Concomitant synthesis of peptides and oligonucleotides in the presence of lipids and lipid vesicles

Dimitri Fayolle

N°d'ordre NNT : 2020LYSE1134

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

Ecole Doctorale N° 206
(Chimie, Procédés Environnement)

Spécialité de doctorat : Chimie
Discipline : Chimie organique

Soutenue publiquement le 09/09/2020 par :
Dimitri FAYOLLE

Synthèse concomitante de peptides et d'oligonucléotides en présence de lipides et de vésicules lipidiques

---

Devant le jury composé de :

Philippe BARTHELEMY	Professeur, Université de Bordeaux	Rapporteur
Carole CHAIX-BAUVAIS	Directrice de Recherches CNRS, Université Claude Bernard Lyon 1	Examinatrice
Eric DEFRANCQ	Professeur, Université Grenoble-Alpes	Examinateur
Martin EGLI	Professeur, Vanderbilt University	Invité
Mélanie ETHEVE-QUELQUEJEU	Professeur, Université de Paris	Rapporteure
Michele FIORE	Maître de Conférences, HDR, Université Claude Bernard Lyon 1	Co-encadrant
Thierry GRANJON	Maître de Conférences, HDR, Université Claude Bernard Lyon 1	Examinateur
Clemens RICHERT	Professeur, Universität Stuttgart	Invité
Pierre STRAZEWSKI	Professeur, Université Claude Bernard Lyon 1	Directeur de thèse

Université Claude Bernard – Lyon 1

Président de l'Université	M. Frédéric FLEURY
Président du Conseil Académique	M. Hamda BEN HADID
Vice-Président du Conseil d'Administration	M. Didier REVEL
Vice-Président du Conseil des Etudes et de la Vie Universitaire	M. Didier REVEL
Vice-Président de la Commission de Recherche	M. Jean-François MORNEX
Directeur Général des Services	M. Damien VERHAEGHE

COMPOSANTES SANTE

Faculté de Médecine Lyon-Est – Claude Bernard	Doyen : M. Gilles RODE
Faculté de Médecine et Maïeutique Lyon Sud Charles Mérieux	Doyenne : Mme Carole BURILLON
UFR d'Odontologie	Doyenne : Mme Dominique SEUX
Institut des Sciences Pharmaceutiques et Biologiques	Directrice : Mme Christine VINCIGUERRA
Institut des Sciences et Techniques de la Réadaptation	Directeur : M. Xavier PERROT
Département de Formation et Centre de Recherche en Biologie Humaine	Directrice : Mme Anne-Marie SCHOTT

COMPOSANTES & DEPARTEMENTS DE SCIENCES & TECHNOLOGIE

UFR Biosciences	Directrice : Mme Kathrin GIESELER
Département Génie Electrique et des Procédés (GEP)	Directrice : Mme Rosaria FERRIGNO
Département Informatique	Directeur : M. Behzad SHARIAT
Département Mécanique	Directeur M. Marc BUFFAT
UFR - Faculté des Sciences	Administrateur provisoire : M. Bruno ANDRIOLETTI
UFR (STAPS)	Directeur : M. Yannick VANPOULLE
Observatoire de Lyon	Directrice : Mme Isabelle DANIEL
Ecole Polytechnique Universitaire Lyon 1	Directeur : M. Emmanuel PERRIN
Ecole Supérieure de Chimie, Physique, Electronique (CPE Lyon)	Directeur : M. Gérard PIGNAULT
Institut Universitaire de Technologie de Lyon 1	Directeur : M. Christophe VITON
Institut de Science Financière et d'Assurances	Directeur : M. Nicolas LEBOISNE
ESPE	Administrateur Provisoire : M. Pierre CHAREYRON

Concomitant synthesis of peptides and oligonucleotides in the presence of lipids and lipid vesicles

Résumé en français

La traduction, synthèse de protéines guidée par les acides nucléiques, est un mécanisme primordial pour toutes les formes de vie connues. Elle repose sur une machinerie biologique complexe, et les chimistes peinent à proposer des alternatives crédibles dans le contexte de l'émergence du vivant. Des travaux récents suggèrent que les peptido-ARN, des conjugués formés d'un peptide et d'un oligonucléotide liés en N→5' par un phosphoramidate, pourraient représenter un exemple simple de synthèse peptidique promue par un ribonucléotide. D'autre part, la vie cellulaire dépend de la formation et de la croissance de compartiments reposant sur des membranes lipidiques auto-assemblées. Dans cette thèse, nous étudions les interactions de peptido-ARNs amphiphiles, formés d'un peptide lipophile et d'un nucléotide hydrophile, avec des membranes lipidiques selon trois axes. Premièrement, nous avons observé la formation de peptido-ARN amphiphiles à séquences aléatoire ou définie en présence de membranes mais sans observer d'influence significative de ces dernières sur la réactivité. Cependant, des peptido-ARN à séquence peptidique membranophile courte ou moyenne interagissent durablement avec des membranes lipidiques et permettent donc un mécanisme de localisation. En raison de la labilité des phosphoramidates, nous envisageons enfin que cette localisation peut être exploitée pour permettre la pénétration de nucléotides et de petits acides nucléiques à travers des membranes sans mécanisme enzymatique. Nos résultats contribuent à l'élaboration d'un monde prébiotique reposant sur une interaction peptide-ARN-lipides.

Mots clés : peptido-ARN ; phosphoramidate ; peptides membranophiles ; vésicules géantes.

Abstract in English

Translation of nucleic acids into proteins is a key mechanism to all known forms of life. It relies on a complex biological machinery, and chemists are struggling to mimic translation under conditions that may be representative of the emergence of early forms of life from a prebiotic world. Recently, it was suggested that peptido-RNA, a family of conjugates in which a peptide is bound to an oligonucleotide through a *N*→5' phosphoramidate, could provide a simple example of ribonucleotide-promoted peptide synthesis. In addition, cellular life is dependent on the formation and growth of compartments primarily based on self-assembled lipid membranes. In this work, we study the interaction of amphiphilic peptido-RNA, in which a lipophilic peptide is bound to a hydrophilic nucleotide, with membranes in three different ways. We have found that amphiphilic peptido-RNA with random or defined sequences could be formed in the presence of lipid membranes but that the latter exerted no significant influence on the reaction networks at play. In contrast, peptido-RNA with short and medium-length membranophilic peptides could durably interact with lipid membranes, providing a localization mechanism. Finally, it is envisioned that, because phosphoramidates are easily cleaved, this localization could be exploited to allow the permeation of nucleotides and short nucleic acids through membranes without the assistance of enzymes. Our results contribute to shed light on a prebiotic peptide-RNA-lipid world.

Keywords: peptido-RNA; phosphoramidate; membranophilic peptides; giant vesicles.

Résumé substantiel en français

Introduction

Les progrès rapides en biologie, notamment cellulaire et moléculaire, mettent régulièrement en évidence de nouveaux phénomènes et structures à l'œuvre dans les êtres vivants. Cet afflux de connaissances génère une curiosité croissante quant à la nature physicochimique de ces processus, aux différences fondamentales distinguant le vivant de la matière non-vivante, et à l'inévitable question de l'abiogenèse : comment et pourquoi la vie peut-elle émerger à la surface d'une planète initialement stérile ? Après s'être majoritairement consacrés, pendant presque deux siècles, à la synthèse et l'étude de molécules isolées, les chimistes organiciens confrontés au vivant doivent adapter leurs modèles et leurs techniques à l'étude de systèmes complexes, hétérogènes et évolutifs. En résulte l'émergence d'un champ disciplinaire récent, baptisé « chimie systémique », dont les applications portent également au-delà de l'étude du vivant.

