecular Systems and Materials*. (Wiley, 2021). doi:10.1002/9783527821990.
4. Grzybowski, B. A., Fitzner, K., Paczesny, J. & Granick, S. From dynamic self-assembly to networked chemical systems. *Chemical Society Reviews* vol. 46 5647–5678 (2017).
5. Tetilla, P. *' et al.* Hydrogen-Bonding Self-Assembly of Multichromophore Structures. *Journal of the American Chemical Society* **112**, 9408–9410 (1990).
6. van Rossum, S. A. P., Tena-Solsona, M., van Esch, J. H., Eelkema, R. & Boekhoven, J. Dissipative out-of-equilibrium assembly of man-made supramolecular materials. *Chemical Society Reviews* vol. 46 5519–5535 (2017).
7. Lee, Y.-C. & Moon, J.-Y. *Introduction to Bionanotechnology*. (Springer, 2020).
8. Silverstein, T. P. The Real Reason Why Oil and Water Don't Mix. *Journal of Chemical Education* **75**, (1998).
9. von Hippel, P. H. Structures in an Aqueous World: *The Hydrophobic Effect* . Formation of Micelles and Biological Membranes. Charles Tanford. Wiley-Interscience, New York, 1973. viii, 200 pp., illus. $12.50. *Science* **184**, (1974).
10. Almeida, P. F. F., Pokorny, A. & Hinderliter, A. Thermodynamics of membrane domains. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1720**, (2005).
11. Walker, D. A., Browne, K. P., Kovalczyk, B. & Grzybowski, B. A. Self-assembly of nanotriangle superlattices facilitated by repulsive electrostatic interactions. *Angewandte Chemie - International Edition* **49**, 6760–6763 (2010).
12. Kuzyk, A. *et al.* DNA-based self-assembly of chiral plasmonic nanostructures with tailored optical response. *Nature* **483**, (2012).
13. Akcora, P. *et al.* Anisotropic self-assembly of spherical polymer-grafted nanoparticles. *Nature Materials* **8**, (2009).
14. Lee, H.-Y. *et al.* Self-Assembly of Nanoparticle Amphiphiles with Adaptive Surface Chemistry. *ACS Nano* **8**, (2014).
15. Walker, D. A., Browne, K. P., Kowalczyk, B. & Grzybowski, B. A. Self-Assembly of Nanotriangle Superlattices Facilitated by Repulsive Electrostatic Interactions. *Angewandte Chemie International Edition* **49**, (2010).
16. Pillai, P. P., Kowalczyk, B. & Grzybowski, B. A. Self-assembly of like-charged nanoparticles into microscopic crystals. *Nanoscale* **8**, (2016).

17. Bundy, F. P. *The P, T Phase and Reaction Diagram for Elemental Carbon, 1979*. *JOURNAL OF GEOPHYSICAL RESEARCH* vol. 85 (1980).

18. Kang, J. *et al.* A rational strategy for the realization of chain-growth supramolecular polymerization. *Science* **347**, (2015).

19. Klajn, R. Immobilized azobenzenes for the construction of photoresponsive materials. *Pure and Applied Chemistry* **82**, (2010).

20. Klajn, R., Stoddart, J. F. & Grzybowski, B. A. Nanoparticles functionalised with reversible molecular and supramolecular switches. *Chemical Society Reviews* **39**, (2010).

21. Klajn, R., Bishop, K. J. M. & Grzybowski, B. A. Light-controlled self-assembly of reversible and irreversible nanoparticle suprastructures. *Proceedings of the National Academy of Sciences* **104**, (2007).

22. Klajn, R., Wesson, P. J., Bishop, K. J. M. & Grzybowski, B. A. Writing Self-Erasing Images using Metastable Nanoparticle "Inks." *Angewandte Chemie International Edition* **48**, (2009).

23. Grzybowski, B. A., Stone, H. A. & Whitesides, G. M. Dynamic self-assembly of magnetized, millimetre-sized objects rotating at a liquid–air interface. *Nature* **405**, (2000).

24. Grzybowski, B. A. & Whitesides, G. M. Dynamic Aggregation of Chiral Spinners. *Science* **296**, (2002).

25. Ng, J. M. K., Fuerstman, M. J., Grzybowski, B. A., Stone, H. A. & Whitesides, G. M. Self-assembly of gears at a fluid/air interface. *Journal of the American Chemical Society* **125**, 7948–7958 (2003).

26. Dambenieks, A. K., Vu, P. H. Q. & Fyles, T. M. Dissipative assembly of a membrane transport system. *Chem. Sci.* **5**, (2014).

27. Leira-Iglesias, J., Sorrenti, A., Sato, A., Dunne, P. A. & Hermans, T. M. Supramolecular pathway selection of perylenediimides mediated by chemical fuels. *Chemical Communications* **52**, (2016).

28. Beatus, T., Shani, I., Bar-Ziv, R. H. & Tlusty, T. Two-dimensional flow of driven particles: A microfluidic pathway to the non-equilibrium frontier. *Chemical Society Reviews* vol. 46 5620–5646 (2017).

29. Merindol, R. & Walther, A. Materials learning from life: Concepts for active, adaptive and autonomous molecular systems. *Chemical Society Reviews* vol. 46 5588–5619 (2017).

30. Novák, B. & Tyson, J. J. Design principles of biochemical oscillators. *Nature Reviews Molecular Cell Biology* **9**, (2008).

31. Gallego, M. & Virshup, D. M. Post-translational modifications regulate the ticking of the circadian clock. *Nature Reviews Molecular Cell Biology* **8**, (2007).

32. Ko, C. H. & Takahashi, J. S. Molecular components of the mammalian circadian clock. *Human Molecular Genetics* **15**, (2006).

33. Field, R. J., Koros, E. & Noyes, R. M. Oscillations in chemical systems. II. Thorough analysis of temporal oscillation in the bromate-cerium-malonic acid system. *Journal of the American Chemical Society* **94**, (1972).

34. Yoshida, R., Takahashi, T., Yamaguchi, T. & Ichijo, H. Self-Oscillating Gel. *Journal of the American Chemical Society* **118**, (1996).

35. Bishop, K. J. M. & Grzybowski, B. A. Localized Chemical Wave Emission and Mode Switching in a Patterned Excitable Medium. *Physical Review Letters* **97**, (2006).

36. Kuhnert, L., Agladze, K. I. & Krinsky, V. I. Image processing using light-sensitive chemical waves. *Nature* **337**, (1989).

37. Hu, G., Pojman, J. A., Scott, S. K., Wrobel, M. M. & Taylor, A. F. Base-Catalyzed Feedback in the Urea−Urease Reaction. *The Journal of Physical Chemistry B* **114**, (2010).

38. Semenov, S. N. *et al.* Rational design of functional and tunable oscillating enzymatic networks. *Nature Chemistry* **7**, (2015).

39. Heuser, T., Weyandt, E. & Walther, A. Biocatalytic Feedback-Driven Temporal Programming of Self-Regulating Peptide Hydrogels. *Angewandte Chemie International Edition* **54**, (2015).

40. Fletcher, D. A. & Mullins, R. D. Cell mechanics and the cytoskeleton. *Nature* **463**, (2010).

41. Pollard, T. D. & Cooper, J. A. Actin, a Central Player in Cell Shape and Movement. *Science* **326**, (2009).

42. Hess, H. & Ross, J. L. Non-equilibrium assembly of microtubules: from molecules to autonomous chemical robots. *Chemical Society Reviews* **46**, (2017).

43. Bornens, M. & Azimzadeh, J. Origin and Evolution of the Centrosome. in *Eukaryotic Membranes and Cytoskeleton* (Springer New York). doi:10.1007/978-0-387-74021-8_10.

44. Pollard, T. D. Regulation of Actin Filament Assembly by Arp2/3 Complex and Formins. *Annual Review of Biophysics and Biomolecular Structure* **36**, (2007).

45. Nogales, E., Wolf, S. G. & Downing, K. H. Structure of the αβ tubulin dimer by electron crystallography. *Nature* **391**, (1998).

46. Kundu, P. K. *et al.* Light-controlled self-assembly of non-photoresponsive nanoparticles. *Nature Chemistry* **7**, (2015).

47. Ueki, T., Shibayama, M. & Yoshida, R. Self-oscillating micelles. *Chemical Communications* **49**, (2013).

48. He, X. *et al.* Synthetic homeostatic materials with chemo-mechano-chemical self-regulation. *Nature* **487**, (2012).

49. https://en.wikipedia.org/wiki/Kinesin.

50. Vale, R. D. & Milligan, R. A. The Way Things Move: Looking Under the Hood of Molecular Motor Proteins. *Science* **288**, (2000).

51. Lodish, H. F., Berk, A., Krieger, M. & Scott, M. P. *Molecular Cell Biology*. (W.H. Freeman, 2008).

52. https://www.encyclopedie-environnement.org/zoom/synthese-atp/.

53. Yoshida, M., Muneyuki, E. & Hisabori, T. ATP synthase — a marvellous rotary engine of the cell. *Nature Reviews Molecular Cell Biology* **2**, (2001).

54. Boyer, P. D. THE ATP SYNTHASE—A SPLENDID MOLECULAR MACHINE. *Annual Review of Biochemistry* **66**, (1997).

55. Berg, H. C. The Rotary Motor of Bacterial Flagella. *Annual Review of Biochemistry* **72**, (2003).

56. Adelstein, R. S. & Eisenberg, E. Regulation and Kinetics of the Actin-Myosin-ATP Interaction. *Annual Review of Biochemistry* **49**, (1980).

57. Sweeney, H. L. & Houdusse, A. Structural and Functional Insights into the Myosin Motor Mechanism. *Annual Review of Biophysics* **39**, (2010).

58. Tian, Y., He, Y., Chen, Y., Yin, P. & Mao, C. A DNAzyme That Walks Processively and Autonomously along a One-Dimensional Track. *Angewandte Chemie International Edition* **44**, (2005).

59. Lund, K. *et al.* Molecular robots guided by prescriptive landscapes. *Nature* **465**, (2010).

60. Klok, M. *et al.* MHz Unidirectional Rotation of Molecular Rotary Motors. *Journal of the American Chemical Society* **130**, (2008).

61. Fletcher, S. P., Dumur, F., Pollard, M. M. & Feringa, B. L. A Reversible, Unidirectional Molecular Rotary Motor Driven by Chemical Energy. *Science* **310**, (2005).

62. Ragazzon, G., Baroncini, M., Silvi, S., Venturi, M. & Credi, A. Light-powered autonomous and directional molecular motion of a dissipative self-assembling system. *Nature Nanotechnology* **10**, (2015).

63. Kudernac, T. *et al.* Electrically driven directional motion of a four-wheeled molecule on a metal surface. *Nature* **479**, (2011).

64. Li, Q. *et al.* Macroscopic contraction of a gel induced by the integrated motion of light-driven molecular motors. *Nature Nanotechnology* **10**, (2015).

65. Foy, J. T. *et al.* Dual-light control of nanomachines that integrate motor and modulator subunits. *Nature Nanotechnology* **12**, (2017).

66. Xiong, Y., Huang, C.-H., Iglesias, P. A. & Devreotes, P. N. Cells navigate with a local-excitation, global-inhibition-biased excitable network. *Proceedings of the National Academy of Sciences* **107**, (2010).

67. van Haastert, P. J. M. & Devreotes, P. N. Chemotaxis: signalling the way forward. *Nature Reviews Molecular Cell Biology* **5**, (2004).

68. Porter, S. L., Wadhams, G. H. & Armitage, J. P. Signal processing in complex chemotaxis pathways. *Nature Reviews Microbiology* **9**, (2011).

69. Wadhams, G. H. & Armitage, J. P. Making sense of it all: bacterial chemotaxis. *Nature Reviews Molecular Cell Biology* **5**, (2004).

70. Iglesias, P. A. & Devreotes, P. N. Biased excitable networks: how cells direct motion in response to gradients. *Current Opinion in Cell Biology* **24**, (2012).

71. Swaney, K. F., Huang, C.-H. & Devreotes, P. N. Eukaryotic Chemotaxis: A Network of Signaling Pathways Controls Motility, Directional Sensing, and Polarity. *Annual Review of Biophysics* **39**, (2010).

72. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **237**, (1952).

73. Kondo, S. & Miura, T. Reaction-Diffusion Model as a Framework for Understanding Biological Pattern Formation. *Science* **329**, (2010).

74. Gurdon, J. B. & Bourillot, P.-Y. Morphogen gradient interpretation. *Nature* **413**, (2001).

75. http://worms.zoology.wisc.edu/frogs/fert/subcortrot/fertsag.html.

76. Lu, X. *et al.* Photophobic and phototropic movement of a self-oscillating gel. *Chemical Communications* **49**, (2013).

77. Dey, K. K. *et al.* Micromotors Powered by Enzyme Catalysis. *Nano Letters* **15**, (2015).

78. Dey, K. K. *et al.* Chemotactic Separation of Enzymes. *ACS Nano* **8**, (2014).

79. Wang, J. & Gao, W. Nano/Microscale Motors: Biomedical Opportunities and Challenges. *ACS Nano* **6**, (2012).

80. Burdick, J., Laocharoensuk, R., Wheat, P. M., Posner, J. D. & Wang, J. Synthetic Nanomotors in Microchannel Networks: Directional Microchip Motion and Controlled Manipulation of Cargo. *Journal of the American Chemical Society* **130**, (2008).

81. Wu, Y., Lin, X., Wu, Z., Möhwald, H. & He, Q. Self-Propelled Polymer Multilayer Janus Capsules for Effective Drug Delivery and Light-Triggered Release. *ACS Applied Materials & Interfaces* **6**, (2014).

82. Wrobel, M. M. *et al.* pH Wave-Front Propagation in the Urea-Urease Reaction. *Biophysical Journal* **103**, (2012).

83. Chirieleison, S. M., Allen, P. B., Simpson, Z. B., Ellington, A. D. & Chen, X. Pattern transformation with DNA circuits. *Nature Chemistry* **5**, (2013).

84. Semenov, S. N., Markvoort, A. J., de Greef, T. F. A. & Huck, W. T. S. Threshold Sensing through a Synthetic Enzymatic Reaction-Diffusion Network. *Angewandte Chemie International Edition* **53**, (2014).

85. Semenov, S. N. *et al.* Ultrasensitivity by Molecular Titration in Spatially Propagating Enzymatic Reactions. *Biophysical Journal* **105**, (2013).

86. Bánsági, T., Vanag, V. K. & Epstein, I. R. Tomography of Reaction-Diffusion Microemulsions Reveals Three-Dimensional Turing Patterns. *Science* **331**, (2011).

87. Horváth, J., Szalai, I. & de Kepper, P. An Experimental Design Method Leading to Chemical Turing Patterns. *Science* **324**, (2009).

88. Baffou, G. *et al.* Deterministic temperature shaping using plasmonic nanoparticle assemblies. *Nanoscale* **6**, (2014).

89. Baffou, G., Polleux, J., Rigneault, H. & Monneret, S. Super-Heating and Micro-Bubble Generation around Plasmonic Nanoparticles under cw Illumination. *The Journal of Physical Chemistry C* **118**, (2014).

90. Vranova, V. *et al.* The significance of D-amino acids in soil, fate and utilization by microbes and plants: review and identification of knowledge gaps. *Plant and Soil* **354**, (2012).

91. https://en.wikipedia.org/wiki/Amino_acid.

92. Pauling, L., Corey, R. B. & Branson, H. R. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences* **37**, (1951).

93. https://chemistry.stackexchange.com/questions/143189/why-are-hydrogen-bonds-in-an-antiparallel-beta-sheet-stronger-than-those-in-para.

94. Prusiner, S. B. Molecular Biology of Prion Diseases. *Science* **252**, (1991).

95. Kumar, S. & Bansal, M. Geometrical and Sequence Characteristics of α-Helices in Globular Proteins. *Biophysical Journal* **75**, (1998).

96. Sievers, A., Beringer, M., Rodnina, M. v. & Wolfenden, R. The ribosome as an entropy trap. *Proceedings of the National Academy of Sciences* **101**, (2004).

97. Radzicka, A. & Wolfenden, R. Rates of Uncatalyzed Peptide Bond Hydrolysis in Neutral Solution and the Transition State Affinities of Proteases. *Journal of the American Chemical Society* **118**, (1996).

98. Rete, C.-V. Libraries of Dynamic Peptides based on reversible native chemical ligation. (2018).

99. Roy, L. & Case, M. A. Recursively Enriched Dynamic Combinatorial Libraries for the Self-Selection of Optimally Stable Proteins. *The Journal of Physical Chemistry B* **115**, (2011).

100. Roy, L. & Case, M. A. Protein Core Packing by Dynamic Combinatorial Chemistry. *Journal of the American Chemical Society* **132**, (2010).

101. Cousins, G. R. L., Poulsen, S.-A. & Sanders, J. K. M. Dynamic combinatorial libraries of pseudo-peptide hydrazone macrocycles. *Chemical Communications* (1999) doi:10.1039/a904091i.

102. Williams, R. J. *et al.* Enzyme-assisted self-assembly under thermodynamic control. *Nature Nanotechnology* **4**, (2009).

103. Williamson, D. J., Fascione, M. A., Webb, M. E. & Turnbull, W. B. Efficient N-Terminal Labeling of Proteins by Use of Sortase. *Angewandte Chemie International Edition* **51**, (2012).

104. Sadownik, J. W., Mattia, E., Nowak, P. & Otto, S. Diversification of self-replicating molecules. *Nature Chemistry* **8**, (2016).

105. Altay, M., Altay, Y. & Otto, S. Parasitic Behavior of Self-Replicating Molecules. *Angewandte Chemie International Edition* **57**, (2018).

106. Bartolec, B., Altay, M. & Otto, S. Template-promoted self-replication in dynamic combinatorial libraries made from a simple building block. *Chemical Communications* **54**, (2018).

107. Ruff, Y., Garavini, V. & Giuseppone, N. Reversible Native Chemical Ligation: A Facile Access to Dynamic Covalent Peptides. *Journal of the American Chemical Society* **136**, (2014).

Chapter 2. The N-methyl-cysteine exchange reaction

1. Introduction

The first step in implementing the desired reversible polymer system described at the end of chapter 1 is to study the basics of the peptide exchange reaction. According to the preliminary modelling, for reaching noticeable out-of-equilibrium behavior there are requirements in terms of reaction enthalpy. Thankfully, with peptide system this can be tuned with the amino acid sequences. As discussed previously, amino acid sequences can fold to create structures. The interactions and folds can be used to stabilize either the product or the reactants of the exchange reaction to swing the reaction enthalpy in the favor of one or the other. For example, α-helixes are stabilized by hydrogen bonding which are estimated to be a $3.4\ kJ.mol^{-1}$ stabilization per bond[1]. Other motifs such as leucine zippers have also been used in conjunction with the N-alkyl-exchange reaction[2]. Finally zinc finger motives are shown to have formation enthalpies in the desired range and cysteines have good affinities for zinc[3].

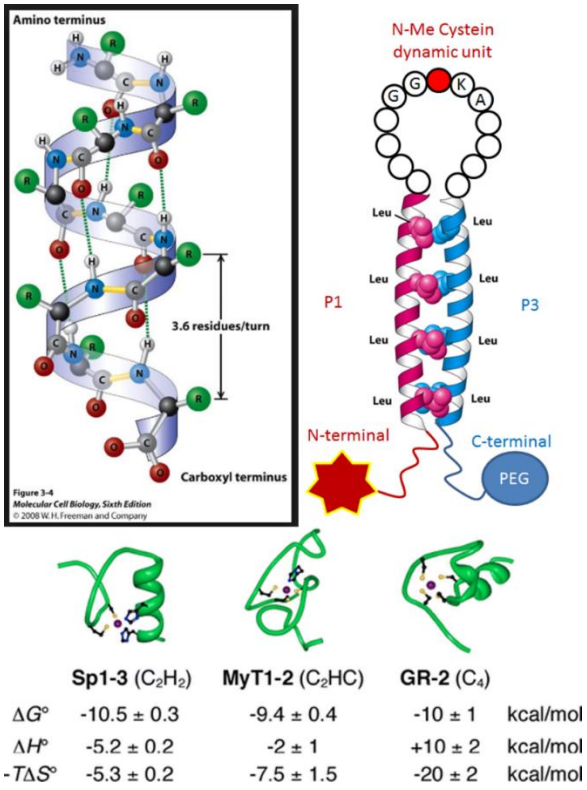


Figure 2-1 Top left alpha-helix motif in proteins[4], top right glycine zipper used in for the creation of artificial inteins with N-methyl-cysteines mediated exchange[2]. Bottom, thermodynamic data for the formation of zinc fingers depending on the complexing amino acids.

For this reason, we wanted to monitor the kinetics of the exchange reaction excluding as much as possible sequence induced interactions which could be added later by modifying the sequences.

2. N-alkyl-cysteine exchange reaction

a. Reaction mechanism

As discussed previously, the peptide is very stable in neutral water. This means that abiotic transamidation is not an option to exchange peptide fragments. The key behind the N-alkyl-cysteine dynamic native chemical ligation (dyNCL) is the introduction of an equilibrium between the peptide amide bond and a thioester bond.

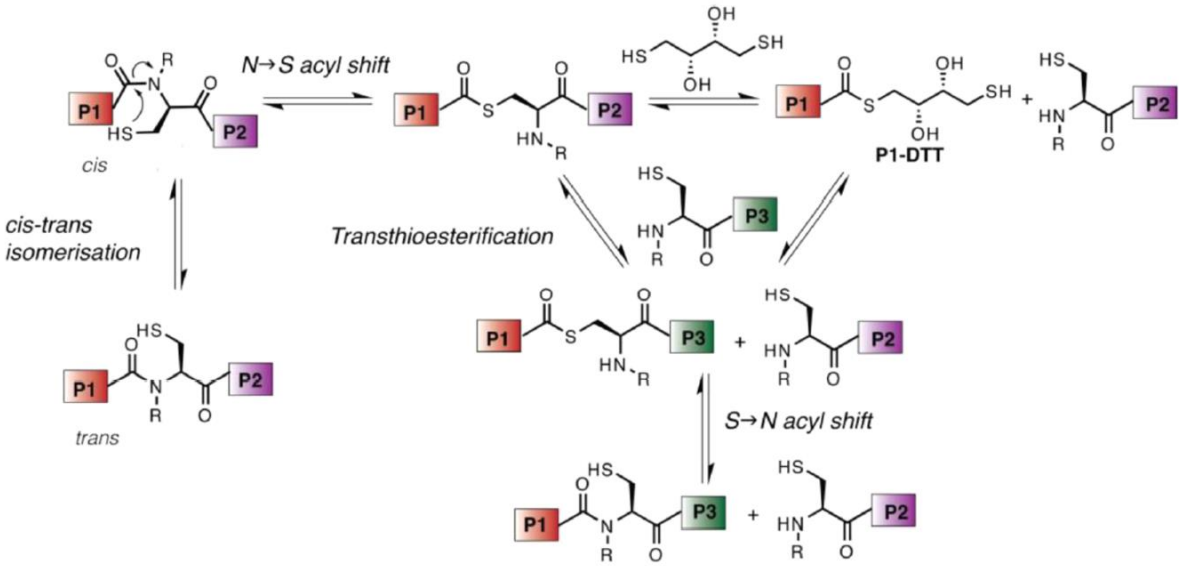


Figure2-2 Proposed mechanism for the N-alkyl-cysteine exchange reaction.[5]

This is permitted by an intramolecular reorganization process called N→S acyl shift (or transfer). The nucleophile sulfur atom will react with the electrophile carbon of the amide bond and release the amine forming a thioester bond. N→S acyl shifts are reversible processes as S→N shifts happen as well. In fact, the reaction is not thermodynamically favorable as the amide bond is more stable, meaning that the peptide bond will be the main state for this bond[6]. However, poor C-S π orbital overlap is thought to increase the reactivity of thioesters making them more reactive than their oxoester counterparts [7]. This in turn makes them good reaction intermediates.

For the acyl shift to take place, the molecule must adopt a favorable geometry for the rearrangement, which in the case of our peptides corresponds to the *cis* configuration. Apart from proline, natural amino acids form secondary amide bonds. The most stable configuration for these peptide bonds is the *trans* configuration, with only 0.5% of *cis* bonds in dipeptides and around 0.15% for longer oligopeptides[8]. This, coupled to the unfavorable N→S shift, leads to low thioester concentration i.e., very few candidate molecules for the fragment exchange reaction. Alkylating the nitrogen, and thus having a tertiary amide, leads to important modifications of the *cis trans* isomerization equilibrium. Studies on the Ala-Gly-Pro-Phe motif show that the proline bond at pH 6 water shows 23% of *cis* conformers[9]. Substitution of the proline by N substituted glycine further increases the ratio of *cis* with as size of the N substituents increase. Switching from N substituted glycine to alanine does not decrease the

ratio drastically with 30% for sarcosine (methyl-glycine) and 26% for methyl-alanine[9]. The alkylation of the nitrogen is thus key for the generation of the thioester intermediates.

Once formed, the thioesters can react with free thiols in solution, either other cysteines or the reducing agent if a thiol is used. The sulfur from the free thiol performs a nucleophilic attack on the carbon and a thiol is released like with transesterification reactions. After transthioesterification, the thioester bond undergoes the favorable S→N acyl shift restoring a peptide bond between the N-alkyl-cysteine and its previous amino acid.

S to N transfers are fairly common. In nature, there are four major processes involving it[6]. The ubiquitination of proteins (marking them to be recycled in proteasomes by the cell) is performed by S to N acyl shift between a thioester enzyme-ubiquitin complex and a lysine residue on the to be recycled protein[10]. Proteins sometimes possess subsequences of amino acids present for early structural and folding properties but are absent from the final structures. Inteins have the particularity of mediating their own excision from the protein sequence. The first steps towards the excision consists of N to S acyl shifts of cysteines[11] (and N to O shifts of Serine) to create the reactive esters that will allow the intein to cyclize as it separates from the finalized protein. Bacteria use them to modify their cell surface protein through sortase transpeptidase enzymes. The membrane proteins are attached via thioesters to the sortase enzyme which will transfer it to free amine on a phospholipid from the bacterial wall by an S to N transfer[12]. And finally, they are used in trans glutamination processes, where transglutaminase enzymes catalyze the formation of isopeptide bonds between a lysine and a glutamine from different proteins or inside a protein sequence. The glutamine side chain is first attacked by the cysteine, creating a thioester bond for the lysine to react with[13].

In synthetic chemistry, the main application for S to N acyl shifts is to create amide bonds by native chemical ligation (NCL) in peptide synthesis. This allows to synthesize long proteins from shorter fragments terminated by thioesters and for mid sequence fragments starting with cysteines (which will either be present in the final sequence, or which can be reduced or modified to yield different amino acids). The fragments are then ligated quantitative proportions and the cysteine perform S to N acyl shifts to yield peptide bonds[14]. Chemists have also been able to replicate ubiquitination processes using NCL[15].

The N-alkyl-cysteine exchange reaction has been developed in our group and has been mainly investigated with the objective of creating DCLs. Particularly, Dr Garavini determined the conditions for the exchange reaction and investigated the properties of N-methyl and N-ethyl-cysteines laying the groundworks for Dr Rete to investigate its ability to create libraries of affibody segments that could combine based on a target placed in the media. Notably it was shown that after 24h the correct arrangement of helices for an analogue of Immunoglobulin G could be observed as opposed to the un templated library. Attempts were also made at the creation of artificial inteins that would dissociate through the presence of two N-methyl cysteines at each end of the intein sequence.

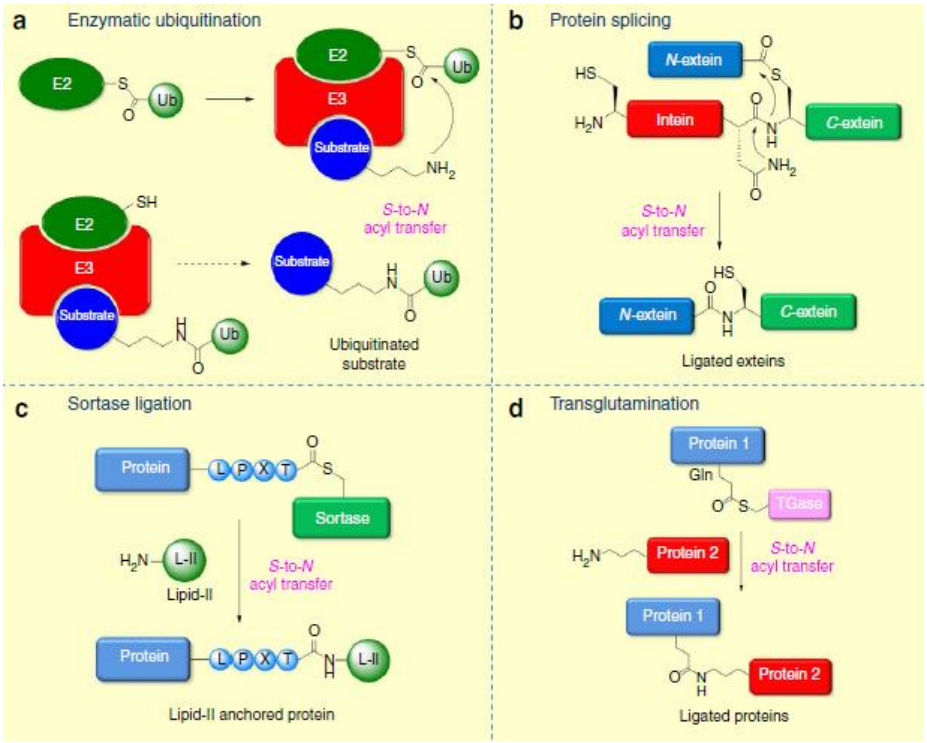


Figure 2-3 S-to-N acyl transfer reactions utilized in nature. (a) Enzymatic ubiquitination where S-to-N acyl transfer forges an isopeptide bond betweena terminal glycine and a lysine residue (b) protein splicing where S-to-N acyl transfer ligates the extein fragments (c) sortase mediated ligation whereS-to-N acyl transfer introduces isopeptide bond between Thr residue of protein and Lys of lipid-II. (d) Transglutamination where S-to-N acyl transfer ligatesproteins via an isopeptide bond. Adapted from ref. [6]

b. Reaction conditions

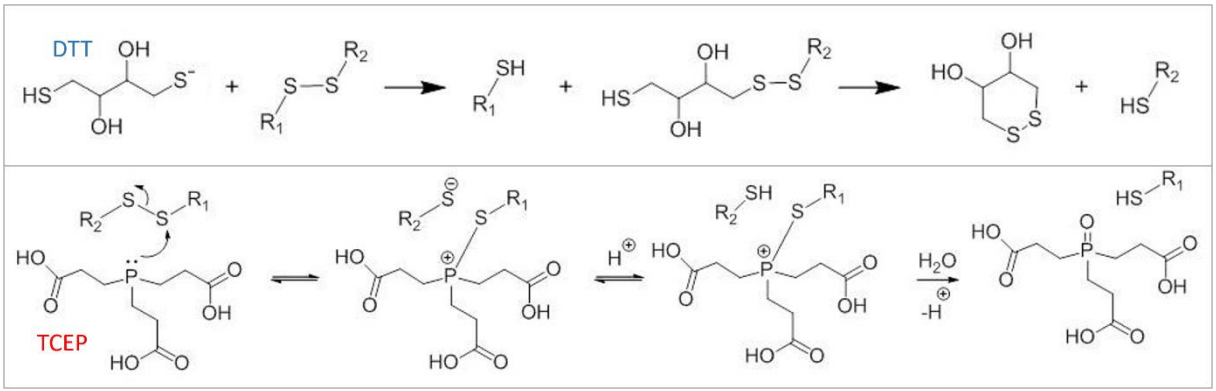


Figure 2-4 Reduction of disulfide bonds by DTT (dithiothreitol) and TCEP (tris(2-carboxyethyl)-phosphine)

The attack of the thioester by a free thiol and the subsequent thiol release is the exchange step in this dyNCL. Thiols however are incompatible with the presence of oxygen as they get oxidized to a di-sulfide state, preventing the nucleophilic attack. Reducing conditions are therefore required. The reaction can be performed in a glove box with inert N_2 atmosphere, but this makes the sampling tedious. The use of a reducing agent was thus preferred. There are many different reducing agents compatible with peptide and thiol chemistry. The most common ones that we used in this work are TCEP (tris(2-carboxyethyl)phosphine) and DTT (dithiothreitol).

DTT, also known as Cleland's reagent, reduces disulfide bonds through a thiol-disulfide exchange reaction. This leads to its oxidation under the form of a six membered ring with an intramolecular disulfide bond. In its first reports, the alkyl-cysteine NCL reaction was performed with DTT. It was observed that DTT also reacted by transthioesterification with the peptides incorporating a N-methyl-cysteine in their sequence. However, the reaction with DTT yields another thioester which can react with cysteine starting fragments in solution and is thus not detrimental to the exchange reaction[5].

TCEP is being more and more common since its introduction twenty years ago. Since it does not contain thiol groups and is mostly unreactive towards thiol-reactive compounds, it is widely used in peptide synthesis. It is more stable than DTT at pH values above 7.5 and can be used over a wider pH range (1.5-8.5) than DTT (7 and upwards, as the thiolate is the reactive species)[16]. Its stability however is heavily affected by buffer composition with phosphate buffer resulting in loss of the reducing power within days [17].

3. Designing the setup for a kinetic assay

a. Peptide sequences and experimental conditions

The objective was ultimately to use the dyNCL exchange reaction to create a polymerizable system. This meant to switch from a fragment exchange to a ligation reaction. This can be done by reducing the C-terminal peptide fragment in the exchange reaction to just the N-methyl-cysteine. Therefore, if another fragment has an N-terminal N-methyl-cysteine, the C-terminal N-methyl-cysteine gets exchanged for the N-terminal N-methyl-cysteine starting fragment as illustrated below. This makes the exchange reaction more akin to a ligation.

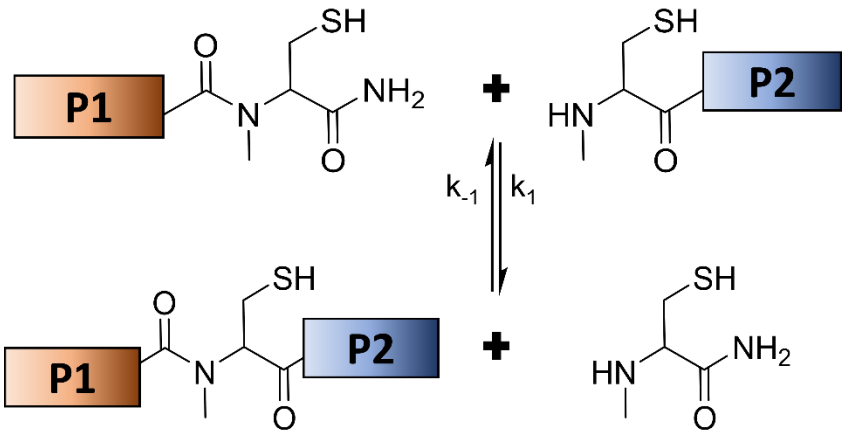


Figure 2-5 Exchange reaction between two peptide fragments. One with a sequence P1 and a C terminal N-methyl-cysteine and another one with an N-terminal N-methyl-cysteine followed by a sequence P2. The exchange yields a peptide of sequence P1-N-methyl-cysteine-P2 and releases a N-methyl-cysteine.

To measure the kinetic parameters of the exchange reaction, we need a way to monitor the concentration of our peptides. Our group has access to a UPLC (Ultra Performance Liquid Chromatography) coupled to electrospray mass spectroscopy (MS) and multiwavelength UV detector. UPLC coupled UV absorption thus seemed to be an easy way to track the kinetics of the reactions, as the MS could allow identification of the new UV peaks of potential side products. We therefore had to make our peptides absorb in UV. There are three aromatic peptides with characteristic UV absorbances: phenylalanine, tryptophane and tyrosine.

Tryptophan has the highest absorption response and would be the best candidate for concentration tracking. However, even if tryptophan would be stable in our aqueous media due to the presence of reducing agents for the cysteine, it is unstable in low and high pH[18] and the elution solvents of our UPLC and HPLC (high performance liquid chromatography, used for peptide purification) have either low or high pH values. Furthermore, it's light induced degradation, even if it is only in the presence of oxygen species, might lead to degradation in the laser heated out of equilibrium setup.

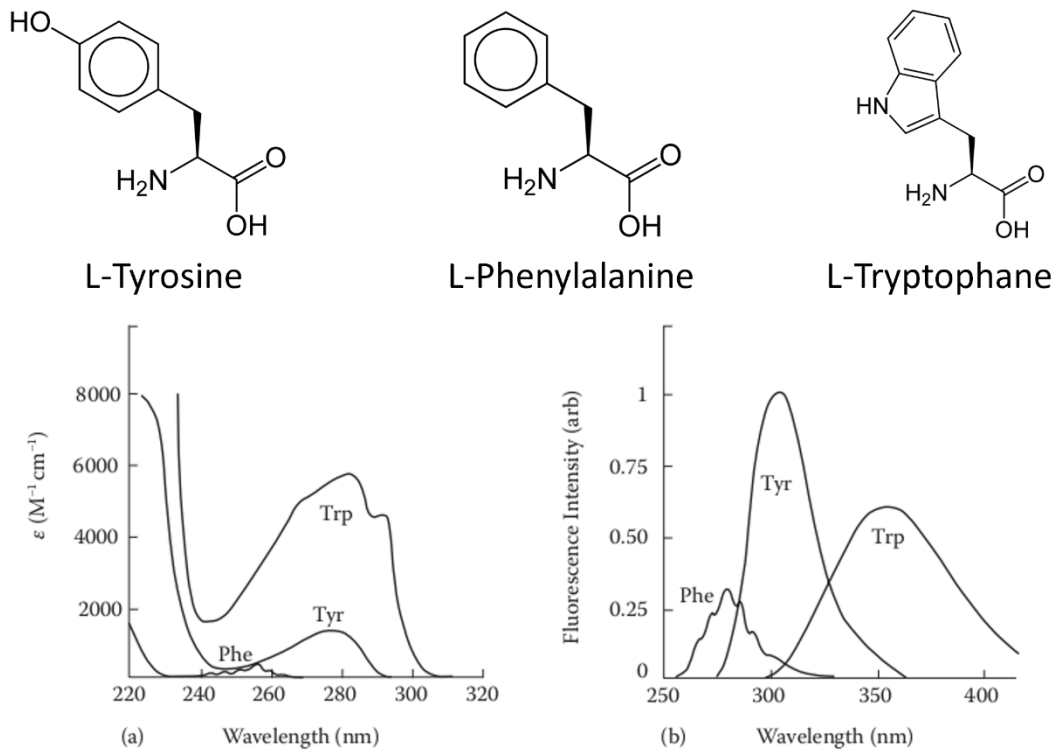


Figure 2-6 The three aromatic amino acids tyrosine, phenylalanine and tryptophane, and their corresponding absorption and fluorescence spectra in neutral water [19]

Tyrosine is more stable in aqueous solution; its absorption coefficient is high enough to allow good concentration measurements at lower concentrations compared to phenylalanine. The phenol side chain also helps with the solubility of the peptide sequences in which it is incorporated compared to phenylalanine. Also, tyrosine hydrogen bonds play an important role in the stability of many proteins, increasing the potential for our peptides to have secondary structure formation or interactions[20].

Phenylalanine is stable in solution but has small absorption coefficients thus complicating concentration measurements.

When thinking of the peptide sequence for our model peptides for the kinetic assay, we had two things in mind. The sequences should be as small as possible to have as less impact as possible from the peptide structure on the kinetics and thermodynamics of the exchange reaction and we wanted the reaction to be relatively quick. This meant removing as much steric hindrance as possible from the thioester bond to allow more reacting attacks from other cysteines. The simplest amino acid is glycine, as its side chain consists of a hydrogen atom. With all these conditions in mind we performed a first kinetic assay with the following peptide sequences and exchange reaction (figure2-7).

In this reaction, the N-methyl-cysteine responsible for the thioester formation is the one of the YG**C** fragment, thus the glycine preceding it. The second fragment, **C**GGY, was chosen longer than the first one to make sure their elution time will be different in liquid chromatography and glycine was chosen as a filler amino acid to prevent possible structural hydrogen bonds between side chains.

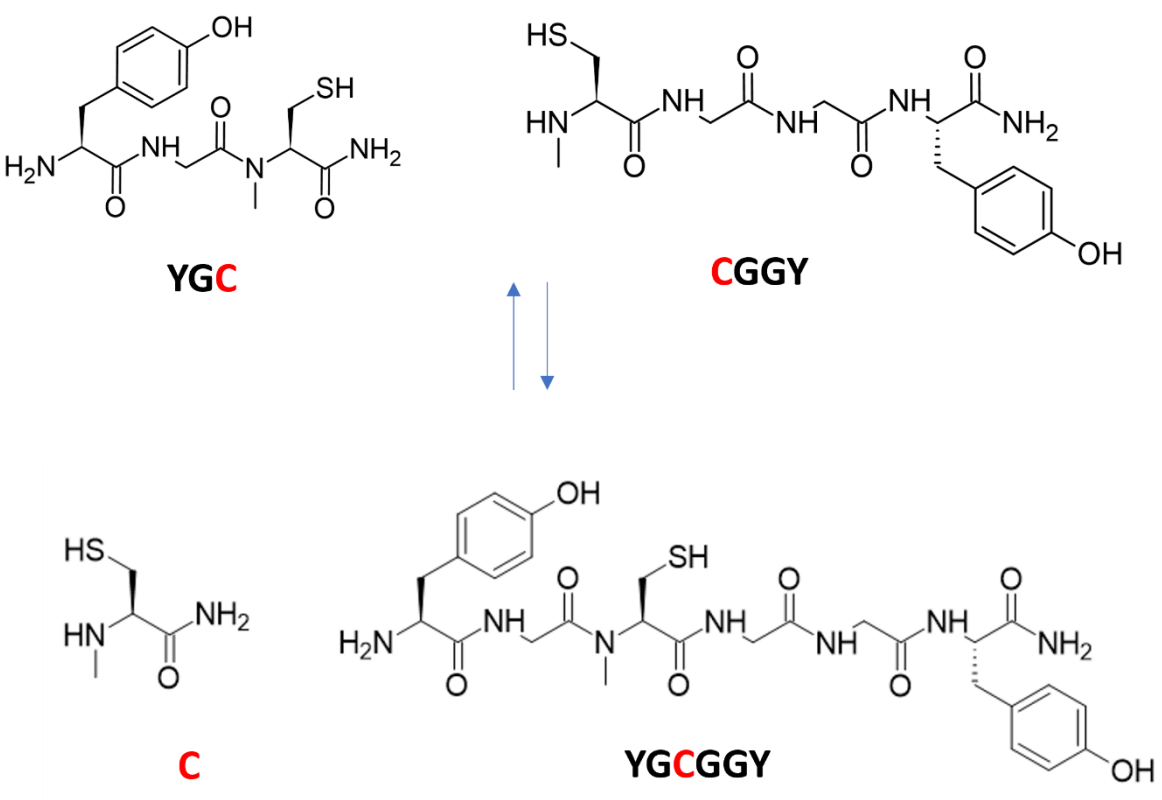


Figure 2-7 Exchange reaction between YGC and CGGY, the red C correspond to N-methylated cysteines

When following the exchange reaction between peptide fragments, we needed an internal standard in our system to account for sampling and measurement issues. To this extent, we first used DMBA (di-methoxy-benzoic acid), but later needed a molecule with longer elution time and used to a protected amino acid, Fmoc-serine-trt-OH.

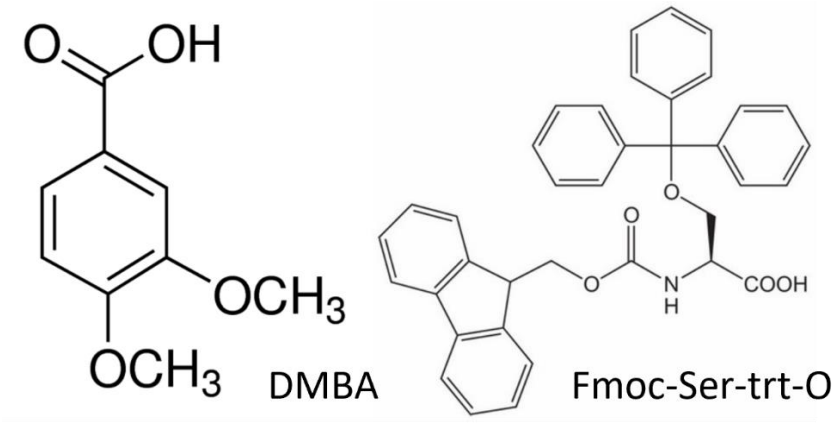


Figure 2-8 UV standards used throughout this study

Both TCEP and DTT were used in our first assay to determine a possible impact of the reducing agent on the exchange reaction. In previous works, concentrations of DTT of 25 mMol, and 100 mMol for TCEP were used for millimolar range exchange reactions. These were therefore going to be the starting concentrations for our first kinetic assays. Finally, TCEP can induce the desulfuration of cysteines[21]. It was shown that ascorbic acid could help protect cysteines from desulfuration and was thus added in concentrations twice those of TCEP to the reaction media[22].

b. N-methyl-cysteine synthesis

To synthesize N-methyl-cysteine we must methylate the amine of a cysteine. This can be done through sodium hydride and iodomethane, but this is incompatible with an unprotected thiol as well as with a free amine as this would lead to N-di-methylated and S-methylated products. We thus used a cysteine with a trityl protected sulfur. To synthesize our peptide fragments for the exchange reactions, we use solid-phase-peptide synthesis. This method works best with Fmoc N-protected amino acids as Boc deprotection takes longer. Fmoc protection, however, is not compatible with the methylation of the nitrogen, whereas Boc is. To this end, we started with Boc protected cysteine for the methylation then removed the Boc protecting group, which also removes the tri-phenyl as both have the same deprotecting conditions, then the thiol was re-protected, and the N-methyl moiety was Fmoc protected. This synthesis was previously described on small scale but in this work, the synthesis was performed with up to 15g of product and the purification methods were optimized, giving an overall yield of 59% (previously reported 47%)[2].

Note that another synthesis path for N-methyl-cysteine was reported, using the reductive amination of the S-trityl-cysteine by formaldehyde [23]. The following purification requires preparative HPLC methods which is incompatible with our gramme scale requirements.

The Four step synthesis scheme for N-methyl-cysteine is shown in figure 2-9 and the protocols can be found in the experimental section.

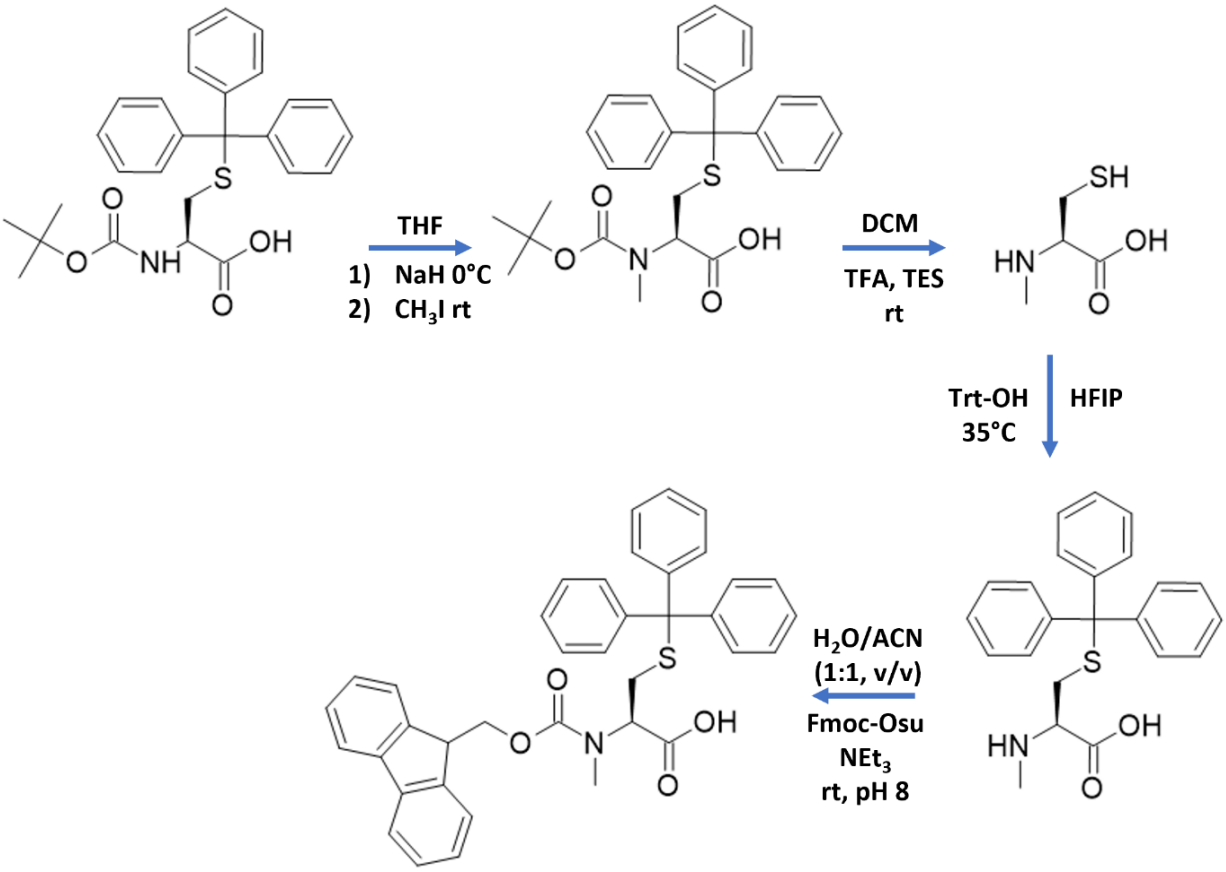


Figure 2-9 Scheme representing the four steps for the synthesis of N-methyl-cysteine.

c. Solid Phase Peptide Synthesis

All the peptides discussed in this work have been synthesized via Solid Phase Peptide Synthesis. SPPS was first designed and set up in 1963[24]. The concept rest on the immobilization of the synthesized peptide sequence on a polymer resin. The resin is functionalized so that an amino acid can react on it through its carboxylic acid moiety. The added amino acid must be protected on its amine and side chain to prevent formation of dipeptides or side chain reactions. In these conditions the only possible reaction will be the amino acid to be grafted on to the resin. The resin is then washed, the α-amine protecting group is then removed and the next protected amino acid in the sequence starting from the C-terminus is added and so on and so forth until the desired sequence is obtained. The main advantage of SPPS is that since the sequence is grafted on the solid, in between each addition and deprotection step, all that is needed is for the resin to be washed, therefore saving on purification time and reagents. Back in the time, SPPS was first developed to use Boc protecting groups for the amine groups of the amino acids which resulted in harsh conditions as Boc cleavage was performed by trifluoroacetic acid (TFA) and the final cleavage from the polystyrene resin was performed by hydrogen fluoride.[25] This led to problematic interactions with amino acids such as tryptophan whose indole ring has poor stability in acidic media[26].

Nowadays, SPPS is performed with the use of Fmoc protecting groups on amino acids α-amines and trityl and tertiobutyl protection on the side chains as this allows for orthogonality in the deprotection. Fmoc protecting groups are removed by weak bases during the synthesis of the sequence, which once finished the side chains protecting groups can be removed as the peptide is cleaved from the resin through TFA.[25]

SPPS can be broken down into four steps. First the deprotection of the resin or grafted amino acid on the resin. In this work, we used H-Rink amide resin. This resin consists of 4-[(2,4-Dimethoxyphenyl)(amino)methyl]phenoxyacetic acid bound to amine functions on the polystyrene beads. The protecting group of the resin or the Fmoc of the amino acid is removed by a base, piperazine. Once the deprotection is complete and the resin is washed to eliminate all traces of protecting groups and more importantly of piperazine that would deprotect the amino acid introduced in the next step and form oligo amino acids. As with most reagents for peptide synthesis this is done with DMF (N,N-Dimethylformamide) which solubilizes well all compounds while being compatible with the required chemistry.[25]

Then, the next amino acid in the sequence starting from the C-terminus, with a protected side chain and α-amine is added in solution. To speed up the reaction between the newly added amino acid two activators are also added in reaction media. A base activator is required to deprotonate the carboxylic group which can then react with the activator acid to exalt the electrophilic properties of the carboxyl carbon of the protected amino acid. In our case we used HBTU 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate as acid activator, which forms an activated leaving group, facilitating the attack of the nucleophilic amine of the grafted amino acid[27]. As a base we used DIEA (di isopropyl-ethyl-amine) which is a very sterically hindered tertiary amine and as such, a good proton acceptor but unlikely to react with the carboxyl. These additives also serve to prevent racemization of the amino acids during the coupling step.

This coupling step can also be associated to microwave irradiation (as was done in our case). This has been shown to tremendously increase the speed and yield of couplings while not showing any increase in racemization of the amino acids[28]. The temperature increase helps with the kinetics of the reaction, the diffusion of the reactants inside the swollen resin matrix. And the microwave irradiation favors the unfurling of peptide chains by dipole rotation, facilitating the access of the N-terminus for reactivity[29].

Finally, the resin is washed again to remove all the free amino acid in prevision for the next deprotecting step. An extra deprotection step is required at the end of the synthesis of a sequence to deprotect the last added amino acid since Fmoc deprotection is orthogonal to the side chain deprotection performed during the cleavage from the resin.

The protocols for the preparation of all the solutions for SPPS and for the synthesis of each peptide can be found in the experimental section.

4. YGC and CGGY assay

a. TCEP vs DTT

A kinetic assay on YG**C** and **C**GGY was performed at four different temperatures: 20°C (room temperature), 40°C, 60°C and 80°C. In the first one DTT was used as the reducing agent in concentrations of 25 mMol in 50 mMol PBS (phosphate buffer saline commercial solution) at pH 7. Figure 2-11 summarizes what we learned from these experiments by representing the evolution for 20°C and 80°C.

The chromatogram showed for 80°C after 9 hours is representative of the fate of all exchange experiments regardless of temperature. Immediately after the addition of the peptides to the reaction media, we see the apparition of a compound at 0.51 elution time. As discussed later, we attributed this peak to a DTT-NMC disulfide form. At high temperatures such as at 80°C, only traces of the mass of the exchange product YG**C**GGY are observed, however at lower temperatures the concentrations of the exchange product are high enough to obtain a UV signal. As showed by the 80°C, 9-hour chromatogram, the reaction ends with no starting material (the lower temperatures still show some **C**GGY signal), high DTT-NMC signal and a compound with 0.32 elution time which is formed faster with temperature.

The DTT-NMC peak indicates that the reverse reaction of the exchange cannot take place as the NMC released from the exchange reaction is captured by DTT, therefore rendering these measurements unsuitable for proper kinetic data on the dynamics of the reaction. The 0.32 peak was attributed to a diketopiperazine (DKP) compound as discussed below, explaining the disappearance of the YG**C** compound. Keeping high temperature experiments running proved challenging. The Eppendorf tubes in which the reactions were ran, where kept at temperature in heating blocks. However even with aluminum foil covering the top, water still evaporated and condensed on the cap, resulting in dried out, non-soluble solids at the bottom of the tubes overnight. Since only traces of the exchange products were observed at 80°C, we stopped using this temperature in our next experiments.

The kinetic assay performed with TCEP was only done at 20°C and 40°C with 50 mM PBS at pH 7 and using 100 mMol of TCEP and 200 mMol of sodium ascorbate for the reducing agents. Again, we see the appearance of the diketopiperazine compound at 0.32 minutes of elution. This means that this reaction is not linked to the presence of DTT or its role in the exchange reaction mechanism discussed previously (i.e., the thioester formed with DTT is not responsible for the DKP formation). The exchange reaction takes place in TCEP as well and the exchange product (YG**C**GGY) is in higher concentration at lower temperature. This might be since temperature accelerates the DKP formation, reducing the amount of YG**C** available for the exchange reaction or because the exchange product can also convert into the DKP compound.

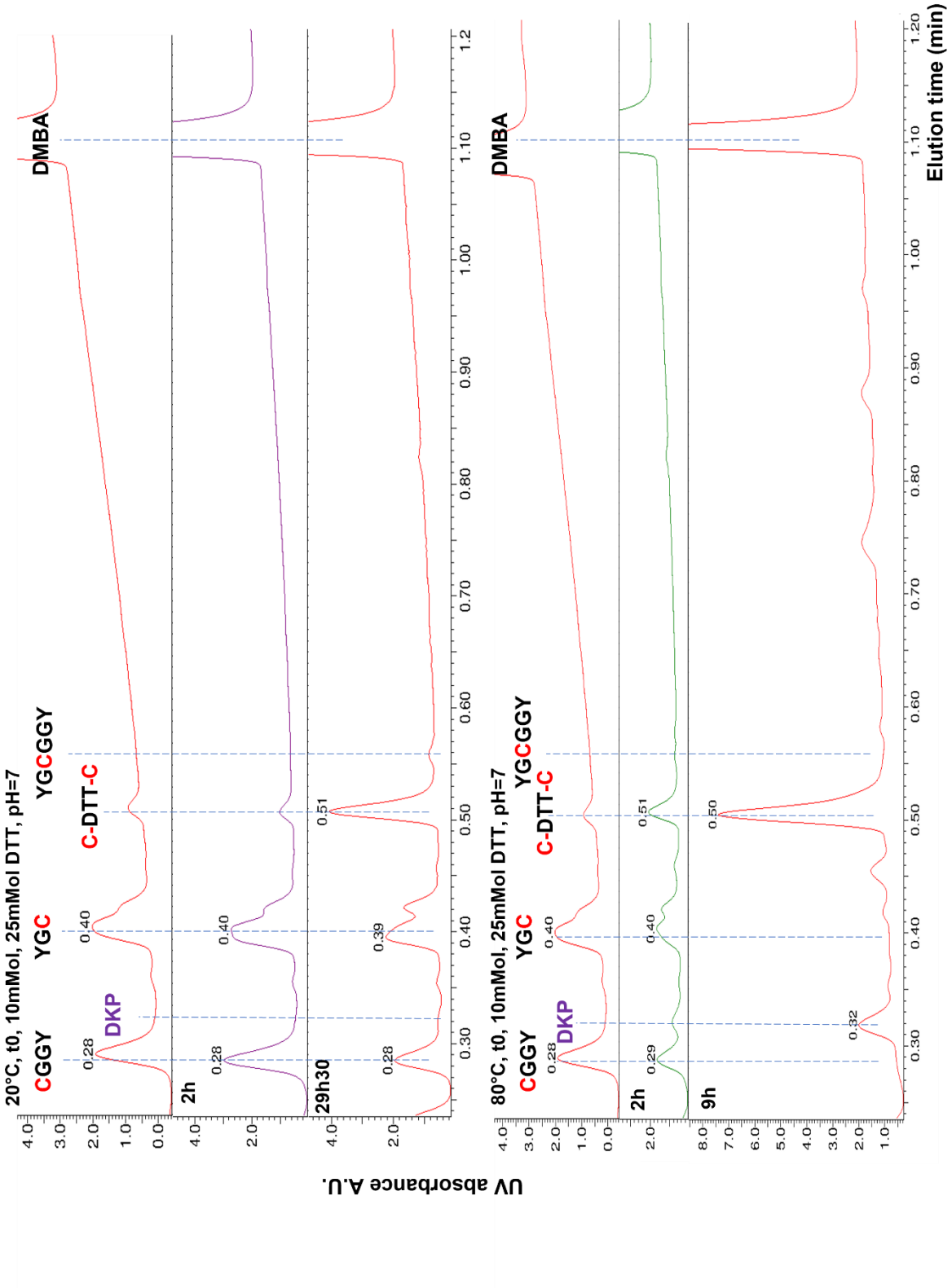


Figure 2-11 Chromatograms summarizing the evolution of the YGC CGGY system at room temperature and at 40°C with DTT as a reducing agent.

TCEP however, leads to cleaner chromatograms and does not seem to interact with NMC like DTT seems to, meaning that apart from the DKP formation, there seems to be no side reactions. This is primarily the reason for which we decided to use TCEP as our reducing agent for further experiments.

b. Side products

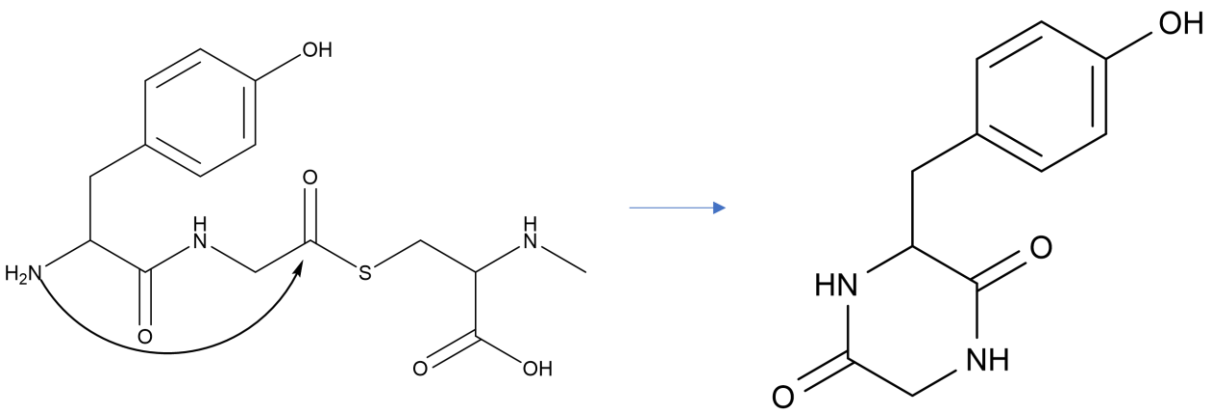


Figure 2-12 Proposed mechanism for DKP formation

The side product with 0.32 min elution time has a mass of 221 according to the mass spectrometer. It is possible to generate a compound with such mass through intramolecular cyclisation with our starting peptide YG**C**. Diketopiperazine compounds are common with short peptides [30] and are very stable meaning that all the DKP formed permanently removes reactant for the exchange reaction. In our case, this reaction is increased by the presence of the N-methyl-cysteine. When the bond between the tyrosine and the NMC is in the thioester form, it is more reactive, therefore facilitating the nucleophilic attack from the amine of the tyrosine. The exchange product YG**C**GGY can also undergo this intramolecular reaction as the internal cysteine also undergoes thioester formation. However, non-terminal cysteines react more slowly, allowing the concentration of the exchange product to increase making it observable.

When looking at the mass spectra of the compound with 0.51 min elution time in the DTT experiments, we see mass signals of 422.5 and 444.5 Da. Due to the isotopic fractioning we know that these are single charged ions. The 22 Da difference is commonly due to the association of sodium when the compound goes through the electrospray meaning that the mass of the compound is likely to be 421.5 Da. There is a mass signal at 135.5 Da which corresponds to the mass of NMC. This means that NMC is generated when ionizing this

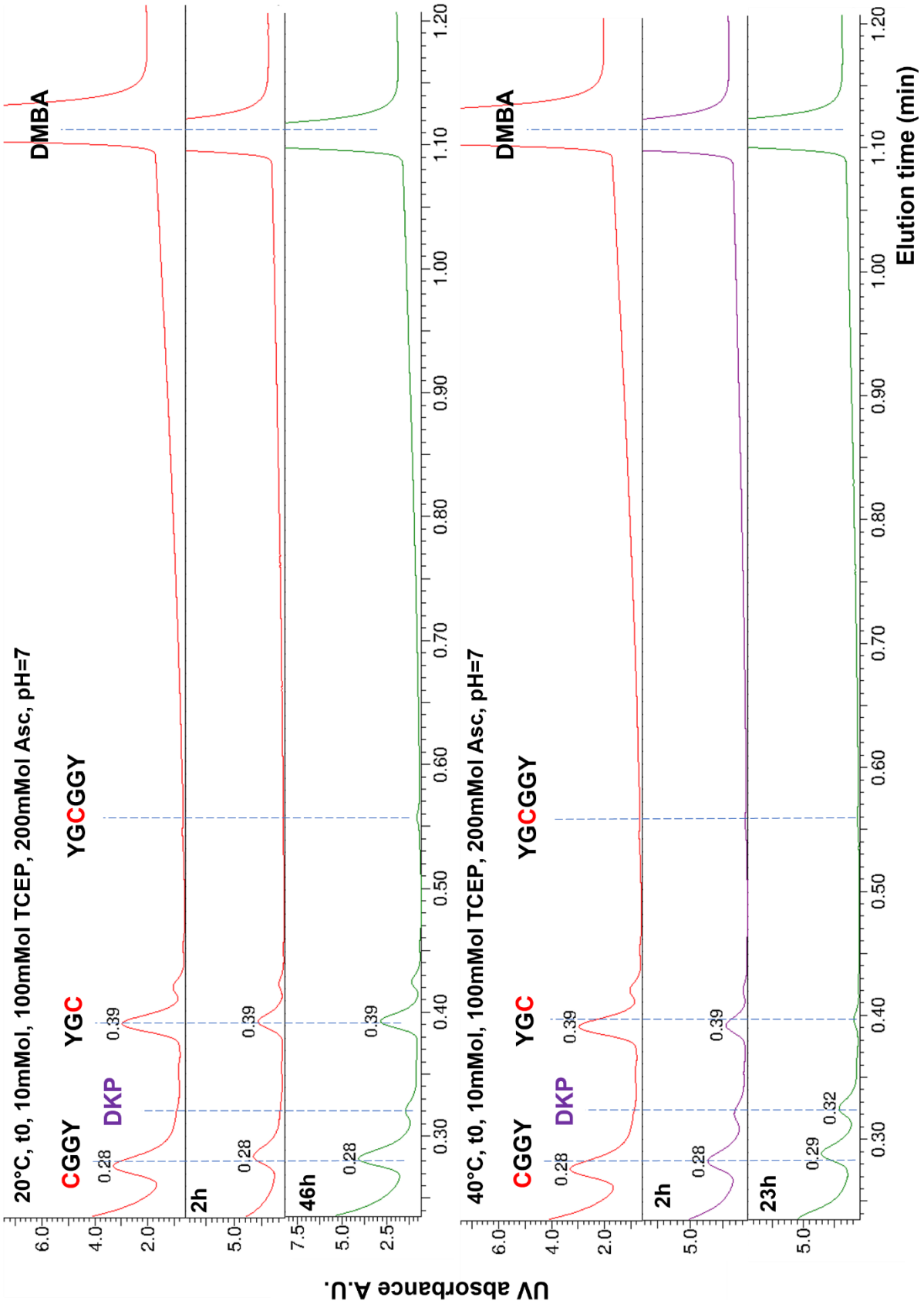


Figure 2-13 Chromatograms summarizing the evolution of the YGC CGGY system at room temperature and at 40°C with TCEP as a reducing agent.

molecule and could be generated by its fragmentation. The UV spectra for this peak is different from tyrosine absorption. The maximum is shifted to 283 nm (275 nm for tyrosine) which is consistent with oxidized DTT UV absorption. We therefore concluded to the association of NMC with DTT through disulfide bonds as the solution oxidizes. However, when generating the structure for this compound, the calculated mass is 420.5 Da which differs by 1 Da from the neutral supposed mass of 421.5 Da for this compound, so this structure might be incorrect even if it fits the rest of the information. Nevertheless, since it only appeared when using DTT, it comforted us in our decision to only use TCEP.

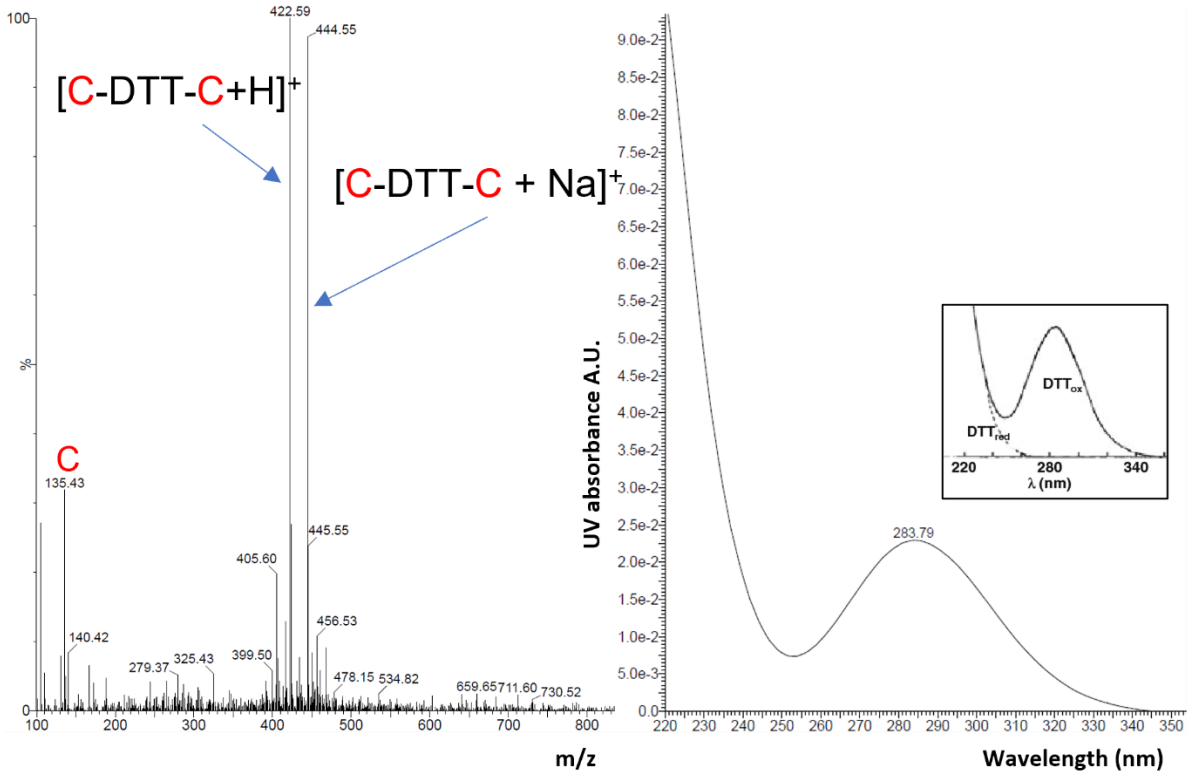


Figure 2-14 Mass spectra and UV spectra for the compound with 0.51 min elution time in the DTT kinetic assay. The UV mass for DTT is from [31]

5. Back to the drawing board

From the first series of experiments, we drew several conclusions. TCEP seemed to be the best reducing agent for the experiments. The YG**C** peptide must be modified to avoid the DKP formation. We increased the length of the sequence by one amino acid reasoning that this might bring further apart the N-terminal amine from the created thioester bond. The DKP compound UV peak appeared right in between the two peptides and in the DTT experiments overlapped with **C**GGY. To decrease the likelihood of having side products overlapping we wanted to have bigger differences in the properties of our peptides, thus leading to different elution times. Since YG**C** was getting longer we decided to add a phenylalanine to the other sequence, this would make it more hydrophobic and increase the elution in the reversed-phase chromatography while keeping the sequence short to avoid secondary structures arising from peptide backbone interactions. When synthesizing the peptide sequence, coupling a phenylalanine to a tyrosine leads to a decrease in yield due to steric hindrance, we therefore interspersed a glycine in between them. The new peptide fragments for our kinetic assay where thus:

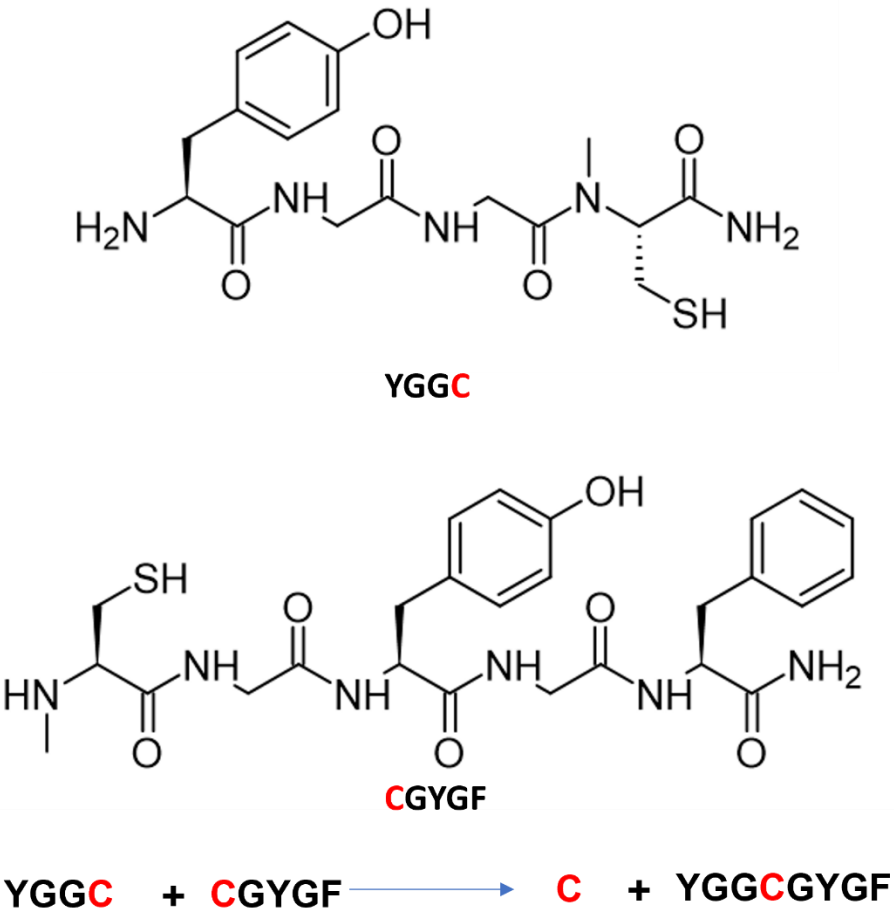


Figure 2-15 Structure of the new peptide fragments, YGGC, CGYGF and their exchange reaction.

6. YGGC and CGYGF assay

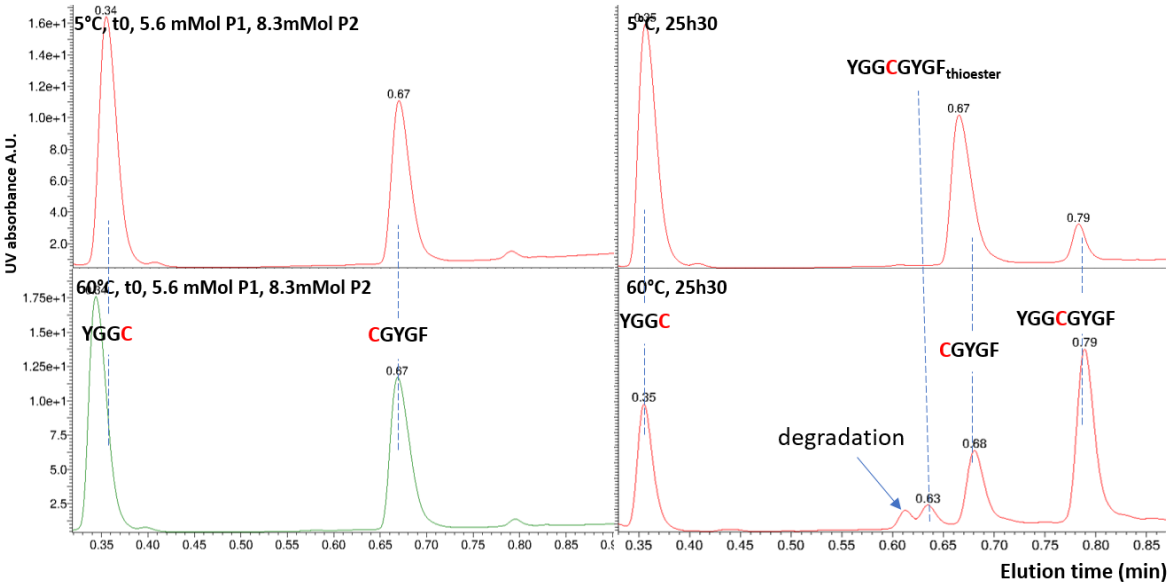


Figure 2-16 Chromatograms summarizing the evolution of YGGC CGYGF system at 5°C and 60°C in TCEP.

After synthesis of the peptide fragments, we performed the kinetic assay between YGG**C** and **C**GYGF in 50 mMol PBS at pH 7 with 100 mMol TCEP and 200 mMol Asc. Chromatograms summarizing the results are shown in figure 2-16. We see the exchange reaction proceed with the increase of the 0.79 min elution time peak, which is already present a few seconds after the reactants are put together indicating a high initial speed. The speed of the reaction is increased by temperature as expected given the difference between peak sizes for YGG**C**GYGF at 60°C and 5°C. We do still have the appearance of side products. The one at 0.63 min elution only appears at 60°C and after 9 hours of reaction time. It seems to only appear at 60°C, only being present in trace amount at 40°C after 50 hours of reaction. The other one at 0.61 elution time eventually appears at all temperatures.

The mass and UV spectra for the high temperature one (0.61 min elution time) can be found on figure 2-17. The isotopic ratio allows us to conclude that the mass of this compound is 1051 Da. If we consider the mass of the exchange product and TCEP we have a mass surplus of 33 Da which could correspond to two hydroxyls or to a thiol. Since the UV spectra of the compound is the same as the one for tyrosine, we suppose that none of the phenol moieties were modified. The exchange product could have linked to TCEP through amidation from the N-terminal amine and the free amine from NMC when the bond is under the thioester form. However, this would probably lead to cleavage of the molecule given the reactivity of thioesters. We therefore propose the structure **a** for this compound, where the exchange product was desulfurized and the sulfur is replaced by TCEP. A similar mechanism was proposed by the group of Professor Li.[32] This compound would also lead through fragmentation during the mass spectrometry to the 800 Da mass signal in the negative electrospray spectra by departure of the TCEP.

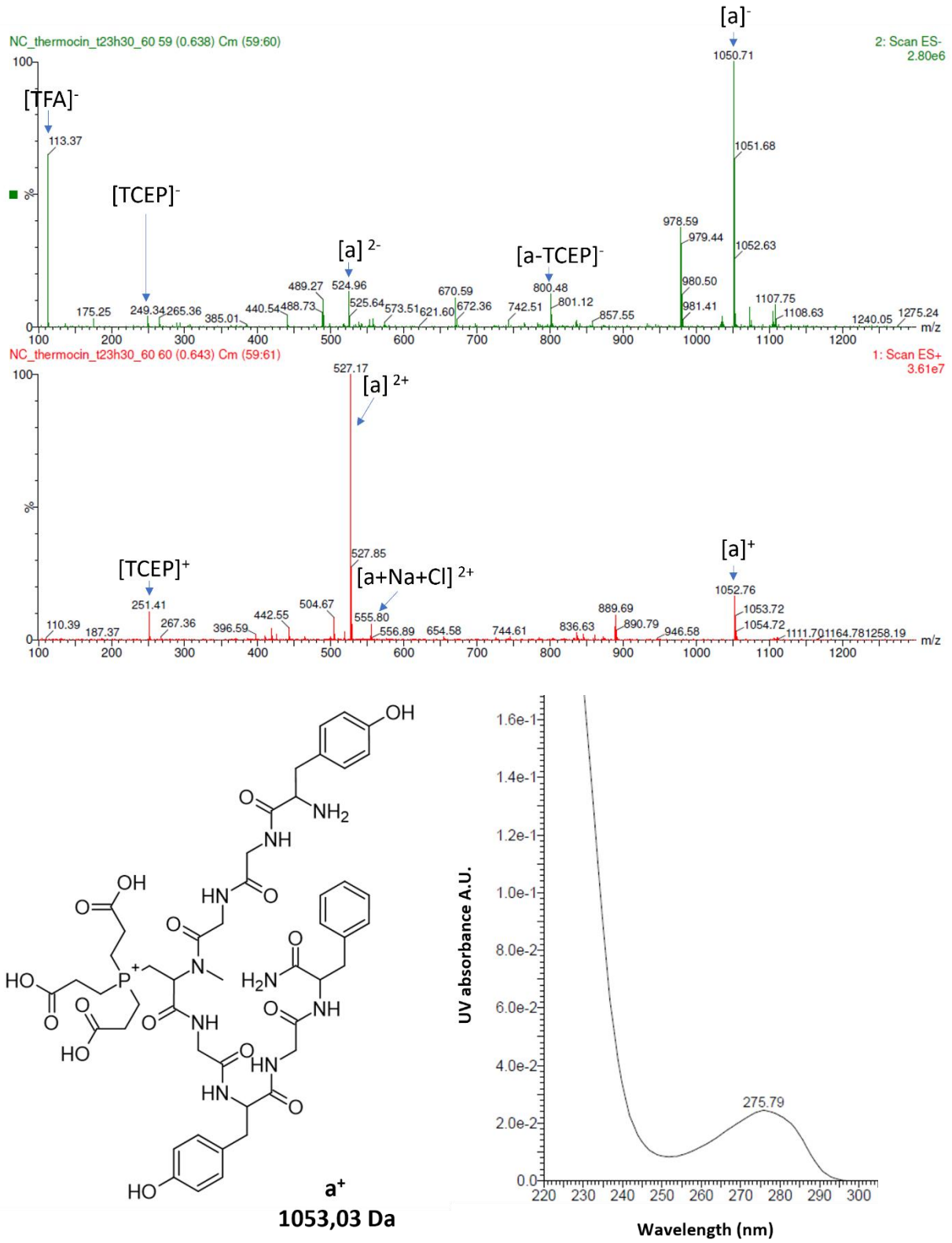


Figure 2-17 Mass spectra from the positive electrospray (ES+) and negative (ES-) for the 0.61 min elution time mass signal. Underneath the proposed structures for the different signals and the UV spectra measure at 0.61 min elution.

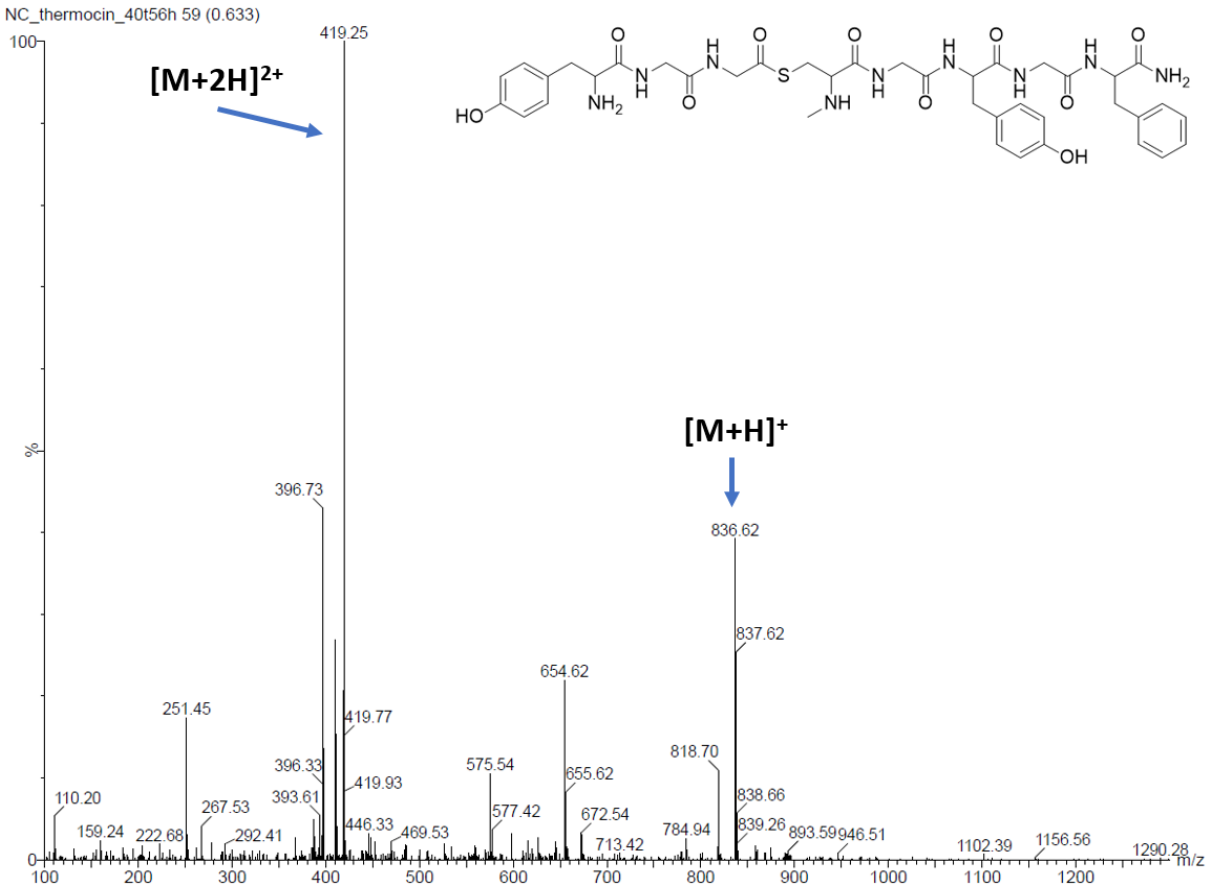


Figure 2-18 Mass spectra for the compound with 0.63 min elution time. Due to its mass, it was attributed to the thioester form of the exchange compound.

The other compound appears at all temperatures and has the same mass as the exchange product. We thus attributed it to the thioester form of the exchange product. This thioester could either be more stable than the ones for the initial peptide fragments, or the presence of the two tyrosine enable a high enough absorbance for it to be seen, explaining why it is only observed for YGGCGYGF.

a. Kinetics of a bimolecular reversible reaction

The UPLC spectra of the reaction never revealed any reaction intermediates except in the case of DTT usage as a reducing agent. This may indicate that the rate limiting state is the N→S acyl-shift and that when formed the thioester directly undergoes transthioesterification or reacts back to the peptide bond state. Based on this hypothesis we decided to consider this reaction as a simple reversible bimolecular reaction. If we consider a general bimolecular reversible reaction, the system is the following:

$$A + B \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C + D$$

The reactants, A and B, react together generating the products, C and D, with a kinetic constant k_1 for the forward reaction, and k_{-1} for the backward one. This system yields the following differential equation:

$$\frac{dA}{dt} = \frac{dB}{dt} = -\frac{dC}{dt} = -\frac{dD}{dt} = -k_1 AB + k_{-1} CD$$

With A, B, C and D representing concentration of the respective species. If we consider the variable E, defined as the distance to equilibrium in concentration:

$$E = A - A_e = B - B_e = C_e - C = D_e - D$$

With A_e, B_e, C_e and D_e the concentrations at equilibrium for the respective species. This means that E also obeys a differential equation:

$$\frac{dE}{dt} = \frac{dA}{dt} = -k_1 AB + k_{-1} CD = -k_1 (E + A_e)(E + B_e) + k_{-1}(C_e - E)(D_e - E)$$

$$\frac{dE}{dt} = -k_1 A_e B_e + k_{-1} C_e D_e - [k_1 (B_e + A_e) + k_{-1}(C_e + D_e)]E + (k_{-1} - k_1)E^2$$

By using the law of mass action, we know the following relationship between the equilibrium concentrations:

$$\frac{C_e D_e}{A_e B_e} = K = \frac{k_1}{k_{-1}}$$

Which means that:

$$k_{-1} C_e D_e = k_1 A_e B_e$$

The differential equation can then be simplified to:

$$\frac{dE}{dt} = -[k_1 (B_e + A_e) + k_{-1}(C_e + D_e)]E + (k_{-1} - k_1)E^2$$

This is a quadratic differential equation of the form $\frac{dy}{dt} = py + qy^2$ with p and q being constants, which can be integrated under the following form:

$$\int_{y_0}^{y} \frac{dy}{py + qy^2} = \int_0^t dt$$

Yielding the solution:

$$y = \frac{py_0}{(p + qy_0)e^{-pt} - qy_0}$$

Going back to our reaction system, we thus have the following solution:

$$E = \frac{-[k_1(A_e + B_e) + k_{-1}(C_e + D_e)]E_0}{[k_{-1}(C_e + D_e) - k_1(A_e + B_e) + (k_{-1} - k_1)E_0]e^{-(k_1(A_e+B_e)+k_{-1}(C_e+D_e))t} - (k_{-1} - k_1)E_0}$$

With E_0 the value of E at the beginning of the reaction. In our case, the reaction starts with only the reactants present. By excluding degradation and side product formation, this means that $C = D$, $E = C_e = D_e$ and $C_0 = D_0 = 0$ meaning that $E_0 = C_e = D_e$ which in turns means $A_e = A_0 - C_e$ and $B_e = B_0 - C_e$. This allows us to write E with only three unknowns: $C_e,\ k_1,\ k_{-1}$.

$$E = \frac{-[k_1(A_0 + B_0 - 2C_e) + 2k_{-1}C_e]C_e}{[2k_{-1}C_e - k_1(A_0 + B_0 - 2C_e) + (k_{-1} - k_1)C_e]e^{-(k_1(A_0+B_0-2C_e)+2k_{-1}C_e)t} - (k_{-1} - k_1)C_e}$$

C_e can further be expressed from the known initial concentrations and the kinetic constants. Indeed, we can rewrite the law of mass action as:

$$\frac{C_e^2}{(A_0 - C_e)(B_0 - C_e)} = \frac{k_1}{k_{-1}}$$

This gives us the second-degree equation in C_e:

$$C_e^2\left(1 - \frac{k_{-1}}{k_1}\right) + (A_0 + B_0)C_e - A_0B_0 = 0$$

the determinant is: $\Delta = (A_0 + B_0)^2 + 4\left(1 - \frac{k_{-1}}{k_1}\right)A_0B_0$

Yielding 2 different solutions for C_e when $\Delta > 0$:

$$C_{e,1} = \frac{-(A_0 + B_0) + \sqrt{\Delta}}{2\left(1 - \frac{k_{-1}}{k_1}\right)} \quad \text{and} \quad C_{e,2} = \frac{-(A_0 + B_0) - \sqrt{\Delta}}{2\left(1 - \frac{k_{-1}}{k_1}\right)}$$

A MATLAB routine was then implemented, which takes as input the experimental concentrations of reactants and products at different time steps. The routine scans different values for the couple (k_1, k_{-1}) and uses the initial reactant concentration to compute what equilibrium concentration C_e that couple yields (non-possible answers such as negative concentrations, or concentrations higher than starting ones, are skipped at this step to increase computation efficiency) and afterwards calculates the values of E at the given experimental timesteps.