Toutes les formes de vie identifiées à ce jour reposent sur un jeu commun de mécanismes fondamentaux, permettant à l'organisme d'assurer ses fonctions et aux espèces de persister sur de grandes échelles de temps malgré la durée de vie individuelle très courte des êtres vivants à l'échelle géologique. Les acides nucléiques sont porteurs d'une information qu'ils propagent en se reproduisant à l'identique. Les peptides et protéines catalysent les réactions chimiques indispensables au fonctionnement de l'organisme. Enfin, un compartiment basé sur une membrane lipidique sépare le système de son environnement et le protège. Les interactions entre ces différentes classes de biomolécules sont également fondamentales. Dans les cellules, les ribosomes dirigent, avec le concours d'une cohorte d'enzymes, la synthèse de protéines à séquence spécifique : la traduction. La stabilité et le fonctionnement de la cellule sont également tributaires de l'encapsulation durable de macromolécules dans des compartiments définis, et du transport régulé de petites molécules : influx de nutriments et efflux de déchets métaboliques.

Contexte et hypothèses

Ces travaux de thèse s'inscrivent dans le cadre d'un projet quinquennal collaboratif international, financé par un appel d'offre de la fondation Volkswagen,[a] retenu sous le nom « Molecular life ». L'objectif central de ce projet est de proposer un modèle dans lequel des petites molécules (acides aminés,

[a] La *VolkswagenStiftung* est une fondation allemande établie en 1961, de droit privé mais opérée sur des fonds publics, aujourd'hui sans lien avec l'entreprise automobile du même nom. Elle se consacre au financement et à la promotion des sciences humaines, sociales et naturelles.

nucléotides et lipides) sont capables de développer des comportements physico-chimiques complexes sous la seule impulsion d'une activation chimique, et sans intervention humaine ou biologique. En particulier, ma contribution au sein de ce projet est d'étudier la coévolution de peptides et d'oligonucléotides dans un environnement riche en membranes lipidiques.

Lorsque des acides aminés et des mononucléotides (nucléoside 5'-monophosphates) sont activés par des agents de couplage de type carbodiimide en solution aqueuse, des co-oligomères nommés peptido-ARN émergent spontanément. Ils présentent une connexion *N*→5' entre le fragment peptidique et l'extrémité phosphorylée du fragment nucléique, sous la forme d'un phosphoramidate. Cette fonction est suffisamment stable en milieu neutre pour synthétiser, isoler et étudier des peptido-ARN, mais facilement clivée en milieu acide. Sous l'effet de cette *N*-phosphorylation réversible, la production de peptides par les carbodiimides est drastiquement améliorée. De plus, lorsque des acides aminés lipophiles sont employés, les peptido-ARN obtenus sont amphiphiles, ce qui en fait de bons candidats pour l'interaction avec des membranes lipidiques. L'hypothèse générale de ce travail est qu'une synergie entre acides nucléiques hydrophiles et peptides membranophiles peut s'établir dans un monde lipidique : les peptides sont maintenus en solution et leur production est améliorée par la conjugaison avec des nucléotides ; les nucléotides sont concentrés à proximité des membranes par la conjugaison avec des peptides fortement membranophiles ; enfin cette conjugaison peut permettre la vectorisation de petits acides nucléiques vers l'intérieur (le lumen) de vésicules lipidiques.

Méthodes et résultats

Un premier volet de mes travaux a cherché à mettre en évidence une influence des lipides sur la production de peptido-ARN amphiphiles. Une réaction modèle entre l'adénosine 5'-monophosphate (AMP) et la valine (Val), un acide aminé lipophile mais relativement soluble dans l'eau, en solution aqueuse concentrée avec l'agent de condensation EDC comme source d'énergie a constitué un premier modèle. Ces conditions, décrites dans la littérature, ont d'abord été simplifiées et revisitées afin d'éviter la formation de sous-produits perturbant les analyses, et de maximiser la formation de peptido-ARN. Des analyses par chromatographie couplée à la spectrométrie de masse à haute résolution, complétées par de la RMN hétéronucléaire (^{1}H et ^{31}P) ont permis d'identifier en détail les produits du réseau complexe de réactions à l'œuvre, et de choisir des conditions optimales. La réaction a ensuite été répétée en présence de divers lipides, parmi lesquels des mélanges dits « prébiotiques » (potentiellement formés par des procédés géologiques avant l'émergence de la vie) et des phospholipides présents dans les membranes cellulaires actuelles. Dans toutes les conditions testées : variations de la nature et de la concentration des lipides introduits, aucune influence significative de la présence de lipides n'a pu être établie par rapport aux expériences contrôle sans lipides. Cette absence d'effet est attribué à la trop grande concentration des réactifs en solution, qui surpasse largement la quantité

de lipide, et à la trop grande efficacité générale de la réaction laissant peu de place pour une amélioration. De plus, la complexité des mixtures de produits obtenus limite les possibilités de quantification et d'études systématiques.

Dans une seconde approche, j'ai étudié la formation d'un peptido-nucléotide défini (Val_2A) à partir de divaline (Val_2) et d'une forme pré-activée d'AMP, en utilisant un suivi HPLC, plus sensible que la RMN et facilitant la mise en place d'études cinétiques. Les premiers tests en présence de lipides ont confirmé les résultats précédents en ne mettant en évidence aucun effet face aux contrôles lorsque la concentration en réactifs était proche ou supérieure à celle des lipides. Des concentrations plus faibles en réactifs ont pu être étudiées par suivi cinétique : lorsque la dilution augmente, la formation de Val_2A ralentit tandis que l'hydrolyse de l'AMP activé n'est pas impactée. Les lipides n'ont pas permis de réduire cet effet. L'hypothèse formulée pour expliquer cette seconde absence d'effet est une faible interaction des réactifs avec les membranes lipidiques. Une réaction impliquant des partenaires très membranophiles est en cours de développement. Elle devrait permettre de quantifier les effets d'un confinement membranaire sur les réactions de couplage.

Afin d'étudier plus en détail l'interaction des peptido-nucléotides avec les membranes, une campagne de synthèse et de purification de molécules de référence a été entreprise. La stratégie de pré-activation, plus efficace et générale que le couplage direct, a été retenue et une petite gamme de peptido-nucléotides amphiphiles a pu être produite. Des conjugués plus polaires ont aussi été produits mais aucune des techniques chromatographiques explorées n'ont fourni des échantillons de pureté satisfaisante. L'association des composés amphiphiles avec des vésicules phospholipidiques a ensuite été étudiée par deux méthodes. La première consiste à séparer les vésicules des molécules restées en solution et de mesurer la quantité de substrat alors fixée aux membranes. Cette méthode est pertinente pour mesurer des phénomènes d'adhérence durables et suffisamment robustes pour résister à la technique de séparation employée. Pour la majorité des composés, une méthode non séparative impliquant la mesure de diffusion par RMN a été employée. Cette technique permet d'observer l'échantillon intact et a permis d'identifier des peptides membranophiles. Malheureusement, les données obtenues ne sont actuellement pas quantitatives en raison de limitations techniques.