For each (k_1, k_{-1}), an error between the experimental and theoretical data is then calculated for each experimental measure and the errors are summed across all points for the reactants and the products (the norms of the differences are summed and used as the error metric). This allows the construction of a 3-dimensional map of the discrepancies for different (k_1, k_{-1}) couples. The minimum of the surface corresponds to the values that fit the experimental data the best. Figure gives an example for the exchange reaction in PBS at 40°C.

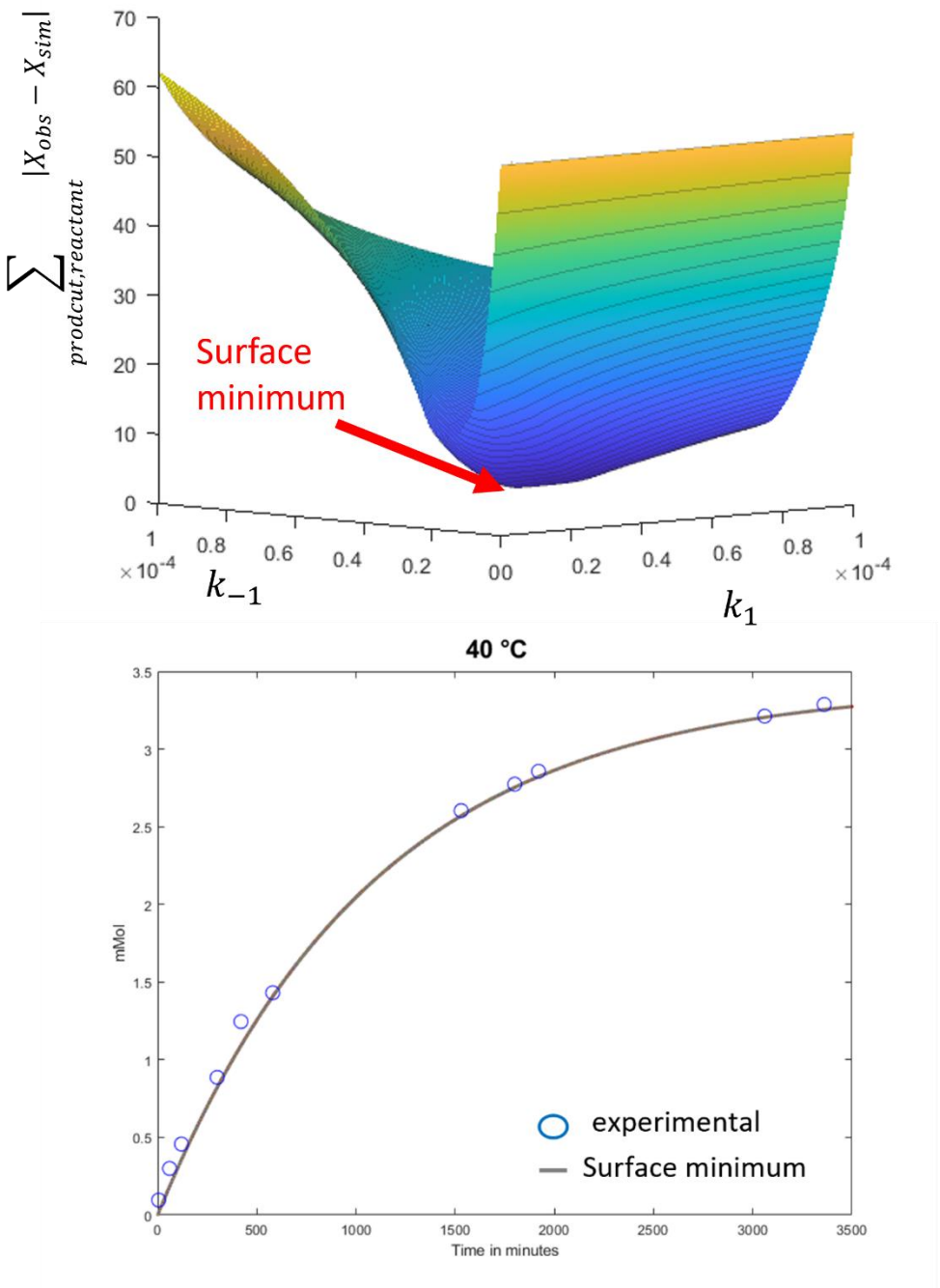


Figure 2-19 Results of the 3D map given for the error for each k_1, k_{-1} guess. The minimum is shown and the plot of the theoretical concentration for this couple against the experimental data is shown underneath.

b. Influence of temperature

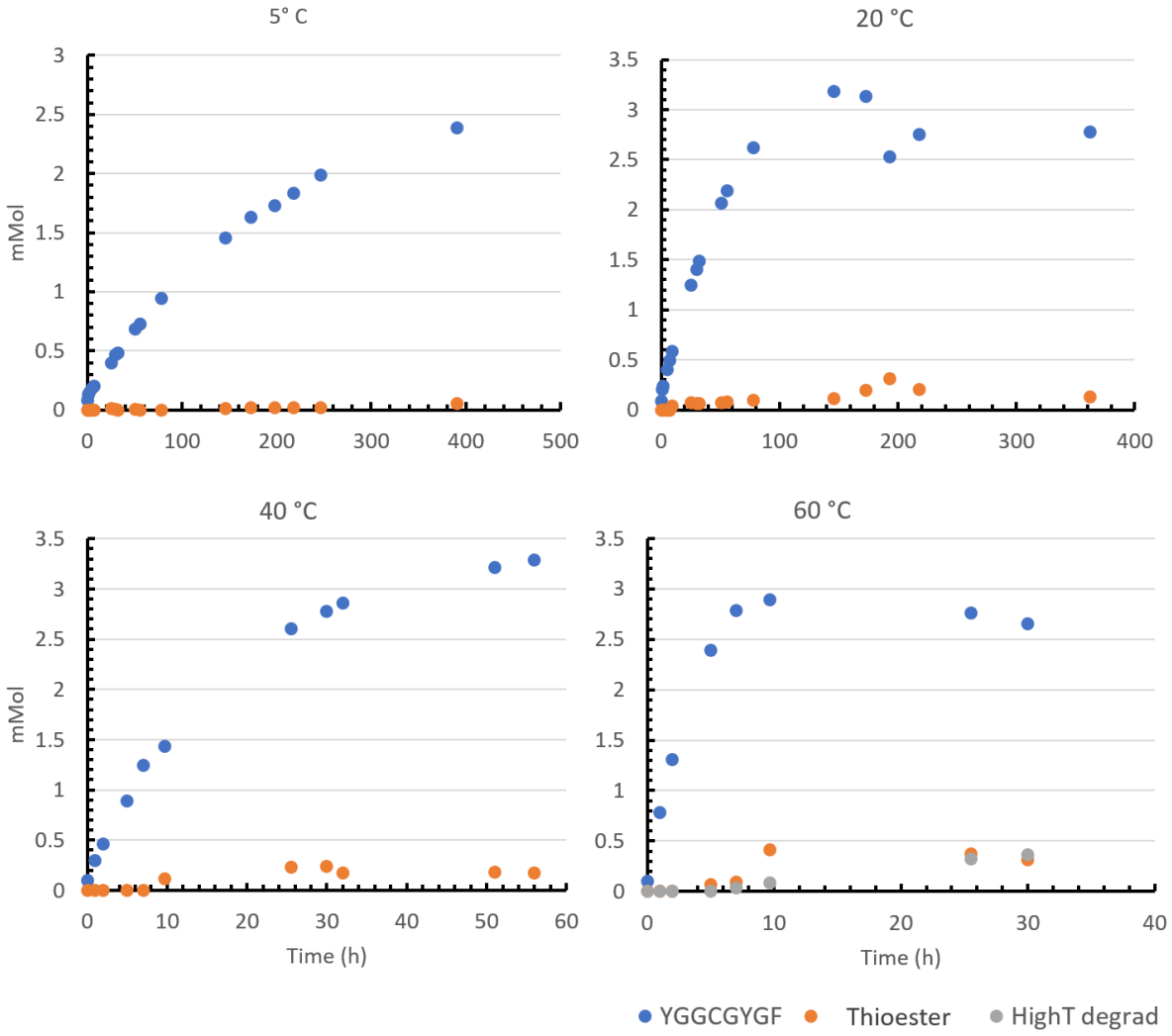


*Figure 2-19 Measured concentrations of exchange product as well as degradation for the kinetic assay of the exchange reaction between YGGC and CGYGF. Initial concentrations were 8.3 mMol YGG**C** and 5.6 mMol **C**GYGF. The reaction was performed at pH 7 with 100mMol PBS, 100 mMol TCEP and 200 mMol Asc*

In this temperature experiment, we can see that the initial kinetics are well defined. However, when the reaction reaches equilibrium, the measurements become less accurate for 20°C as this is correlated with long reaction times, higher side product content and the reaction media is no longer in complete reducing conditions ultimately resulting in the appearance of oxidized forms as precipitate in the Eppendorf tubes, affecting the equilibrium. Notably at 5°C, the reaction media became oxidized before the equilibria could be reached. In the case of 60°C we do see that the exchange product starts decreasing as the high temperature side product (named HighT degrad) increases in concentration, comforting us in our structure hypothesis.

The kinetic data extracted from these measurements is summarized in Table 1. Because the side products appear in non negligeable concentrations at 60°C after 9 hours, only the first 9 hours of measures were used in the kinetic model to obtain the data.

Temperature (°C)	k_1 (L.mol^{-1}.s^{-1})	k_{-1} (L.mol^{-1}.s^{-1})	$t_{1/2}$ (hours)	K (k_1/ k_{-1})
5°C	8.43 10^{-5}	1.58 10^{-4}	117	0.53
20°C	3.67 10^{-4}	4.83 10^{-4}	39.6	0.76
40°C	1.15 10^{-3}	1.07 10^{-3}	13.1	1.08
60°C	4.83 10^{-3}	4.0 10^{-3}	2.8	1.21

*Table 1 Kinetic data calculated for the exchange reaction between YGG**C** and **C**GYGF in PBS at pH 7*

We do see a clear effect of temperature on the speed of the reaction. If we look at the time needed for each reaction to reach half of the maximum exchange product concentration measured, we observe a times four factor for every 20°C. It seems that the equilibrium shifts towards the exchange product as temperature increases between 20°C and 40°C, the kinetic data predicts that this behavior would also be the case for 60°C but with the side products interfering with the reaction we do not measure the equilibrium representative of only the exchange reaction.

With temperature increasing the amount of side products as well as favoring the reaction between TCEP and our exchange product, we stopped doing kinetic experiments at 60°C. However, this temperature does allow for correct exchange and going back to the gradient experiment setup, could be used as the high temperature extreme of the gradient.

Keeping the reaction media in reduced conditions appeared to be problematic for the kinetic experiments as the media oxidized faster with temperature but low temperature experiments take longer as well, inevitably, the kinetic measures had to be stopped as a precipitate formed, and the chromatograms were no longer relevant as almost all peptide signal was lost. We thus decided to investigate using a different buffer system from PBS, which was previously used in studies featuring this exchange reaction, as TCEP is less stable in phosphate buffers.

c. Influence of the buffer system

We wanted to use a non-phosphate buffer that would allow exploration of acidic and basic pH without need of composition modification. Citric acid is a common buffer in biochemistry, notably for reversed phase liquid chromatography of proteins. It is usually used for applications between pH 3 and 6.2. However, its last pKa of 6.4 does confer it some buffering power at pH 7. Ammonium on the other hand with a pKa of 9.2 offers some buffering power in the 8-10 pH range. We therefore settled on using commercially available di-ammonium citrate as our pH exploration buffer.

We were able to monitor the kinetics for the same duration as for PBS before precipitation would remove the peptide signal. The exception is for 10°C where we were able to follow the kinetics for up to 37 days. This reaction was sampled less frequently which means that the oxygen introduced in the Eppendorf when sampling seems to be the reason behind the oxidation of the system.

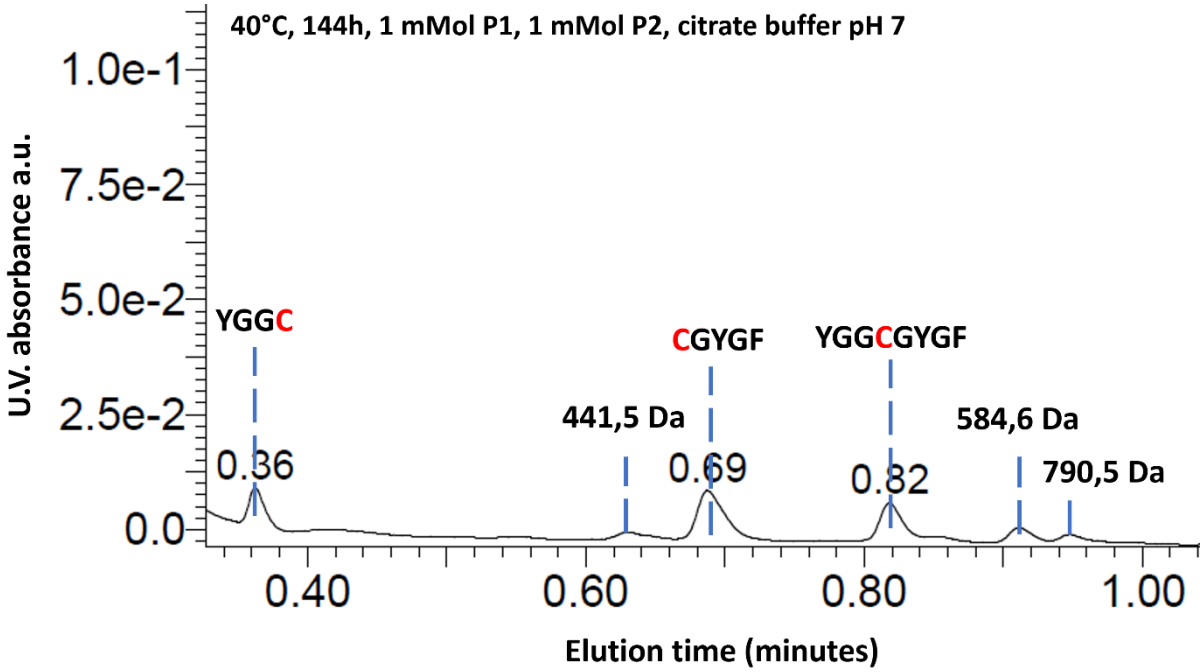


*Figure 2-20 Chromatogram for the exchange reaction performed at 40°C after 144h in 50 mMol citrate buffer at pH 7 with 100 mMol of TCEP and 200mMol of Asc. Starting cooncentrations of YGG**C** and **C**GYGF were 1 mMol.*

Since the reaction was not performed at 60°C we did not see the high temperature side product identified previously however the cleavage of the N-terminal NMC of **C**GYGF was observed again. Three new side products were observed. The one at 0.96 min elution time (figure 2-19) was attributed to the reaction between TCEP and **C**GYGF. Unfortunately, the concentrations are too low for a UV spectrum to be measured to conclude on a potential modification of the absorption of tyrosine, but the mass matches the proposed structure.

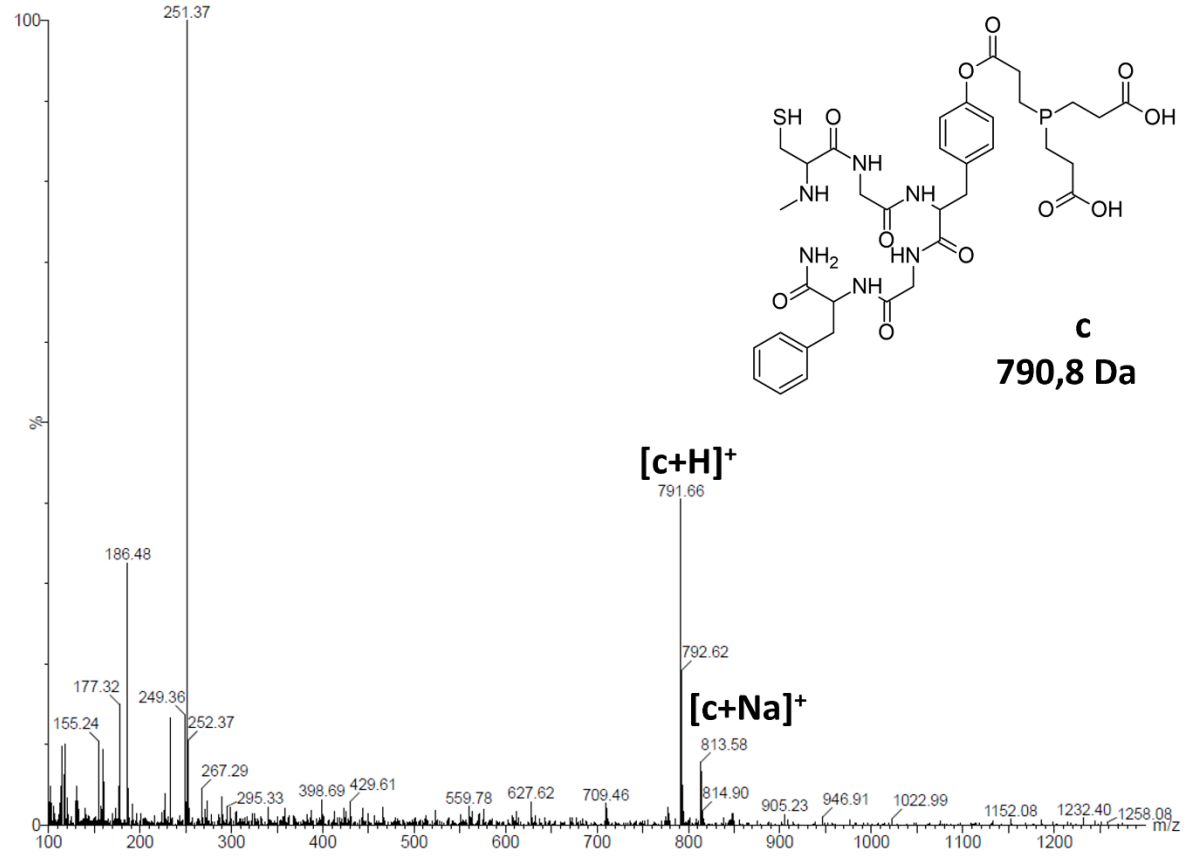


Figure 2-19 Mass signal obtained for the 0.96 min elution time product and the proposed structure.

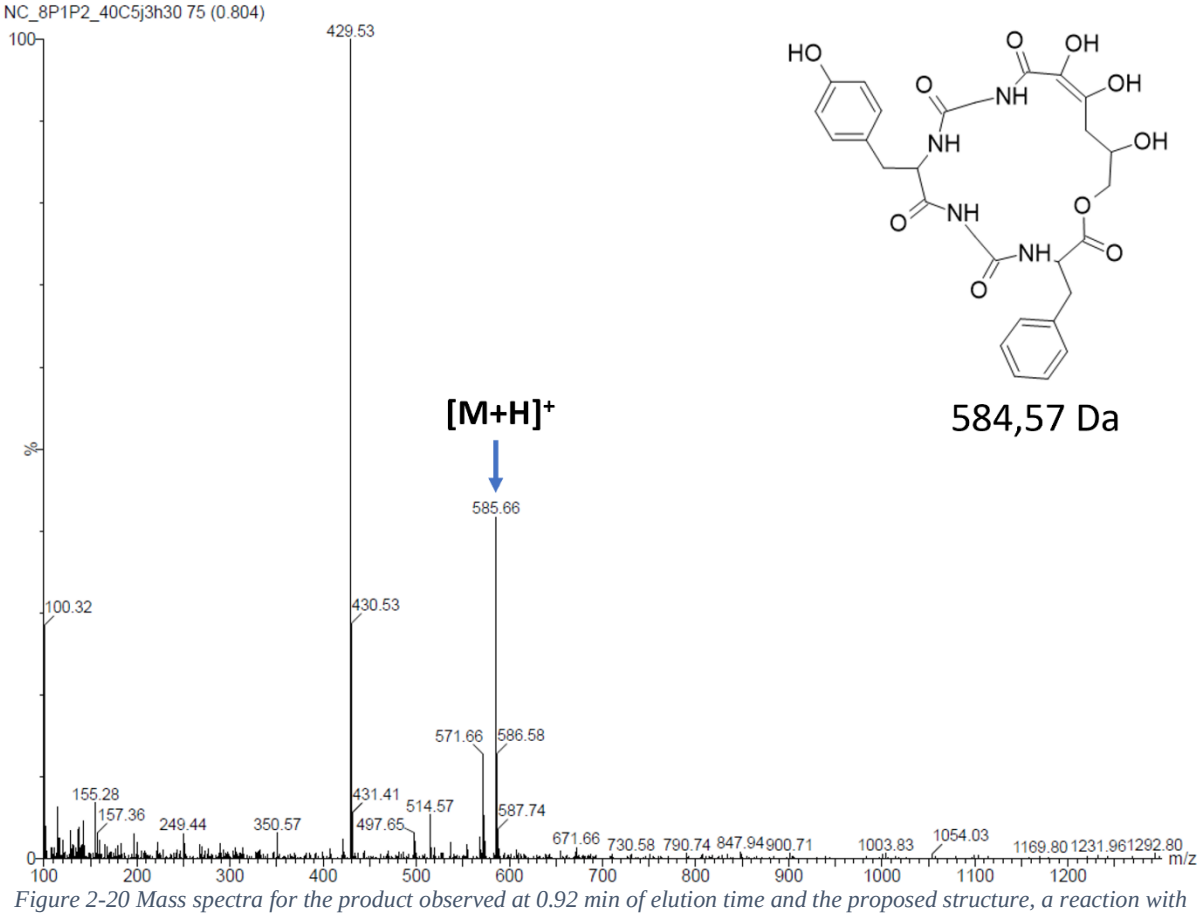


*Figure 2-20 Mass spectra for the product observed at 0.92 min of elution time and the proposed structure, a reaction with **C**GYGF with loss of the N-methyl-cysteine and esterification at the C-terminus and ring opening amination of the Asc.*

The other observed compound at 0.92 min elution time has either a mass of 428.5 Da or 584.6 Da due to the observed isotopic ratios. We could not explain a mass of 428.5 Da that would through interactions to yield a mass of 585 Da when ionized, however if **C**GYGF loses its N-terminal cysteine, reacts with Asc through esterification at its C-terminus and performs a ring opening on the ascorbate with the amine from the N-terminal glycine we obtain a cyclic compound of 584.57 Da.

We were unable to observe the thioester in citrate buffer but observed another side product in its stead. This side product only appears in the room temperature experiment after one week of reaction. We hypothesize that it results from the removal of the N-terminal cysteine of the CGYGF fragment. A side reaction that has been observed for cysteine containing peptides in solutions with TCEP.[33]

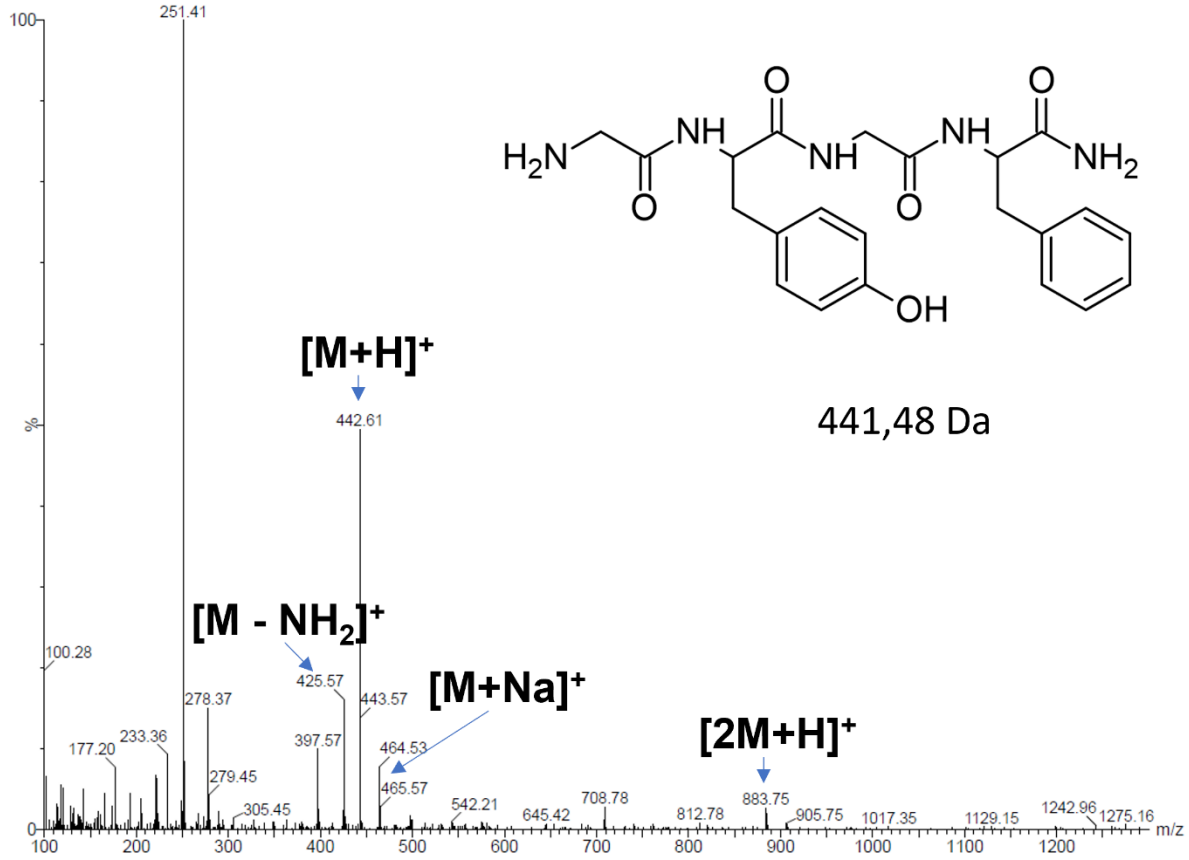


*Figure 2-21 Mass spectra for the product observed with 0.55 min elution time, the mass was attributed to the removal of the N-terminal N-methyl-cysteine from **C**GYGF according to literature reports.*

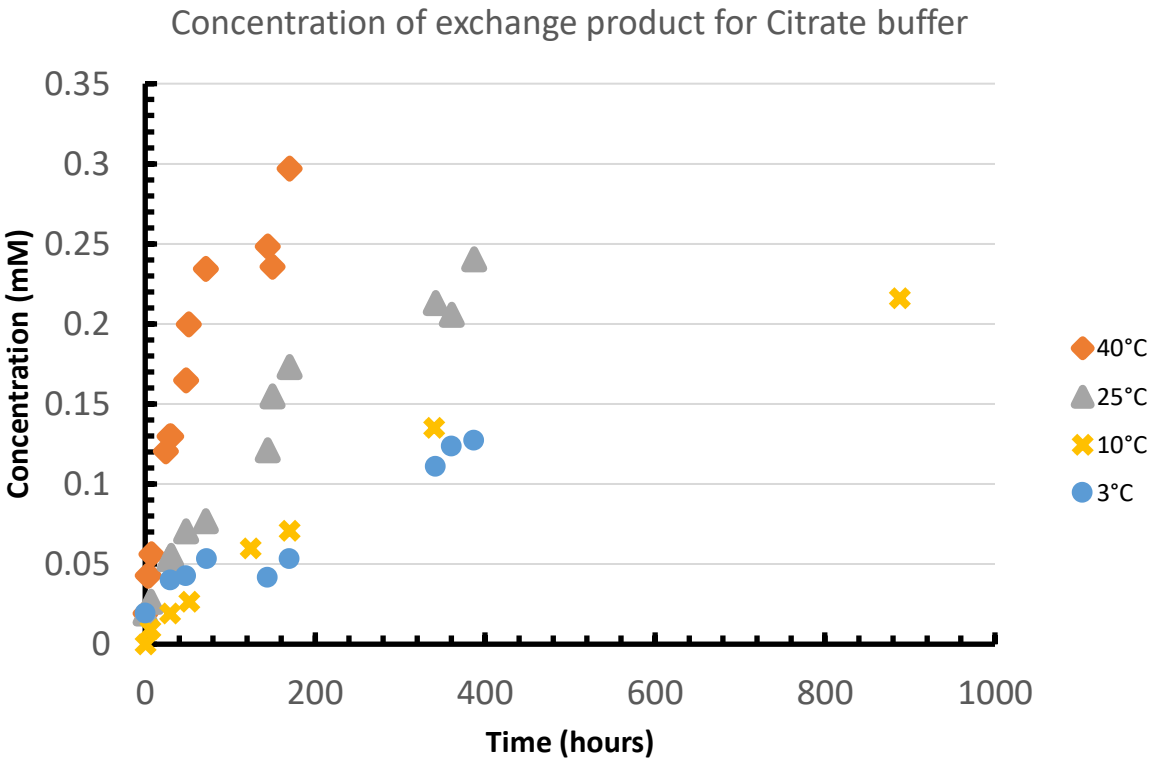


Figure 2-21 Concentration of exchange product YGGCGYGF measured for the YGGC CGYGF exchange reaction with initial concentrations of 1 mMol in 50 mMol di-amonium citrate buffer at pH7 with 100 mMol of TCEP and 200 mMol of Asc.

As can be seen on figure 2- 21, the reactions did not reach equilibrium. This is due to the increase in side-products, and on the decreased starting concentrations slowing the kinetics of the exchange. It does seem like the equilibrium shifts in favor of the exchange product as the temperature increases but this cannot be known for certain. However, this assay can still be used to determine the kinetic constants of the forward reaction of the exchange.

Temperature (°C)	k_1 $L.mol^{-1}.s^{-1}$
3°C	$1.19\ 10^{-4}$
10°C	$1.33\ 10^{-4}$
20°C	$3.77\ 10^{-4}$
40°C	$1.53\ 10^{-3}$

Table 2 Kinetic constants for exchange product formation in citrate buffer pH 7

The reaction seems to be faster at low temperatures, as the kinetic constant for 5°C is twice higher than in PBS. For 20°C and 40°C are the same. This could mean that the citrate has an effect at low temperature, since water is more structured close to 0°C, the buffer might modify it and impact the kinetics. This could also be linked to uncertainties in the measurement of the kinetic constants. But apart from at low temperatures, the buffer does not impact the kinetic of the forwards reaction.

d. Influence of pH on the system

To be able to design a system later on and have it potentially evolving at different pH, we wanted to document the effect that a pH change would have on the exchange reaction. To this end the exchange was performed at pH 6 and 8 in the citrate buffer system at 40°C.

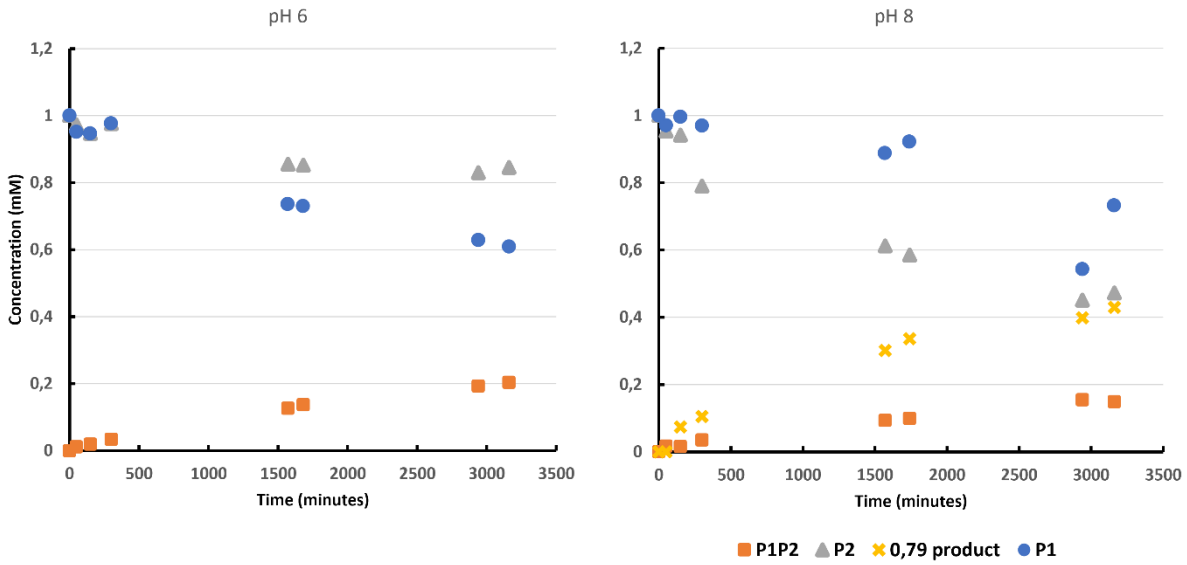


Figure 2-22 Exchange reaction between YGGC and CGYGF performed at pH 6 (left) and pH 8 (right) in the citrate buffer system with starting concentrations of 1 mMol in each peptide at 40°C.

As can be seen in figure 2-2, the exchange proceeds at both pH but side reactions take place. At acidic pH it is the hydrolysis of the thioester forming YGG**C** at the C-terminus in its thioester form which comes to perturb the reaction, the hydrolysis product is too hydrophilic and co-eluted with the buffer and salt peak so its concentration could not be monitored. At pH 8, the reaction is disturbed by the previously discussed 585 Da side product with 0.79 min elution time, whose formation from **C**GYGF seems favored at higher pH, whereas the hydrolysis of YGG**C** is almost negligible.

The 442 Da side product attributed to the departure of the N-terminus N-methyl-cysteine departure from **C**GYGF is absent at pH 6 but is in higher concentrations at pH 8, an effect consistent with the literature. The concentrations of this side product stay low compared to the rest of the species present.

Unfortunately, due to some technical problems with our UPLC the end of the kinetics could not be monitored

The side reactions affect the reaction too much at basic pH for the kinetics to be relevant to only our exchange reaction. At pH 6 we observed the similar kinetics as for pH 7 at 40°C and the modeled $k_1 = 1{,}03\ 10^{-4}\ L.mol^{-1}.s^{-1}$ is slightly lower than for pH 7. This is accord with previous works[22] which showed an increase in reaction speed and an optimum with pH around 8.

7. Thermodynamics of the exchange reaction

Going back to the definition of free energy given in the introduction, for a given process we have:

$$\Delta G = \Delta H - T\Delta S$$

Standard free energy is also linked to the equilibrium constants of chemical reactions through:

$$\Delta_r G^0 = -RT \ln K$$

Therefore, giving us a direct relation between the equilibrium constant and the standard reaction enthalpy and entropy.

$$-RT \ln K = -T\Delta_r S^0 + \Delta_r H^0$$

By plotting $-RT \ln K$ as a function of temperature we can extract the values for $\Delta_r S^0$ and $\Delta_r H^0$, which are respectively the slope of the plot and the intersection between the plot and the intersection with the ordinate axis. Using the equilibrium constants calculated for the exchange experiments performed in PBS at different temperatures, we obtain:

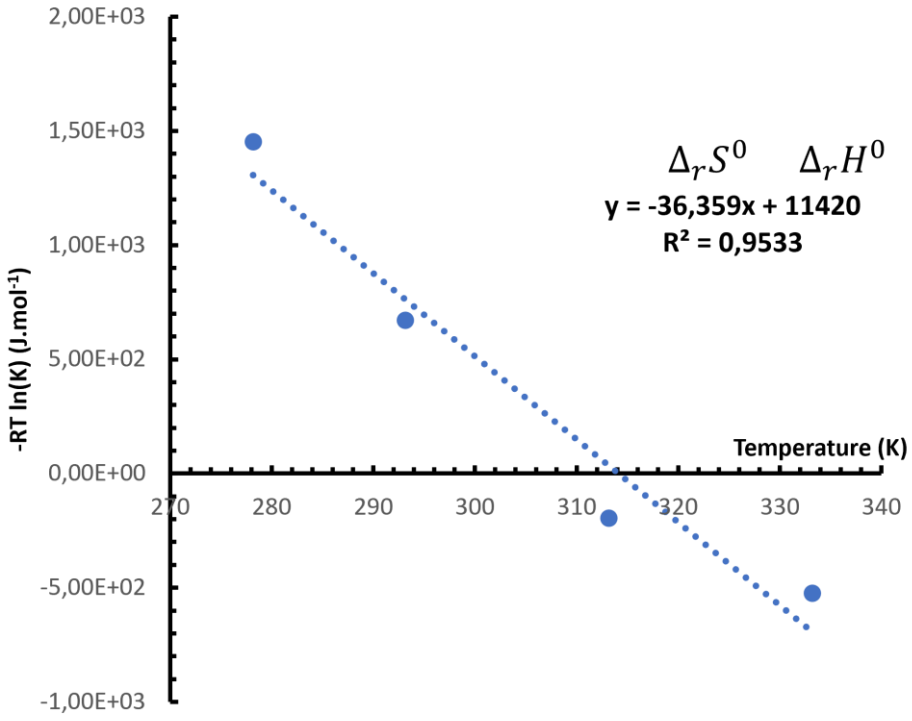


Figure 2-23 linear regression obtained for the plot of the thermodynamic data for the exchange reaction between YGGC and CGYGF in the PBS TCEP system.

This gives us a reaction enthalpy of 11.4 $kJ.mol^{-1}$ and a reaction entropy of 36 $J.mol^{-1}.K^{-1}$. The reaction enthalpy is above the 10 $kJ.mol^{-1}$ which were shown in the preliminary simulation to allow noticeable equilibrium displacement in a temperature gradient for just the exchange reaction. This makes this exchange reaction a good candidate for the creation of an out-of-equilibrium system.

Another information that could be extracted from the kinetic constants is the activation energy for the reaction. We can apply the Arrhenius equation which gives us:

$$k = Ae^{-\frac{\mathrm{Ea}}{RT}}$$

This can be linearized in:

$$R \ln\, k = -Ea \cdot \frac{1}{T} + R \ln A$$

With k the reaction constant, T the temperature, R the gas constant, Ea the activation energy for the reaction and A a pre-exponential factor. By plotting this for the different values of kinetic constants measured at different temperatures, we have for k_1 and k_{-1}:

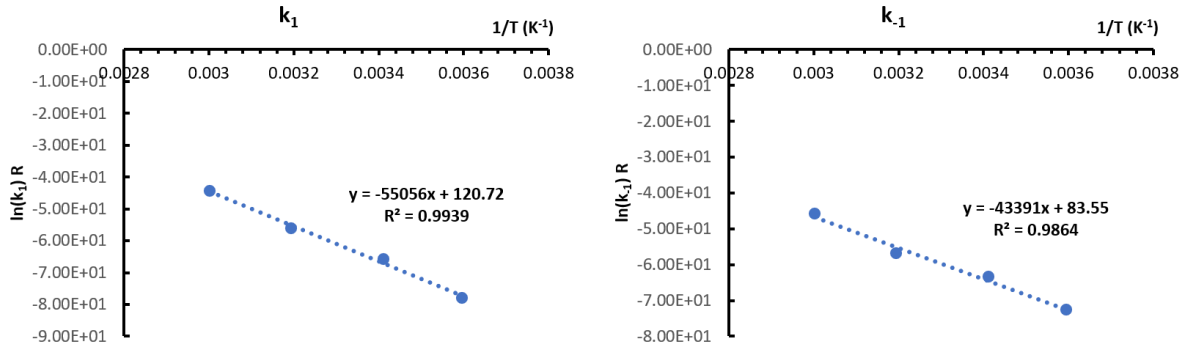


Figure 2-24 Arrhenius plots for the kinetic constants of the equilibrium of the exchange reaction between YGGC and CGYGF.

$\Delta_r H^0$	$11.4\ kJ.mol^{-1}$
$\Delta_r S^0$	$36\ J.mol^{-1}.K^{-1}$
Ea_1	$55\ kJ.mol^{-1}$
Ea_{-1}	$43\ kJ.mol^{-1}$

Table 3 Summary of the thermodynamic data for the NMC exchange reaction, obtained through the PBS kinetic experiments

The calculated values for the activation energies are shown in table 3 along with all the thermodynamic data obtained for the NMC exchange reaction. It seems that the ligation reaction requires overcoming a higher energy barrier than the dissociation reaction but both values are similar in magnitude.

8. Conclusion

We have studied the N-methyl-cysteine exchange reaction and calculated its kinetic parameters adapting the theory from bimolecular reversible reactions. From the obtained data we were able to calculate the thermodynamic values for the reaction and found out that in terms of reaction enthalpy the reaction met the criteria imposed by the preliminary modelling.

We did not observe any significate effect from the pH on the reaction, apart from the fact that more basic pH values seem to be detrimental for the creation of side products. However, the reactions did not reach equilibrium, so we do not know about potential effects on the proportions between reactants and products at equilibrium.

The phosphate buffer system allows for a "clean" exchange reaction, but less stable TCEP leads to faster oxidation of the system. The citrate buffer system yields to the formation of side products, but citrate has not been found to be part of any of them in our proposed structures and therefore these side products might have appeared for longer reaction times in the PBS system. These side products do remain in negligeable amounts apart from pH 8.

With the data for the ligation and dissociation acquired, we now turn towards the design of a dynamic polymerizable system which could be used in the out-of-equilibrium experiment.

9. References

1. Dragan, A. I. & Privalov, P. L. Unfolding of a leucine zipper is not a simple two-state transition. *Journal of Molecular Biology* **321**, 891–908 (2002).

2. Rete, C.-V. Libraries of Dynamic Peptides based on reversible native chemical ligation. (2018).

3. Rich, A. M. *et al.* Thermodynamics of Zn^{2+} Binding to Cys_2His_2 and $Cys_2HisCys$ Zinc Fingers and a Cys_4 Transcription Factor Site. *Journal of the American Chemical Society* **134**, (2012).

4. Lodish, H. F., Berk, A., Krieger, M. & Scott, M. P. *Molecular Cell Biology*. (W.H. Freeman, 2008).

5. Ruff, Y., Garavini, V. & Giuseppone, N. Reversible Native Chemical Ligation: A Facile Access to Dynamic Covalent Peptides. *Journal of the American Chemical Society* **136**, (2014).

6. Burke, H. M., McSweeney, L. & Scanlan, E. M. Exploring chemoselective S-to-N acyl transfer reactions in synthesis and chemical biology. *Nature Communications* **8**, (2017).

7. Yang, W. & Drueckhammer, D. G. Understanding the Relative Acyl-Transfer Reactivity of Oxoesters and Thioesters: Computational Analysis of Transition State Delocalization Effects. *Journal of the American Chemical Society* **123**, (2001).

8. Wawra, S. & Fischer, G. Amide *Cis - Trans* Isomerization in Peptides and Proteins. in *cis - trans Isomerization in Biochemistry* (Wiley-VCH Verlag GmbH & Co. KGaA, 2006). doi:10.1002/9783527609338.ch9.

9. Kern, D., Schutkowski, M. & Drakenberg, T. Rotational Barriers of *cis* / *trans* Isomerization of Proline Analogues and Their Catalysis by Cyclophilin. *Journal of the American Chemical Society* **119**, (1997).

10. Kerscher, O., Felberbaum, R. & Hochstrasser, M. Modification of Proteins by Ubiquitin and Ubiquitin-Like Proteins. *Annual Review of Cell and Developmental Biology* **22**, (2006).

11. Eryilmaz, E., Shah, N. H., Muir, T. W. & Cowburn, D. Structural and Dynamical Features of Inteins and Implications on Protein Splicing. *Journal of Biological Chemistry* **289**, (2014).

12. Hendrickx, A. P. A., Budzik, J. M., Oh, S.-Y. & Schneewind, O. Architects at the bacterial surface — sortases and the assembly of pili with isopeptide bonds. *Nature Reviews Microbiology* **9**, (2011).

13. Murthy, S. N. P., Lukas, T. J., Jardetzky, T. S. & Lorand, L. Selectivity in the Post-Translational, Transglutaminase-Dependent Acylation of Lysine Residues. *Biochemistry* **48**, (2009).

14. Kent, S. Origin of the chemical ligation concept for the total synthesis of enzymes (proteins). *Biopolymers* **94**, (2010).