Dans le cadre du projet collaboratif « Molecular life », nous avons étudié l'interaction d'un peptido-ARN à séquence plus longue avec des membranes phospholipidiques par microscopie confocale. Le conjugué a été préparé par l'équipe du Pr. Richert (Stuttgart, Allemagne) et porte un fluorophore permettant sa visualisation microscopique. Des résultats préliminaires montrent que lorsque ce peptido-ARN est ajouté à des vésicules géantes phospholipidiques, il s'accumule rapidement et spontanément à proximité des membranes.

Conclusions

Les résultats de ce travail suggèrent que des nucléotides et oligonucléotides pourraient être immobilisés à la surface de vésicules lipidiques via une conjugaison covalente mais réversible avec des petits peptides membranophiles à séquence aléatoire. Inversement, des peptides peu solubles dans l'eau à cause de leur tendance à l'agrégation sont rendus solubles par la conjugaison à des nucléotides.

Ces travaux ont permis la mise au point de méthodes simples de synthèse et purification de peptido-nucléotides amphiphiles, et le développement de méthodes analytiques pour caractériser leurs interactions avec des vésicules, ainsi que suivre l'évolution de mélanges complexes de conjugués. Ils s'inscrivent durablement dans l'étude de l'invasion de vésicules lipidiques par des biomolécules et les interactions peptide-ARN en milieu interfacé.

Financement et contributions

Ces travaux de thèse ont été accomplis au sein de l'ICBMS (UMR 5246), Université Claude Bernard Lyon 1, et sont financés par la fondation Volkswagen dans le cadre du projet « Molecular life » (Az 92850). Ils ont donné lieu à 3 publications scientifiques (auxquelles s'ajoutent 2 manuscrits en cours d'évaluation, dont une revue de la littérature), 5 participations à des congrès (2 posters et 3 communications orales) et un mois de mobilité internationale (août 2019) dans l'équipe du Pr. Richert (université de Stuttgart, Allemagne) financée par le DAAD (Deutscher Akademischer Austauschdienst) après sélection d'un projet de recherche. L'auteur a également co-encadré deux stagiaires (Master 2 international, programme Erasmus +, 5 mois, et Master 1 local, 2 mois).

Remerciements

Si les mythes fondateurs ont pu guider des sociétés vers la lumière, le mythe du savant fou errant seul dans sa tour d'ivoire a bien failli me faire passer à côté d'une vocation. Que cette note de remerciements rétablisse la vérité quant à la place du collectif dans la réalisation de ces travaux de thèse.

Avant de plonger dans presque quatre années de laboratoire, je souhaite adresser toute ma reconnaissance et ma considération aux membres de mon jury, ayant accepté de porter un regard nouveau et indépendant sur ces travaux.

A Michele, je dois mon entrée dans le monde de la recherche alors que je n'étais qu'un étudiant agité et quelque peu insubordonné. Tu m'as donné à la fois la confiance et les coups de pied dont j'avais besoin, et je te dois bon nombre des aventures que j'ai vécues des dernières années. Pierre m'a transmis une curiosité sans bornes et une solide éthique scientifique. J'ai pu mener cette recherche avec autonomie, tout en bénéficiant du soutien individuel et scientifique dont j'avais besoin. Je garderai un souvenir durable de nos (très) longues discussions dont je suis toujours sorti instruit. Je remercie chaleureusement ce tandem de m'avoir fait confiance et de m'avoir ouvert toutes les portes dont ils avaient les clés.

Après deux ans de solitude au bout d'un couloir sombre, c'est avec une joie non dissimulée que j'ai accueilli Augustin et Carolina au laboratoire. Je vous souhaite de trouver dans votre propre thèse toute l'excitation que j'ai ressentie en préparant la mienne. En particulier, je suis fier d'avoir participé à la formation de Carolina puis à l'obtention de sa place en thèse. *Congratulazioni* !

Je veux dire encore tout mon amour à mes collègues du LCO2, pour les bons et les mauvais moments, le café du matin comme les tâches ingrates du laborantin. Merci à Yoann et Nico qui m'ont accompagné dans mes premiers pas, merci à Baptiste, Bastien, Fanny, Idris, Lei, Maxime, Yann pour ces bonnes années, à tous nos stagiaires et visiteurs : Gabrielle, Pierre Neff, Betty et tous ceux qui se reconnaitront. Bien entendu, tous mes remerciements à nos chercheurs permanents David, Peter, Seb et Thierry pour leur disponibilité et leurs conseils. J'ai eu le plaisir d'évoluer pendant ces années dans un environnement privilégié, entouré de personnes que je respecte et affectionne.

Je tiens à remercier également les membres de l'institut avec qui j'ai pu échanger, apprendre, qui m'ont prodigué des conseils ou confié du matériel. Et parce que nous ne sommes pas là uniquement pour la science, je salue la bonne humeur de Nono, Willy, Aurélie, Youssef, Alexander, Marie-No, Chérif, Fra, Kévin, Clara, Paolo et tant d'autres, grâce à qui les coups durs devenaient plus faciles à surmonter.

J'ai toujours aimé bricoler des machines compliquées. Malgré tout, j'étais en arrivant au laboratoire d'une incompétence crasse pour tout ce qui concerne les sciences analytiques, qui allaient pourtant prendre une place importante dans mes travaux. Je suis infiniment redevable au personnel du CCSM et du CCRMN

pour leur travail quotidien acharné et leur patience. En particulier j'aimerais remercier Alexandra pour m'avoir patiemment accompagné lors de mes débuts en chromatographie, Emmanuel, avec qui j'ai fait mes premiers pas en RMN, et Anne, qui a pris le relais. Le temps et l'attention que vous avez pu m'accorder (et avec le sourire en prime) continue de me surprendre, et cette transmission de savoir s'est révélée très précieuse.

Je tiens à remercier Clemens Richert, Martin Egli et Pradeep Pallan pour nos discussions très instructives lors de nos réunions annuelles à propos du projet Molecular Life. J'ai également eu le plaisir d'être accueilli au sein du Richert Group en août 2019 et je salue Peter, Eric, Jennifer, Maximilian, Dejana, Carolina et le reste de l'équipe pour leur accueil chaleureux et leurs conseils.

Enfin, je tiens à saluer toutes les personnes, non chimistes pour la plupart, qui ont égayé mon quotidien hors du laboratoire. Merci à mes amis, mes collègues de l'orchestre de Neuville, et à ma famille : Catherine, Pascal, Mathias (bonne chance pour ton propre doctorat !), Noémie, Dadou, Nanou, Vincent, Corinne, Gabin et tous les autres, pour les soirées, les week-ends, les vacances, et tout ce qui fait le sel de la vie.

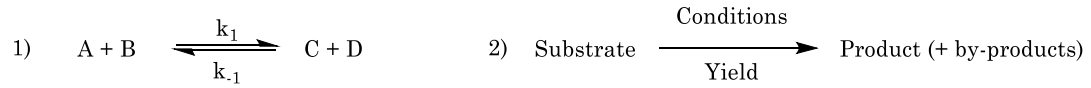
Dimitri

Table of contents

The tables of contents and the cross-references in the text are all clickable.

1. Introduction

Molecular chemistry is a relatively recent science. Physics laws as old as Newton's universal gravitation[1] are still considered cornerstones more than three hundred years after their publication. In contrast, Wöhler's synthesis of urea in 1828,[2] classically considered the first example of the artificial preparation of a natural organic compound, would look outdated in a modern chemistry textbook. The wide acceptation of modern atomic and molecular models was only achieved during the twentieth century and much of the organic chemistry currently taught to students was developed after the Second World War. With the industrial revolution in the western world and the suddenly wide availability of petroleum, organic chemistry mainly grew through the prism of synthesis, driven by lucrative industrial applications. The predominance of synthetic organic chemistry persisted to this day, and our textbooks are still filled with organic transformations of one molecule into another, that usually take one of the two "reagents-arrow-products" forms (Scheme 1.1). This typography is actually representative of the way chemists think of chemical processes: either by meticulously characterizing one process in isolation, or by focusing on the transformation of a substrate of interest with a certain disregard for the other "sacrificial" reagents. While this target-oriented approach of chemistry is particularly efficient when dealing with the preparation of chemicals for specific applications, many phenomena cannot be conveniently described in such formalism or studied with the associated methods.



Scheme 1.1 Typical presentation of chemical equations in current textbooks.

For about fifty years, supramolecular chemistry has been developing new tools to approach non-covalent interactions, the folding of macromolecules in space, cooperative interactions between molecules and many dynamic or non-covalent behaviors that are poorly described with a synthetic chemistry approach. This task was made possible by the considerable improvements of analytical chemistry, that now permits the routine measurement of affinity, diffusion, or energetic parameters, as well as the study of complex mixtures. This last point in particular is significant, as chemists tend to prefer pure compounds over mixtures, although many natural systems are actually constituted of thousands of substances.[3]

The artificial synthesis of molecules of biological interest from inorganic or man-made precursors evidenced that living organisms were built from the same matter as inanimate objects. Ultimately, the vitalistic models of life that prevailed before the nineteenth century had to be rejected. Yet, this did not solve the questions about the nature of life, nor about the mechanisms behind its emergence at the surface of a formerly sterile planet: abiogenesis. Indeed, no absolute definition of life is agreed upon to this day, although some processes that distinguish living organisms from inanimate systems have been evidenced.[4]

When chemists first tried to reproduce the synthesis of key molecules of life under conditions that they deemed geologically plausible on a young planet Earth, they often faced the production of intractable solutions[5] or thick brown tars.[6] They had discovered the hard way that life is not just a juxtaposition of chemical reactions. When living organisms first emerged from the inanimate (abiogenesis), mechanisms that enabled the selection of pertinent products among a soup were highly needed, as were chemical pathways mild and selective enough to limit the complexity of the soup itself in the first place.[7] It is also pertinent to question the variety of conditions that were necessary to obtain the bricks of living organisms. In the laboratory, the chemist purifies each synthetic intermediate and chooses an optimized set of conditions for the next reaction. Such prospect is unrealistic for any spontaneous geological event.[8] Spontaneous purification may only be achieved by simple processes such as crystallization or rain-down of volatile compounds, a number of times low enough to be reasonably plausible. These considerations stressed the fact that chemical networks producing all, or at least many, of the components necessary for life were more promising than individual, contradictory sets of conditions.[7,9]

There is now a good consensus that the dominant form of life, if not the only one, to have arisen from abiogenesis was cell-like.[10] Such protocellular organism must have featured a support for heredity (the genome), functionalities (metabolic pathways) encoded in the inherited genome (translation) and the possibility to produce faithful copies of organisms at an exponential, or at least parabolic, rate (nucleic acid replication and membrane growth and division). The encapsulation of the biochemical machinery responsible for this prodigy was essential to let fit systems grow at the expense of weaker ones (evolutionary selection),[11] selectively harvest small nutrients from the outer environment and eliminate waste, while preserving the order and local concentration of key biomolecules (Figure 1.1). Of all possible compartments, lipid vesicles are the simple models closest to current cell membranes and are by far the most represented in synthetic biology studies.[12]

The prospect of synthesizing cells[13,14] seems to be an unavoidable quest in science, that would suffer no question about both its fundamental relevance and the plethora of applications that may arise from such technology. The overwhelming number of steps and challenges that were successively faced since the beginning of this enterprise has mobilized skills and investigations in an unusually high variety of fields. Chemistry continues to play a prominent role among these. However, the traditional toolbox developed under the influence of synthetic chemistry is insufficient to describe complex life-like systems. One reason is that constructs more complex than chemical reactions in solution are involved: molecular interaction, aggregation, self-assembly are so many phenomena that must be considered. Another is the inherent property of living organisms to keep themselves far from the thermodynamic equilibrium (death) by exploiting continuous environmental variations[15] or wasting fuel.[16] Taken together, these aspects brought about the recent elaboration of a new theoretical and experimental framework: systems chemistry.[17–19]