15. McGinty, R. K., Kim, J., Chatterjee, C., Roeder, R. G. & Muir, T. W. Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature* **453**, (2008).

16. Getz, E. B., Xiao, M., Chakrabarty, T., Cooke, R. & Selvin, P. R. A Comparison between the Sulfhydryl Reductants Tris(2-carboxyethyl)phosphine and Dithiothreitol for Use in Protein Biochemistry. *Analytical Biochemistry* **273**, (1999).

17. *TCEP•HCl.*

18. Bellmaine, S., Schnellbaecher, A. & Zimmer, A. Reactivity and degradation products of tryptophan in solution and proteins. *Free Radical Biology and Medicine* **160**, (2020).

19. Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*. vol. 65 (Springer, 2006).

20. Pace, C. N. *et al.* Tyrosine hydrogen bonds make a large contribution to protein stability. *Journal of Molecular Biology* **312**, (2001).

21. Wang, Z., Rejtar, T., Zhou, Z. S. & Karger, B. L. Desulfurization of cysteine-containing peptides resulting from sample preparation for protein characterization by mass spectrometry. *Rapid Communications in Mass Spectrometry* **24**, (2010).

22. Garavini, V. Native Chemical Ligation for the Design of Dynamic Covalent Peptides. (2015).

23. Hojo, H., Onuma, Y., Akimoto, Y., Nakahara, Y. & Nakahara, Y. N-Alkyl cysteine-assisted thioesterification of peptides. *Tetrahedron Letters* **48**, (2007).

24. Merrifield, R. B. **Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide**. *Journal of the American Chemical Society* **85**, (1963).

25. Fields, G. B. Introduction to Peptide Synthesis. *Current Protocols in Protein Science* **26**, (2001).

26. Bellmaine, S., Schnellbaecher, A. & Zimmer, A. Reactivity and degradation products of tryptophan in solution and proteins. *Free Radical Biology and Medicine* **160**, (2020).

27. Valeur, E. & Bradley, M. Amide bond formation: beyond the myth of coupling reagents. *Chem. Soc. Rev.* **38**, (2009).

28. Erdélyi, M. & Gogoll, A. Rapid Microwave-Assisted Solid Phase Peptide Synthesis. *Synthesis* (2002) doi:10.1055/s-2002-33348.

29. Nishiuchi, Y. *et al.* Chemical synthesis of the precursor molecule of the Aequorea green fluorescent protein, subsequent folding, and development of fluorescence. *Proceedings of the National Academy of Sciences* **95**, (1998).

30. Carpenter, K. A., Weltrowska, G., Wilkes, B. C., Schmidt, R. & Schiller, P. W. Spontaneous Diketopiperazine Formation via End-to-End Cyclization of a Nonactivated Linear Tripeptide: An Unusual Chemical Reaction. *Journal of the American Chemical Society* **116**, (1994).

31. Seo, A., Jackson, J. L., Schuster, J. v. & Vardar-Ulu, D. Using UV-absorbance of intrinsic dithiothreitol (DTT) during RP-HPLC as a measure of experimental redox potential in vitro. *Analytical and Bioanalytical Chemistry* **405**, (2013).

32. Jin, K., Li, T., Chow, H. Y., Liu, H. & Li, X. P−B Desulfurization: An Enabling Method for Protein Chemical Synthesis and Site-Specific Deuteration. *Angewandte Chemie International Edition* **56**, (2017).

33. Liu, P. *et al.* A tris (2-carboxyethyl) phosphine (TCEP) related cleavage on cysteine-containing proteins. *Journal of the American Society for Mass Spectrometry* **21**, (2010).

Chapter 3. The living polymer system

1. Introduction

With the study of our model exchange reaction revealing that its thermodynamics would allow it to display interesting out-of-equilibrium effects. We decided to create a polymeric system based on this reaction. This chemistry does present the limitations of requiring reducing conditions to occur and seems hardly compatible with basic pH, but the presence of cysteine brings interesting properties.

For instance, glutathione (GSH) is a peptide composed of γ-glutamic acid, cysteine, and glycine. It is used by cells to maintain their redox potential, with the ratio between GSH and its oxidized form GSSG determining if the cell is under oxidative stress.[1] Oligomers of glutathione have been observed in plants, fungi, nematodes, and algae, and are known as phytochelatins[2]. They range in between dimers and undecamers and are key in heavy metal detoxification.

In addition, poly-cysteine-arginine (PolyCR) peptides have shown cell penetrating properties.[3] These were further accentuated for a specific polymer, the cyclic tetramer, which was able to amplify the intake of an attached fluorescent probe by cells 16 to 20-fold depending on the fluorescent probe as opposed to the other polymer lengths and forms. This shows that for cysteine repeating oligomers even small sequence and structural changes can lead to different properties. In the framework of our project, immobilized liposomes could represent a target within a DCL experiment in a specific part of the temperature gradient to capture only the oligomers with cell penetrating properties.

Thiols also interacts readily with gold, making a polymer system including N-methyl-cysteine particularly attractive for nanoparticle interaction studies. Such nanoparticles could be envisioned as targets for the selection of compounds in a N-methyl-cysteine created DCL as described in the first chapter.

Finally, the presence of cysteines also means that the “survival assay” can be conducted according to another parameter, the reducing agent, and select species based on their capacity to avoid oxidation.

2. Designing the “living polymer” system

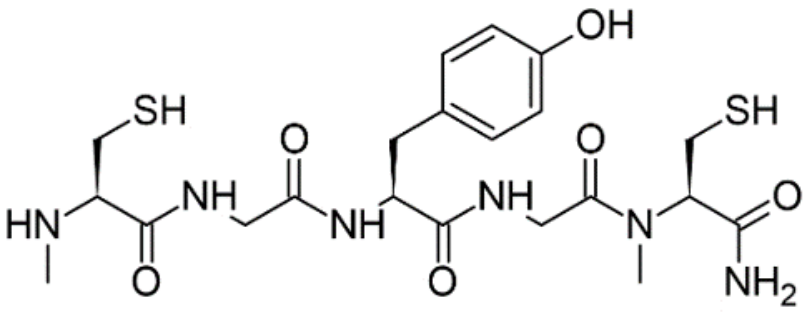


*Figure 3 -1 **C**GYG**C**, the monomer used to create the living polymer system.*

As shown previously to switch from a fragment exchange reaction to something more akin to a ligation reaction we need a peptide with a C-terminal N-methyl-cysteine which will react with an N-terminal N-methyl-cysteine. For a system to be able to polymerize, we need the sequence to be able to react on both ends. This is done by having both an N-terminal and C-terminal N-methyl-cysteine. Just as pointed out in the previous chapter, to be able to follow the kinetics and oligomers formed by UV coupled liquid chromatography, we need a UV absorbing amino acid. Also, for kinetic purposes the amino acid preceding the C-terminal N-methyl-cysteine must be a glycine to facilitate the attack on the thioester bond and to increase SPPS coupling yields we included a glycine between the N-terminal N-methyl-cysteine and the tyrosine giving us the structure shown in figure 3-1.

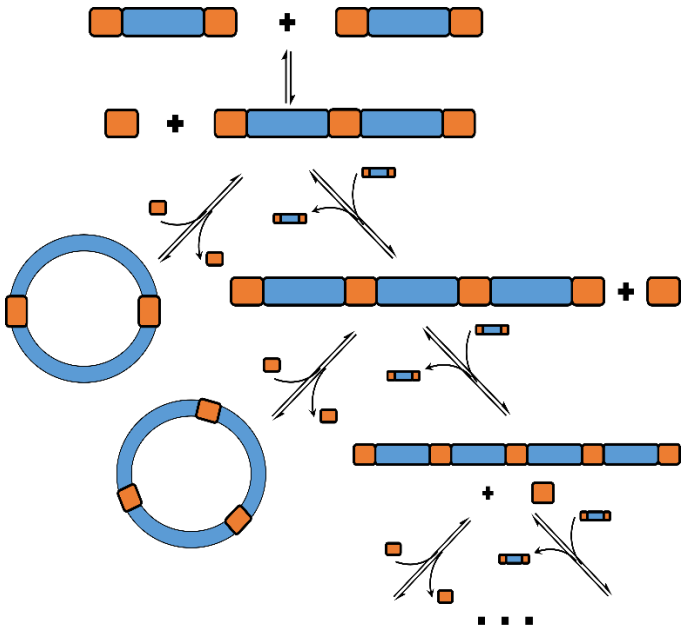


*Figure 3 -2 Living polymer system formed by the sequence **C**GYG**C**. The orange blocks represent the N-methyl-cysteines, and the blue blocks represent the GYG motif*

CGYGC is then able to polymerize, releasing an N-methyl-cysteine in the process, which can in return attack the thioester bonds of oligomers to dissociate and regenerate the monomers, giving us a dynamic polymer system. The N-terminal cysteine of an oligomer can also react with its C-terminal cysteine, yielding cyclic polymer structures of different sizes which can have interesting properties and interactions as displayed by the cell penetrating PolyCR and the fibers created by the Otto group[4–6]. In this system, there is no termination reaction (apart for undesirable side products), and since no activation is required, no chain transfer reactions. These similarities with living polymerization processes, and the parallels to population distribution of species led us to name it: "the living polymer system". There is however no requirement for activation as the oligomers are reactive by themselves through their thioester forms. And the rate of propagation is not necessarily slower than the rate of initiation in contrast with classic living polymerization.

3. Assessing the diversity of structures

A solution of 10mM of **C**GYG**C** with 100 mM of TCEP, 200 mM of Asc and 50mMol of citrate buffer was left at room temperature to explore all its possible structures. We can see on the chromatogram, especially in the end many different species. The first peak at 0.27 min of elution time, only appears after 50 hours of reaction and corresponds to a compound of 412.2 Da. It is attributed to the hydrolysis of the C terminal cysteine of the monomer as was observed for YGG**C** in the previous exchange experiment.

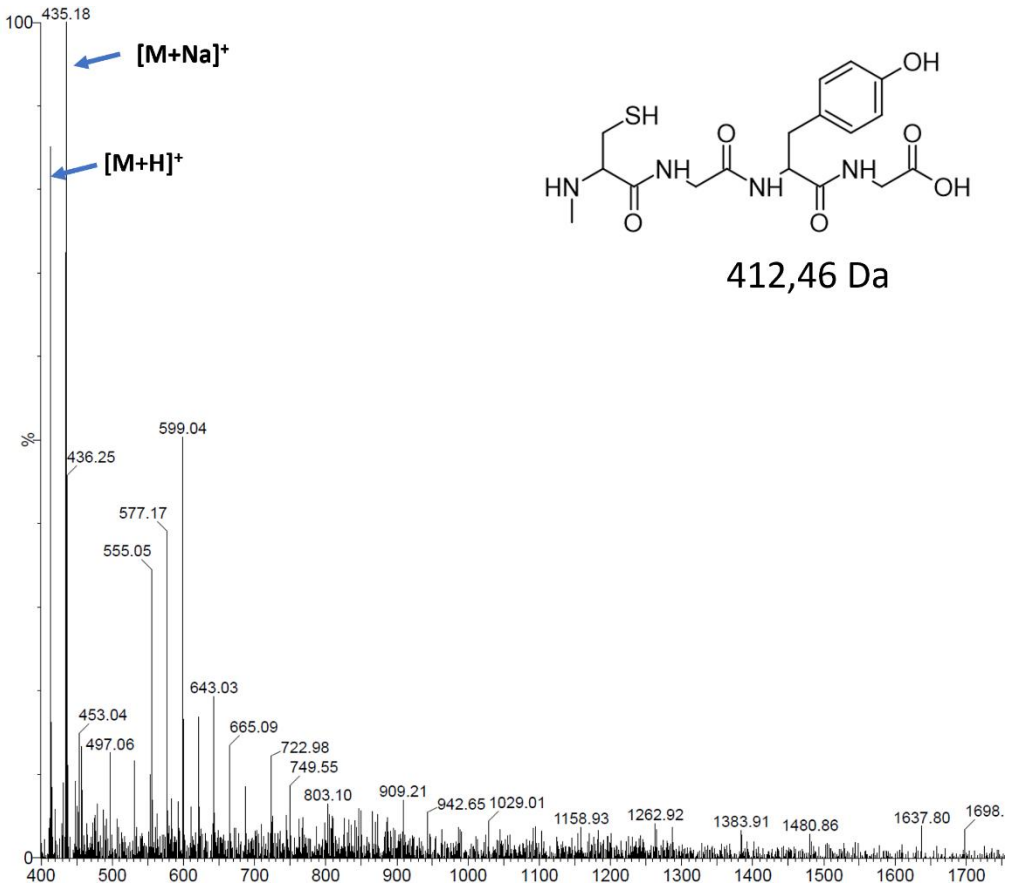


Figure 3-3 Mass signal for the compound at 0.27 min elution time, attributed to the monomer with hydrolysis of the C-terminal cysteine

Then at 0.40 min elution comes the monomer followed by a small peak, absent at the beginning, merged with the monomer peak after the first couple of days, and which separates after 6 days of reaction. The mass of this compound appears to be 687 Da as the 529 m/z signal comes from the proximity to the monomer peak. This mass corresponds to the mass of the monomer 528 Da with the mass of ascorbic acid 176 Da with the loss of a water molecule. If Asc reacted to create an ester at the C-terminal NMC, the released compound would be ammonia and the mass would be 687.6 and 685.7 for the oxidized form of Asc, both differ from 1 Da from the observed mass. As in the previous chapter we propose a ring opening of the Asc by the amine (even though it is secondary) to explain the observed mass. Maybe the reaction is facilitated by hydrogen bonds formed between the ascorbate and the side chains of the monomer, stabilizing the transition state.

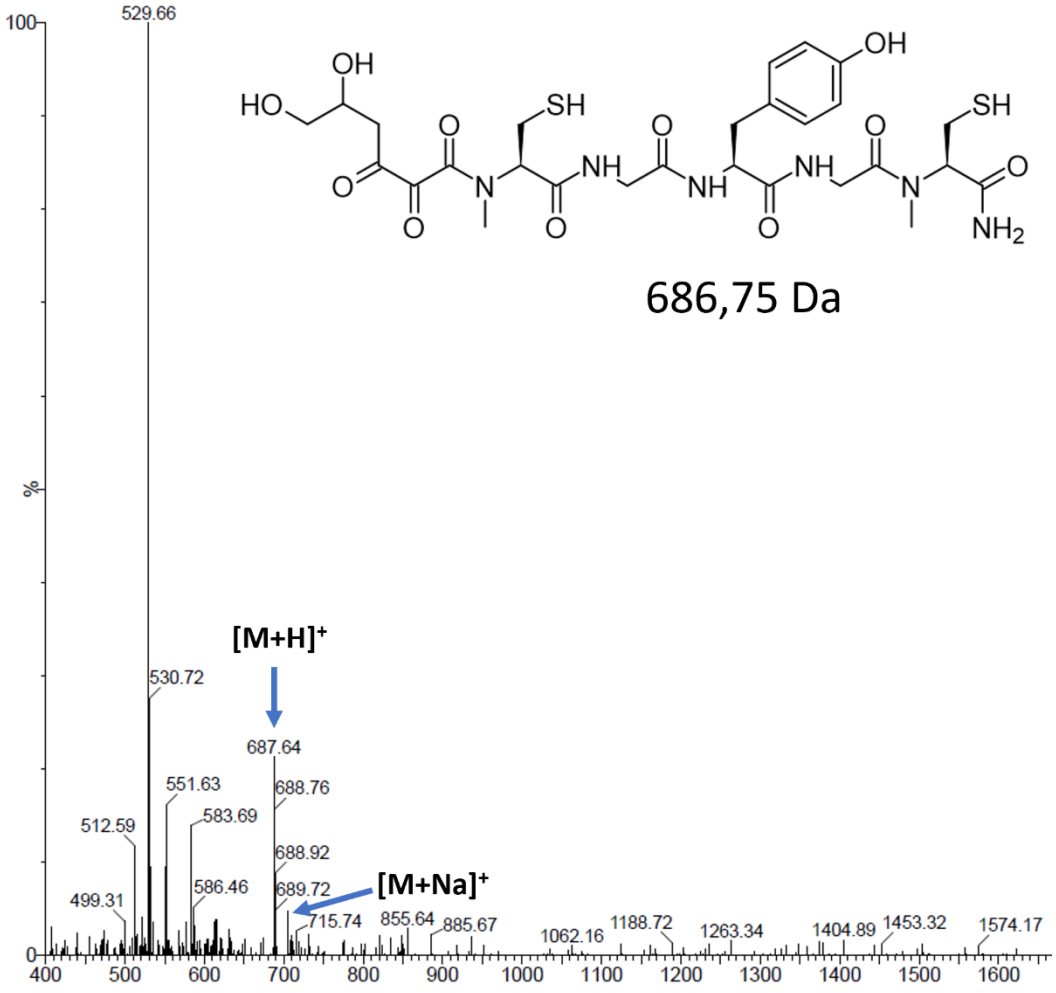


Figure 3-5 Mass signal for the compound with 0.43 min elution time which is attributed to a reaction with Asc and loss of a water molecule, we proposed the amination of Asc by the monomer as a structure for the compound

At 0.53 min comes a third side product in our system whose mass was identified as 510.5 Da as the presence of a signals at 533.5, 511.5 and 494.5 Da are characteristic of our peptides that have a plus sodium and a deamination or de-hydroxylation signal. This mass corresponds to the loss of a water molecule in the monomer. Since the peptides are aminated on their C terminus, this means that the only loss of water could come for the tyrosine hydroxyl group. The concentrations are however too small to measure a UV spectrum to see if the tyrosine absorption is modified and support this structure. The signal from this specie is constant

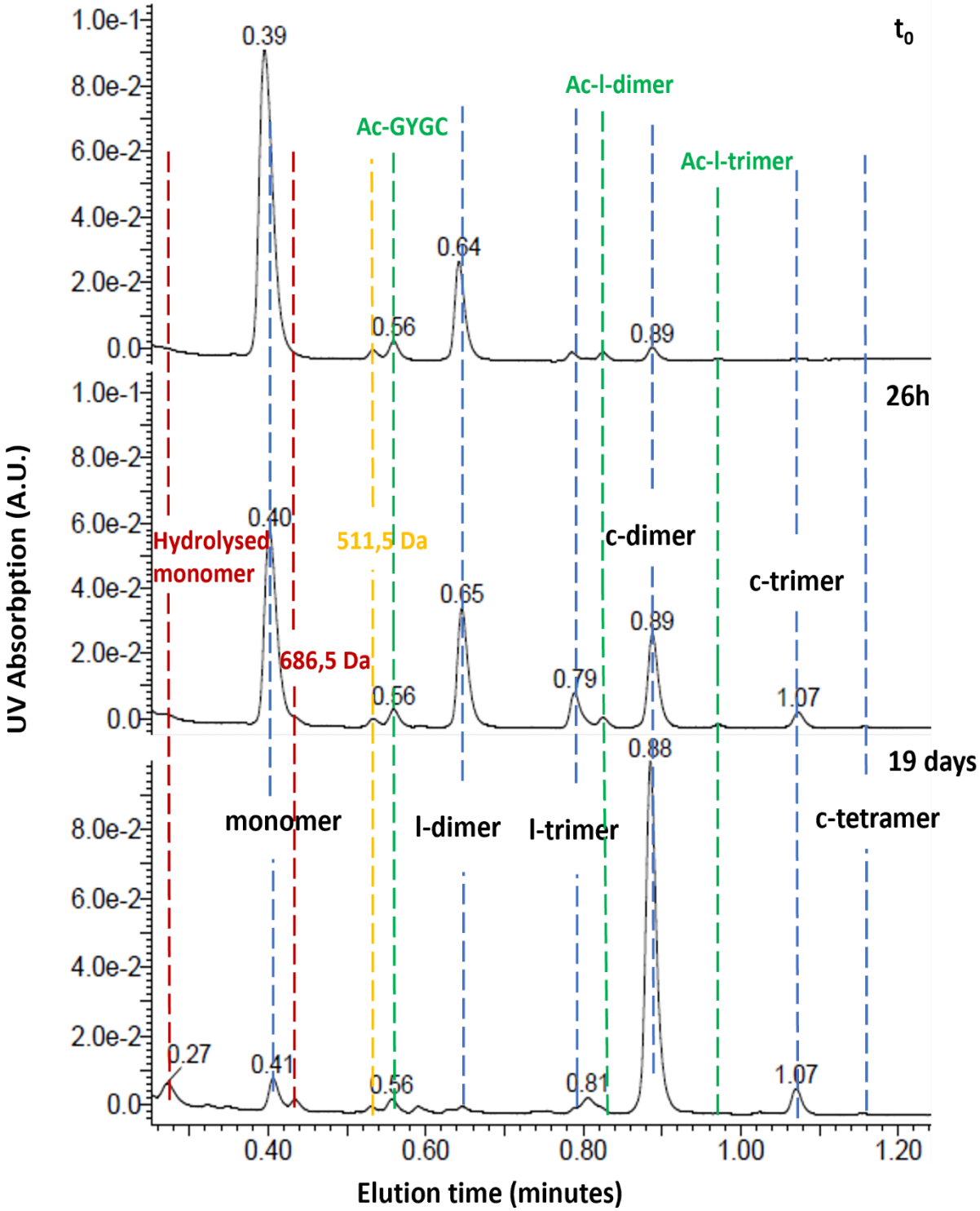


Figure 3-4 UPLC Chromatograms summarizing the evolution of the living polymer system in 100 mMol of TCEP, 200 mMol of Asc and 50 mMol of citrate buffer at pH 7 at room temperature for a starting concentration of 10 mMol in monomer. The species in blue are the oligomers of the living polymer system, those in red are byproducts resulting from the degradation of the monomer, the specie in yellow is suspected to be impurity initially present in the system with no interaction with it and the species in green are the compounds formed with the Ac-GYGC compound which are in a dynamic equilibrium with the system.

across the entire experiment. We suppose that this structure does not interact with our system. The signal at 0.56 min comes from a purification problem. During the last step of the N-methyl-cysteine synthesis, acetic acid is used as an additive for silica column purification. The N-methyl-cysteine used for SPPS still contained some acetic acid with reacted in the stead of the N-methyl-cysteine and the resulting product could not entirely be removed through HPLC purification. This compound corresponds to acetylated GYG**C** (Ac-GYG**C**) and can perform the exchange reaction as well as the polymer mass minus 76 Da can also be observed for a dimer at 0.82 min elution time and for a trimer at 0.97 elution. However, since this compound lacks an N-methyl-cysteine at both extremities no corresponding cyclic compounds linked to this molecule are observed.

From the evolution of the signal for these different compounds and their supposed structures, not all of them interact with our system in similar ways. The hydrolysis or cleavage of terminal cysteines as well as the supposed interaction with Asc constitute irreversible reactions which drain species from our living polymer system. On the other hand, the 511.5 Da is constant with time meaning that this byproduct is probably inherited from purification but does not interact with the system. Finally, the Ac-GYGC compounds are at equilibrium with the polymer system, they simply cannot form cyclic species.

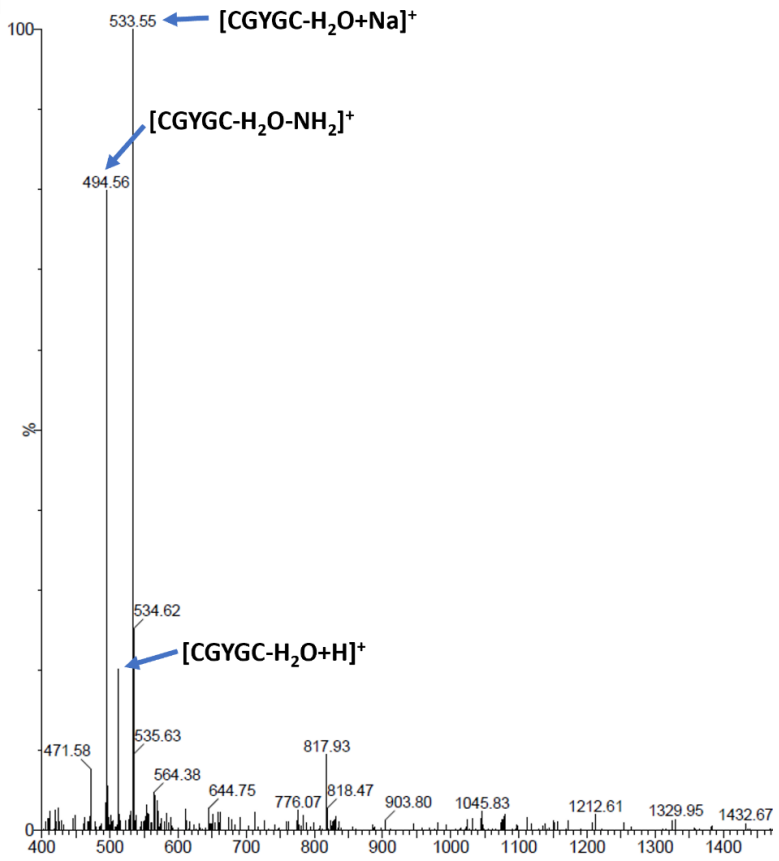


Figure 3-6 Mass signal for the compound with 0.53 min elution time. The mass signal corresponds to the monomer with the loss of a water molecule.

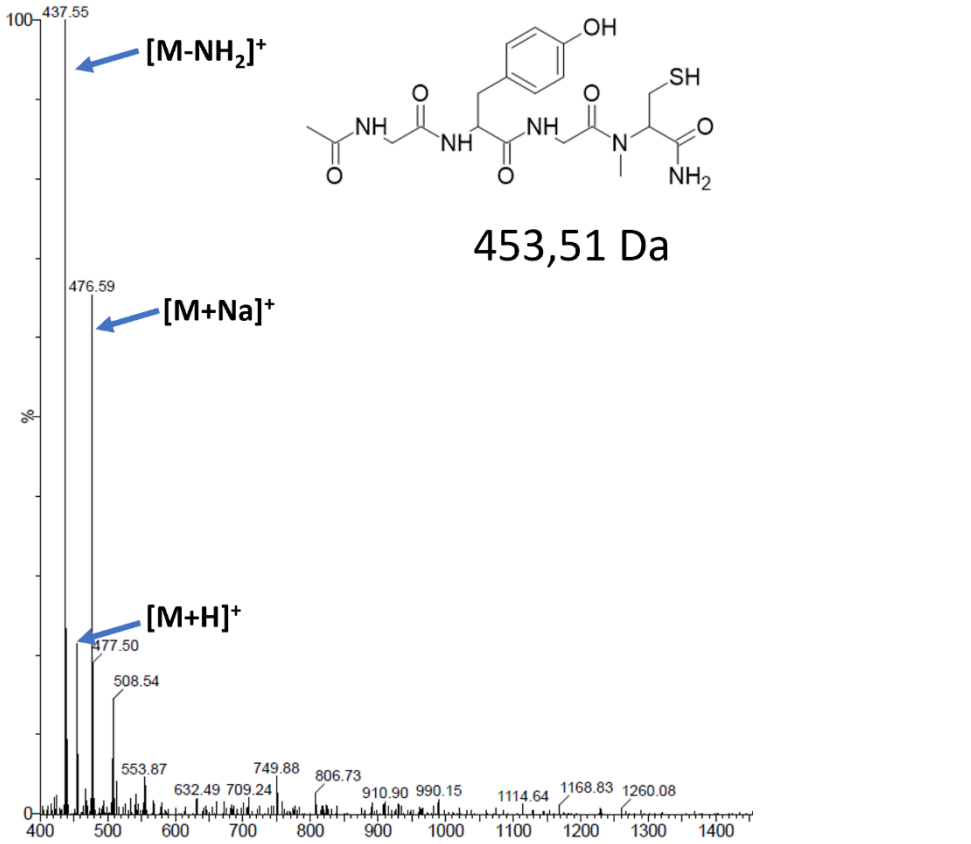


Figure 3-7 Mass signal for the compound with 0.56 min elution time. The mass signal corresponds to a side product from the SPPS with partially pure N-methyl-cysteine

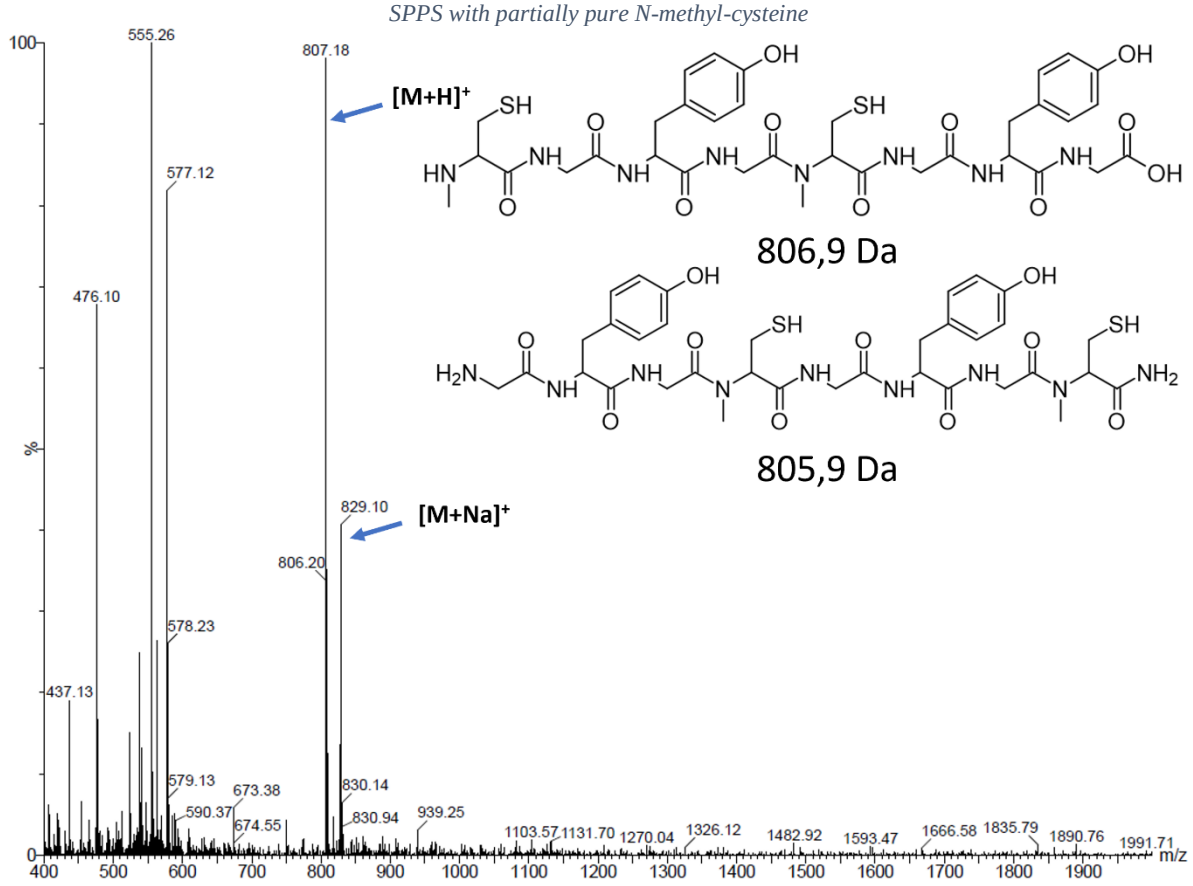


Figure 3-8 Mass signal for the compound with 0.61 min elution time. This signal corresponds to a dimer with a hydrolyzed C-terminal cysteine.

A similar pattern is observed for the hydrolysis product. Linear dimers with a hydrolyzed cysteine are observed at 0.61 min elution time. This can either be the direct result of the hydrolysis at the C-terminus of a dimer, or the result of an exchange reaction with a monomer with a hydrolyzed cysteine through its N-terminal one. It could also be the N-term NMC which was removed through a TCEP side reaction discussed previously[7], as the masses are similar.

The last side product observed corresponds to the signal which appears after three days of reaction between the linear trimer and the N-acetylated dimer at 0.81 min elution time.

To prevent interferences between the peak of our reference compound and our peptides, we switched from DMBA which had a retention time of 1.11 min with Fmoc-serine-(trt)-OH. We can see that the polymerization reaction starts very quickly as just minutes after the beginning we already have high concentrations of linear dimers and cyclic dimers are starting to be formed. Then the trimeric species, linear and cyclic appear as the monomer and linear dimer species are slowly consumed. We also manage to observe cyclic tetramers, but linear tetramers are not observed in solution, suggesting that they cyclize rapidly after their formation. Finally, the monomer, linear dimer, and linear trimer concentrations diminish, until the remaining species in solution are mostly cyclic dimers with small concentrations of monomers and cyclic trimers. This means that the information on the concentration of linear trimers gets lost as its concentration gets low and the one of the 0.81 min side product increases.

Soon after the apparition of cyclic tetramers, we observe the formation of a precipitate in the Eppendorf tubes. The analysis of the precipitate reveals that the system does explore longer sequences. At the end of the 10 mMol experiment at room temperature, the precipitate was washed and solubilized in methanol with 100 mMol of TCEP and analyzed by HPLC in colbaoration with Jean-Marc Strub from the Laboratoire de Spectrométrie de Masse Bio-Organique.

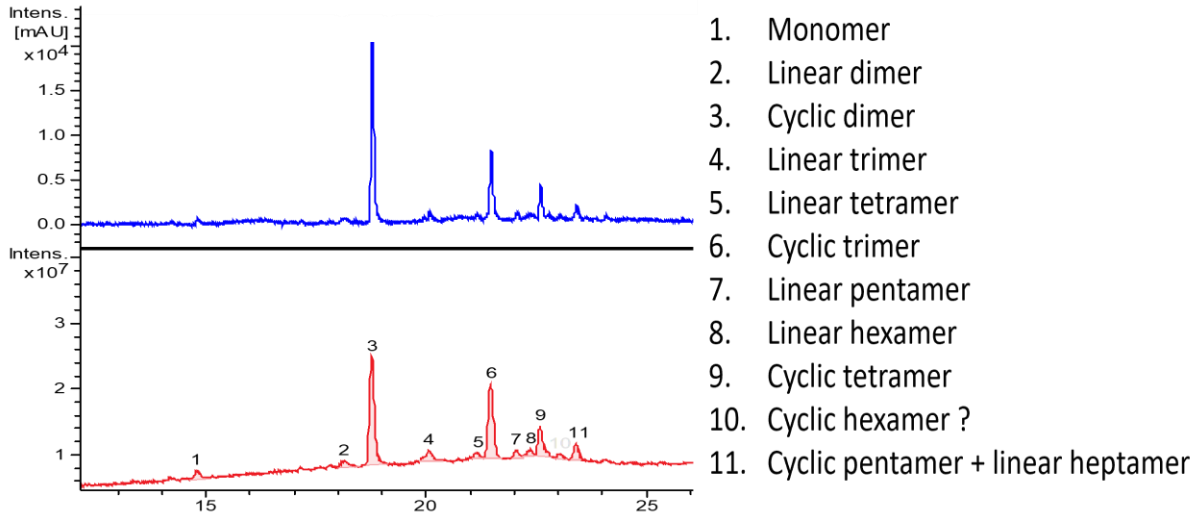


Figure 3-9 HPLC chromatogram of the precipitate from the living polymer system from 10 mMol initial concentration at room temperature. The blue curve corresponds to the UV absorption at 280 nm and the red curve is the total measured m/z signal both according to elution time.

When looking at the mass signals, we observe mixtures with the mass of the cyclic oligomer appearing during the peak attributed to the linear form. However, since those signals are later

observed on their own without their respective linear mass signals, we concluded that the cyclization could be performed during the mass spectrometry.

As we can see the system is able to generate many oligomers with cyclic ones seemingly more stable. Here the longest oligomer observed is a linear heptamer, which along with the four cyclic structures observed brings the diversity of structures generated to eleven.

4. The influence of temperature and concentration

The equilibrium concentrations for the system at different temperatures and concentrations are important for its implementation in the temperature experiment gradient. A first step would be to study the reached equilibrium so that newly observed behavior in the temperature gradient can be attributed to the out-of-equilibrium chemistry. The effect of concentration may help tune working concentrations depending on the objective. To this end the living polymer system was allowed to evolve starting from different monomer concentrations 1 mM, 5mM and 10 mM at three different temperatures, 10°C, 20°C and 40°C. The evolution of each oligomer is shown in figure 3-10. The linear trimer is not shown since as discussed previously the concentration information is lost when its concentration decreases, and the 0.81 elution time side product appears.

We can see overall that the system behaves very similarly regardless of concentration and the patterns observed at 40°C seem to be reproduced at 20°C albeit slower. The linear dimer is a good example for this. It is formed quickly at the beginning then reacts to form the other structures making it a kinetic product in this system.

There are however a few differences. It seems that the cyclic trimer is more stable in solution at 10°C for 1 mMol than it is for 20 and 40°C as it managed to reach proportions similar to the 5 and 10 mMol experiments. Interestingly, for the cyclic trimer at all temperatures, higher proportions are observed for 5 mMol of starting monomer. On the other hand, higher concentrations of starting monomer led to higher cyclic dimer proportions, a dynamic that seems to be accentuated at low temperatures whereas at 40°C the c-trimer and c-dimer proportions are almost the same at 5 and 10 mMol.

The cyclic tetramer is only observed in solution for concentrations of 5 mMol and higher, indicating that linear trimers cyclize before being able to react with monomers at lower concentrations, however, we observe higher proportions of cyclic tetramer for 5 mMol at 10°C, whereas at room temperature they are the same for 5 and 10 mMol and at 40°C it is 10 mMol that shows the highest proportion of cyclic tetramers.

Eventually, all that are left in solution are the cyclic species, with a high preference for the cyclic dimer, with even the concentrations of cyclic trimer and tetramer decreasing in favor of the cyclic dimer. This could be due to multiple reasons. Either, as the precipitate forms, the tetramer precipitates with it as we saw previously that it was present in the 20°C 10 mM one.

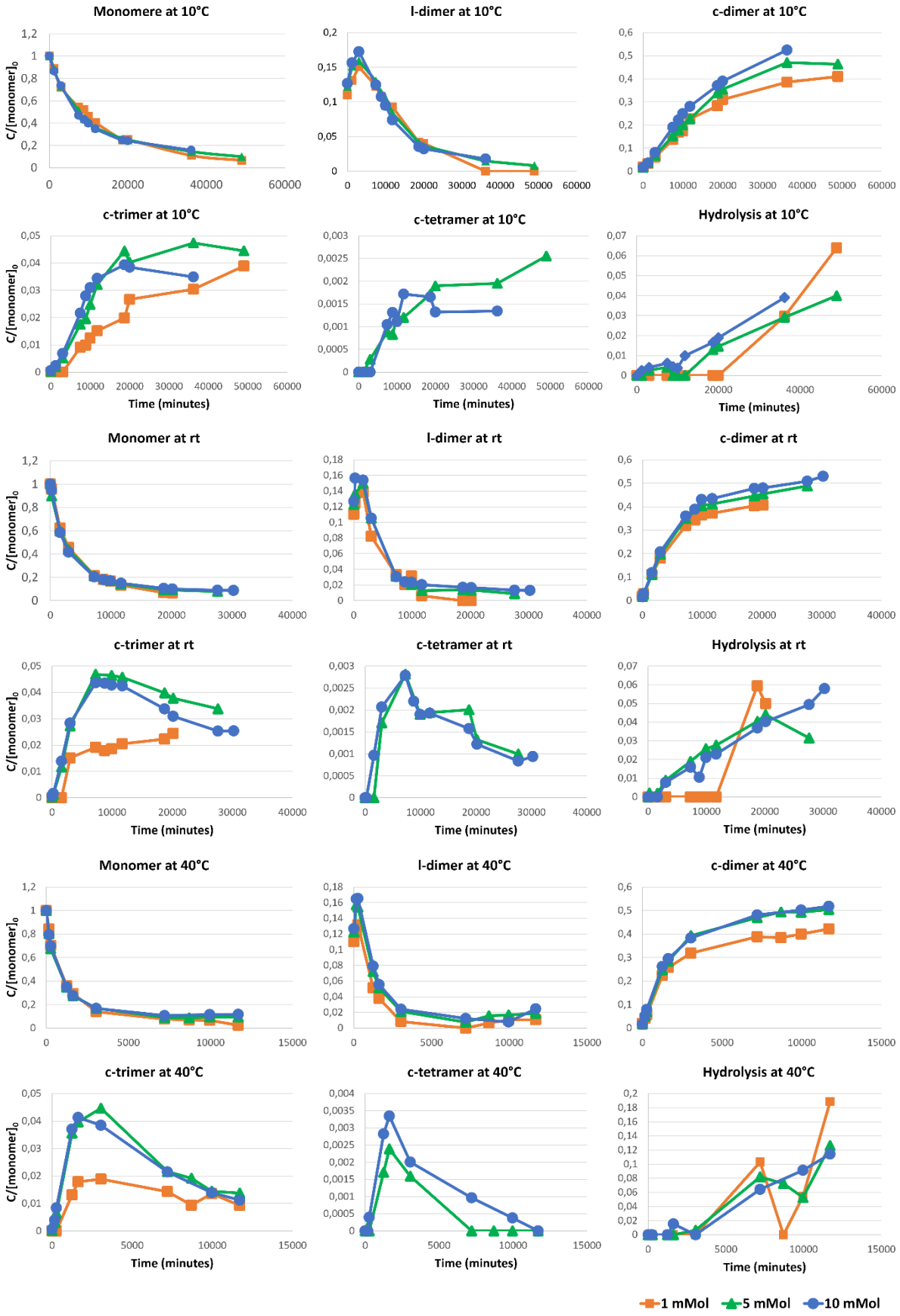


Figure 3-10 Evolution of different oligomer concentrations for different temperatures and starting monomer concentrations. The concentration for each oligomer is normalized by the starting monomer concentration for comparison. The experiments were performed at pH 7 with 100 mMol TCEP, 200 mMol Asc, 50 mMol citrate. Fmoc-Serine(trt)-OH was used as UV standard.

Or the cyclic dimer has the capacity to template or favor its own creation at the detriment of the other oligomers.

Finally, the hydrolysis at the C-terminus of the monomer eventually happens at all temperatures but is favored by temperature with the 40°C experiment reaching higher proportions of hydrolysis for similar system states.

The precipitates formed by each of those experiments was then resolubilized in methanol with TCEP to analyze its composition. The summary of these results is presented in figure 3-12. We see that our UV reference is present which means it coprecipitates with our oligomers. The most abundant species in the precipitates is still the cyclic dimer one. However, we do observe species that were absent in solution, such as the linear tetramer. It seems that temperature favors the cyclic dimer seeing its proportions for the 40°C whereas slower temperatures allow the system to explore longer configurations as shown by the 20°C one and the presence of cyclic pentamer in the 5°C one. Given the kinetics of the exchange reaction obtained previously, the higher the temperature, the more the exchange products are favored meaning that we should expect longer oligomers for 40°C than for 20°C, however this is not the case. It could be that the lower temperature allows of interactions between peptides stabilizing longer oligomers whereas at 40°C those interactions are broken by thermal agitation and only the stable cyclic dimer remains.

Another thing to note is that the amount of precipitate depends on the temperature and concentration. Higher concentrations will give more precipitate but higher temperatures as well even if their precipitate is composed mostly of cyclic dimer which would be thought to be more soluble.

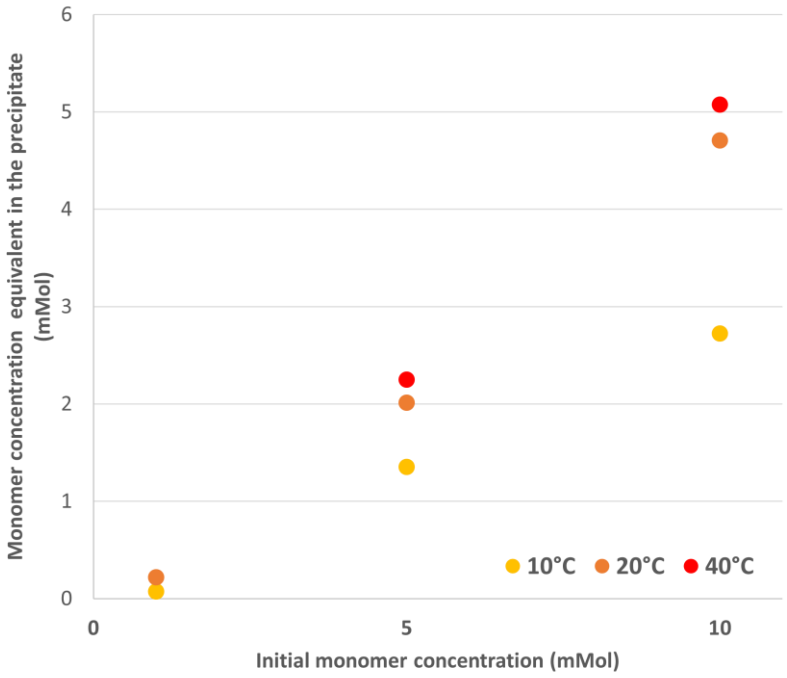


Figure 3-11 Equivalent of initial monomer concentration present in the precipitate.

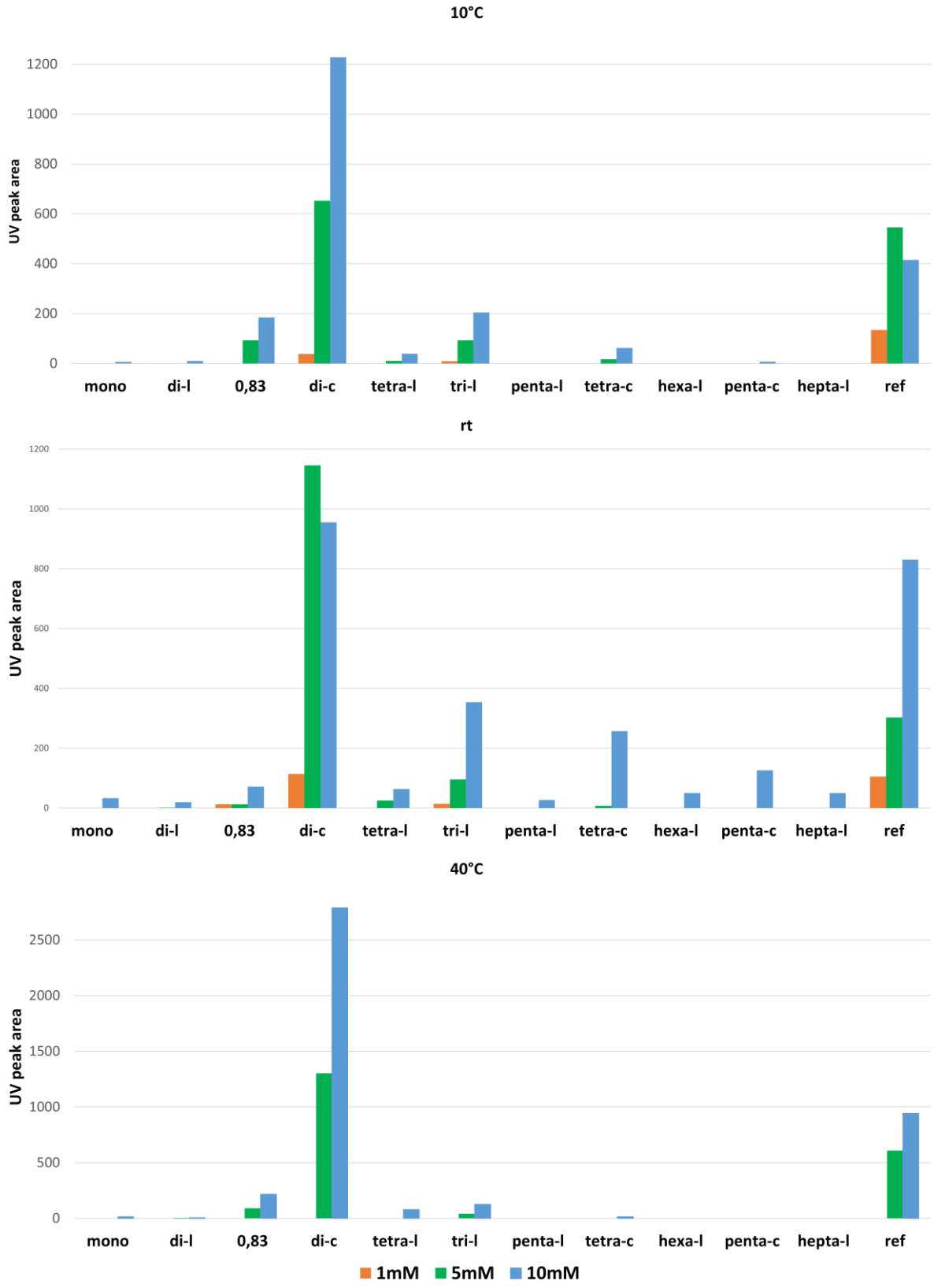


Figure 3-12 Composition of precipitate obtained at different temperatures and different concentrations. The value plotted for each compound is the area of its UV peak on the UPLC chromatogram.

5. The influence of pH

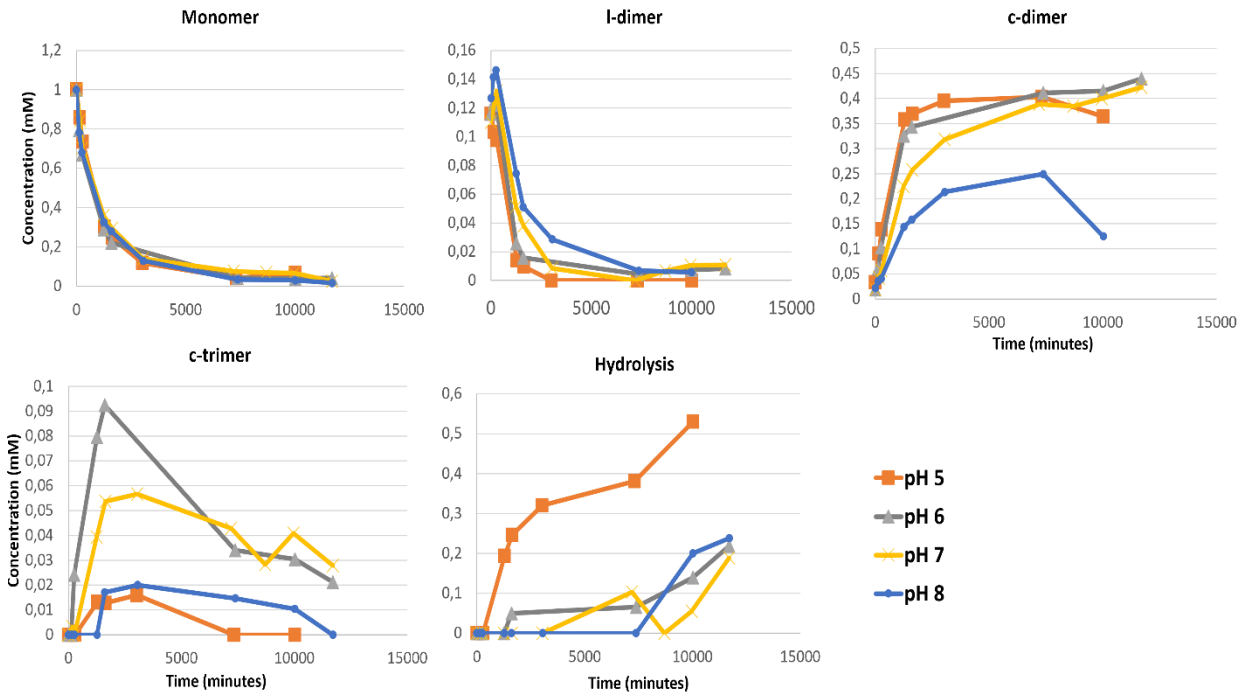


Figure 3-13 Evolution of the concentration of different oligomers at different pH. The experiment was performed at 40°C with 1 mM of starting monomer concentration in 50 mM citrate buffer with 100 mMol TCEP and 200 mMol Asc

With our main objective being to drive these systems out of equilibrium, being able to modify the pH could help in different ways. In the DCL experiment, different pH might enable the use of an increased diversity of targets for the system which would otherwise be unstable at others. pH also plays a key role in protein folding, with most enzymes displaying huge losses of activity if not total if they are not at their optimum pH conditions. Adding a pH gradient on top of the temperature gradient in the out-of-equilibrium experiment might help explore new structures and properties for the system and could be done by having a controlled release of protons throughout the solution through a photoacid such as malachite green mentioned in the introduction.

To this end, we tested the response of the developed living polymer system to a change of pH. We tested the pH values of 5, 6, 7, 8 and 9. At basic pH a degradation of the product occurs leading to big losses in dynamic species, the mass of the side product is 428 Da. This could be explained by the hydrolysis of the C-terminal N-methyl-cysteine and the formation of a sulfenic acid, a common oxidative degradation of cysteine[8], possibly stabilized by hydrogen bonding with the tyrosine. As this happened within a few hours at pH 9 we did not show the values alongside the others. This also explains the low total concentration of species at pH 8 as the system gets slowly degraded. We do note however that the experiment at pH 8 showed the highest concentrations for linear dimers hinting to an increased stability of dimer or reactivity of the monomer to create the dimers. We also observe a transitory increase in cyclic trimers at pH6 compared to pH 7. The experiment at pH 5 exhibits fast hydrolysis with almost 60% of the system being hydrolyzed

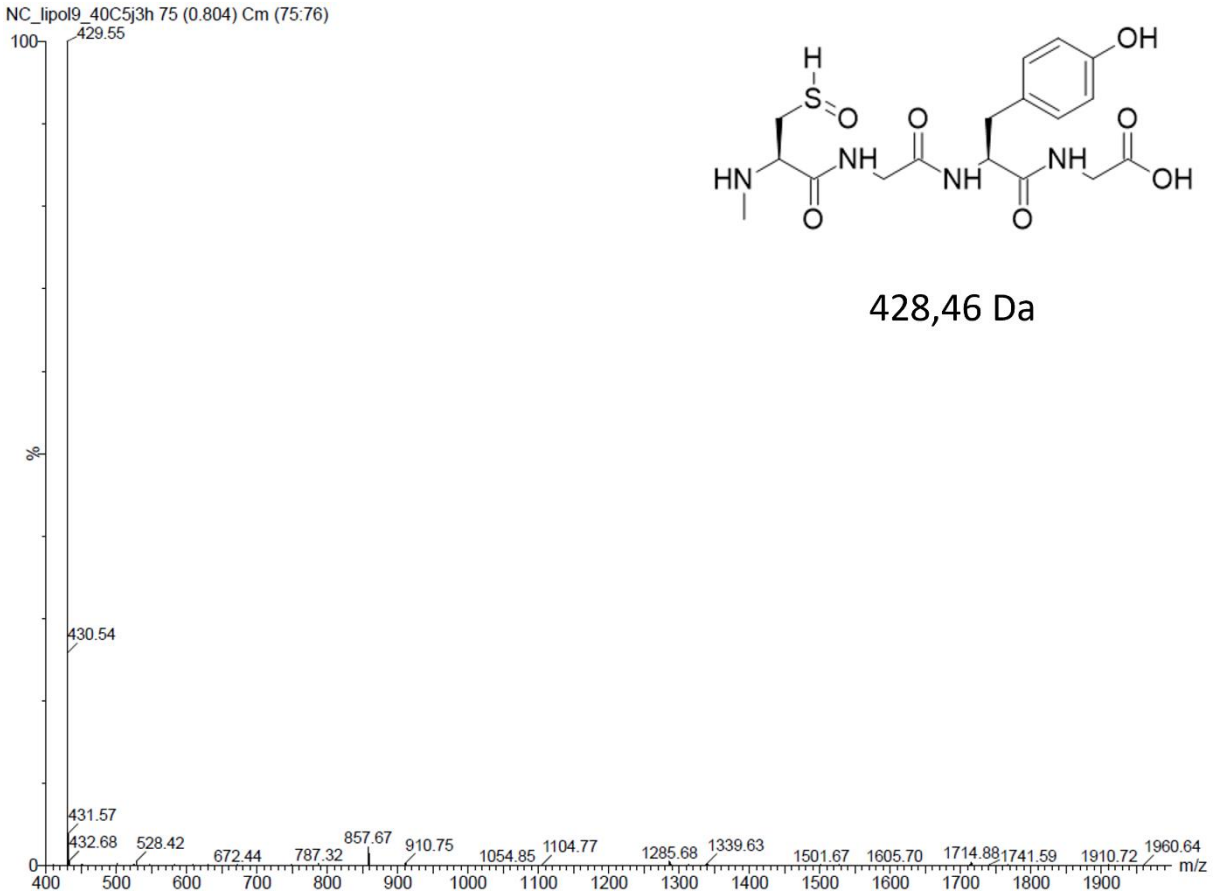


Figure 3-14 Positive electrospray m/z signal for the degradation product found in quantitative amounts at pH 8, this was almost the only signal observed after 5 hours at pH 9. The proposed structure is the hydrolysis of one of the NMCs and the creation of a sulfenic acid.

after a week. Because of the hydrolysis at acidic pH and side product at basic pH there was no precipitate in the Eppendorf tubes to analyze.

These results imply the exchange reaction must mostly be run between pH 6 and pH 7 with only minor differences in behavior between both, mainly an increase in cyclic trimers at the beginning and a slight decrease of linear dimers for pH6 compared to pH 7.

6. SEM-EDX analysis of the precipitate

The precipitates obtained in our living polymer experiments we observed with scanning electronic microscopy and the elemental composition of the obtained solids was analyzed through energy dispersive X-ray spectroscopy. We observed a material that seemed to form fiber-like (figure 3-15). These structures displayed a sulfur signal, characteristic of the presence of cysteine. It did however display a high phosphor signal implying that the reducing agent TCEP also participated in the formation of such material. Furthermore, we know from the precipitate analysis that the UV reference compounds, Fmoc-serine-(trt)-OH is also inside the precipitate and could be part of these structures as well.

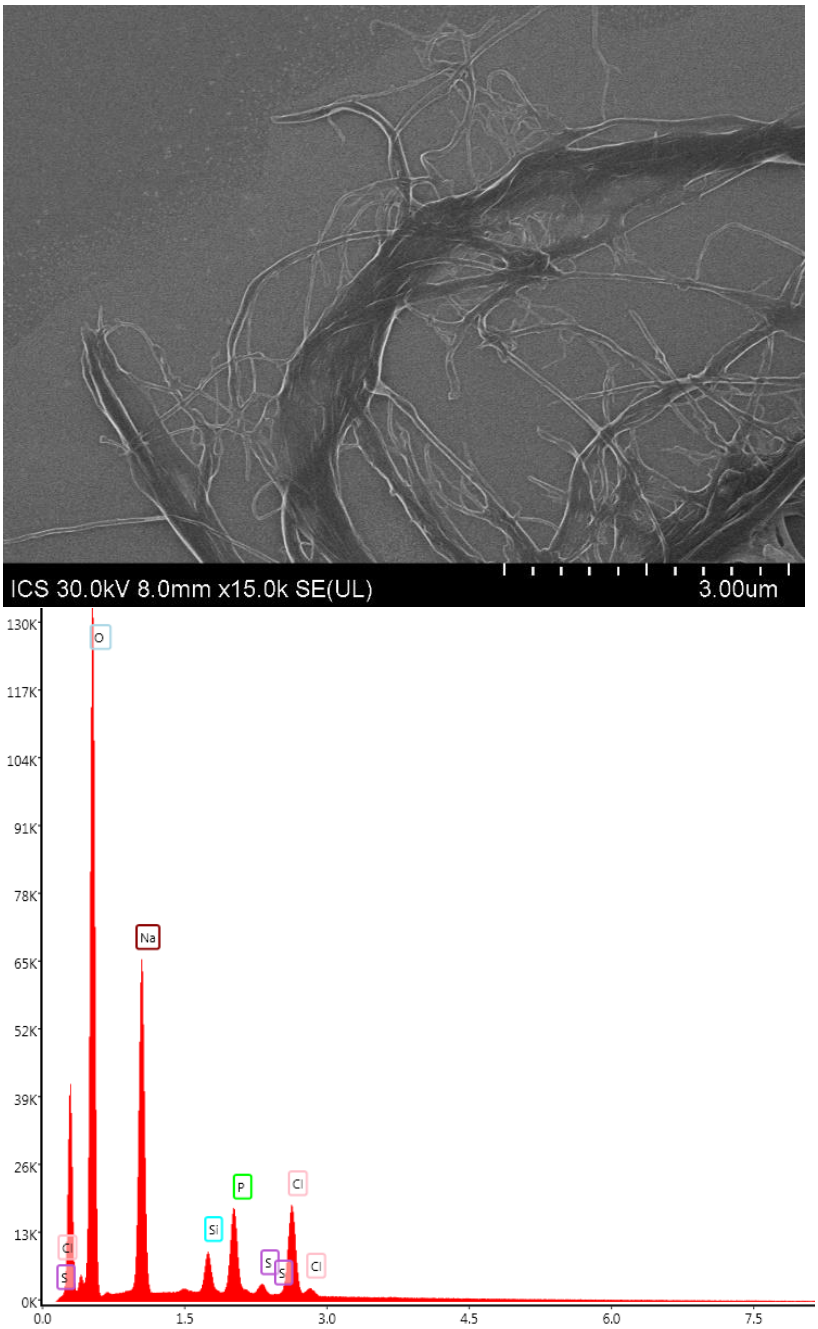


Figure 3-15 (left) amorphous solid observed in the precipitate by SEM (right) Interpreted EDX spectrum for the amorphous solid. This spectrum was measured with a voltage tension of 8 keV.

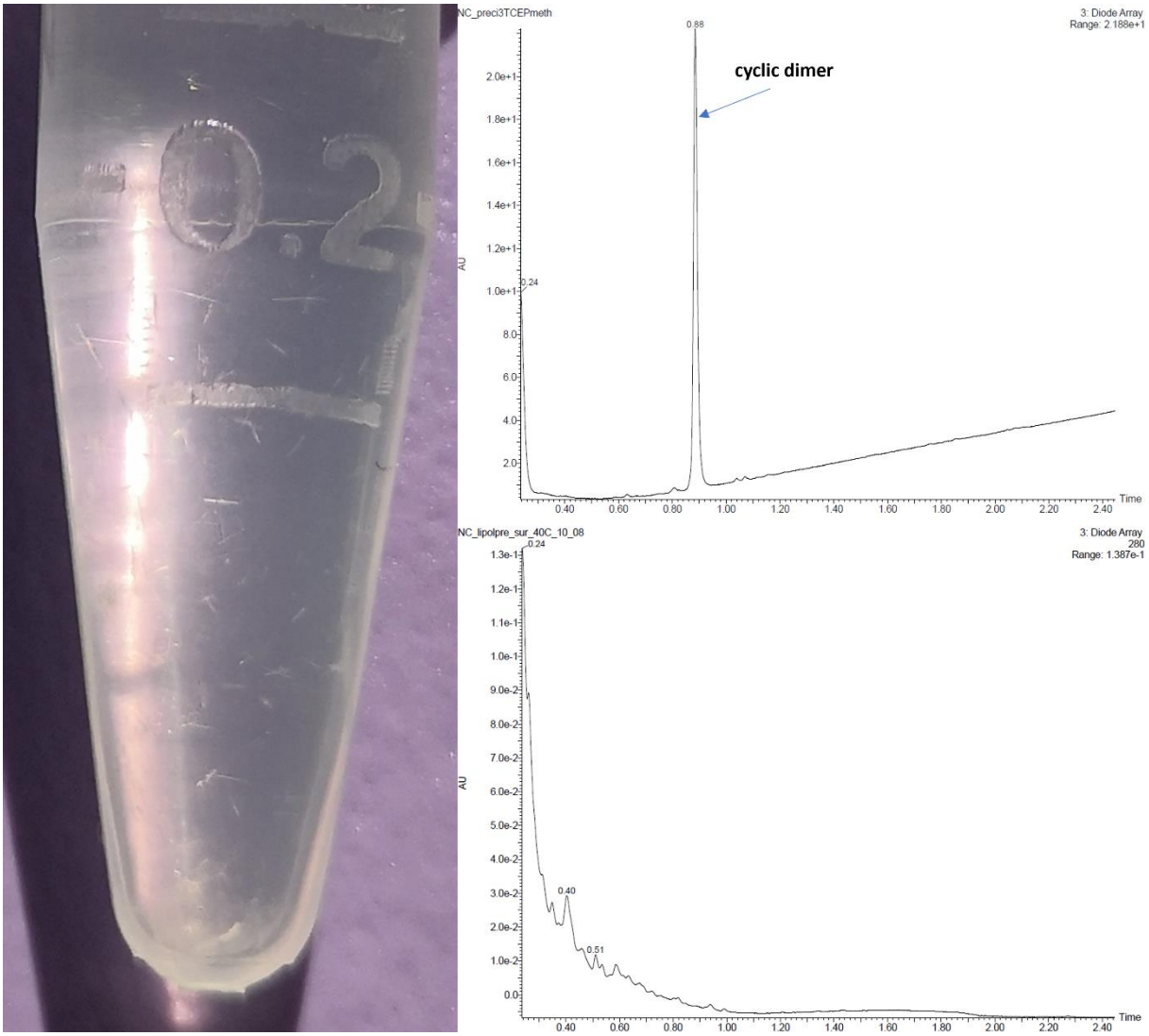


Figure 3-16 (left) Photography of the Eppendorf at 40°C with the living polymer system with 5mMol starting concentration of monomers in citrate buffer at pH 7 after 3 weeks at 40°C without a UV standard. Small needles can be observed in the solution. (Top right) chromatogram obtained after re-diluting the needles in methanol and TCEP (bottom right) chromatogram of the supernatant, devoid of any tyrosine containing peptide.

We reconducted a living polymer experiment at pH 7 at 10°C and 40°C with 5 mMol of starting monomer concentration to further study the precipitate. We did not include any UV reference in the reaction media. No precipitate has appeared after five weeks in the 10°C Eppendorf, however no sampling was performed. After four weeks, the 40°C system displayed the formation of small needle-like crystals (figure 3-16). The needles could be redissolved with TCEP, meaning that they resulted from the oxidation of our system. The solubilization of the needles in TCEP and methanol showed that they were composed uniquely of cyclic dimers (chromatogram showed on figure 3-16 top right), whereas the super natant was completely devoid of any tyrosine containing specie (3-16 bottom right), this showed a quasi-perfect self-selection of the cyclic dimer. SEM-EDX elemental analysis such as elemental mapping (figure 3-17), were performed on one of such crystals whose size was measured to be 45 µm in width. As can be seen in figure 3-16, the crystal does not contain any salt or phosphor limiting the possibilities for coprecipitation to sodium ascorbate, citrate, and ammonia.

The apparition of structures is not uncommon for peptide systems. As discussed previously, the system from Otto creates fiber-like structures whose appearance are modified through the addition of aliphatic amines.[5] Cyclic peptide has initially been described by Ghadiri's group[9], in that case, the cyclic peptide was composed of amino acids of alternating chirality (L & D) which could stack, creating nanotubes. Other cyclic peptide stacking has been observed with $\beta, \delta, \varepsilon$ amino acids.[10–12] If the needles result from the stacking of our α, L amino acids, it would be a great example. Other possibilities for peptide self-assemblies are nanosheets[13], but are much smaller than the amorphous material observed.

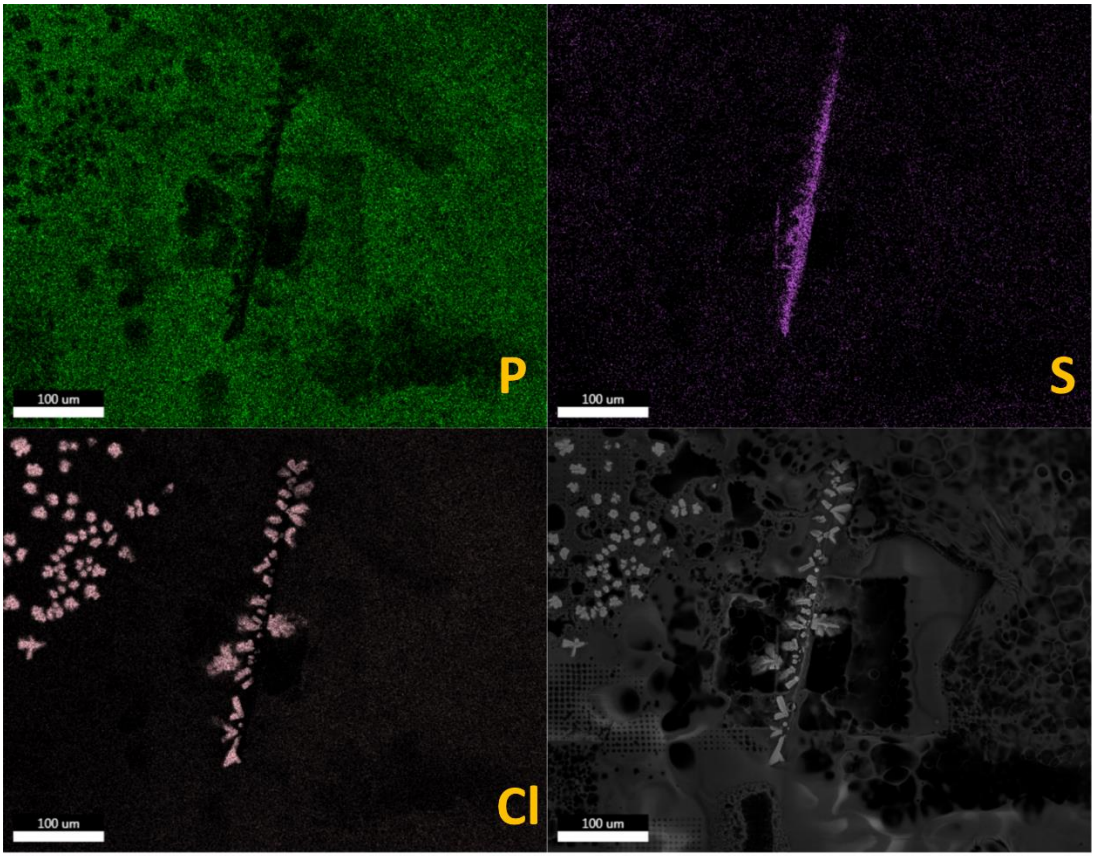


Figure 3-17 Top SEM image of a needle. The amorphous material comes from the remaining reaction media that was sample with the needle. Bottom, elemental mapping of the image realized by EDX showing that the needle was the only structure containing sulfur and that it does not contain any phosphor, therefore no TCEP.

7. Needle structure:

The observed needles being crystalline in nature, X-ray diffraction experiments were performed in order to access the crystalline structure. These experiments were performed in collaboration with Nathalie Gruber (Service de Radiocristallographie, Fédération de chime Le Bel).

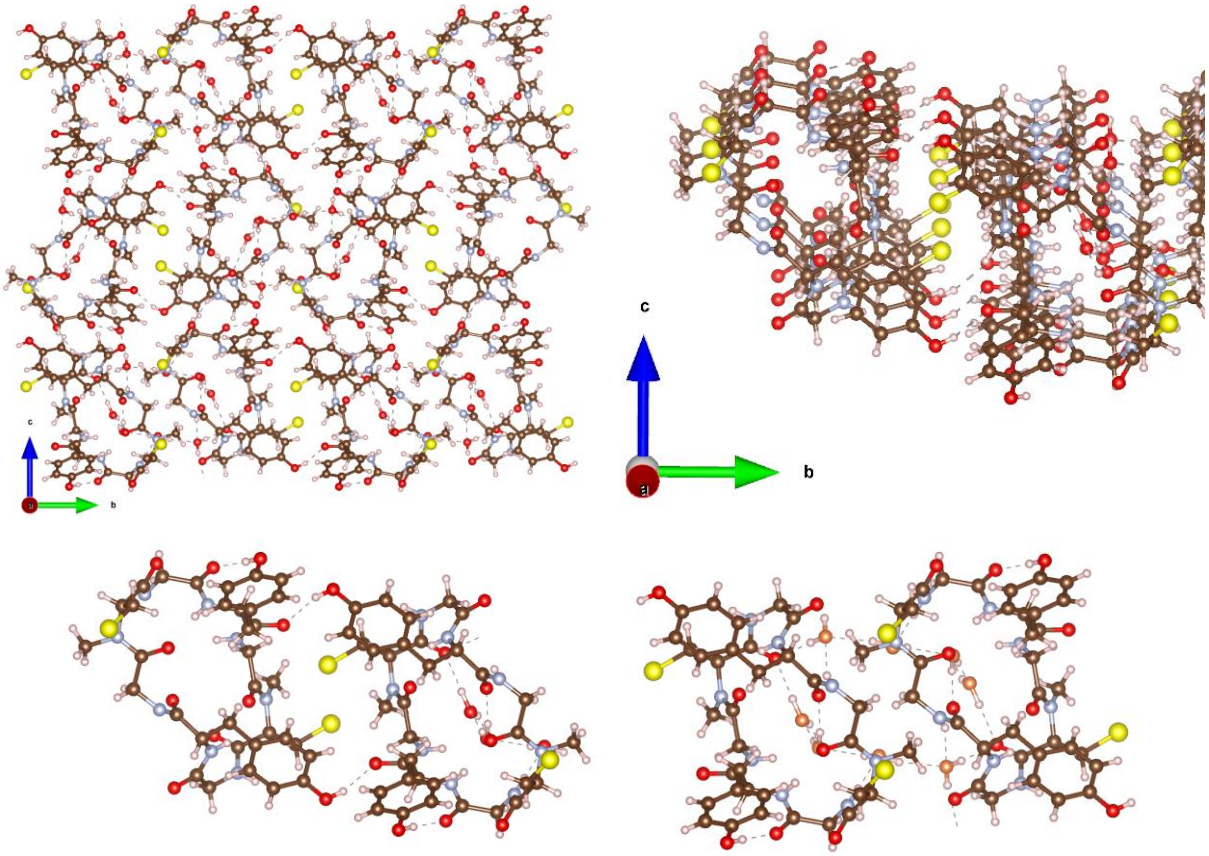


Figure 3-18 Crystal structure obtained from Xray diffraction experiments on the needles. Brown carbons correspond to carbon, white to hydrogen, yellow to sulfur, blue to nitrogen, red corresponds to oxygens from the cyclic dimer whereas orange is the oxygens from water molecules in the structure. Dashed lines represent hydrogen bonds. Top left, view along the a axis, top right, view along the a axis tilted to show the stacking, bottom left and right represent two different kinds of interactions between cyclic dimers along the b/c plane.

The crystal is composed of stacks of cyclic dimers in the fashion of nanotubes in the a direction as illustrated in the top right of figure 3-18 as well as some water molecules incorporated inside the structure. The cycles interact in two different ways along the b/c plane. One is through hydrogen bonding between the tyrosine hydroxyl group and the oxygens from the peptide back bone. The other is through the interaction of the back bone and the water molecules trapped inside the structure, bridging the dimers together.

When looked at from the b axis as shown in figure 3-19, we clearly observed the stacking of the cycles in the a direction leading to the formation of the nanotubes. We also noticed that the dimers are tilted from one stack to another as showcased in the interpretative drawing. The distance between two sulfur atoms in this structure is of 5 Å, which is too long to be considered a disulfide bond which averages around 2.05 Å.[14]

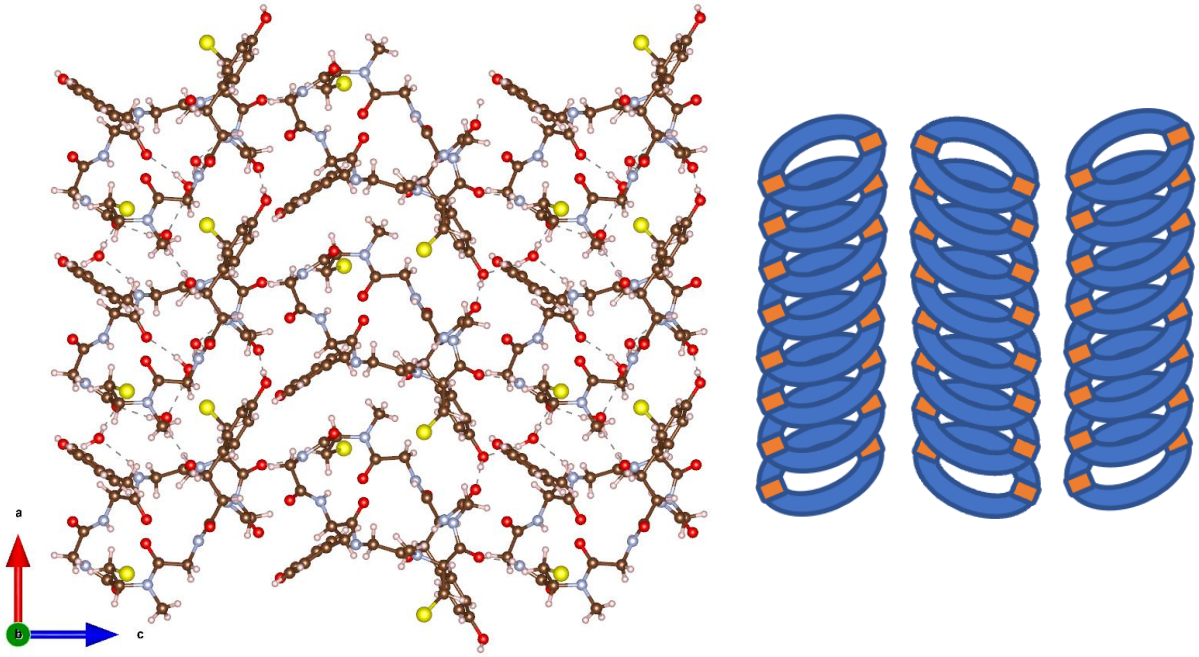


Figure 3-19 View of the crystal along the b axis and an interpretative scheme

8. Conclusions

We managed to design a polymer system from the previously studied exchange reaction. This system has been demonstrated to explore multiple lengths (up to the heptamer) and structures (cyclic and linear). However, the system does not show a preference for long sequences at high temperatures and monomers at lower ones as the kinetics of the exchange reaction hinted. There is still however a modification of the system regarding temperature as the precipitates show different compositions. It is hard to fully understand the behavior of the system at cold temperatures. Indeed, the system is limited by the reducing capacities of the reducing agent and even switching from phosphate buffer to citrate did not allow the system to reach equilibrium while sampling it. Longer experiments should be performed on several tubes that could be stopped upon sampling maybe with months as intervals to understand the nature of the equilibrium at low temperature.

We observed that, at room temperature (20°C in our laboratory), the system explores many lengths and structures whereas at 40°C only the cyclic dimer seems to thrive. It could be that the cyclic dimer can amplify its formation through interactions strong enough to be present at 40°C but which could be overcome to create longer structures at 20°C. The creation of needles of cyclic dimer supports the fact that they interact in specific ways as shown by the crystal structure of the needles, through a hydrogen bond network which necessitates the intervention of water molecules. Interestingly enough, the thiols in the crystal do not appear to be in an oxidation state and might be protected from disulfide formation due to their arrangement.

There seems to be a difference in precipitate formation depending on the temperature. This may be due to the solubility of the formed structures or their capacity to interact with one another, resulting in more precipitate at 40°C as the cyclic dimer is favored, or could be explained by the fact that higher temperatures lead to a faster oxidization of the system.

The living polymer system already shows promising diversity with just a single monomer. With our current design, switch from structure exploration to sequence exploration can be easily done by modifying the second amino acid in the sequence if the tyrosine is kept. The first experiments in the previous chapter showed that creating try peptides is not a possible alternative as it would lead to DKP compounds but creating monomers with **C**XG**C** or **C**XYG**C,** with X any amino acid and Y depending on whether the tyrosine needs to be kept for UV monitoring or not, would allow in this system a wide variety of sequences.

9. References

1. Lodish, H. F., Berk, A., Krieger, M. & Scott, M. P. *Molecular Cell Biology*. (W.H. Freeman, 2008).
2. Vatamaniuk, O. K., Bucher, E. A., Ward, J. T. & Rea, P. A. A New Pathway for Heavy Metal Detoxification in Animals. *Journal of Biological Chemistry* **276**, (2001).
3. Shirazi, A. N. *et al.* Cysteine and arginine-rich peptides as molecular carriers. *Bioorganic & Medicinal Chemistry Letters* **26**, (2016).
4. Altay, M., Altay, Y. & Otto, S. Parasitic Behavior of Self-Replicating Molecules. *Angewandte Chemie International Edition* **57**, (2018).
5. Bartolec, B., Altay, M. & Otto, S. Template-promoted self-replication in dynamic combinatorial libraries made from a simple building block. *Chemical Communications* **54**, (2018).
6. Sadownik, J. W., Mattia, E., Nowak, P. & Otto, S. Diversification of self-replicating molecules. *Nature Chemistry* **8**, (2016).
7. Liu, P. *et al.* A tris (2-carboxyethyl) phosphine (TCEP) related cleavage on cysteine-containing proteins. *Journal of the American Society for Mass Spectrometry* **21**, (2010).
8. Allison, W. S. Formation and reactions of sulfenic acids in proteins. *Accounts of Chemical Research* **9**, (1976).
9. Ghadiri, M. R., Granja, J. R., Milligan, R. A., McRee, D. E. & Khazanovich, N. Self-assembling organic nanotubes based on a cyclic peptide architecture. *Nature* **366**, (1993).
10. Horne, W. S., Stout, C. D. & Ghadiri, M. R. A Heterocyclic Peptide Nanotube. *Journal of the American Chemical Society* **125**, (2003).
11. Leclair, S. *et al.* Micrometer-Sized Hexagonal Tubes Self-Assembled by a Cyclic Peptide in a Liquid Crystal. *Angewandte Chemie International Edition* **43**, (2004).
12. Seebach, D. & Matthews, J. L. β-Peptides: a surprise at every turn. *Chemical Communications* (1997) doi:10.1039/a704933a.
13. Hamley, I. W., Dehsorkhi, A. & Castelletto, V. Self-assembled arginine-coated peptide nanosheets in water. *Chemical Communications* **49**, (2013).
14. Sun, M. *et al.* Prediction of reversible disulfide based on features from local structural signatures. *BMC Genomics* **18**, 279 (2017).

Chapter 4. Modelling the living system

1. Introduction

We now have access to the thermodynamic data for the exchange reaction and designed a system that could explore different lengths and structures. Even if it is the same exchange reaction governing the living polymer system than in kinetic assay discussed in chapter2, it is likely that due to the nature of peptides and amino acids, the reaction constants are subject to change as interactions arise. We see that certain species are favored whereas others are but transitory, and that the species created change with temperature with the system exploring longer sequences at 20°C. To understand some mechanisms, we turned towards modeling. Modeling the system as it evolves towards equilibrium is also a first step towards setting up predictive algorithms for its out-of-equilibrium evolution which would then just have to consider the diffusion between species across the temperature gradient.

In a chemical system, two reactants interacting with each other to yield products is a random event occurring with a certain probability. The evolution over time of the concentrations of the different components then corresponds to a stochastic process, a family of random variables corresponding to the punctually occurring reactions. When the system includes multiple reactions linked with one another through their involved species, we use the term stochastic reaction network. Such is the case for the living polymer system, where all our reactions are reversible, and every polymer can react through N-methyl-cysteine mediated exchange with any other.

In this chapter, we will first give an overview of the modeling of reaction networks by discussing the two main types of modelling algorithms, stochastic and deterministic. In order to choose between the two, we then examine the complexity of the living polymer system by looking at the number of species it may create as well as the number of reactions a model algorithm must consider to have some insight on the scalability to longer species before the system even possess different monomers. A model based on the Euler method is then established for the living polymer and discussed. Finally, some modelling tools to obtain thermodynamic data in proteins which were implemented are briefly addressed.

All the algorithms written as part of this chapter were done in MATLAB can be accessed at the following link: https://github.com/Snorana/Programs-in-Designing-a-peptide-system-for-out-equilibrium-chemistry-exploration.git

2. Forwards stochastic kinetic algorithms

The generation of the living polymer system from the monomer, is but a series of stochastic events. A monomer will eventually react with another one creating a dimer, the dimer will react with another monomer and form a trimer and the N and C terminus might react together to form the cyclic trimer. All those events are fully described by the set of reactions that govern the system, the exchange reactions. Let us consider all the species that in a chemical system and number them from 1 to N. If we note the concentration of a compound i X_i then, if we consider the vector $\boldsymbol{X}(t) = [X_1(t), X_2(t), \ldots . X_N(t)]^T$ as the representative of our system at the time t, if a reaction j occurs and modifies our system then the new system will be :

$$X(t) = X(t^-) + \nu_j$$

With ν_j the stochiometric vector of the reaction j and t^- the time immediately before the reaction took place. ν_j can be obtained by subtracting the stochiometric coefficients of the reactants and adding those of the products. As an example, if I consider a simple system where there are 4 species and the following reactions between them:

$$\begin{array}{lcl} R_1 & \mathrm{A + B \rightarrow C} & k_1 \\ R_2 & \mathrm{C \rightarrow A + B} & k_2 \\ R_3 & \mathrm{C + A \rightarrow D} & k_3 \end{array}$$

Then the stochiometric vector for the reactions are:

$$\begin{array}{c|ccc} & \nu_1 & \nu_2 & \nu_3 \\ \mathrm{A} & -1 & 1 & -1 \\ \mathrm{B} & -1 & 1 & 0 \\ \mathrm{C} & 1 & -1 & -1 \\ \mathrm{D} & 0 & 0 & 1 \end{array}$$

If we then consider the capacity of the system to change by a specific reaction, then the probability of such a reaction to happen is:

$$P(reaction\ i\ occurs\ in\ [t, t+dt]) \propto dt \times combinations\ in\ \boldsymbol{X}(t) for\ reaction\ j$$

We then define the *propensity function* $a_j(X(t))$ so that:

$$a_j\big(X(t)\big) = constant\ \times combinations\ in\ \boldsymbol{X}(t) for\ reaction\ j$$

Which gives us:

$$P(reaction\ i\ occurs\ in\ [t, t+dt]) = a_j\big(X(t)\big)dt$$

The constant is the kinetic rate, the probability that this process will take place. The total combinations correspond to the possibilities for the reaction, if the reaction involves only one species it corresponds to the concentration of this species as each one could react, if it involves two it is the product of both concentrations as each molecule from reactant one has the number of molecules of reactant B combinations.

We can then write the propensity functions, going back to our previous simple system:

$$a_1 = k_1[A(t)][B(t)]$$
$$a_2 = k_2[C(t)]$$
$$a_3 = k_3[A(t)][C(t)]$$

From this ruleset we can describe the evolution of our system over certain time interval $[0, T]$. There are different approaches to this, we will describe two of them.

The exact stochastic approach consists in identically following the probabilities of evolution of the system. To this end, the system must be monitored very often, i.e., a very small timestep dt, such that multiple reactions are not likely to take place in such a short time. This allows the reactions to be separate punctual occurrences with independent probabilities.[1] This means, considering a system with M reactions, the probability of a reaction occurring between time t and the time t and our small dt is the sum of the probability all the reactions:

$$P(a\ reaction\ occurs\ in\ [t, t+dt]) = P(reaction\ 1\ occurs\ in\ [t, t+dt]) + P(reaction\ 2\ occurs\ in\ [t, t+dt]) + \cdots + P(reaction\ M\ occurs\ in\ [t, t+dt])$$

Which going back to the definition of our propensity functions is:

$$P(a\ reaction\ occurs\ in\ [t, t+dt]) = \sum_{i=1}^{M} a_i\big(X(t)\big)\, dt$$

Let us define a_0 such as:

$$a_0\big(X(t)\big) = \sum_{i=1}^{M} a_i\big(X(t)\big)$$

We then have:

$$P(a\ reaction\ occurs\ in\ [t, t+dt]) = a_0(X(t))dt$$

If we know that a reaction happens in our dt interval, then we get:

$$P(a\ reaction\ occurs\ in\ [t, t+dt]) = 1 = a_0\big(X(t)\big)dt$$

Which gives:

$$dt = \frac{1}{a_{0(X(t))}}$$

This means that the probability for reaction i, to be the one occurring during this interval is:

$$P(reaction\ i\ occurs\ in\ [t, t+dt]) = \frac{a_i(X(t))}{a_0(X(t))}$$

We then just must choose a timestep, Δt for which we know one reaction will occur and then update the system with a reaction chosen with the previously described probability.

The following demonstration was given by Gillespie in ref.[2] If we consider the probability of having the next reaction occur in the small timestep $[t+\tau, t+\tau+d\tau]$ knowing the state of the system at time t

If we note $P_0(\tau)$, the probability that no reactions occur in the interval $[t, t+\tau]$. Then the probability that no reaction takes place in $[t, t+\tau]$ and then a reaction takes place in the following time interval $[t+\tau, t+\tau+d\tau]$ is:

$$P(a\ reaction\ occurs\ in\ [t+\tau, t+\tau+d\tau]) = P_0(\tau)a_0(X(\tau))d\tau$$

If we then consider:

$P_0(\tau + d\tau) \equiv$ no reaction occur in the interval [t,t+τ+dτ]
$P_0(\tau + d\tau)$
$= P($no reaction occur in the interval [t,t+τ] ∩ no reaction occur in the interval [t+τ,t+τ+dτ])
$P_0(\tau + d\tau) = P_0(\tau) \times P_0(d\tau)$

We defined previously the opposite probability from $P_0(d\tau)$:

$$\overline{P_0(d\tau)} \equiv a\ reaction\ occurs\ in\ [t+\tau, t+\tau+d\tau]$$
$$\overline{P_0(d\tau)} = a_0(X(\tau))d\tau$$

$$\text{thus} \quad P_0(d\tau) = 1 - a_0(X(\tau))d\tau$$

This gives us:

$$P_0(\tau + d\tau) = P_0(\tau)(1 - a_0(X(\tau))d\tau)$$
$$\frac{P_0(\tau + d\tau) - P_0(\tau)}{d\tau} = -P_0(\tau)a_0(X(\tau))$$
$$P_0(\tau) = \exp(-\tau a_0(X(\tau)))$$

Which in turn means that:

$$P(a\ reaction\ occurs\ in\ [t+\tau, t+\tau+d\tau]) = a_0(X(\tau)) \exp\left(-\tau a_0(X(\tau))\right) d\tau$$

We can therefore see that τ follows an exponential distribution of parameter $a_0(X(\tau))$.