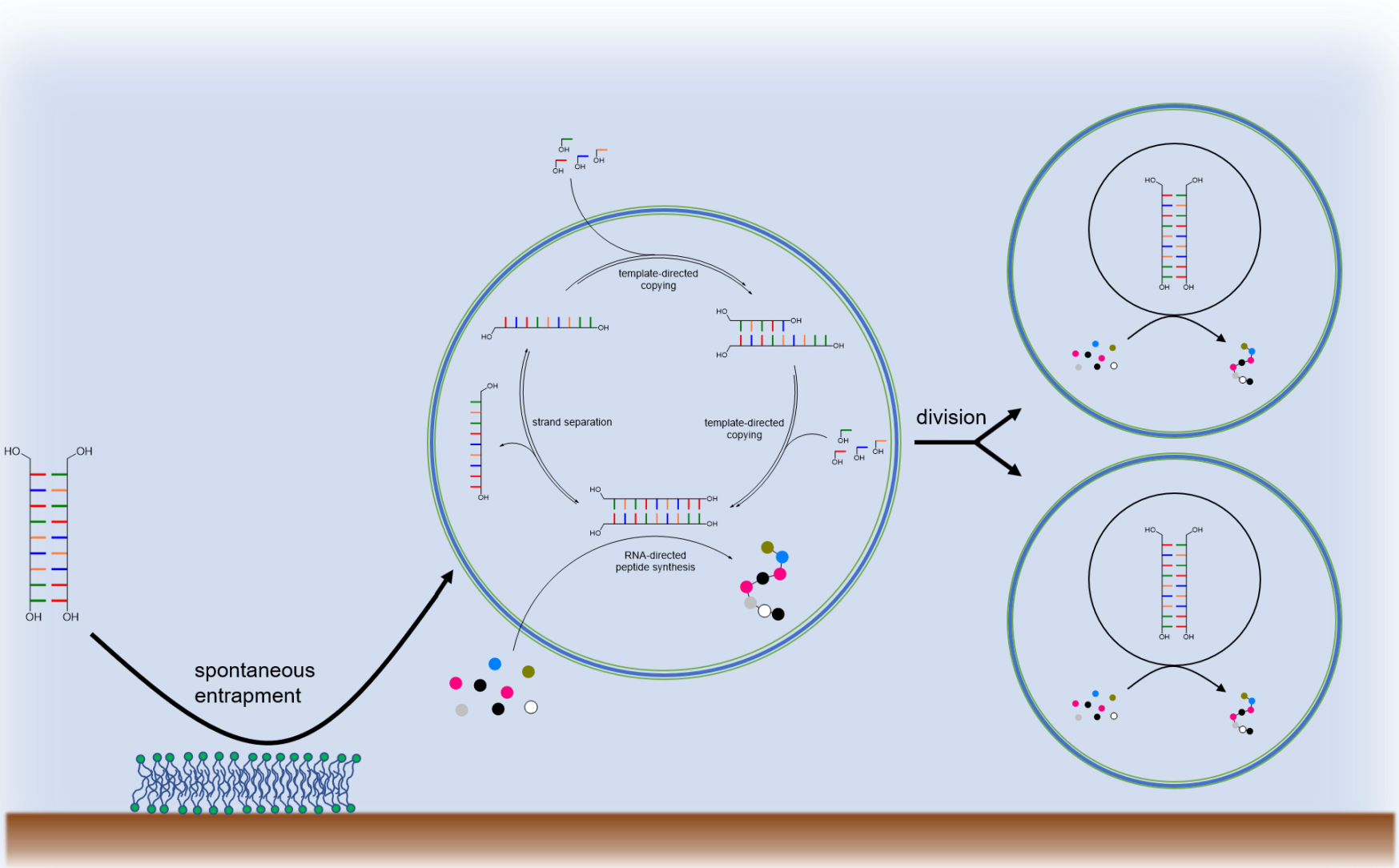


Figure 1.1 Schematic representation of a possible mechanism for the emergence of protocells.

As it will be reviewed later on, it is generally assumed that protocells were formed by the spontaneous entrapment of essential oligomers, typically RNA and peptides, in the aqueous core of lipid vesicles (the lumen)[14] or through anchoring to their membrane.[13,20] The selective permeation of small molecules through the membrane then established a continuous influx of nutrient and efflux of waste. This scenario required that lipids have been available very early on Earth, which is believable from a chemist's point of view. In the absence of metabolic pathways, the many C-C bonds of the long hydrocarbon chains required to build amphiphiles were likely produced under strong thermal conditions that would degrade most other molecules.[21] It must then have happened very early. This consideration should motivate the study of prebiotic oligomer formation and physicochemical behavior in lipid-interfaced environments, a topic that knew relatively few developments, perhaps because of the underlying practical difficulties. The examples are mostly limited to the formation of very hydrophobic compounds.

Investigating the hypothesis of a membrane-promoted formation of oligomers in a relatively general context requires a model oligomerization reaction. In most prebiotic studies, peptide and RNA formation and elongation were treated separately under quite different conditions. However, the crucial role of translation in life stresses the need for an early emergence of a joint peptide-RNA chemistry, in which nucleic acids were able to produce peptides in a controlled manner before the evolution of functional ribosomes and enzymes. Indeed, progress toward the prebiotic access to aminoacyl RNA were made.[22] In addition to this come some physicochemical preoccupations. Prebiotic peptides were probably dominated by hydrophobic residues and insoluble aggregates, that would typically accumulate on the ground and remain inert. At the opposite, the soluble, mostly membranophobic nucleic acids would be inevitably washed away and diluted before they could serve any purpose. The early formation of amphiphilic peptide-RNA conjugates[23] between lipophilic peptides and hydrophilic nucleic acids lets us envision a win-win situation[20,24] in which peptides were made soluble enough to react, while they retained their covalently bound nucleic acid through the formation of aggregates,[25] or through anchoring to membranes.[26]

Recently, the formation of peptide-RNA conjugates bound through a phosphoramidate function received a renewed interest.[27] Under simple conditions, using chemical activation provided by carbodiimide coupling agents, the so-called peptido-RNA formed from amino acids and mononucleotides, with notably a peptide growth much more efficient than from amino acids alone under the same conditions.[28] This system is thus a good example both of integrative chemistry, capable of forming several useful chemical bonds of different nature under a single, simple set of conditions, and of ribonucleotide-enhanced production of peptides, a step toward translation. Another interesting feature is that phosphoramidates have an intermediate stability. They are robust enough to be synthesized and studied more conveniently than the extremely labile aminoacyl nucleotides, but are somewhat sensitive to hydrolysis, and therefore

prone to release their peptide chain,[29] which is critical to achieve fuel-driven dynamic features: make and break.[16]

To this day, the formation and behavior of peptido-RNA has never been studied in the presence of lipid membranes. RNA is inherently polar and short RNA strands typically do not interact with natural, anionic membranes. Quite long sequences have been found to adhere to the surface of vesicles through very specific aggregation patterns, but without providing any permeation.[30] Our hypothesis is that amphiphilic peptido-RNA featuring a membranophilic peptide residue may be localized to membranes, as already demonstrated with medium-length stable analogues of peptidyl-RNA.[26] Moreover, forming peptido-RNA from mixtures of nucleotides and lipophilic amino acids in the presence of lipid aggregates may give rise to a form of selectivity, due to the expectedly varying interaction of these species with the lipids. The opposite benefit is also expected. Many peptides are poorly soluble in water, either because of their lipophilicity or because they aggregate, which is a major obstacle in reactivity. Reversible phosphoramidate conjugation to the extremely soluble oligonucleotides may be a way of providing water solubility to peptides. Interestingly, long peptide chain growth has been observed on peptido-RNA, even with typically aggregating residues such as alanine, leucine, or phenylalanine.[31]

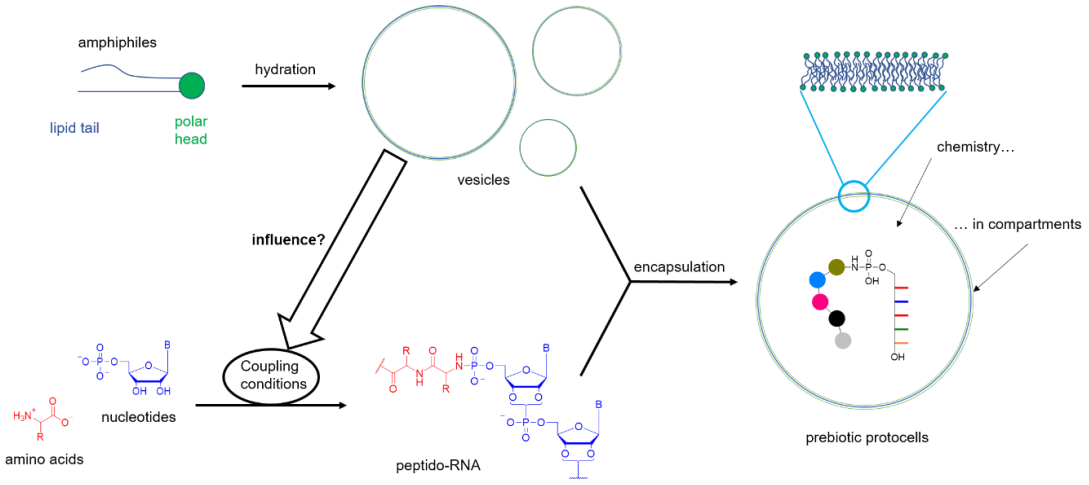


Figure 1.2 Overview of the postulated peptide-RNA-lipid system studied in this work.