Considering a system with M reactions, and x_0 the state vector with the initial concentrations of each species, the exact stochastic simulation algorithm would then be:

1) Initialize, t=0 and state vector $X = x_0$
2) calculate propensities: $a_1(X) \dots a_M(X)$ and $a_0(X(t)) = \sum_{i=1}^{M} a_i(X(t))$
3) Generate the time step towards next reaction: $\Delta t \approx \exp(-a_0(X(t)))$
4) If $t + \Delta t > T$ the simulation is over

5) Choose a reaction with probabilities: $P(reaction\ i) = \frac{a_i(X(t))}{a_0(X(t))}$
6) Update the system state: $X = X + \nu_i$ and $t = t + \Delta t$ and go back to 2)

This method, also named the *Gillespie direct method*, is the most exact method of calculating the evolution of a stochastic system. There are some other possibilities such as the *next reaction method* [3] which computes a time between reactions for each reaction and actualizes the system every time a reaction timestep is reached therefore removing the random event selection. But these stochastic exact methods become computationally intractable for complex systems with many reactions as the timesteps become very small to account for all the possible reactions occurring.[1]

Improving the computation time of more complex stochastic systems comes at the cost of accuracy. Instead of evolving the system one reaction at a time, the system can be evaluated at a discrete timestep τ and all the events which would have occurred in between steps $[t, t+\tau]$ are actualized. The punctual events corresponding to a reaction happening described previously, follow an exponential distribution. This means that these events form a Poisson process in time, and the number of events in a time interval of a Poisson process follow a Poisson distribution.[4] In other words, Y_j the number of times reaction j occurred in $[t, t+\tau]$ follow a Poisson distribution of parameter $a_j(X(t))\tau$; $Y_j \sim Po(a_j(X(t))\tau)$. This allows to dissect the time interval modelled into a chosen amount of timesteps. The *tau-leaping method*[5] for a system with M reactions, and x_0 the state vector with the initial concentrations of each species, would then be:

1) Initialize, t=0 and state vector $X = x_0$
2) If $t + \tau > T$ the simulation is over
3) calculate propensities: $a_1(X) \dots a_M(X)$
4) Calculate how many times each occurred: $Y_j \sim Po(a_j(X(t))\tau)$ for $j = 1,2, \dots M$
5) Update the system $X = X + Y_1\nu_1 + \cdots + Y_M\nu_M$ and $t = t + \tau$ and go back to 2)

Choosing which type of modeling algorithm to use then depends on the complexity of the system. There are other possibilities such as solving the chemical master equation as was done in chapter 2 for the bimolecular reversible case. But for bigger systems more often than not it is unsolvable.[1]

3. Forwards deterministic modelling

Contrary to stochastic modelling, deterministic models always yield the same result by always considering the average amount of reactions taking place during each time step and updating the system. This yields to approximation errors as in the case of the τ-leaping algorithm if the time step is taken too big, but also tends to erase the effects observed for small concentrations of molecules inside large networks. The main advantage of deterministic modelling lies in its low resource requirement compared to stochastic modelling as probability distributions need not to be computed for each reaction step, saving computation time. Here we give the example of the Euler method to showcase deterministic models. If we consider a chemical reaction

$$A + B \rightarrow AB \text{ with constant } k_1$$

Then the evolution of the system through this reaction is the following:

$$\frac{d[A]}{dt} = -k_1[A][B]$$

$$\frac{d[B]}{dt} = -k_1[A][B]$$

$$\frac{d[C]}{dt} = k_1[A][B]$$

This means that if we consider a small enough time step, we can modify the concentrations of each constituent in the system by a small quantity created:

$$[A](t + \Delta\mathrm{t}) = [A](t) + d[A] \approx [A](t) + \Delta A = [A](t) - k_1[A](t)[B](\mathrm{t})\Delta t$$

$$[B](t + \Delta t) = [B](t) - k_1[A](t)[B](\mathrm{t})\Delta t$$

$$[C](t + \Delta t) = [C](t) + k_1[A](t)[B](\mathrm{t})\Delta t$$

We can see here that the number of occurrences for the reaction in our Δt time step is $k_1[A](t)[B](\mathrm{t})\Delta t$ which is the propensity of the reaction times the timestep, the parameter for our Poisson distribution in τ-leap. A random variable following a Poisson distribution of parameter λ has an expectancy of λ, making the parallel with an averaging of the τ-leap method.

Now there could be in our system multiple reactions that influence a single species, for example lets add to the previous reaction another one so that it becomes:

$$A + B \rightarrow AB \quad k_1$$
$$B + AB \rightarrow BB + A \quad k_2$$

The B species in consumed through two different processes and A is created through one and consumed by another making their variations with time:

$$\frac{d[B]}{dt} = -k_1[A][B] - k_2[AB][B]$$

$$\frac{d[A]}{dt} = -k_1[A][B] + \mathrm{k}_2[AB][B]$$

$$[B](t + \Delta t) = [B](t) - k_1[A](t)[B](\mathrm{t})\Delta t - k_2[AB](t)[B](\mathrm{t})\Delta t$$

$$[A](t + \Delta t) = [A](t) - k_1[A](t)[B](\mathrm{t})\Delta t + k_2[AB](t)[B](\mathrm{t})\Delta t$$

But instead of calculating the entire modification at a time, this can be done in steps going through one reaction at a time thanks to the commutativity of additions. We can thus actualize the concentration change of A through reaction 1, store this new value, calculate the change of A through reaction 2, add it to the stored value for reaction one and then add the result to the concentration of A to obtain the concentration at the new timestep.

$$[A](t + \Delta t) = [A](t) + d[A]_{reaction\,1} + \mathrm{d[A]}_{\mathrm{reaction\,2}}$$

4. Assessing the complexity of our system

After having discussed the two main "classes" of algorithms for the modeling of chemical reaction networks, we now turn towards the complexity of the problem (i.e. the number of computation step required by our program for it to run), if the complexity is low enough then the precise stochastic algorithms would be favored whereas a high complexity would mean the less resource-hungry deterministic approach is better suited. In order to explore the complexity of our system we wanted to know the number of different chemical species and reactions occurring. We will make the following hypothesis: **a polymer can only "attack" another polymer with its C-terminal N-methyl-cysteine**. This means that we consider that the sulfur of the N-methyl-cysteines inside the sequence cannot lead to transthioesterification with other polymers.

In its current state, our system only has one monomer, which can elongate, cyclize, and dissociate. Let us first consider the total number of species. In our case if the maximum observed length of a polymer is M, then the maximum number of compounds is 2M-1 since our system possesses each length but the monomer under two possible states, linear or cyclic.

Now if we consider the mechanism behind the formation of the longest observed linear polymer. In our case it is the heptamer. This polymer can be created through many different reactions that we will divide in two categories: end condensation reactions and constructive separations. End condensation is a classic of polymer chemistry, two polymers of length a and b react with each other from their ends, in our case C-terminal N-methyl-cysteine with the N-terminal from the other to create a polymer of length a+b. It is noteworthy that a reacting with b is not the same as b reacting with even though it leads to the same result. Indeed, in one case we have the N-terminal cysteine of an a-mer react with the C-terminal one of a b-mer, and because those molecules are peptides, they might not have the same structure and thus reactivity.

Going back to the formation of M, it could be the (M-1)-mer reacting with a monomer, or the dimer with the (M-2)-mer and so on as the (M-1)-mer reacting with the monomer is not the same reaction as the monomer reacting with the (M-1)-mer. If we take M = 5 this gives us the tetramer with the monomer, the trimer with the dimer, the dimer with the trimer and the

monomer with the tetramer. This means that the end condensation reactions in the system gives us M-1 reactions able to create the maximum length M polymer, as each (M-i)-mer reacts with the i-mer to form M.

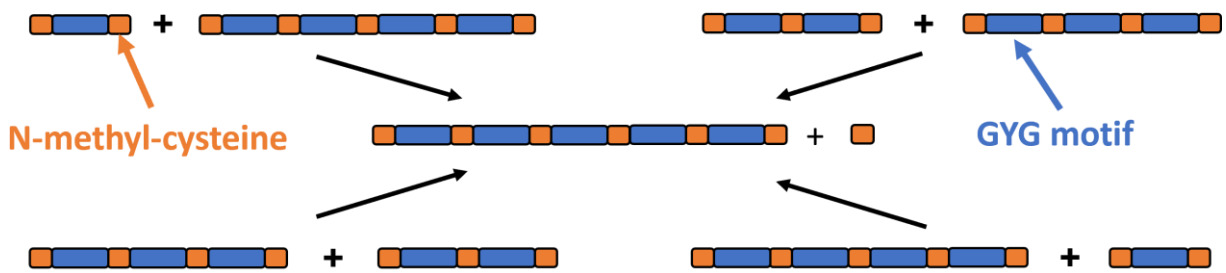


Figure 4-1 End condensation reactions responsible for the pentamer formation

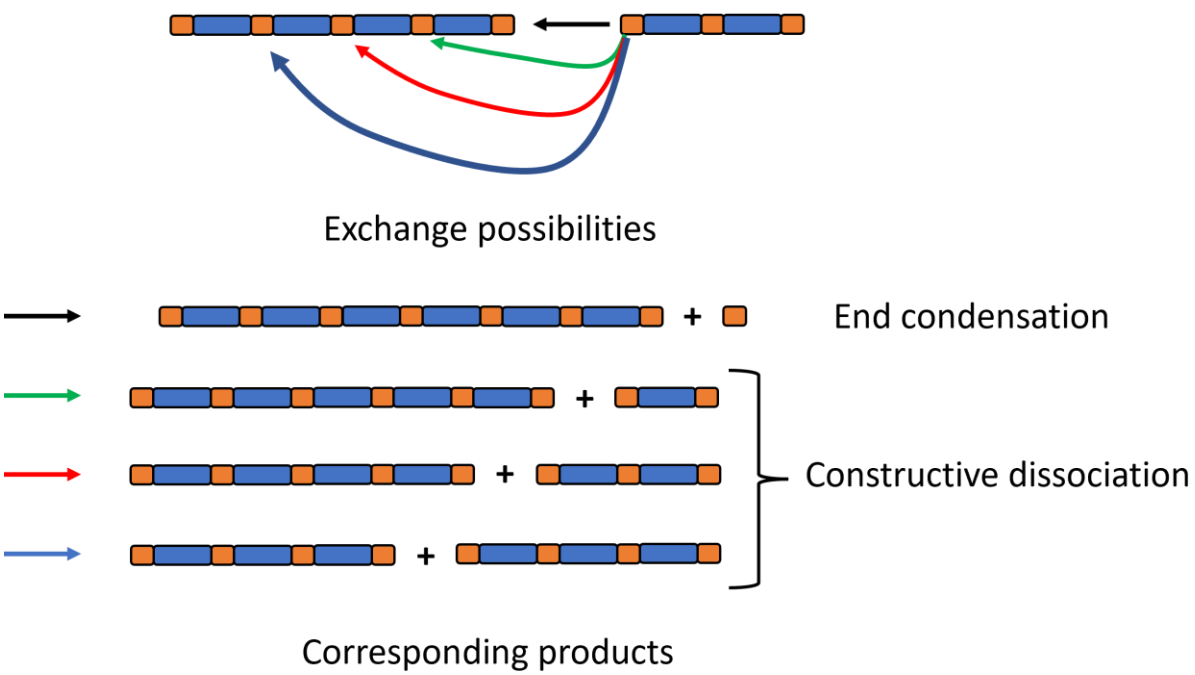


Figure 4-2 Example of possible exchange reactions between the linear dimer and the linear tetramer depending on which N-methyl-cysteine bond gets attacked by the dimer.

In the living polymer system, the bond between monomer subunits is reactive. This means that they can react with other linear polymers in the system. For example, a tetramer can react with a dimer through end condensation to create the hexamer. But if it reacts at the bond between the two subunits in the dimer. This would dissociate the dimer and elongate the tetramer, thus creating a pentamer and releasing a monomer. If we consider hexamer formation this means that the pentamer can form it with every specie in the system (including another pentamer). The tetramer can form it with himself, and a trimer by ligating to two subunits and releasing respectively a dimer and a monomer. To generalize each i-mer can create an M-mer by reaction with a k-mer if $k + i \geq M$ by either constructive separation if $k + i > M$ or by end

condensation if $k + i = M$, giving us $M - i \leq k \leq M$. By summing over all the lengths up to length M-1 we have considering only linear forms:

$$\sum_{i=1}^{M-1} \sum_{k=M-i}^{M} 1 = \sum_{i=1}^{M-1} i + 1 = \frac{M(M-1) + 2M - 2}{2}$$

For a system going only up to a trimer this yields 4. As the possible reactions are the monomer with the dimer, the dimer with the monomer, the dimer with the dimer (releasing a monomer) and the dimer with a trimer (releasing a dimer).

If we now consider the cyclic forms the system can adopt, then the species of length M can only be formed if a cyclic (M-i)-mer reacts with the i-mer opening the cycle and the cyclic M-mer can react with an N-methyl cysteine opening the cycle and generating the linear M-mer. This means that there are an extra M-1 reactions (as the monomer cannot cyclize) that can generate the M-mer, giving us a total number of reactions of:

$$\frac{M(M-1) + 4M - 4}{2}$$

This gives us 7 possible ways to create the linear trimer if the trimer is the longest species observed as depicted below:

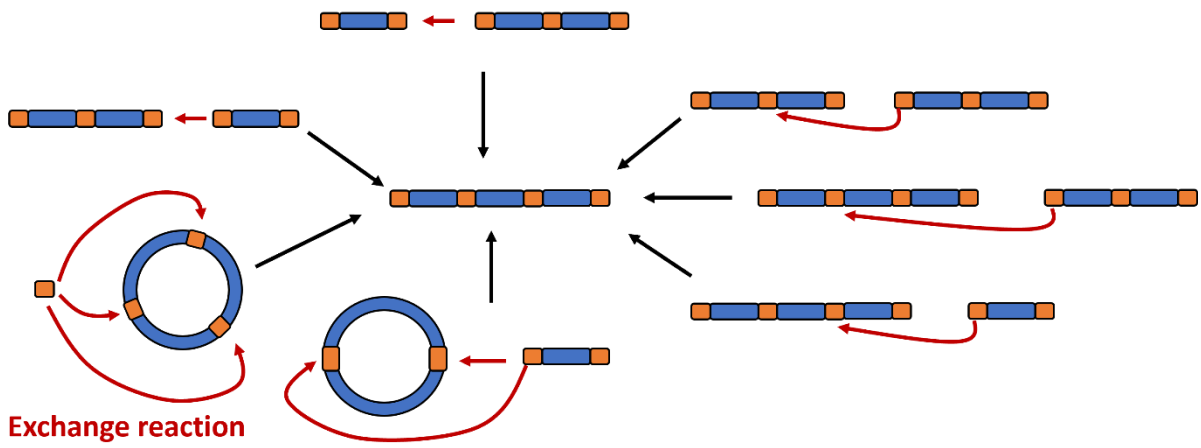


Figure 4-3 The seven possibilities for creating the linear trimer if it is the longest species observed in the system.

If we now consider a smaller polymer, an N-mer with $N < M$. Let us once again first focus on only the linear polymers. There are now three ways to create the N-mer. It can either be created through elongation of species (both end condensation of constructive dissociation), or dissociation of longer species who get attacked by other polymers and the released sequence is of length N.

If we consider the i-mers, $i < N$, the polymers shorter than N which can elongate to create N. Then they can react with any other k-mer in the system so long as both have enough subunits to create N, i.e., $i + k \geq N$. We then have:

$$\sum_{i=1}^{N-1}\sum_{k=N-i}^{M} 1$$

For longer species than N-mers which will dissociate, $i > N$, there are two types of dissociations which can take place in our system. First a free N-methyl-cysteine can dissociate a sequence at any given subunit junction. This means that for each i-mer there are two possibilities of dissociation from N-methyl-cysteine which yields an N-mer. As an example, a pentamer can yield a tetramer by either dissociation after the first subunit, or before the last one. The corner case being for polymers of length 2N which can only react with N-methyl-cysteine in its middle. We then have:

$$2(M-N) - \delta_{N\leq\frac{M}{2}}$$

We $\delta_{N\leq\frac{M}{2}}$ as being 1 if the 2Nmer exists and 0 if not.

Lastly, the N-mer can be created by dissociation of longer i-mer species, $i > N$, by other ones, k-mers. The main constraint for this process is that the k-mer which dissociates the i-mer cannot get elongated past a length of M as M is the longest species observed, thus $i - N + k \leq M$. This means that number of such processes are:

$$\sum_{i=N+1}^{M}\sum_{k=1}^{M+N-i} 1$$

We cannot, however, simply add these three contributions as if we look at the first contribution and the last ones some reactions have been counted twice. If we consider a system where M =4, the trimer can be created by the dimer elongating into a trimer by reacting with a tetramer, considered by contribution 1, but the tetramer can be dissociated into a trimer by the dimer, considered by contribution 3, and those are the same reaction, a dimer dissociating the tetramer. This case happens when an i-mer reacts with a k-mer and $i + k = 2N$. If we look at the values which I and k can take:

$$N + 1 \leq i \leq M$$
$$1 \leq k \leq M + N - i$$

Meaning that for $k + i$ we have:

$$N + 2 \leq k + i \leq 2M + N - i$$

Since the maximum value I can achieve is M:

$$N + 2 \leq k + i \leq M + N$$

If for I there is a value of k, for which $k + i = 2N$ we then must prove that:

$$N + 2 \leq 2N \leq M + N$$

Since $N \leq M$, the right part is always true. Which gives us that if $N \geq 2$ there is for each i, a k for which $k + i = 2N$, meaning that for each i we are counting one extra reaction. We must therefore subtract one for each i, which means that considering only linear species so far, the number of reactions which can produce the N-mer for $N \geq 2$ are:

$$\sum_{i=1}^{N-1} \sum_{k=N-i}^{M} 1 + 2(M-N) - \delta_{N \leq \frac{M}{2}} + \sum_{i=N+1}^{M} \sum_{k=1}^{M+N-i} 1 - (M-N)$$
$$= \frac{M^2}{2} - \frac{M}{2} - N^2 + N + MN - 1 - \delta_{N \leq \frac{M}{2}}$$

If we use the previous result to consider the number of ways, we have of forming M, we obtain:

$$\frac{M^2}{2} - \frac{M}{2} - M^2 + M + M^2 - 1 = \frac{M^2 + 3M - 2}{2} = \frac{M(M+1) + 2M - 2}{2}$$

Which is what we obtained previously.

When we consider the number of cycle-opening reactions which can form the N-mer. For each cycle of length i with $2 \leq i \leq N$ there is only one species of length $N - i$ which can perform this reaction. Therefore, this process adds N-1 reactions.

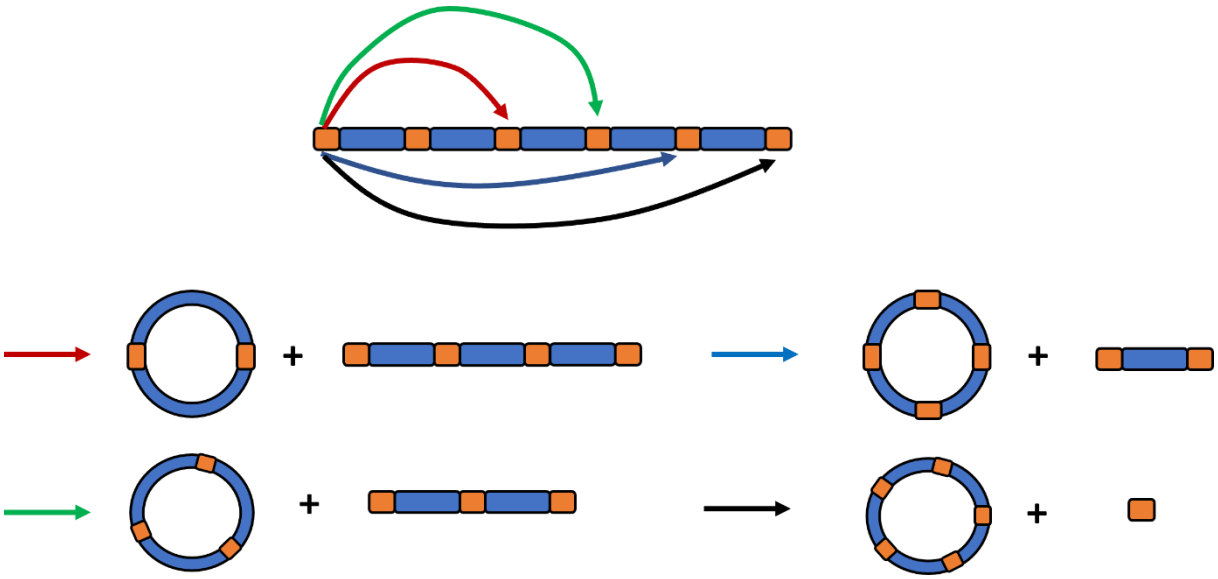


Figure 4-4 Possible cyclization reactions for the pentamer.

Linear N-mers can also be created by cycle formation by expulsion of a linear fragment. This process is possible only for i-mers with $i > N$. Since the smallest cycle observed in our system is of length 2. This means that $2 \leq i - N < M - N$ meaning that $2 + N \leq i \leq M$. This yields M-N-1 possibilities. We then have the number of ways to create a linear N-mer $(N \geq 2)$ in the living polymer system, and add back the M=N case which is subtracted one too many times:

$$\frac{M^2}{2} + \frac{M}{2} - N^2 + N + MN - 3 - \delta_{N \leq \frac{M}{2}} + \delta_{M=N}$$

Finally, if we consider the number of reactions which can create a circular specie of length N. This could be a linear polymer of length N, or a linear polymer of length i, with $i \geq N$, cyclizing and releasing an $(i - N)$-mer. This means that only linear species of length greater or equal to N can form a cyclic N giving us:

$$M - N + 1$$

If we consider the total amount of reactions taking place in a living polymer system with M being the longest species observed:

$$\frac{M^2}{2} + \frac{7M}{2} - 5 + \sum_{i=2}^{M} \left\{ \frac{M^2}{2} + \frac{M}{2} - N^2 + N + MN - 3 - \delta_{N \leq \frac{M}{2}} \right\} + \sum_{i=2}^{M} \{M - i + 1\}$$

With the first part representing the contributions for the monomeric specie.[1]

5. The simple Euler method for the living polymer

Given the number of possible reactions in our system, it seemed that a deterministic approach suits the system best. Since every specie in our system can react with any other specie the structure of our algorithm is the following:

1) Initialize the system:
    a. longest observed specie is M
    b. Run time is T
    c. L= [mono, l-di, l-tri, …l-M] initial linear species concentrations
    d. C = [c-di, ...c-M] initial cyclic species concentration
    e. NMC the initial concentration in N-methyl-cysteine
2) If $t + \Delta t > T$ stop the program
3) Create copies of C, L and NCM: C^* ; L^*; NMC^*
4) For each i specie in L, go through each j specie in C and L (for L, $j \geq i$), calculate the evolution caused by these reactions:
    a. $L_i + L_j$ can yield all polymers from $L_{i+1} + L_{j-1}$ to L_{i+j} or $L_M + L_{j-(M-i)}$ and in the case $i + j$ release an NMC, all concerned concentrations of species are incremented by $k_1 L_i L_j \Delta t$ and $L_i, L_j$are decreased by this amount times the number of possible reactions in $L^*; NMC^*$. If $i = j$ then increment half of the amount (as the reaction will be counted twice by the algorithm).
    b. $L_i + C_j$ yields L_{i+j+1} [2] or does not react if $i + j > M$ and the concentration of the linear specie L_{i+j} is incremented by $j k_1 L_i L_j \Delta t$ as there are j different

[1] Use of the formula to create linear N-mers with N=1 without substracting the (M-N) due to reactions not being counted twice for the monomer.

[2] The +1 comes from the fact that the index of C is one less compared to the size of the species it represents as the first (j=1) is the dimer, the second (j=2) is the trimer. The same goes for the -1 in the following c) step.

possible junctions for attack and the concentrations of L_i and C_j are decreased by that amount in $L^*; C^*$

c. L_i can then undergo cyclization with a constant k_{C_i} so it is decreased by $k_{C_i} L_i \Delta t$ and NMC and C_{i-1} are increased by that amount in $L^*; C^*; NMC^*$

5) The NMC reacts with all species in L and C, with constant k_{-1} :
   a. $NMC + L_j$ giving $L_{i-k} + L_k$ with k going from 1 to $i-1$, the concentrations are respectively decreased and increased of $k_{-1} NMC L_j \Delta t$ for each reaction in $L^*; NMC^*$
   b. $NMC + C_j$ gives L_{j+1} so the concentrations are modified accordingly with k_{-1} as the kinetic rate in $NMC^*; C^*; L^*$
6) Update the system: $C = C^*; L = L^*; NMC = NMC^*$ and go back to step 2)

This algorithm can then model, the behavior of the system based on specific inputs:

- The initial concentrations
- The longest specie observed
- The kinetic constants for the exchange k_1 and dissociation k_{-1}
- The cyclization constant of the specie k_{C_i}

The assumption is made here that the exchange kinetics are always the same for all polymers and all structures.

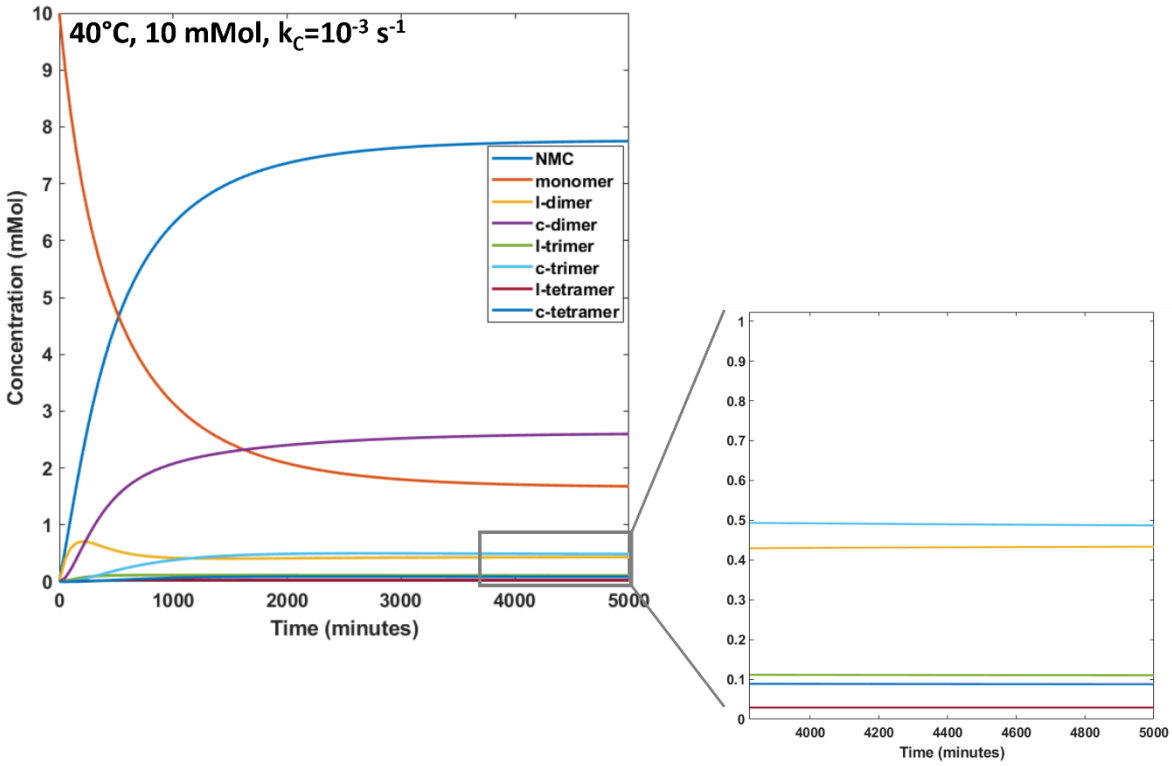


Figure 4-5 Evolution of the living polymer generated for 10mMol and values of k_1, k_{-1} obtained for the exchange reaction at 40°C. The values of k_{C_i} used are $10^{-3} s^{-1}$ and $\Delta t = 1\ s$.

The Euler model can reproduce the abundance of cyclic dimers in concentrations close to the ones observed, here 2.8 mMol against the 4 mMol observed at 5000 minutes. This is closely linked to the values of the cyclization constants, as is the evolution of the linear dimer. With the shown trajectories, we observe the transitory increase in linear dimer and the following decrease as time advances. The height of this peak increases as the cyclization constant for the cyclic dimer decreases but doing so yields to lower concentrations of cyclic dimer in longer times. This can either be because of auto amplifying effects for the cyclic dimer or because the kinetic constants for the exchange reaction do not apply directly to the living polymer system. However, now that a predictive model has been set up it can be used to infer the kinetic parameters with comparison to the experimental data.

6. Kinetic inference algorithms

Now that we have seen how to model the evolution of a system, the forwards problem, we will focus on how to infer the kinetic data from observations, the backwards problem. The experimental data obtained of the system $Y_{obs} = [Y(t_1), Y(t_2), \dots Y(t_n)]$ at different discrete points in time $t_1, t_2, \dots t_n$, are never entirely accurate and subjective to uncertainty and noise ξ but they are linked to the observation of the state vector[6–8] :

$$Y(t) = AX(t) + \xi$$

A is a $K \times N$ matrix here and represents the fact that not all the species are always observable. Sometimes only a few or a combination of them are. In our case, we will not consider noise as we have no model for it and, apart from the linear trimer, all species can be accurately observed in solution for early timesteps.

If we consider $\theta = [k_1, k_2, \dots k_n]$ the vector containing values for the kinetic parameters of our reactions, then the value we are interested in is $p(\theta|Y_{obs})$ which is the probability of the having this vector of kinetic constant for our experimental data. Through the Bayes theorem this gives us:

$$p(\theta|Y_{obs}) = \frac{p(Y_{obs}|\theta)p(\theta)}{p(Y_{obs})}$$

$p(\theta|Y_{obs})$ is also referred to as the posterior distribution; $p(\theta)$ is the probability of obtaining this parameter configuration before taking the data into account and is user defined, it depends on how we are choosing our θ candidates, it is also referred to as the prior distribution; $p(Y_{obs}|\theta)$ is the likelihood, the probability of obtaining our experimental data given our θ candidate; and finally $p(Y_{obs})$ is the evidence, the probability of observing our experimental data over all possible parameters and acts as a normalization constant.

If the chemical master equation can be solved, then the likelihood can be calculated. However, in most cases this is not possible and likelihood-free methods must be used. ABC methods are the intuitive and one we used for the bimolecular reversible system. It consists of measuring a discrepancy between the modeled data and the experimental data, the parameters yielding small discrepancies are kept as the best solutions. In other words, we define an acceptance parameter ϵ, such that if a given metric ρ between the observed data Y_{obs} and simulated data S_{obs} is

underneath the acceptance $\rho(Y_{obs}, S_{obs}) \leq \epsilon$ then the kinetic guess used to create the simulated data is kept as a candidate. The above equation then becomes:

$$p(\theta|Y_{obs}) \approx p(\theta|\rho(Y_{obs}, S_{obs}) \leq \epsilon) = \frac{p(\rho(Y_{obs}, S_{obs}) \leq \epsilon|\theta)p(\theta)}{p(\rho(Y_{obs}, S_{obs}) \leq \epsilon)}$$

Different metrics can be used, the one we used in the bimolecular reaction being:

$$\rho(Y_{obs}, S_{obs}) = \sum_{i=1}^{n} |Y(t_i) - S(t_i)|$$

Simpson et al.[1] propose:

$$\rho(Y_{obs}, S_{obs}) = \left[\sum_{i=1}^{n} \left(Y(t_i) - S(t_i)\right)^2\right]^{\frac{1}{2}}$$

The algorithm to infer the parameters then becomes:

1) initialize index $i = 0$;
2) generate a candidate $\theta^* \sim p(\theta)$
3) Generate the simulated data, $S^*_{obs}(\theta^*)$
4) if $\rho(Y_{obs}, S^*_{obs}) \leq \epsilon$ then accept $\theta_\epsilon^{i+1} = \theta^*$ and $i = i + 1$
5) if $i = m$ terminate else go back to step 2

This enables to create m guesses within the input estimate. The distributions of the guesses can then be used to calculate the parameters. The starting values for epsilon can be chosen high at first when the distribution for generating the guesses is very broad. Then the distribution obtained can be used as the guess generator for smaller values of epsilon, refining the guesses in the process.

The algorithm for the kinetic inference through this ABC process for the living polymer is almost finished at the time of writing. Optimization is however necessary when it comes to giving initial guesses to the kinetic parameters which are so far considered to be $k_1; k_{-1}; k_{C_i}$.

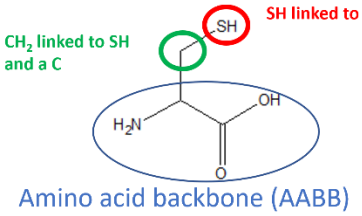
7. Determining energy of unfolded peptide sequences

Ultimately, the living polymer system will feature different monomers with different sequences of amino acids in between the N-methyl-cysteine in order to generate on top of multiple lengths and structures, multiple sequences. We then developed tools to anticipate certain properties of the different structures synthesized. Since we do not have any information on the possible folds for our structures, these were investigated for unfolded proteins.

At the core of these tools, lies the principle of group additivity. Meaning that the thermodynamic values of the entire molecule correspond to the sum of the values of its constituents[9,10]. If we consider a thermodynamic value Ξ such as $\Delta G_f^\circ, \Delta H_f^\circ, S_{P_r,T_r}^\circ$. Then we can consider that:

$$\Xi_{molecule} = \sum_{i\ in\ [constituents]} n_i \Xi_i$$

With n_i corresponding to the amount of time the constituent i is present in the molecule. An example for cysteine is given in figure 4-. The cysteine can be broken into three different constituting groups. An amino acid backbone (denoted AABB), and the sidechain divided in a CH_2 linked to a carbon on one side and a sulfur on the other, and a sulfur linked to carbon and a hydrogen. The values for the amino acid backbone are the ones calculated in ref. and the ones taken for the side groups come from ref [11].



$$\Delta G^o_{f\ cystein} = \Delta G^o_{f\ S-(C)H} + \Delta G^o_{f\ C-(H)_2(C)(S)} + \Delta G^o_{f\ AABB} = -269\ kJ.mol^{-1}$$
$$\Delta H^o_{f\ cystein} = \Delta H^o_{f\ S-(C)H} + \Delta H^o_{f\ C-(H)_2(C)(S)} + \Delta H^o_{f\ AABB} = -377\ kJ.mol^{-1}$$
$$\Delta S^o_{f\ cystein} = \Delta S^o_{f\ S-(C)H} + \Delta S^o_{f\ C-(H)_2(C)(S)} + \Delta S^o_{f\ AABB} = 446\ J.mol^{-1}.K^{-1}$$

Figure 4-6 Estimation of thermodynamic values for the cysteine amino acid through group additivity. The values are taken from ref.[10,11]

This can be extended from amino acids to peptides. If we consider a peptide composed of n amino acids, then the group additivity becomes:

$$\Xi_{peptide} = \Xi_{AABB} + (n-1)\Xi_{PBB} + \sum_{i=1\rightarrow n} \Xi_{SC}$$

With SC referring to the amino acid side chain contribution (i.e. the amino acid minus the backbone). If we consider the peptide sequence there is half of an amino acid group at the N-terminus and the other half of the amino acid group at the C-terminus, there are then $(n-1)$ peptide bones, denoted as protein backbone (PBB) between the n amino acids and then the properties of each sidechain are added. An algorithm for this method has been implemented.

In our case we have slightly modified equations since our peptides incorporate N-methyl-cysteines. This means that one of the PBB value gets modified to an N-methyl-cysteine bond value and that the AABB gets modified as at the N-terminus the amine is methylated and we have amine terminated peptides. To account for this, to the value of AABB we subtract the

contribution of a carboxylic acid group and add the corresponding amide contribution. We approximate that the methylation does not affect the amine part of the AABB apart from bringing its own contribution. This leads to the following equation for the monomer:

$$\Xi_{\mathbf{C}\mathrm{GYG}\mathbf{C}} = \Xi_{AABB} + \Xi_{-CH_3} - \Xi_{-COOH} + \Xi_{CONH_2} + 3\Xi_{PBB} + \Xi_{NMC-bond} + \sum_{i\in[C,G,Y,G,C]} \Xi_i$$

In the previous equation all the values are documented[10,11] apart for the NMC bond. If we only use this equation the calculate reaction values between NMC containing species, this is not an issue as the contributions in reactants and products will cancel out and the only thing that will remain is relative stability in solution of the products relative to the reactants (which is mostly based on pH as will be discussed afterwards). The same applies when comparing species with the same number of N-methyl-cysteine links in them (the cyclic dimer with a linear trimer for example). A possibility would be to consider an equilibrium between the peptide bond and the thioester bond with a conversion percent μ. We would then have:

$$\Xi_{NMC-bond} = (1-\mu)\left(PBB + \Xi_{cysteine}\right) + \mu(\Xi_{C-(O)(S)(CH)} + \Xi_{O=C} + \Xi_{S-(CO)(CH_2)} + \Xi_{C-(H)_2(S)(CH)} + \Xi_{C-(H)(NH)(CH_2)(C)} + \Xi_{N-(H)(CH_3)(CH)} + \Xi_{CH_3-(NH)})$$

Another property which was investigated is the charge of peptides given the pH. If we look at the peptide sequence, charge can again be computed through the group additivity principle:

$$Z_{net} = \sum_i n_i \alpha_i Z_i$$

With Z_{net} the total charge, Z_i the charge of group I and α_i being the proportion of ions for this species. Through the Henderson equation, we can link the pH and pKa values to the concentration of species:

$$pH = pKa + \log\left(\frac{[A^-]}{[AH]}\right)$$

$$\frac{[A-]}{[AH]} = 10^{pH-pKa}$$

$$\alpha = \frac{[A^-]}{[AH]+[A-]} = \frac{1}{1+\frac{[AH]}{[A^-]}} = \frac{1}{1+10^{pKa-pH}}$$

Given the pH, the charge for each sidechain and for the protein backbone can be computed to yield the overall charge of the protein. A MATLAB routine calculating charge values for given peptide sequences has also been implemented using the pKa values for amino acids from ref.[12]

8. Conclusion

The living polymer system is a very complex one, not specifically through the number of different compounds that are produced so far with only one monomer but through the number of reactions taking place in the system as all entities can react with one another and in many ways. This has not been a, obstacle so far to simulations as the modelling for the system at 40°C with tetramers as the longest observed species only requires 2 seconds to model the evolution of 3 days and 6 seconds for 20°C with heptamer formation. This makes it so that inference algorithms may allow us obtain values in reasonable amounts of time.0 So far, the only considered approach is similar to the bimolecular reversible reaction and is a uniform probability of intervals around the measured kinetic constants from chapter 2, this may lead to inference algorithms running for a few days to accumulate enough guesses especially at initial low tolerances. This may allow the estimation of certain values especially for the cyclization constants of the polymer species but different kinetic constants for terminal and internal cysteines could also be envisioned.

The group additivity method allows for some informed first guesses when it comes to certain properties of our system to come. Especially when multiple monomers are added in the living polymer setup, this could allow us to forecast different behaviors such as mixing or possible monomer segregation of sequence selection. The charge is also a crucial factor in determining possible interactions whether between the polymers themselves are high charges of the same sign will deter the peptides from interactions but will favor interactions with other opposite charged species in the system which for the target experiment of the temperature gradient setup is a key parameter.

9. Reference

1. Warne, D. J., Baker, R. E. & Simpson, M. J. Simulation and inference algorithms for stochastic biochemical reaction networks: from basic concepts to state-of-the-art. *Journal of The Royal Society Interface* **16**, (2019).

2. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* **81**, (1977).

3. Anderson, D. F. A modified next reaction method for simulating chemical systems with time dependent propensities and delays. *The Journal of Chemical Physics* **127**, (2007).

4. Stirzaker, D. Advice to hedgehogs, or, constants can vary. *The Mathematical Gazette* **84**, (2000).

5. Gillespie, D. T. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics* **115**, (2001).

6. Finkenstädt, B. *et al.* Reconstruction of transcriptional dynamics from gene reporter data using differential equations. *Bioinformatics* **24**, (2008).

7. Golightly, A. & Wilkinson, D. J. Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. *Interface Focus* **1**, (2011).

8. Schnoerr, D., Sanguinetti, G. & Grima, R. Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. *Journal of Physics A: Mathematical and Theoretical* **50**, (2017).

9. Schmidt, C. L. A. Proteins, Amino Acids and Peptides as Ions and Dipolar Ions (Cohn, Edwin J.; Edsall, John T.). *Journal of Chemical Education* **20**, (1943).

10. Dick, J. M., LaRowe, D. E. & Helgeson, H. C. Temperature, pressure, and electrochemical constraints on protein speciation: Group additivity calculation of the standard molal thermodynamic properties of ionized unfolded proteins. *Biogeosciences* **3**, (2006).

11. Domalski, E. S. & Hearing, E. D. Estimation of the Thermodynamic Properties of C-H-N-O-S-Halogen Compounds at 298.15 K. *Journal of Physical and Chemical Reference Data* **22**, (1993).

12. Lide, D. R. *CRC handbook of chemistry and physics*. (CRC press, 1991).

General conclusions and perspectives

We set up to design a reversible polymeric peptide system for out of equilibrium experiments in a temperature gradient. To this end, we first studied the exchange reaction permitted by N-methyl-cysteines. We found that the reaction could be performed over the time scale of days which is not incompatible with the gradient experiment setup, and the reaction enthalpy of $11.4\ kJ.mol^{-1}$ was in good accord with the desired value from the preliminary modelling. It seems that TCEP is the best candidate for reducing agents as it keeps the system clean and if the reaction is to be performed at pH 7 and for short periods of time (over a week) then phosphate buffer saline is a good candidate for pH buffering. The gradient should not go too high as desulfuration by products are observed for temperatures of 60°C. A temperature range of 5 to 50 °C is advised. Unfortunately, the presence of ascorbic acid leads to byproducts, but their concentration should not pose problems if the exchange is executed at pH 7. Further work to understand the nature of the equilibrium at low temperatures should be pursued. Having different tubes which are only open and therefore oxidized after different long periods of time (one every 2 weeks over 2-3 months) is the envisioned process so far, as the sampling seems to be the main reason behind the oxidation of the system. Other buffer systems could also be investigated.

The N-methyl-cysteine exchange reaction was successfully incorporated to create a dynamic polymer system that we dubbed the "living polymer" system. It displays a diversity of lengths and structures which vary with temperature although the variation can only be observed in the precipitate formed. It seems that the system can self-assemble at 40°C leading to the creation of needles. This assembly appears to be linked to an auto-selection mechanism, transforming the distribution of polymer species into an exclusive cyclic dimer population. The exact arrangement of the cyclic-dimers in these needles must be characterized by X-ray diffraction as macrocycle stacking would constitute a great example for α-L-amino acids and may hint as other possibilities for cyclic oligomers of greater size. The monomer can also be modified on its second amino acid to add on top of structure exploration, a sequence exploration for the system.

Theoretical tools have been developed for the living polymer system. The complexity of reactions performed and to be modelled has been quantified and a finite element algorithm capable of representing the evolution of the different oligomers has been developed. Additional work on the inference algorithms may allow the obtention of the kinetic constants for multiple reactions in the system, notably the cyclization reactions. The latter can be linked to flexibility properties of the polymer chains and probabilities for ends to meet. The evolution of kinetic constants over the course of an experiment may also highlight auto-catalytic processes as might

be the case for the cyclic dimer at 40°C. Initial work has begun to forecast the properties and energies of different structures and sequence in anticipation of systems with multiple monomers. Comparison between predicted distributions and observed will correspond to interactions between sequences or compounds in the system. The finite elements model can also be incorporated as is in a diffusion model for a temperature gradient. It would solve the chemistry at each location for each time step, while a diffusion code would account for the transfer processes. This also implies the need to characterize the diffusion constants for all the species in our polymer system or a way to estimate it.

With the living polymer system in its current state, we believe we are close to having a temperature gradient ready system. What needs to be addressed is the apparition of a precipitate. This can help fractionate the system at different temperatures as we saw that the nature and amount of precipitate was temperature dependent. Another monomer with more hydrophilic amino acids such as lysine could also be synthesized and might polymerize further without precipitation.