Studying complex mixtures of products in heterogenous samples is a difficult task that cannot be achieved with a single analytical strategy. Methods to survey peptido-RNA formation using ^{31}P NMR spectroscopy were already well-established by others at the beginning of this project[31] and were progressively refined and enriched with ^{13}C-based techniques in the last years.[32,33] As a complement, we propose to use liquid chromatography hyphenated with high-resolution bidimensional mass spectrometry to get a deeper view of the chemical variety of the species and by-products arising from peptido-RNA chemistry. Crude lipid vesicles may be imaged by optical microscopy, giving qualitative information. The use of fluorescent labeling allows for richer microscopic studies, sometimes quantitative. Model lipid membranes prepared by well-controlled methods may

be used in spectroscopic assays and are handled and isolated more conveniently than crude vesicles. We propose to quantify the binding of small peptido-nucleotides to lipid vesicles either in situ by the use of NMR spectroscopy, or after separation of substrate-enriched vesicles from the non-encapsulated solutes.

Finally, the synthesis of defined peptido-nucleotides will be a necessary support to many experimental studies. Relatively few methods are known for the preparation of nucleotide phosphoramidates, and none directly yield amphiphilic conjugates in good yields on scales of 0.1-1 mmol. Refining and adapting literature procedures is therefore necessary to efficiently produce the library of compounds necessary for this work. Interestingly, the observations made during the development of synthetic methods should also provide useful insight into the reactivity, stability and mechanisms of formation and hydrolysis of peptido-nucleotides.

2. State of the art

2.1. Prebiotic chemistry: synthesizing the bricks of life under abiotic conditions

Little reliable information remains about the exact conditions under which life arose on Earth. Although geological studies and comparisons with observations made on other planets (and on stellar objects in general) provide hints, a large part of the so-called "prebiotic chemistry" is partially based on speculations. We may never identify the conditions and pathways that actually led to the contemporary living organisms that we know, and the coexistence of multiple believable pathways should not be excluded either. Furthermore, unraveling the fundamental physicochemical mechanisms that allowed known life to emerge and persist over inanimate matter would be a remarkable achievement, even without the definitive certitude of every individual event or the exact geological source of every chemical. The experimental demonstration of such abiogenesis of a living organism, even different from "canonical" ones (a fancy and definitive term to designate the only form of life that we know), would be equally impressive. One must also consider the difficulties inherent to modeling natural processes in a chemistry laboratory. Geological time scales are impractically long, and chemists need to find a balance between plausibility and feasibility. For example, concentrations and temperatures may be increased above the values expected in a true geological setting in order to reach manageable reaction times, provided that no other effect is triggered.

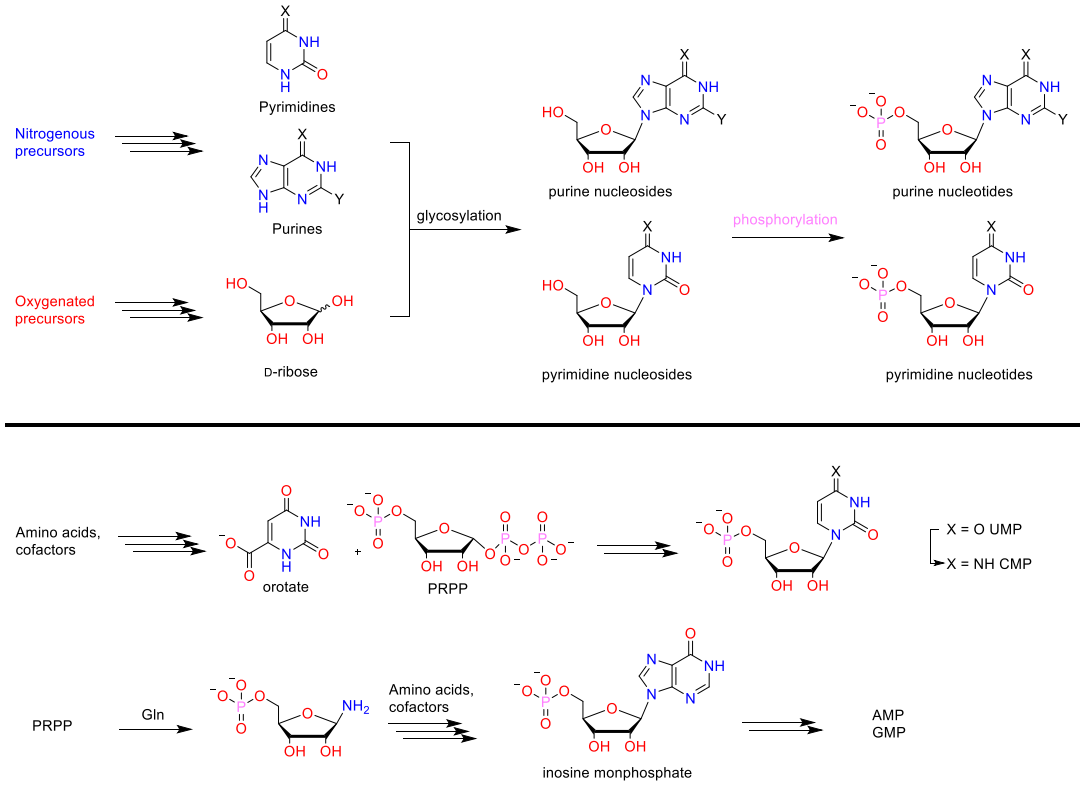
Contemporary life chose the cell as a robust construct, that may be adapted to a plethora of specific environments by fine-tuning the physical characteristics of its constituents.[34] Cells evolved to a spectacular complexity by progressively incorporating features, but primitive organisms may have been more simple. On a (macro)molecular level, the structures conserved across all known forms of life include nucleic acids, carbohydrates, peptides, and lipids, that naturally became the targets of many synthetic attempts from prebiotic chemists. Non-canonical variations of these building blocks are also regularly discovered. Some of them are found in contemporary living organisms while others may have played an early role before being replaced by optimized components through an evolutionary process.

The abiotic, sometimes prebiotic, synthesis of nucleic acids, amino acids and lipids has motivated a very large number of studies, that have been extensively reviewed.[10] We do not wish to give an exhaustive overview of such syntheses but rather to highlight the important development, in the last 10 to 20 years, of increasingly intricate and unified pathways to the bricks of life, suggesting that physicochemical phenomena connected chemical reactions before the emergence of proper living organisms.

2.1.1. Classical approaches

2.1.1.1. Nucleic acids

The most intuitive retrosynthesis of ribonucleotides involves the separate formation of the nucleobases and ribose, that are then joined together through a glycosylation step giving nucleosides, and phosphorylated to nucleotides (Scheme 2.1, top). This is similar to the current biosynthetic pathway to pyrimidine nucleotides. In contrast, purine nucleotides are elaborated progressively from an early ribosylated scaffold. In both cases, amino acids provide most of the atoms for the nucleobases and ribose comes in an activated form as an anomeric pyrophosphate (Scheme 2.1, bottom).



Scheme 2.1[a] Top: intuitive abiotic retrosynthesis of nucleotides. Ribose (or an alternative sugar scaffold) and the nucleobases are formed separately. A glycosylation step yields nucleosides that are further 5'-phosphorylated. Bottom: simplified biosynthesis of nucleotides. In the pyrimidine pathway, orotate is first synthesized then glycosylated and derived to either uridine or cytidine monophosphate (UMP and CMP). In the purine pathway, the ribose scaffold is built first in a hemiaminal form, sourcing nitrogen from glutamine (Gln). Inosine monophosphate (IMP) is elaborated in situ then derived to adenosine or guanosine monophosphate (AMP and GMP).