The interaction between the living polymer and different compounds is also ready to be investigated. Gold nanoparticles are a good candidate to evaluate the impact of thiol interaction on the polymerization of the system. The thiols might interact too much locking the bonds in their peptide bond state and removing the possibility of polymerization if the monomer concentration is not high enough. On the opposite the sequences with highest interactions would stick to the nanoparticles while others desorb and react preferentially leading to efficient coating selection assays.

We believe that the designed system brings a robust way to create DCCs and explore sequences and structure for proteins. The only drawback is the number of incorporated cysteines in the sequence created, but monomer design can help reduce the cysteine content of create oligomers. Since N-methylation is also observed in biological post translation processes to increase the resistance of proteins to proteases, the generated structures for different targets could result in many peptide sequence candidates for selected biological applications.

Experimental section.

1. Materials and instrumentation

a. Solvents and Reagents

Reagent	Supplier
Triethylsilane	Sigma Aldrich
Triisopropylsilane	Sigma Aldrich
Iodomethane	Sigma Aldrich
Sodium Hydride (60% in mineral oil)	Sigma Aldrich
Triphenylmethanol	Sigma Aldrich
Triethylamine	Sigma Aldrich
Phosphate Buffer Saline (PBS)	Sigma Aldrich
Trifluoroacetic acid (TFA)	Fluorochem
1,1,1,3,3,3-Hexafluoro-2-propanol (HFIP)	Fluorochem
Fmoc and Boc protected amino acids	Iris Biotech
tris(2-carboxyethyl)phosphine, HCl	Iris Biotech
dithiothreitol	Fisher Scientific
Sodium Ascorbate	Alfa Aesar
Diammonium citrate	Alfa Aesar
Sodium sulfate	Alfa Aesar
HPLC TFA	Fisher Scientific
UPLC Formic acid (FA)	Fisher Scientific

Table 1 Suppliers for the reagents used in this work.

All reagents and solvents were used without any further purification except for tetrahydrofuran (THF) for the synthesis of N-methyl-cysteine, which was dried a first time on a double column *SolvTech* purification system before use. All glassware was oven dried at 80°C. Column chromatography purification was performed with silica gels of 40-63µm particle size, 230-400 mesh, supplied by Sigma Aldrich, Steinheim, Germany. Thin layer chromatography was performed on 200 µm silica plates of 60 Å-F_{254}. Unless specified, when solvent was removed by rotatory evaporation, the water bath was set at 25°C (notably as to prevent side reactions for amino acids and peptides). When compounds were left to dry overnight, they were connected

to a belt-drive oil vacuum pump with an acetone/dry ice trap to remove remaining volatile compounds (solvent mostly).

b. Instrumentation

Solid phase peptide synthesis was microwave assisted and performed using a CEM Liberty 1^{TM} automated Microwave peptide synthesizer with the standard 25 mL reaction vessel supplied by CEM. ^{1}H and ^{13}C NMR spectra were collected with a Bruker Avance 400 spectrometer respectively at 400 and 100 MHz and the residual hydrogenated solvent was used as reference.

Analytical and Preparative High-Performance Liquid Chromatography. The purification of peptides was performed using a Waters AutoPurify system equipped with a dual wavelength UV detector (set at 220 nm and 275 nm), a 3100 mass spectrometer, reverse phase columns (Waters, Sun Fire Anal C_{18} 5.0 µm 4.6x150 mm, Sun Fire Prep C_{18} 5.0 µm, 19 × 150 mm) running with a 0.08% TFA MilliQwater/acetonitrile gradient as eluent, and the MassLynx 4.1 – XP software. For preparative HPLC, the fractions are collected automatically according to a preset trigger (Coup threshold for m/z signal or UV absorption threshold) and are separated in different tubes represented on the spectra allowing the user to theoretically know the composition of each tube. The mass detector had several configurations but was only used in its 150-1500 m/z configuration.

Time (minutes)	Water (volume %)	Acetonitrile (volume %)
0	80	20
5	80	20
12	50	50
20	50	50
25	95	5
29	80	20
30	80	20

Table 2 Method A for Analytical HPLC for peptide purification, 2mL/min flow

Time (minutes)	Water (volume %)	Methanol (volume %)
0	80	20
5	80	20
12	50	50
20	50	50
25	95	5
29	80	20
30	80	20

Table 3 Method B for Analytical HPLC for peptide purification, 2mL/min flow

Time (minutes)	Water (volume %)	Acetonitrile (volume %)
0	95	5
10	80	20
18	80	50
29	20	80
37	5	95
38	100	0
45	80	20

Table 4 Method A for preparative HPLC for peptide purification, 23.28 mL/min

Time (minutes)	Water (volume %)	Methanol (volume %)
0	80	20
18	80	20
29	20	80
37	5	95
38	100	0
45	80	20

Table 5 Method B for preparative HPLC for peptide purification, 23.28 mL/min

Ultra-Performance Liquid Chromatography coupled to Mass Spectroscopy (UPLC-MS) were carried out on a Waters Acquity UPLC-SQD apparatus equipped with a PDA detector (190–500 nm, 80 Hz), using reverse phase columns (Waters, BEH C_{18} 1.7 µm, 2.1mm x 50 mm and Waters), and the MassLynx 4.1 – XP software with a gradient (MilliQwater-acetonitrile + 0.1% formic acid (FA)) as eluent. The mass detector has two configurations, one for signals between 100 and 1300 m/z and another one for 400-2000 m/z. Higher mass compounds have been identified through their n-charged compounds recognized by the shift in isotopic ratio. The gradient used for acquiring the chromatograms is:

Time (min)	% Water (0.1% FA)	% Acetonitrile (0.1% FA)
0	95	5
3.5	5	95
4	5	95
4.1	95	5
5.1	95	5

Table 6 Gradient used for the UPLC RP 18 analysis. The evolution of solvent composition between two given times is linear.

The analytical HPLC-MS analysis of the 20°C precipitate was performed using a Water Alliance 2695 HPLC system coupled to a Bruker microTOF-Q mass spectrometer. The sample was injected on a Waters XBridge® BEH C18 column (130 Å, 3.5 µm, 2.1 x 150mm column). The solvent system consisted of 0.1% trifluoroacetic acid in water and 0.08% trifluoroacetic acid in acetonitrile. Elution was performed at 60°C at a flow rate of 0.25 mL/min. ESI-MS spectra were acquired using a Bruker Daltonics microTOF-Q directly connected to the output of the HPLC column. which was operated in positive mode with a capillary voltage of -4500V. Acquisitions were performed on the mass range 300-3000 m/z with a 1 sec scan time. Calibration was performed using the singly charged ions produced by an ESI-TOF tuning reference (Agilent Technologies, Paolo Alto, USA) in the mass range of 322-2722 m/z. Data analysis was performed using Bruker Compass DataAnalysis 4.3.

The pH of solutions and buffers were measured using a Mettler Toledo Seven Compact pH mete preemptively calibrated.

MilliQ water was obtained through an Advantage Milli-Q.

2. Solid Phase Peptide Synthesis procedures

Resin Loading. H-Rink amide ChemMatrix® resin (0.47-0.54 mmol/g loading) was swollen in DMF (15 mL) for 30 minutes and then washed with DCM (3 x 5 mL) and DMF (3 x 5 mL). Double coupling of the first amino acid was carried out using protocols recommended by CEM. Typically, the Fmoc-amino acid (0.2 M in DMF, 4 eq.), HBTU (0.5 M in DMF, 4 eq.) and DIEA (2 M in NMP, 8 eq.) were added in sequence to the resin, which was then irradiated for 5 minutes at 25 W, allowing a maximum temperature of 72 °C.

Single Coupling. Typical recommended CEM protocol for single coupling consisted of sequentially adding the Fmoc-amino acid (0.2 M in DMF, 4 eq.), HBTU (0.5 M in DMF, 4 eq.) and DIEA (2 M in NMP, 10 eq.) to the resin and coupling using a one-step microwave program (70 °C, 35 W, 5 min). This program was used the previous amino acid incorporated in the sequence was a glycine.

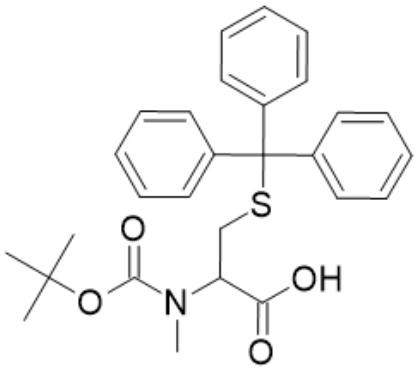
Double Coupling. Typical recommended CEM protocol for double coupling consisted of sequentially adding the Fmoc-amino acid (0.2 M in DMF, 4 eq.), HBTU (0.5 M in DMF, 4 eq.) and DIEA (2 M in NMP, 10 eq.) to the resin and coupling using a two-step microwave program. This program was used when the previous amino acid in the sequence was a Tyrosine, Phenylalanine, or an N-methyl-cysteine

N-methyl Cysteine Coupling. N-methyl cysteine were attached to the sequences by preparing the activated carboxylic acid manually: Fmoc-N-methyl-Cys(Trt)-OH (4 eq.), HBTU (4 eq.), DIPEA (6 eq.) in DMF (4 mL). This mixture was then added to the reaction vessel and coupled using a two-step microwave program (50°C, 35 W, 5 min).

Deprotection. Following the coupling step, the Fmoc protecting group was removed by 5% w/v piperazine in DMF containing 0.1 M HOBt using a two-step microwave program (a. rt, 0 W, 30 sec b. 75 °C, 55 W, 3 min).

Cleavage. On completion of the peptide sequence, the product was cleaved from the resin after incubation in a mixture of TFA/H_2O/TES (55/40/5/v/v/v) for three hours at room temperature under inert atmosphere. The cleavage cocktail was then drained and concentrated on a rotatory evaporator. The concentrated reaction mixture was then diluted is cooled diethyl ether (-78°C), to precipitate the deprotected peptide. The solution is then centrifuged at 4000 RPM for 10 minutes, the supernatant was removed, and the process is repeated2 additional times to remove residual protecting groups from the precipitated peptide. The peptide is then dried by rotatory evaporation and either lyophilized and kept at -80°C for later purification or solubilized in aliquots of 1 mL of a 50/50 water/methanol mixture to be purified by preparative HPLC.

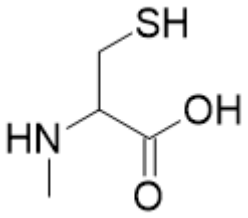
3. Chemical synthesis of N-methyl cysteine



Compound 1, Boc-N-Methyl-L-cysteine-trt-OH

Molecular thieves were placed under vacuum overnight at 200°C then, once cooled used to dry the pre-dried THF from our purification system over two days. This THF was then used a solvent for the synthesis. This synthesis requires two Schlenk flasks which were in preparation an hour before the reaction placed under vacuum and heated with a heat gun then filled with argon. 3.1g (77 mmol) of NaH (60% in mineral oil) were added under argon flow in one of the flasks then dissolved and stirred in 45 mL of syringe injected dry THF, this flask was cooled to under 0°C with salted ice bath. 15g (33 mmol) of Boc-L-Cys(trt)-OH were added to the other Schlenk flask under argon flow, and then dissolved and stirred in 30mL of syringe injected dry THF. This solution was then added dropwise to the NaH containing flask and the cooling was maintained all throughout the addition. After 30 minutes of stirring and cooling, 15 mL (243 mmol) of iodomethane were added dropwise to the reaction mixture. After the addition the cooling was stopped, and the reaction was left at room temperature under stirring for 16 hours. The reaction media was then quenched with 150 mL of 200 mM phosphate buffer at pH 7 and the THF was removed by rotatory evaporation (bath at 40°C). The obtained aqueous solution is then acidified to pH using 1M HCl and then extracted with 100 mL of dichloromethane (DCM) four times. The combined organic phases were then washed with saturated NH_4Cl and dried over anhydrous Na_2SO_4 and filtered before being concentrated by rotatory evaporation in prevision of purification by silica gel column chromatography in DCM with 0.5% MeOH. Affording pure compound 1 which was dried overnight (14.5g, 30 mmol, 91% yield).

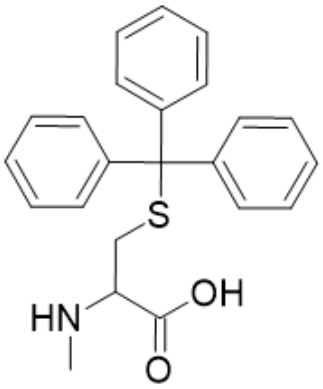
^{1}H NMR (400 MHz, $CDCl_3$, 25 °C) δ = 7.41 (d, J = 7.6 Hz, 6H), 7.33 (t, J = 7.3 Hz, 6H, overlapping with solvent residual peak), 7.22 (t, J = 7.0 Hz, 3H), 3.85 (q, J = 5.2 Hz, 0.6H), 3.64 (q, J = 5.2 Hz, 0.6H), 2.82-2.64 (m, 5H), 1.43 (s, 5H), 1.32 (s, 4H); **^{13}C NMR (100 MHz, $CDCl_3$, 25 °C)** δ = 175.8, 175.3, 156.3, 155.0, 144.7, 129.8, 128.2, 127.0, 81.3, 67.2, 60.6, 34.3, 31.8, 28.5; UPLC retention time 2.25 min; MS (ESI+) calcd. for $C28H31NO4Na^+$ $[M+Na]^+$ m/z = 500.60, found 500.27.



Compound 2, N-methyl-L-cysteine

TIPS (20 mL, 98 mmol) and TFA (50 mL) were added to solution of BocMeCys(Trt)-OH (14.5 g, 30 mmol) in DCM (8.3 mL) cooled at 0 °C (be careful as the removal of trt is very exothermic and releases triphenyl vapors). After addition, the reaction was stirred for 3 hours at rt. The solvents were removed by rotary evaporation. The solid was then dissolved in 50 mL DCM and extracted with distilled water (5x25 mL). The aqueous solution was then lyophilized yielding the pure compound **4** (4 g, 29 mmol, yield 97%).

1H NMR (400 MHz, D2O, 25 °C) δ = 3.96 (t, *J* = 4.4 Hz, 1H), 3.17 (dd, *J* = 15.2, 2.4 Hz, 1H), 3.04 (dd, *J* = 15.2, 2.4 Hz, 1H), 2.75 (s, 3H). **13C NMR (100 MHz, D2O, 25 °C)** δ = 170.9, 63.2, 31.5, 22.9. **UPLC** retention time 0.28 min; **MS** (ESI+) calcd. for C4H10NO2S+ [M+H]+ *m/z* = 136.04, found 136.31.

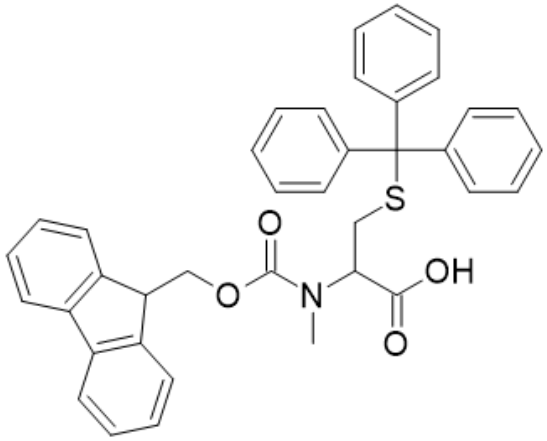


Compound 7, N-methyl-cysteine(trt)-OH

Triphenyl methanol (9.5 g, 37 mmol) was added to a solution of L-Me-Cys-OH (4 g, 29 mmol) in HFIP (80 mL) under inert atmosphere. The reaction was stirred at 35 °C for 4 hours. The solvent was removed by rotatory evaporation and the solid was redissolve in a few milliliters of DCM in prevision of purification by silica gel column chromatography with pure DCM. Affording pure compound 3 which was dried overnight (9.8g, 26 mmol, 90% yield).

1H NMR (400 MHz, CDCl3, 25 °C) δ 7.40 (d, *J* = 7.8 Hz, 6H), 7.22 (t, *J* = 7.8 Hz, 6H), 7.15 (t, *J* = 7.3 Hz, 3H), 3.04 (t, *J* = 5.98 Hz, 1H), 2.76 (dd, *J* = 13.4, 5.7 Hz, 1H), 2.68 (dd, *J* = 13.4, 5.7 Hz, 1H), 2.45 (t, *J* = 7.1 Hz, 3H); **13C NMR (100 MHz, CDCl3, 25 °C)** δ = 167.8, 144.1,

129.1, 128.1, 126.8, 77.4, 66.2, 61.3, 32.0. **UPLC** retention time 1.45 min; **MS** (ESI-) calcd. for C23H22NO2S- [M-H]- *m/z* = 376.14, found 376.46.



Compound 8, Fmoc-N-methyl-L-cysteine

Fmoc-OSu (8.8 g, 26 mmol) was added to a mixture of MeCys(Trt)-OH (9.8 g, 26 mmol) in H2O:ACN (200 mL, 1:1, v/v) at 0 °C and the resulting suspension was titrated with NEt_3 until pH 8. After 15 h at room temperature, the organic solvent was removed *in vacuo*. The aqueous layer was acidified with aqueous HCl 1 M until pH 3 and extracted with ethyl acetate (3 x 100 mL). The organic phase was washed with HCl 1 M (3x100 mL), dried on Na_2SO_4 and filtered. Further purification by column chromatography was carried out in two steps. A first silica column chromatography was performed with 20% ethyl acetate and 1% acetic acid in cyclohexane (Rf = 0.24) and the combined mixed phases containing the product were then purified a second time through silica column chromatography using pure DCM (Rf = 0.3) affording compound 4 as a white foam which was left to dry overnight (11.7 g, 19.5 mmol, yield 75%).

1H NMR (400 MHz, CDCl3, 25 °C) δ 7.73 (d, *J* = 7.6 Hz, 2H), 7.50 (d, *J* = 7.6 Hz, 2H), 7.40 (d, *J* = 7.6 Hz, 4.5H), 7.36-7.27 (m, 4H), 7.25 (t, *J* = 7.2, 6H, overlapping residual solvent peak), 7.19-7.14 (m, 5H), 4.52-4.37 (m, 1.3H), 4.20 (t, *J* = 6.4 Hz, 0.7H), 3.31-3.22 (m, 1H), 3.15-3.11 (dd, *J* = 6.8, 2.9 Hz, 0.9H), 2.94-2.88 (m, 2.2H), 2.73 (t, *J* = 6.9 Hz, 3H); **13C NMR (100 MHz, CDCl3, 25 °C)** δ = 175.1, 156.7, 144.6, 144.0, 129.8, 129.7, 128.2, 127.9, 127.0, 125.3, 125.1, 120.1, 77.4, 68.1, 67.9, 67.3, 59.3, 47.3, 33.2, 31.1, 30.9, 27.1; **UPLC** retention time 2.74 min; **MS** (ESI+) calcd. for C38H33NO4SNa+ [M+Na]+ *m/z* = 622.19, found 622.67

4. Synthesis and purification of peptide sequences

All synthetic peptides were stored at -80°C after purification in tubes with airtight caps.

YGC: This peptide was synthesized on H-Rink Amide ChemMatrix® resin at a 0.25 mmol scale. The N-methyl-cysteine was incorporated according to the double coupling described previously. Glycine was double coupled since it followed an N-methyl-cysteine and tyrosine was added through single coupling. The analytical HPLC-MS (method B) showed a retention time of 0.99 and 1.95 min (first is the thioester). The peptide was then purified by preparative HPLC-MS (method B) 2.03 min retention time with a purity of after lyophilization 94%.

CGGY: This peptide was synthesized on H-Rink Amide ChemMatrix® resin at a 0.25 mmol scale. Tyrosine being the first amino acid of the sequence was attached to the resin through a double coupling and the following glycine was double coupled as well. The third glycine was added through single coupling and the N-methyl-cysteine through its own described coupling method. The analytical HPLC-MS (method B) showed a retention time of 0.73 min. The peptide was then purified by preparative HPLC-MS (method B) 2.12 min retention time yielding after lyophilization a purity of 85%.

YGGC: This peptide was synthesized on H-Rink Amide ChemMatrix® resin at a 0.25 mmol scale. The N-methyl-cysteine was incorporated according to the double coupling described previously, followed by a double coupled glycine. The remaining glycine and tyrosine were added through single coupling. The analytical HPLC-MS (method A) showed a retention time of 0.92 min. The peptide was then purified by preparative HPLC-MS (method A) 4.77 min retention time yielding after lyophilization a purity of 97%.

CGYGF: This peptide was synthesized on H-Rink Amide ChemMatrix® resin at a 0.25 mmol scale. In this peptide all the amino acids we double coupled, with the N-methyl-cysteine using its own double coupling method. The analytical HPLC-MS (method A) showed a retention time of 12.87 min. The peptide was then purified by preparative HPLC-MS (method A) 11.57 min retention time yielding after lyophilization a purity of 95%.

CGYGC: This peptide was synthesized on H-Rink Amide ChemMatrix® resin at a 0.25 mmol scale. In this peptide all the amino acids we double coupled, with the N-methyl-cysteine using its own double coupling method. The analytical HPLC-MS (method A) showed a retention time of 1.14 min. The peptide was then purified by preparative HPLC-MS (method A) 10.77 min retention time yielding after lyophilization a purity of 88%.

5. Preparation of buffers and reaction media

Peptide stock solutions were prepared by diluting the lyophilized peptide solutions in the desired amount of milliQ water. The solutions were prepared very concentrated (typically 50 mMol peptide concentrations) and the aliquots for the experiments were diluted to reach 150µL before adding them to the reaction media. After sampling the solution was kept at - 80°C for further use.

Stock PBS buffer was prepared by dissolving the commercially available PBS tablets in 15 mL of milliQ water. To 4.5 mL of this solution are added 430 mg of TCEP,HCl ,300 mg of sodium ascorbate and 10 mg of DMBA or 5 mg of Fmoc-serine-(trt)-OH depending on the UV reference used. The pH was then adjusted to 7 using 5M NaOH with the 0.5 remaining mL and once the pH adjusted the solution was topped to 5mL using milliQ water. Yielding a solution of 150 mMol phosphate buffer with 40mMol potassium chloride and 2 Mol sodium chloride, 300 mMol of TCEP,HCl 300 mMol of sodium ascorbate at 25°C

Stock citrate buffer, 430 mg of TCEP,HCl, 300mg of sodium ascorbate, 170mg of di ammonium citrate and 10mg of DMBA or 5mg of Fmoc-Serine-(trt)-OH were dissolved in 3 mL of milliQ water. The pH was then adjusted the desired pH using 10M NaOH and once reached the volume was topped to 5mL using milliQ water. This yields a solution of 300 mMol TCEP,HCl, 300mMol Sodium ascorbate, 150 mMol diammonium citrate.

Preparation of reaction media. Typical reaction volumes realized were 450µL in Eppendorf tubes. For the kinetic assay, we added 150µL of the desired buffer system to the tube and let it equilibrate at the desired temperature for the experiment. Then 150µL of each peptide stock solution was added to the reaction media. For the living polymer system, 150µL of milliQ water were added before thermal equilibration. Then 150 µL of monomer stock solution was added.

6. Sampling and Preparation of samples for imaging and analysis

Reaction media sampling. The reaction media was sampled at the different timesteps according to this protocol. 15µL of reaction media were sampled and added to a UPLC vial and diluted with 30µL of milliQ water (mostly to decrease the salt content to avoid precipitation in the UPLC column and to make the first UV peak smaller and not encompass the hydrolyzed monomer peak). The vial was then used for UPLC analysis which realizes 2 µL injections.

The first SEM images of the precipitate. The reaction media was centrifuged at 9000 RPM for 10 minutes. The supernatant was then removed and 300µL of milliQ water were added. The Eppendorf was then sonicated for 5 minutes and centrifuged at 9000RPM for 10 minutes. The water was again removed and 300µL of methanol were added. After vigorous shaking, some of the media was sampled and deposited on a copper TEM grid with carbon film which was then imaged and analyzed. For the EDX analysis, the TEM grid was placed on a silicon stub (explaining the silicon signal)

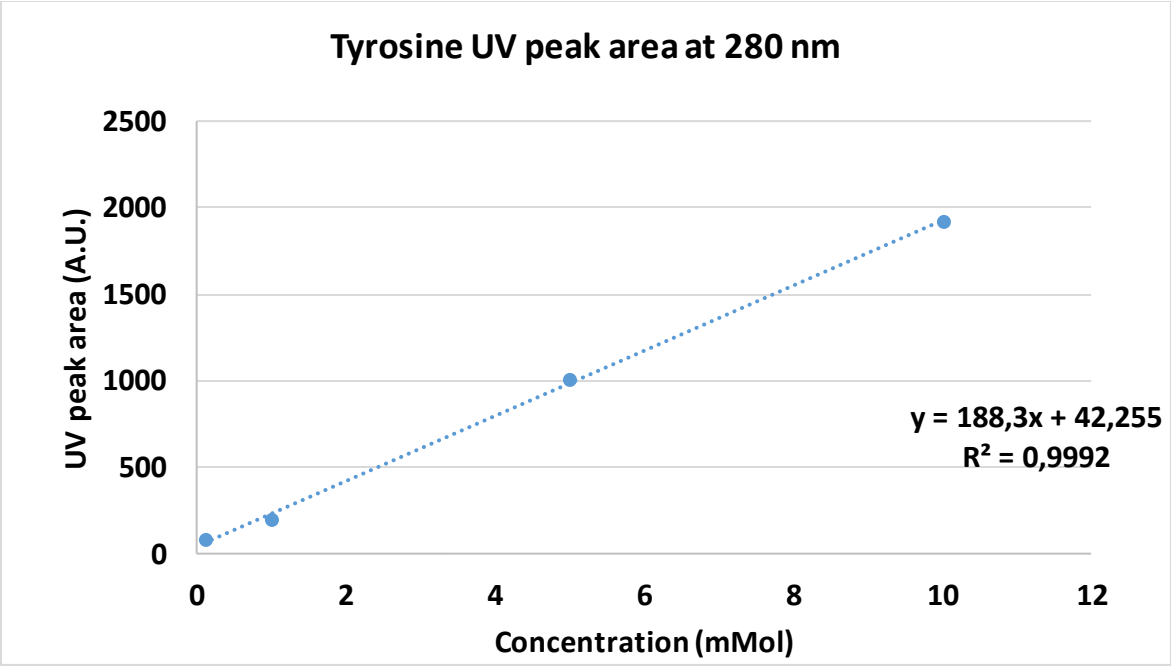
SEM imaging of the needles. Centrifuging the reaction media did not manage to dislodge the needles so the reaction media was sampled and left to dry under vacuum on a silicon stub after optical microscopy confirmed the presence of a needle. The dried sample was then imaged and analyzed by SEM-EDX. This explains the amorphous mass surrounding the needle resulting from the buffer and reducing agents.

Precipitate analysis. The reaction media was centrifuged for 10 min at 9000RPM. The supernatant was removed and 300µL of milliQ water were added followed by another 10 minutes of centrifugation and removal of the supernatant. 86 mg of TCEP,HCl were diluted in 3 mL of methanol and 200µL of this solution were added to each precipitate. The precipitate was then sampled for analysis (40µL in UPLC vials).

7. Integration of the UV signal

The UV signal measured by the UPLC was integrated using the integration method of the MassLynx 4.1 – XP software. This method however does not deconvolute the signal and if multiple peaks superpose yields to approximate results and sometimes the peaks are not properly separated and considered as a single one. This is particularly problematic when it comes to determining the concentration of the linear trimer compound.

The Area signal was then converted to concentration using the following calibration curve obtained by diluting tyrosine and measuring the UV signal at 280 nm for different concentrations:



The area of the UV standard at 280nm was then considered at t_0 to normalize the concentration. This meant that the concentration was obtained through:

$$[tyrosine\ containing\ species] = 188 \times UV\ peak\ area \times \frac{UV\ standard\ peak}{UV\ standard\ peak\ at\ t_0}$$

Appendix: UPLC Spectra of pure peptides

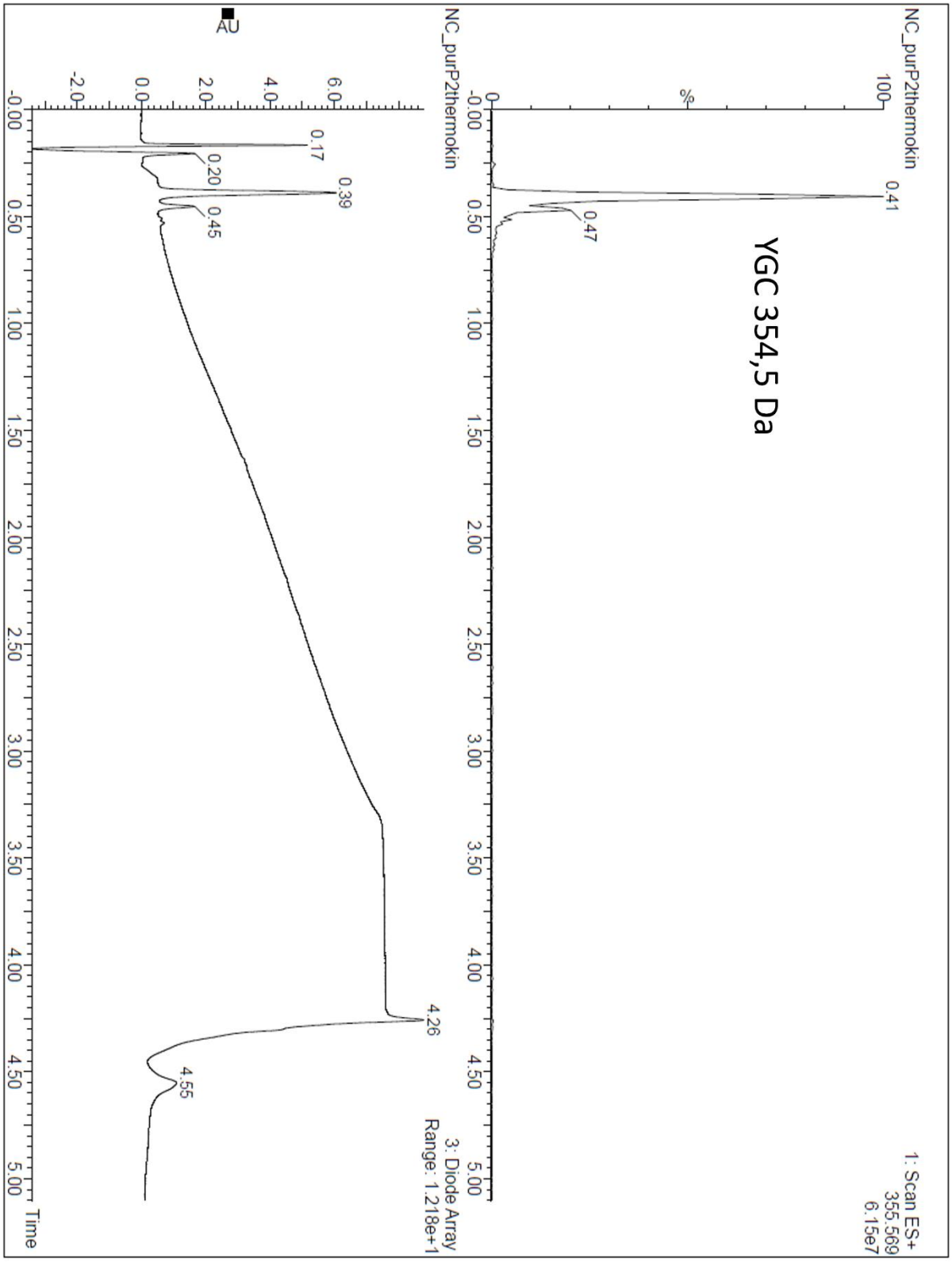


Table 9 Positive electrospray signal for 355.5 m/z and UPLC chromatogram for purified YGC. The two peaks are atributed to the peptide and its thioester form. The discrepency at 0.17 corresponds to the elution front of the solvent.

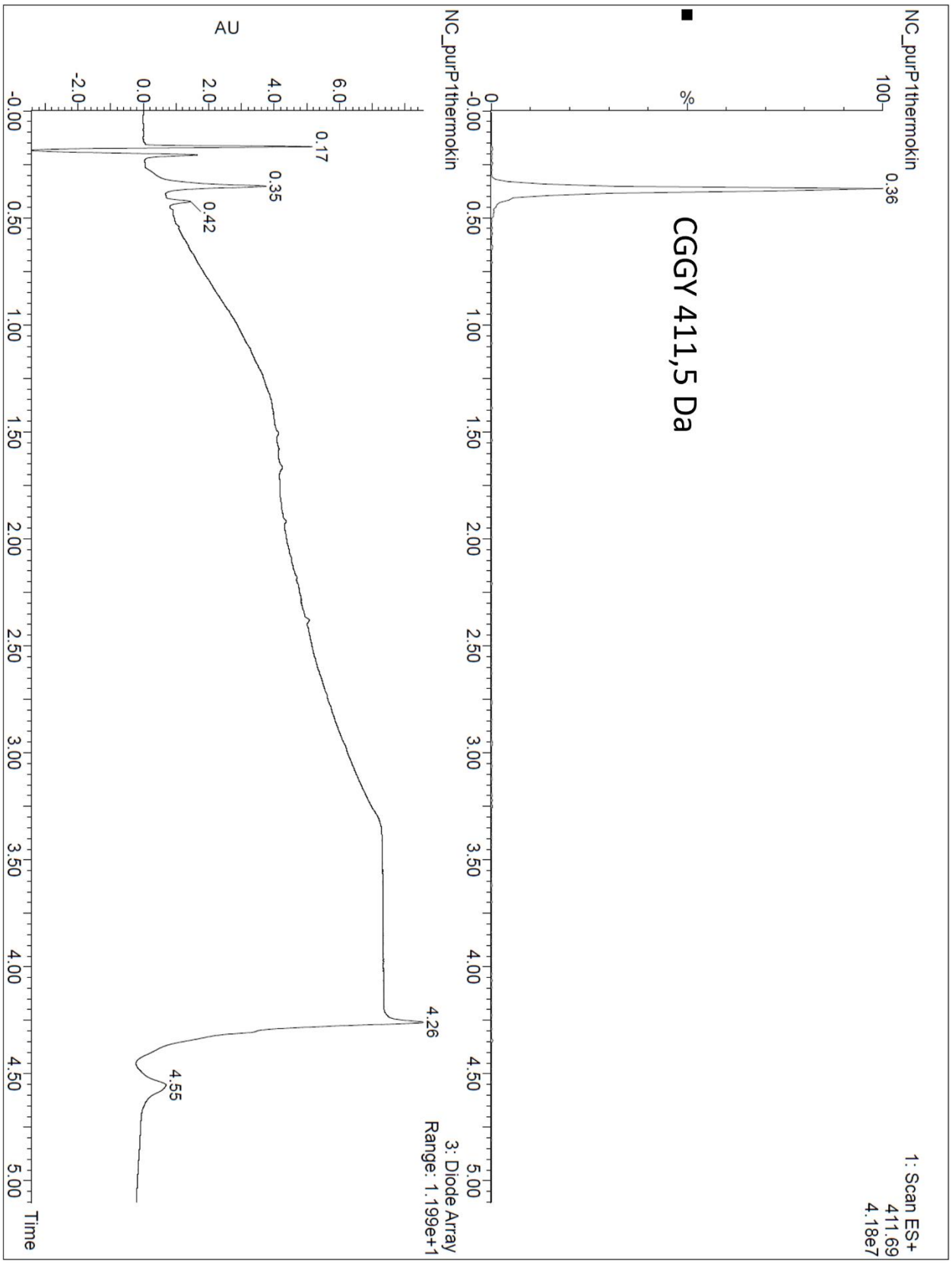


Table 10 Positive electrospray signal for 411.7 m/z and UPLC chromatogram for purified CGGY. The discrepency at 0.17 corresponds to the elution front of the solvent.

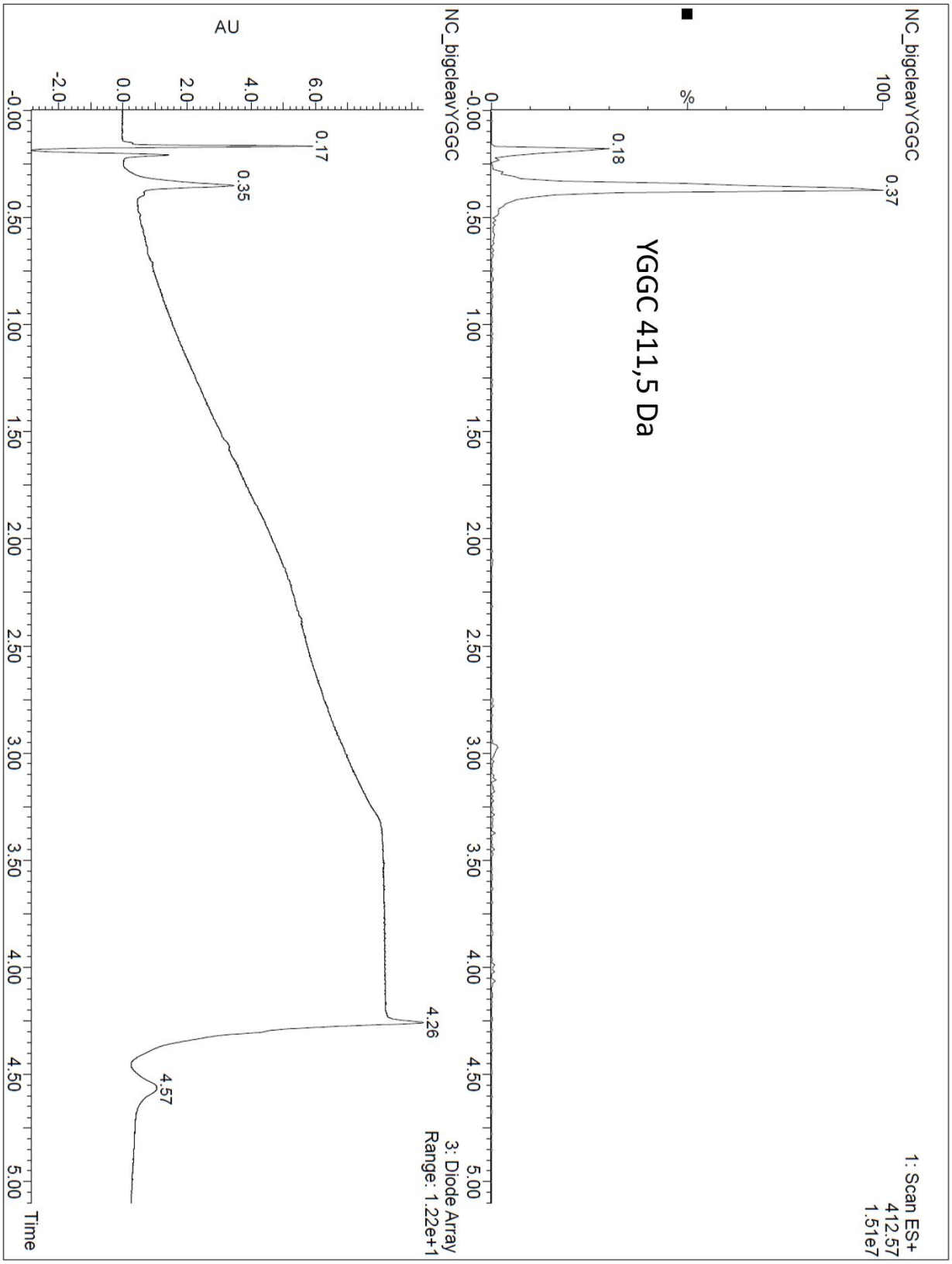


Table 11 Positive electrospray signal for 412.5 m/z and UPLC chromatogram for purified YGGC. The two peaks are attributed to the peptide and its thioester form. The thioester seems to have coeluted with the solvent front.

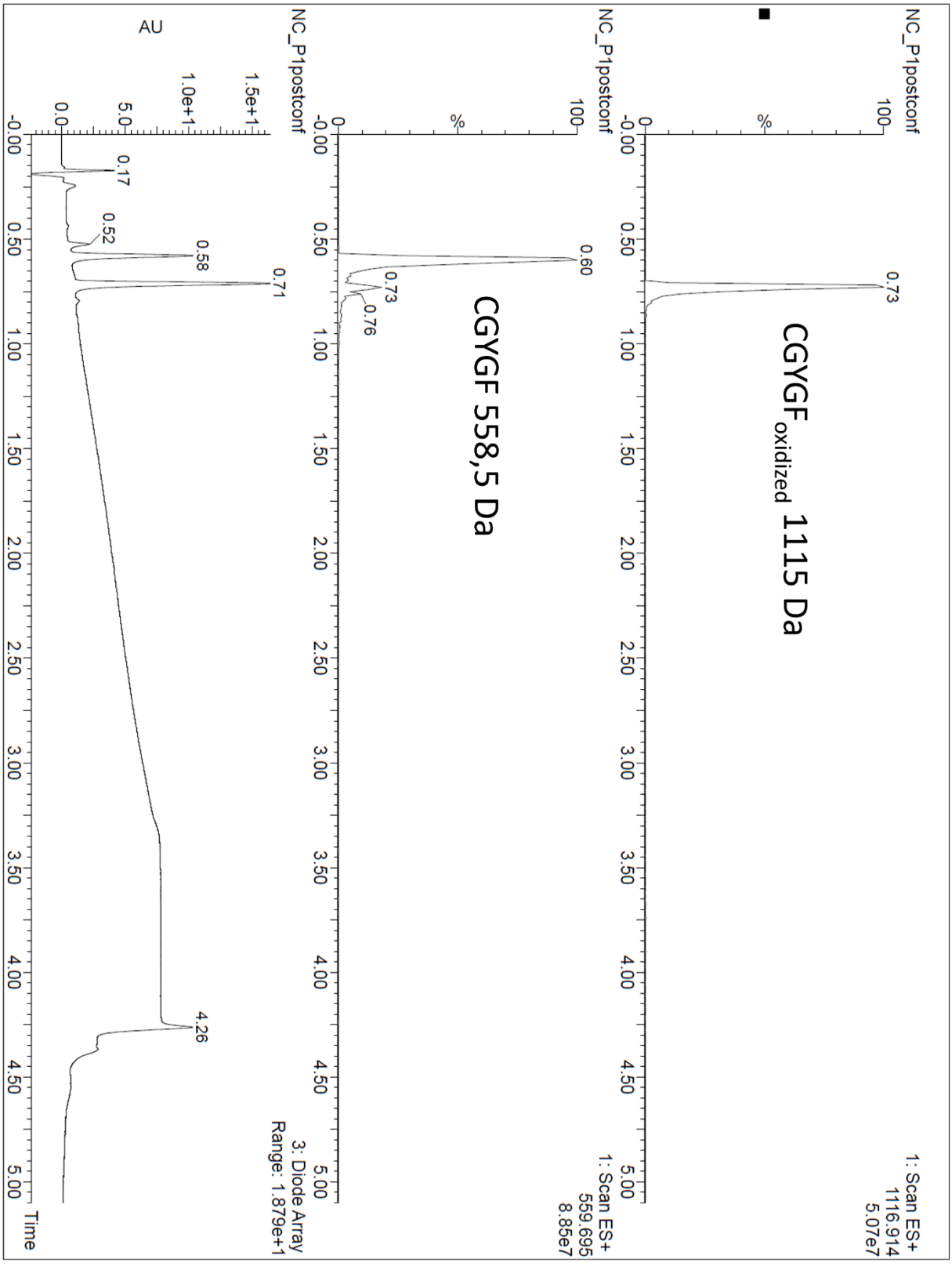


Table 12 Positive electrospray signal for 559.5 and 1116.9 m/z and UPLC chromatogram for purified CGYGF. The two peaks correspond to the peptide and its oxidized form. The UPLC was not performed right away after purification but after 3 months in the freezer explaining the amount of oxidized product and emphasizing the need for reducing agent. The thioester seems to have coeluted with the solvent front.