Several chemical pathways to nucleobases have been proposed very early. Ammonium cyanide condenses to purines,[35] while cyanoacetylene was recognized to be the key building block to pyrimidines.[36] Uracil was also obtained from urea.[37]

[a] This manuscript features a large number of ionic species. The exact states of protonation depend on the precise pH that may remain elusive in some cases. Furthermore, counter-ions are not represented unless specifically relevant, as they may not be precisely defined, especially in solution.

More recently, the versatile building block formamide was also investigated as a source of nucleobases under thermal conditions.[38] The perspective of a very straightforward access to nucleobases from simple precursors is attractive and has been widely accepted but is considerably hampered by the matters of yields and selectivity. Indeed, under all these conditions, complex mixtures of small molecules and heterocycles are obtained, in which canonical bases are often very minor. This is a problem when developing networks that must behave autonomously, without the intervention of a chemist.[8]

Ribose is formally a pentamer of formaldehyde, as are all canonical pentoses. Boutlerow's formose reaction,[6] later refined by others under various conditions[39,40] illustrated that heating formaldehyde under basic conditions provided indeed a very complex mixture of sugars, from which ribose was unfortunately a minor constituent that was unlikely to stand out in any further chemical process. The stabilizing role of some minerals, especially borate, has been suggested[41] but, to date, it remains unclear how ribose could have been selected from a tar to give rise to ribonucleotides, without extensive formation of other glyconucleotides.

Further concern was raised by the very difficult glycosylation step that should join nucleobases and ribose together. Indeed, only purines could be glycosylated under currently known conditions, with a very low yield, and not with the desired regioselectivity. Indeed, it appeared that glycosylations exploiting exocyclic nitrogen atoms of canonical and noncanonical bases were more efficient than the desired glycosylations at endocyclic nitrogen(s).[42,43] This observation placed the intuitive retrosynthesis of nucleosides in contradiction with the selection of the current canonical set of bases.

2.1.1.2. Amino acids

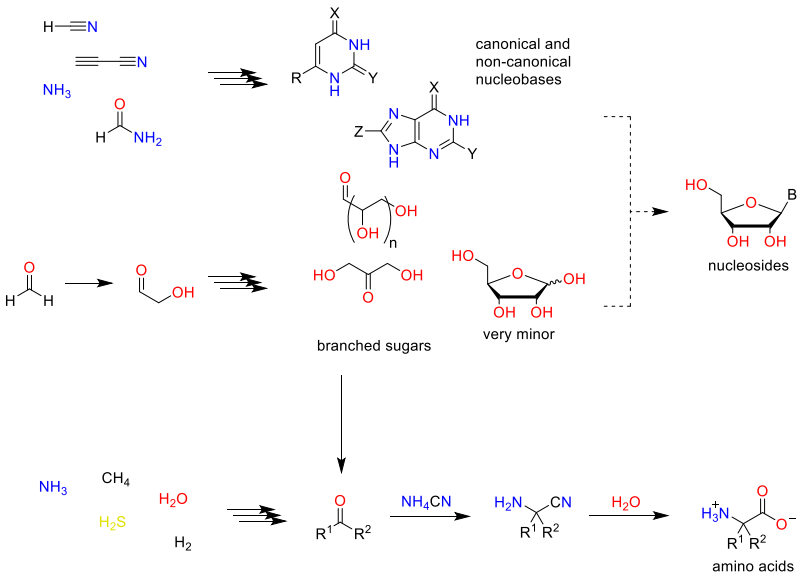
In a similar pattern, the abiotic synthesis of some proteogenic amino acids was demonstrated early in a now historical experiment using simple gases (NH_3, H_2O, H_2 and CH_4) as precursors and repeated sparks to provide activation energy.[5,44] The inclusion of hydrogen sulfide in more recent experiments expanded the scope of products to sulfur-containing amino acids.[45] Similarly to the oligomerization of HCN, many small molecules other than amino acids were also obtained under these conditions.

Amino acids can also be generated in a more specific manner by the Strecker synthesis. It involves the reaction of imines with cyanide to yield α-aminonitriles, that are further hydrolyzed to α-amino acids (Scheme 2.2).[46] This strategy can also be used to produce increasingly large carbohydrates by sequential homologations, and is known in this context as the Fischer-Kiliani sugar homologation.[47,48] An aldose with n carbon atoms first reacts with cyanide ions to generate a cyanohydrin ($n+1$ carbon atoms) that is reduced to a hemiacetal (a diastereomeric mixture of $n+1$ sugars). In a prebiotic context, a contradiction arises from the fact that the strongly basic conditions of the formose reaction must be switched to a somewhat acidic environment to perform both the reduction and hydrolysis steps (see section 2.1.2.4 for a more complete discussion of prebiotic reductions). The easy formation of cyanohydrins also raises the concern that the

oxygenous, strongly alkaline chemistry of the formose reaction must be performed separately from the acidic cyanide chemistry that is classically believed to lead to nucleobases and amino acids.[49] Furthermore, aldehydes and ketones are equally sensitive to cyanide addition, so that applying the Strecker synthesis to a crude aldol mixture will lead to both α-mono- and α,α-disubstituted amino acids, only the first ones being proteogenic.[50]

2.1.1.3. Overview

Many of the now classical reactions that were long considered to provide the molecular bricks to primitive life (summarized in Scheme 2.2) are derived from ancient organic reactions. Chemists were probably attracted by their "rustic" aspect, when compared to the current standards of organic synthesis. As discussed above, the strict likeliness of each individual process and substance on the early Earth may never be strictly proven, but it must be pointed out that the so-called "three pillars of prebiotic chemistry"[7] are mainly unified by the concerns that they raise. All of these reactions yield complex mixtures of myriads of similar products, with little to no selectivity toward the universal components of contemporary life. Moreover, the poor compatibility of the chemical conditions between the different routes require strict spatial or time separations that are unlikely. As the most intuitive and, at first glance, the simplest pathways failed to deliver the key molecules of biology, a fresh approach was required.



Scheme 2.2 Overview of the classical prebiotic syntheses of nucleosides, carbohydrates and amino acids. The limited interactions between the different domains (nitrogenous and oxygenous chemistry) highlight the lack of chemical compatibility between the routes.