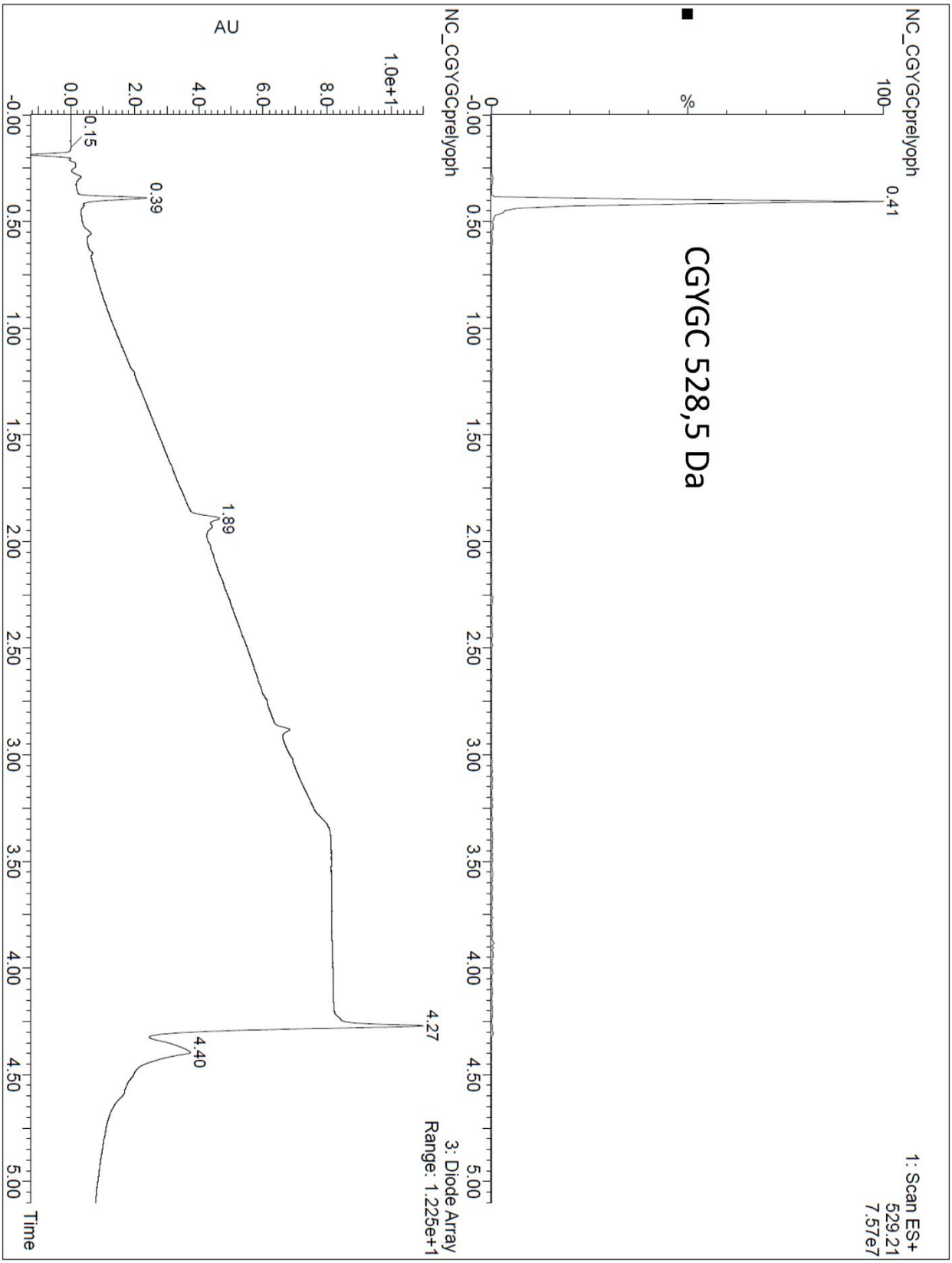


Table 13 Positive electrospray signal for 5529.2 m/z and UPLC chromatogram for purified CGYGC. The peaks observed at 1.89 are impurities left on the column by previous samples and do not come from the purification of the monomer.

Appendix: UPLC m/z Spectra of living polymer oligomers:

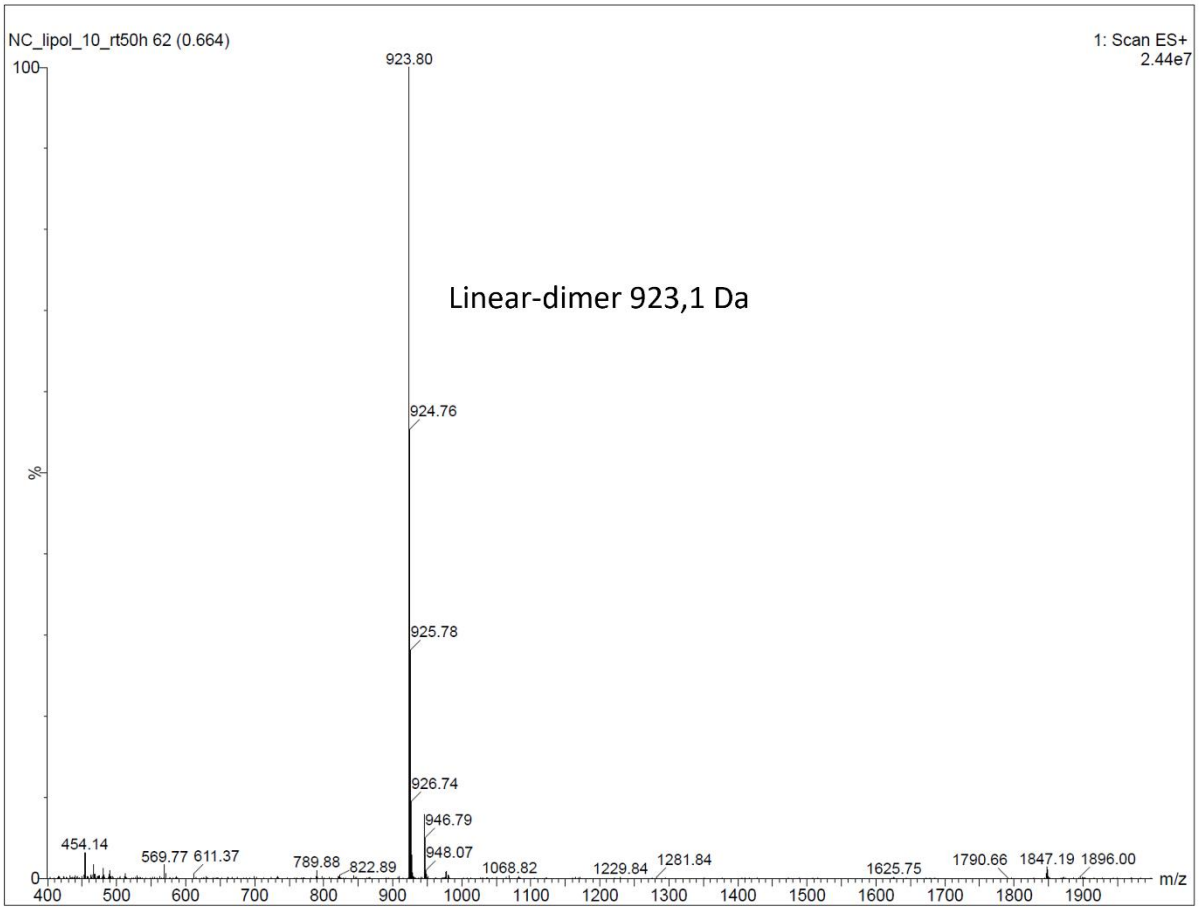


Table 14 Positive electrospray signal attributed to the linear dimer and calculated mass.

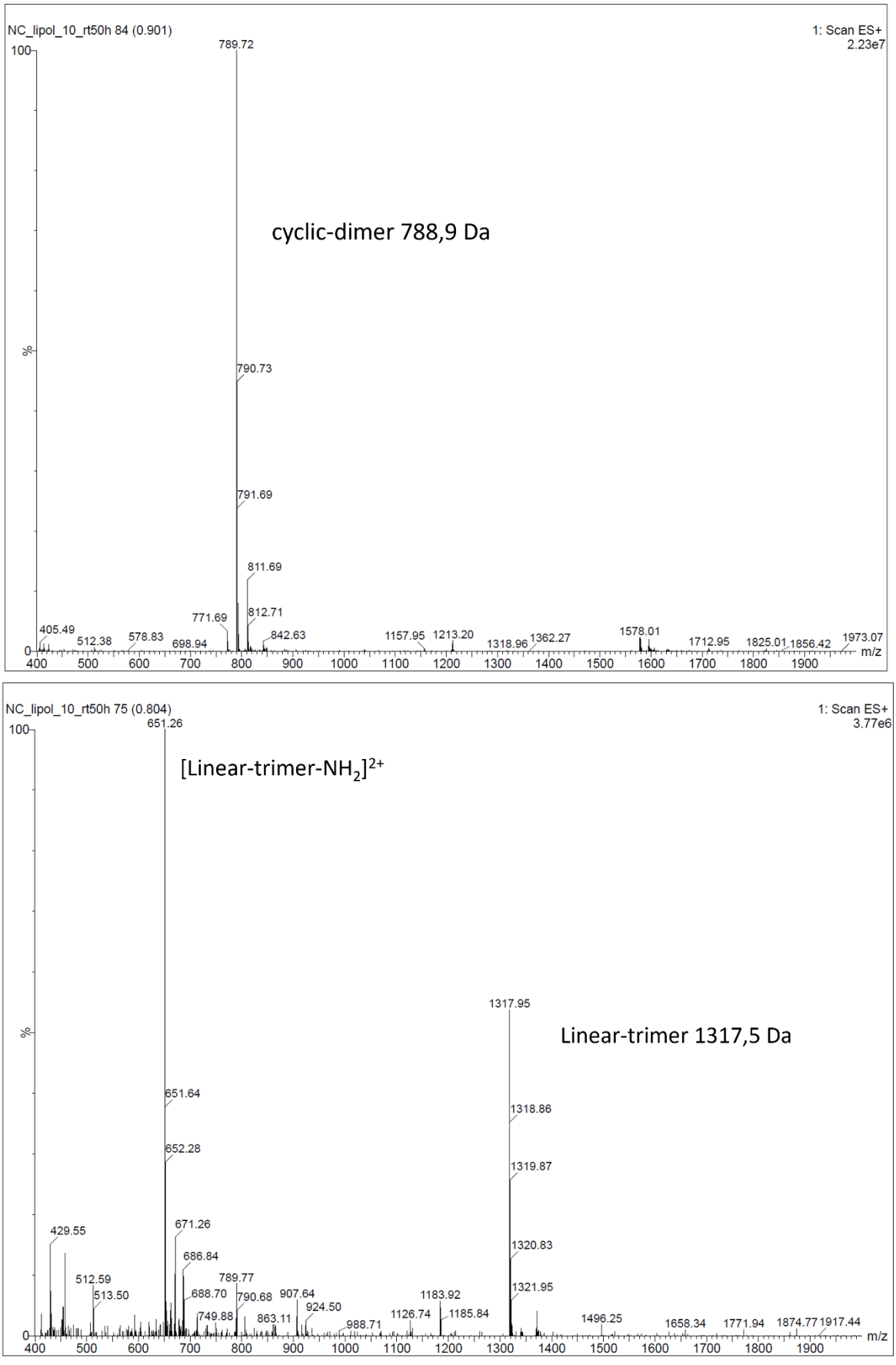


Table 15 Top, Positive electrospray signal attributed to the cyclic- dimer and calculated mass. Bottom, Positive electrospray signal attributed to the linear trimer and calculated mass.

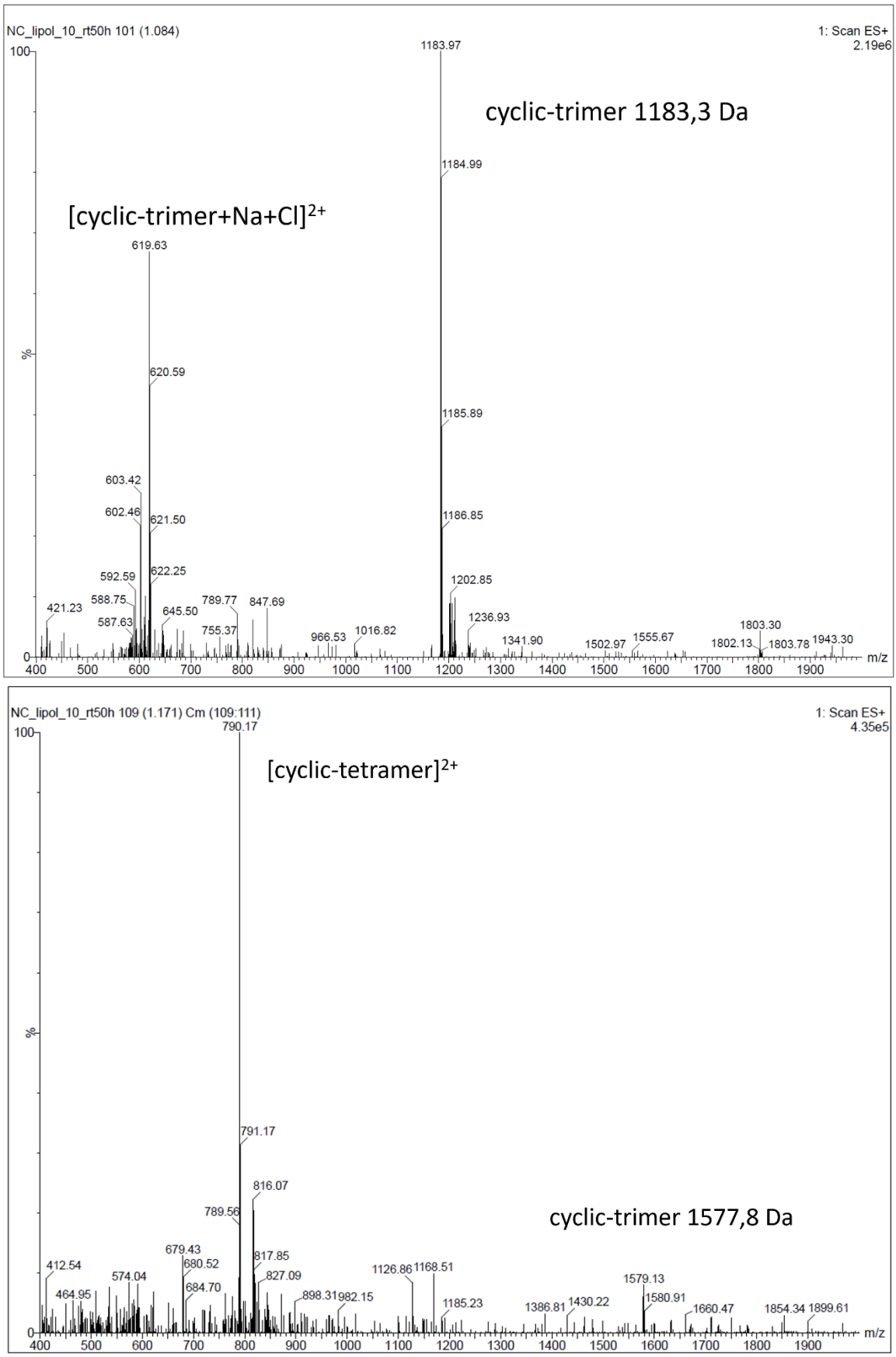
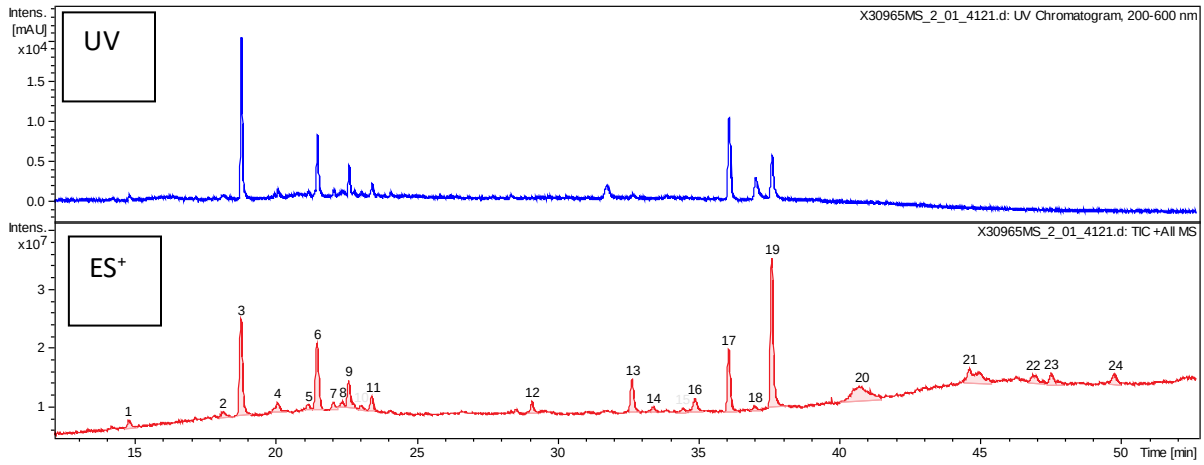


Table 16 Top, positive electrospray signal attributed to the cyclic trimer and calculated mass. Bottom, positive electrospray signal attributed to the cyclic tetramer and calculated mass.

Appendix: HPLC spectra of the 20°C precipitate:

Chromatogram LC-UV-MS



Positive electrospray mass spectra

A. Pic 1

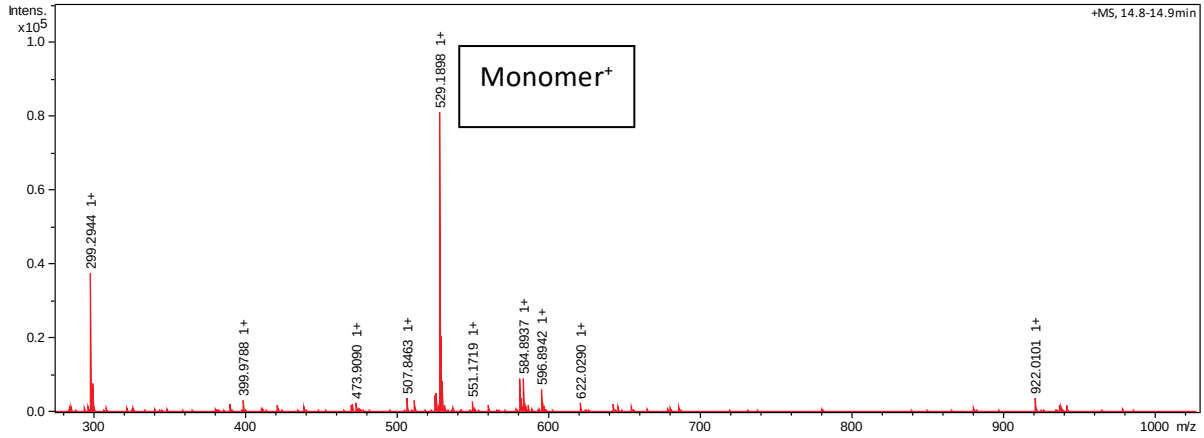


Table 17 Monomer

B. Pic 2

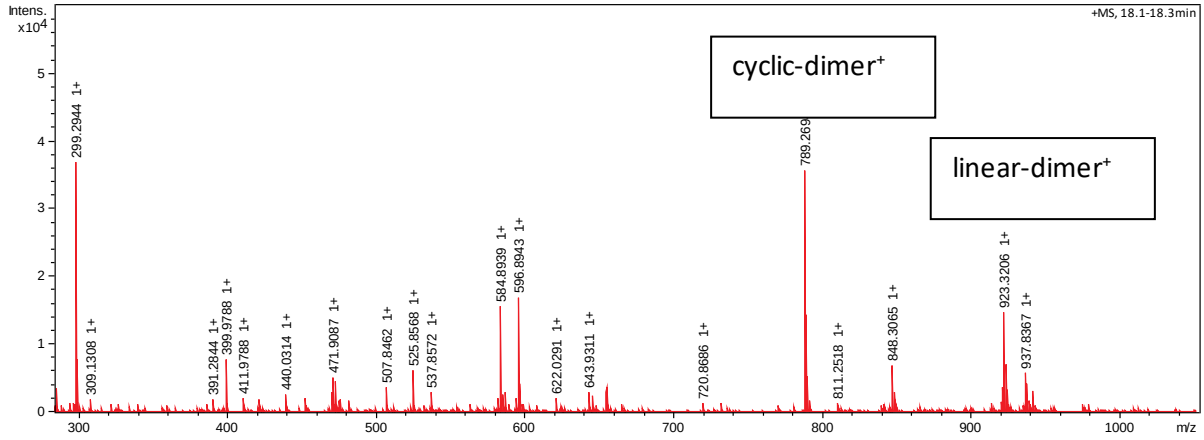


Table 18 Linear dimer with cylcization during the elution

C. Pic 3

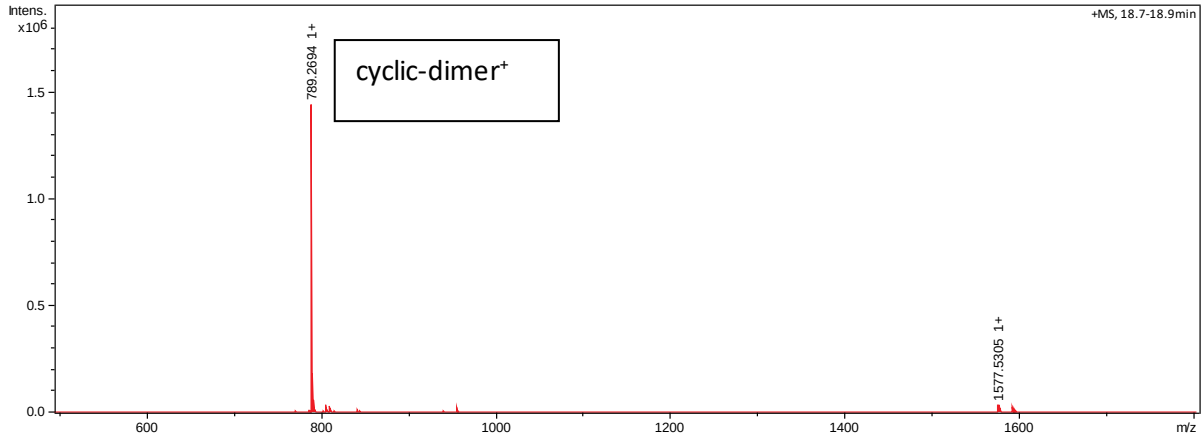


Table 19 Cyclic dimer

D. Pic 4

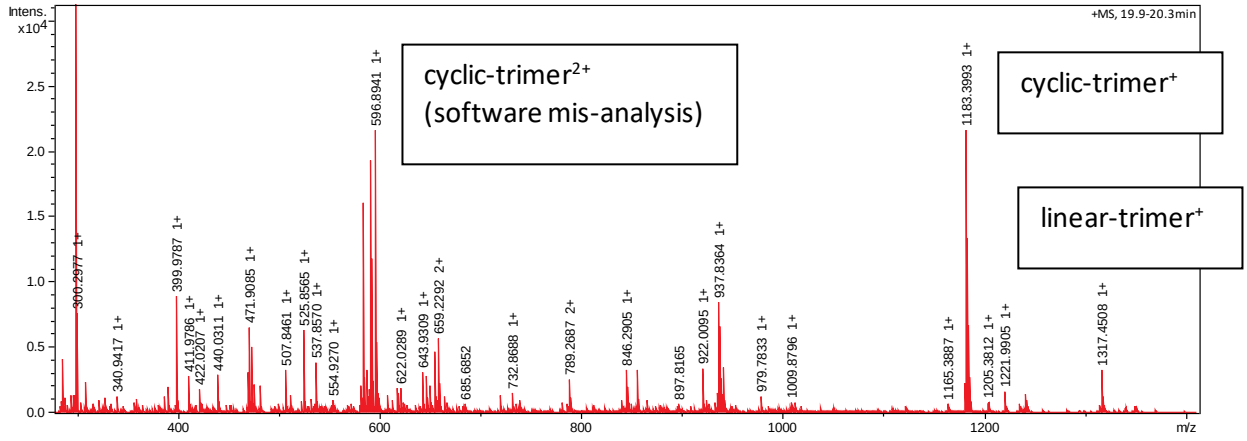


Table 20 Linear trimer with cyclization during elution

E. Pic 5

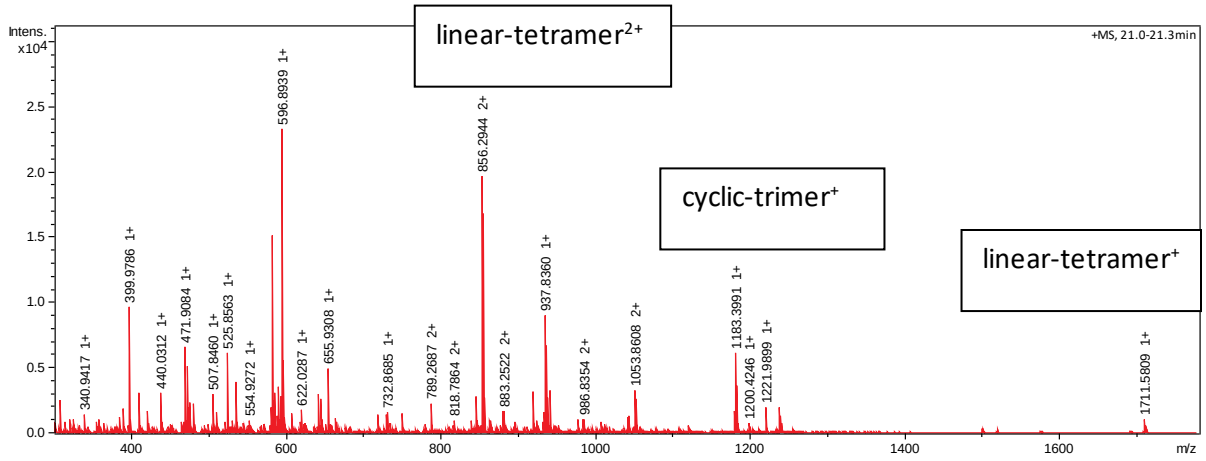


Table 21Linear tetramer

F. Pic 6

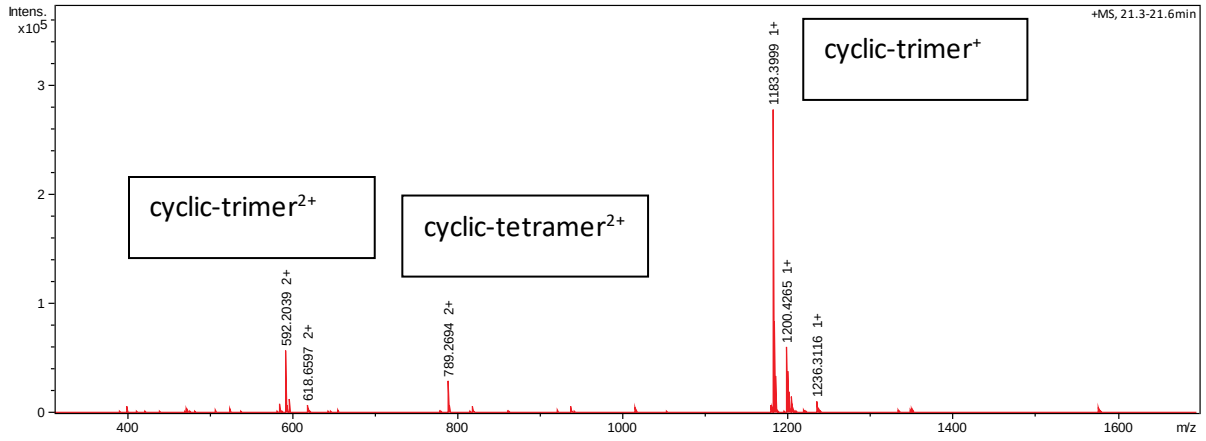


Table 22 Cyclic dimer

G. Pic 7

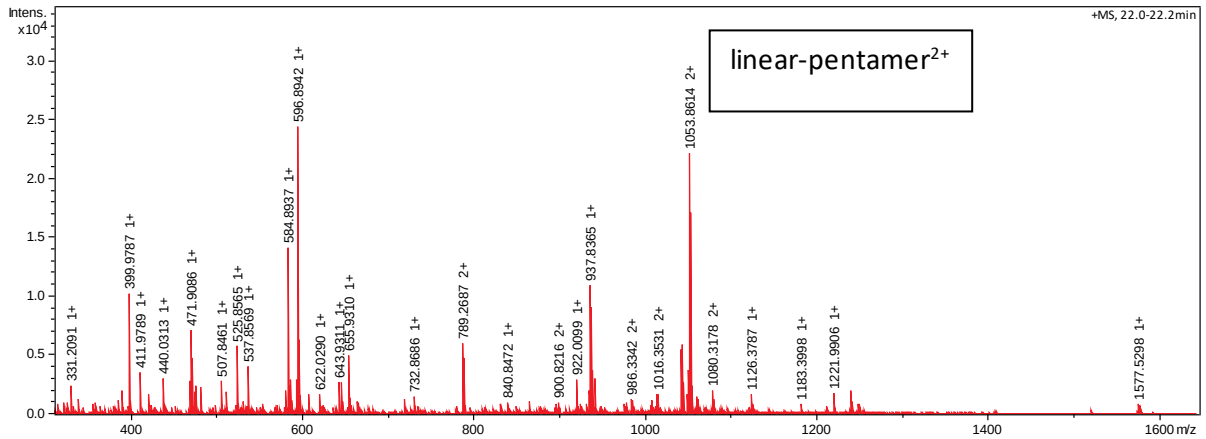


Table 23 Linear pentamer

H. Pic 8

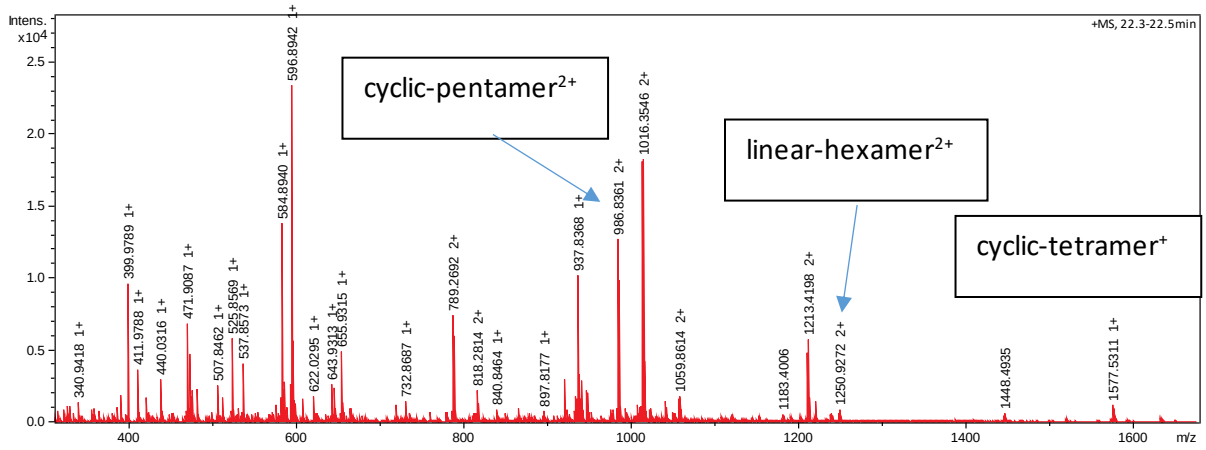


Table 24Mixture of different compounds which could be formed during the ellution and seperate from their respective phases.

Pic 9

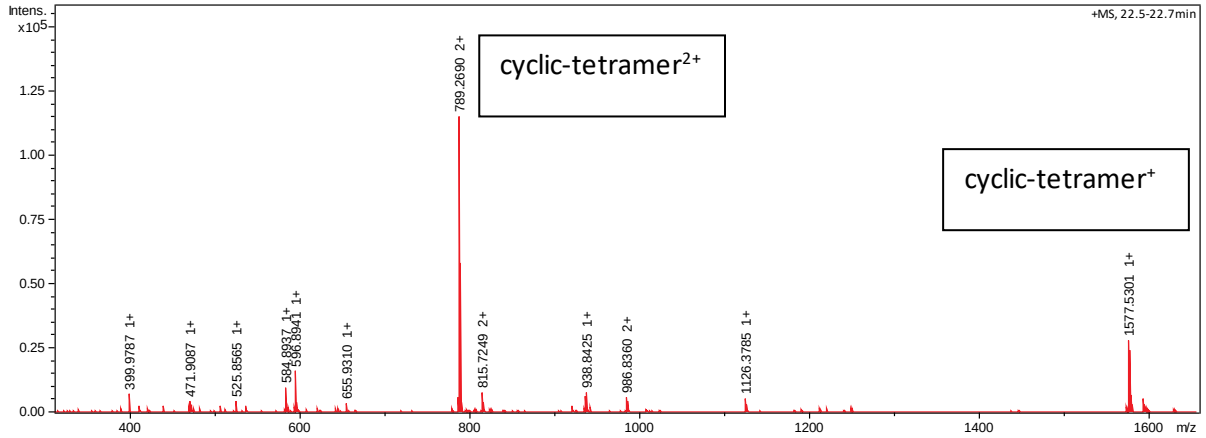


Table 25 Cyclic tetramer

I. Pic 10

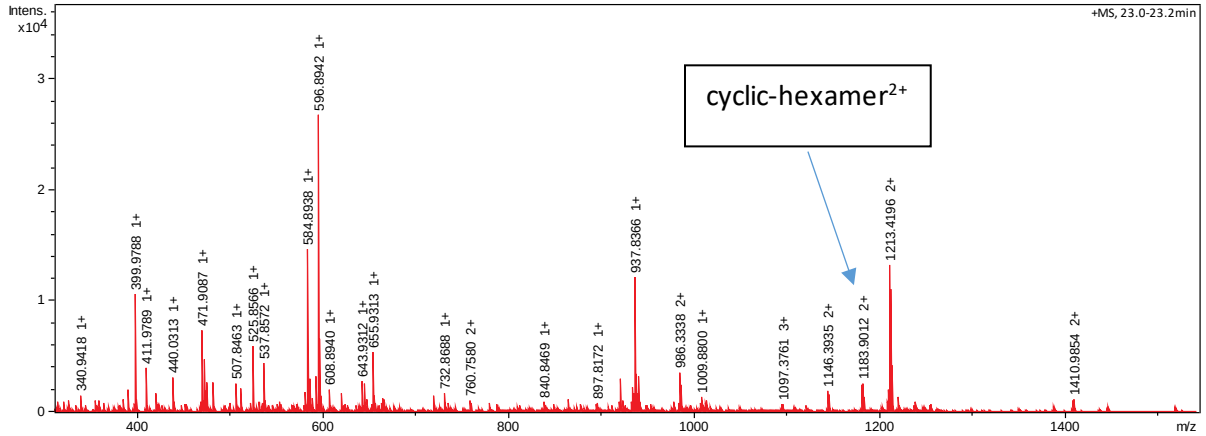


Table 26 There is a weak signal corresponding to the cyclic hexamer. The 1213 m/z could also be the hexamer interacting with NaCl

J. Pic 11

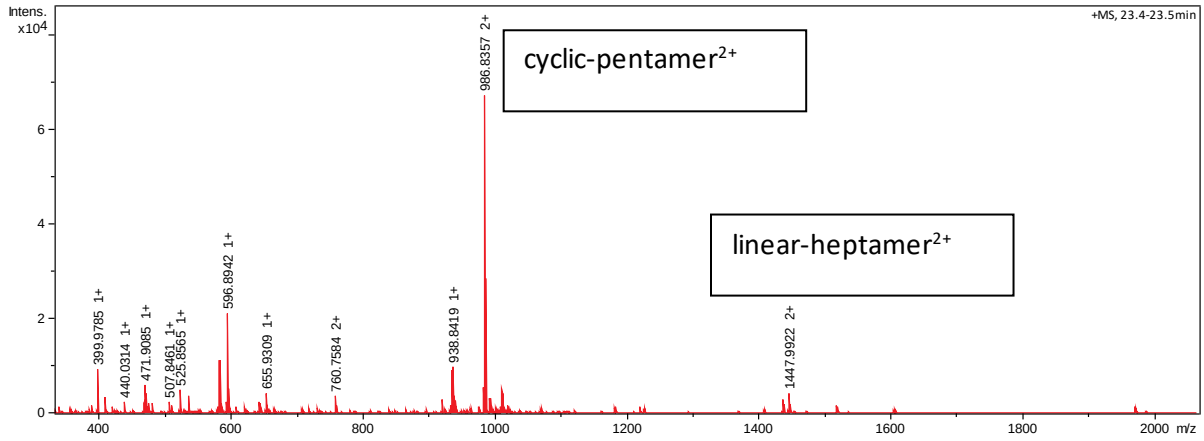


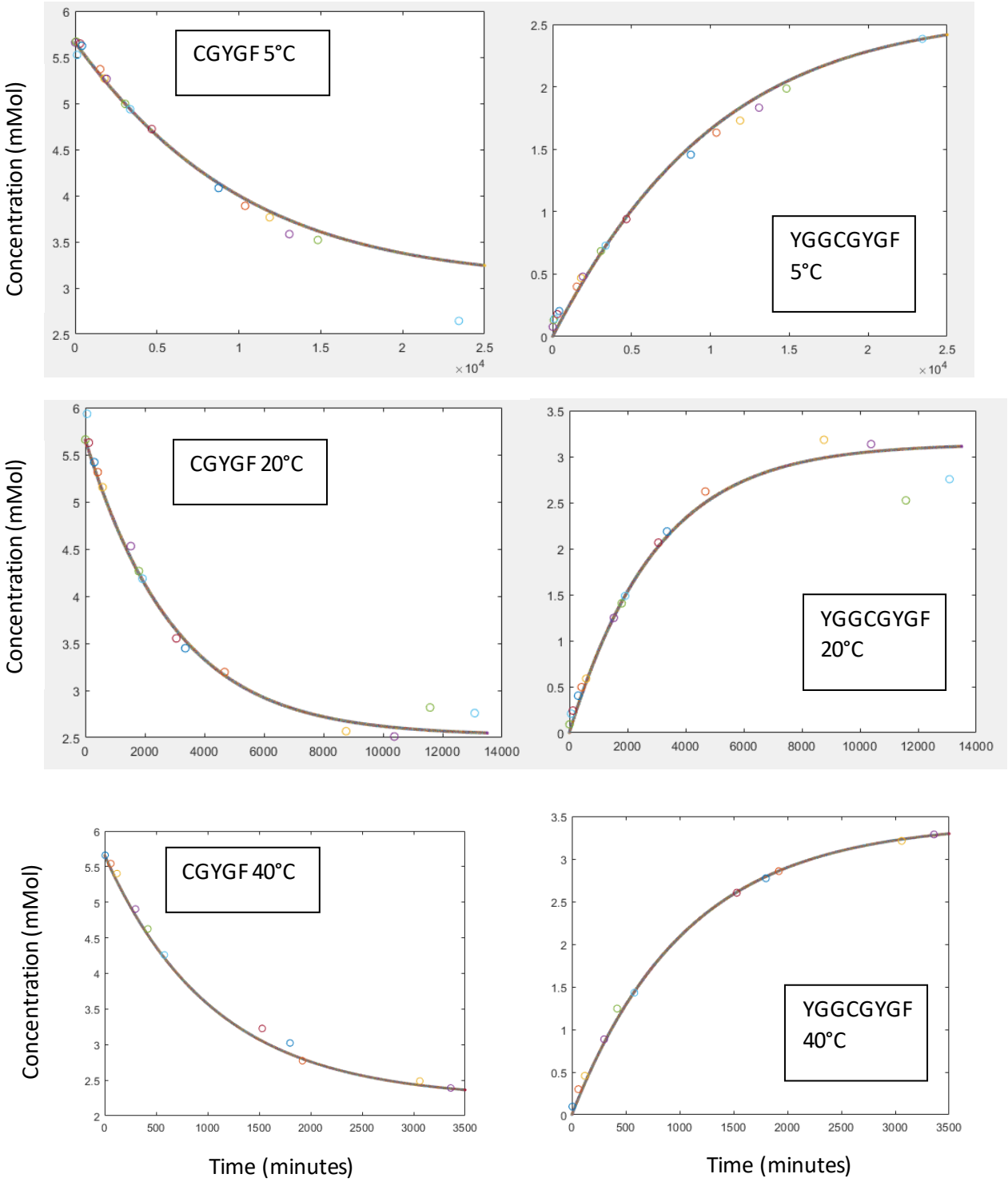
Table 27 In this mass spectra both signals of cyclic pentamer and linear heptamer are present.

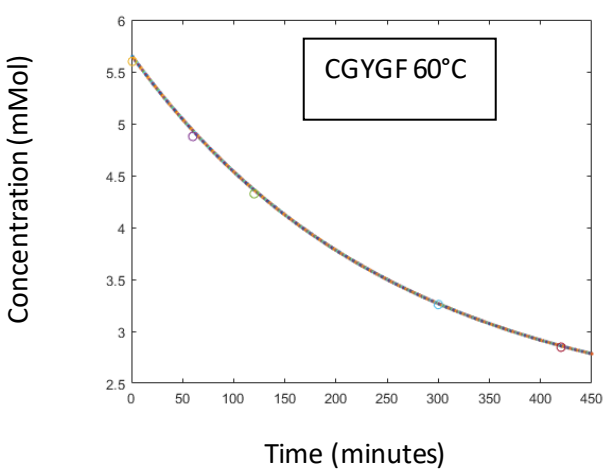
The other peaks are not related to our polymer system and are not shown.

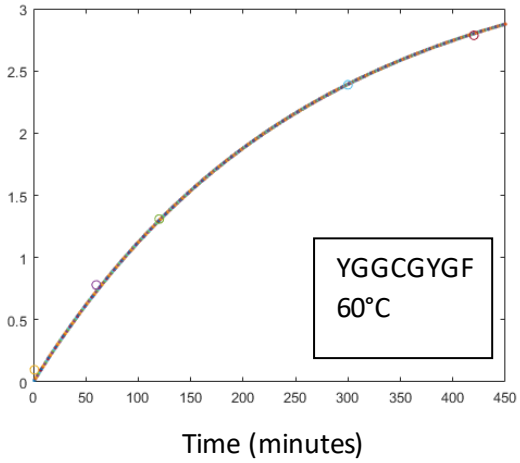
Polymer	Mass (Da)
monomer	528.6
Linear dimer	923.1
Cyclic dimer	788.9
Linear trimer	1317.5
Cyclic trimer	1183.3
Linear tetramer	1712.0
Cyclic tetramer	1577.8
Linear pentamer	2106.4
Cyclic pentamer	1972.2
Linear hexamer	2500.89
Cyclic hexamer	2366.69
Linear heptamer	2895.34
Cyclic heptamer	2761.1

Appendix: Kinetic fits by the bimolecular reversible model

The circles represent the experimental data, the line the modeled evolution with the parameters given in chapter 2.




Concentration (mMol)
CGYGF 60°C
Time (minutes)


YGGCGYGF
60°C
Time (minutes)

"Don't adventures ever have an end? I suppose not. Someone else always has to carry on the story"

-Bilbo Baggins, The Fellowship of the Ring, J. R. R. Tolkien

Nicolas CAPIT

Designing a reversible peptide system for out of equilibrium chemistry exploration

Résumé

Les êtres vivants sont des systèmes chimiques maintenus en état de hors équilibre en dissipant l'énergie qu'ils prélèvent dans leur milieu. Ceci leur confère des capacités remarquables telles que le mouvement, la capacité de communication, des structures dynamiques et le maintiens de conditions propices à leur fonctionnement. Si l'on souhaite que les matériaux de demain possède de telles capacité alors il convient d'étudier et d'apprendre à élaborer des systèmes dissipatifs hors équilibre. C'est l'objectif que ce fixe le présent travail de thèse. La mise au point d'un système polymérisable de façon réversible qui conduirait à un système hors équilibre en présence d'un gradient de température par phénomènes de réaction diffusion. Le système polymérisable choisi est le système peptidique dont la liaison est rendue dynamique via l'incorporation de N-méthyl-cystéine au sein de la séquence d'acides aminés. Pour ce faire les données thermocinétiques de la réaction de ligation/dissociation sont obtenues et un système dynamique surnommée « polymère vivant » est mis au point. Son évolution dans différentes conditions de températures et concentrations sont évalué ainsi que l'évolution sur la distribution des espèces générées. Une approche théorique est ensuite réalisée afin d'obtenir un modèle cinétique simple du système.

Chimie hors équilibre – synthèse peptidique – chimie supramoléculaire – polymères dynamiques

Résumé en anglais

Living organisms are chemical systems self-maintained in an out of equilibrium state through the dissipation of energy taken from its environment. This allows for some remarkable properties such as movement, the ability for communication, dynamic structures and the upkeep of conditions vital for the survival of those systems. If the materials of the future are to possess such properties, then the study and creation of out of equilibrium dissipative systems is of the uttermost interest. This is the main goal of this thesis work. The design of a polymerizable reversible system which would to an out of equilibrium system in a temperature gradient by reaction diffusion processes. The chosen polymerizable system is the peptide system and the peptide bond is made reversible through the incorporation of an N-methyl-cysteine in the amino acid sequence. To this end, the thermokinetic data for the ligation/dissociation reaction were obtained and a dynamic system dubbed "living polymer" was designed. The evolution of such system in different temperature and concentration conditions is then monitored to track the evolution of the distribution of species. A theoretical approach of the system then leads to the creation of a simplistic kinetic model to predict its behavior.

Out of equilibrium chemistry – peptide synthesis – supramolecular chemistry – dynamic polymers