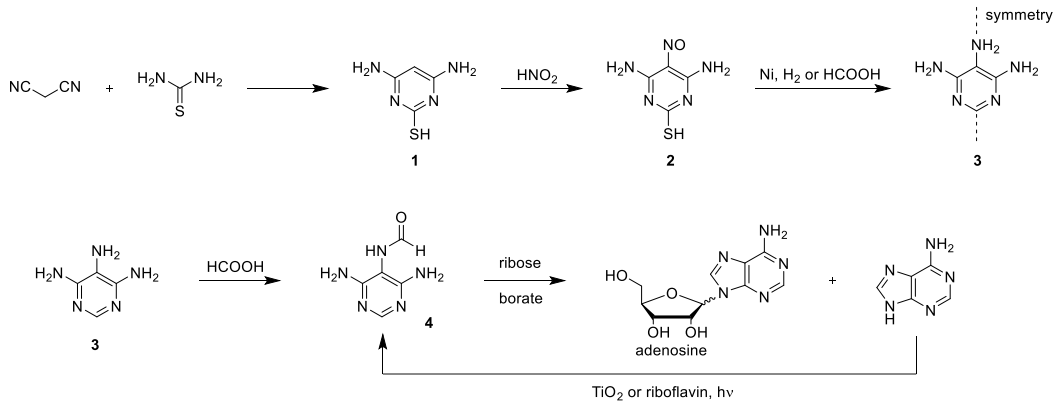
2.1.2. The emergence of unified synthetic pathways

To take over the challenges enumerated above, new approaches to the abiotic synthesis of the bricks of life were proposed during the last decade and are under

active development. These new pathways are somewhat less intuitive, but they avoid some notoriously problematic steps such as the ribosylation of nucleobases. Another remarkable feature is the emergence of the so-called "systems approach", in which chemists try to connect their reactions and intermediates through dense, vastly interconnected networks. The core concepts of these systems are that most molecules of life must have arisen from a common set of reagents and conditions, and that the intrinsic reactivity of the intermediates must have provided the driving force toward the desired products, with the best possible exclusion of side products. Furthermore, the matters of accumulation and purification of the intermediates through natural processes such as crystallization, selective precipitations, evaporation or sublimation followed by rain-down are progressively incorporated, leading to more plausible networks where far less experimental intervention is required. Prebiotic synthesis is thus paving the way to systems chemistry.

2.1.2.1. Regioselective pathways to natural canonical and non-canonical nucleosides

The direct glycosylation of purines and pyrimidines with ribose is poorly efficient and yields a mixture of regioisomers. This difficulty was overcome by Carell and coll. (exploiting ancient observations of Traube[51]) by performing the glycosylation with partially formed nucleobases that were ring-closed in a later step. They first described the synthesis of adenosine by this strategy,[52] starting from malononitrile and thiourea. The two compounds were condensed to yield thiopyrimidine **(1**, Scheme 2.3**)** that was then nitrosylated to **(2)**, and reduced and desulfurated to **(3)**. Compound **(3)** is symmetrical along a C2-C5 axis, and because of the different basicities of the 5- and the 4 and 6 amino groups, the 5-amino group could be selectively formylated in formic acid to give the formamidopyrimidine (FaPy)[a] **(4)**. **(4)** is also symmetrical along the same axis and could react with ribose and concomitantly cyclize to give adenosine without regioselectivity concerns (Scheme 2.3).

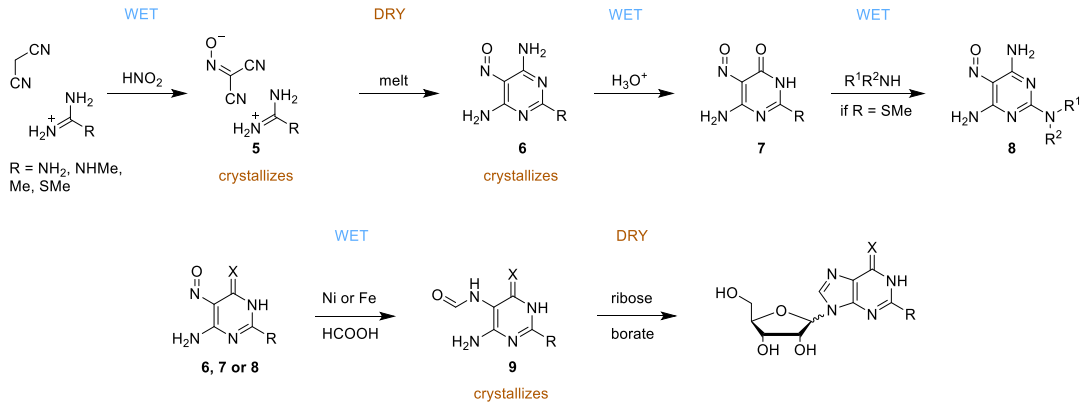


Scheme 2.3. Synthesis of adenosine by the formamidopyrimidine (FaPy) pathway.

[a] A list of abbreviations is available in the appendices.

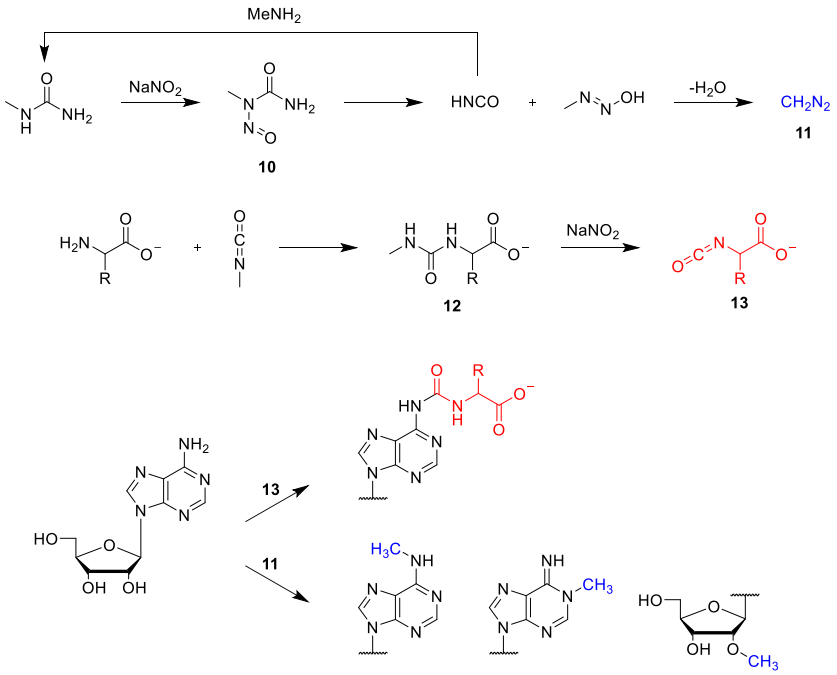
Interestingly, the reaction also performed when a crude mixture of sugars, prepared from glycolaldehyde and glyceraldehyde under basic conditions, was used instead of pure ribose. In this case, many products were obtained, among which threonucleosides and the desired ribonucleosides, in the pyranose and furanose forms, were detected and quantified by liquid chromatography hyphenated to mass spectrometry (LC/MS).

The authors later refined the route,[53] exploiting alternations of aqueous and dehydrative conditions both as a synthetic driving force and as a means to concentrate certain intermediates by selective evaporative crystallization (Scheme 2.4). The careful choice of pH allowed for the selective nitrosylation at C2 of malononitrile to take place in the presence of amidines, giving co-crystals of amidinium nitrosylates **(5)**. When molten by heating, the crystals reacted to give 5-nitrosopyrimidides **(6)** that precipitated spontaneously from water. The by-products could then be washed away. At this step, some structural diversity could be introduced. The 4-amino group could be hydrolyzed to give guanine precursors **(7)**. Moreover, if a methyl sulfide were present on the position 2 of **(6)** (R = SMe, using *S*-methylisothiourea as a precursor), it could be substituted by nucleophiles such as ammonia or amines to give derivatives **(8)**. This convenient platform resonates with the presence of several non-canonical nucleotides in contemporary DNA and RNA, that serve various purposes, and may have originated simultaneously to canonical nucleobases.[54] As in the previous strategy, nitroso derivatives **(6)**, **(7)** or **(8)** were all reduced and selectively formylated using Fe or Ni powder in formic acid. Further ribosylation provided the corresponding nucleosides.



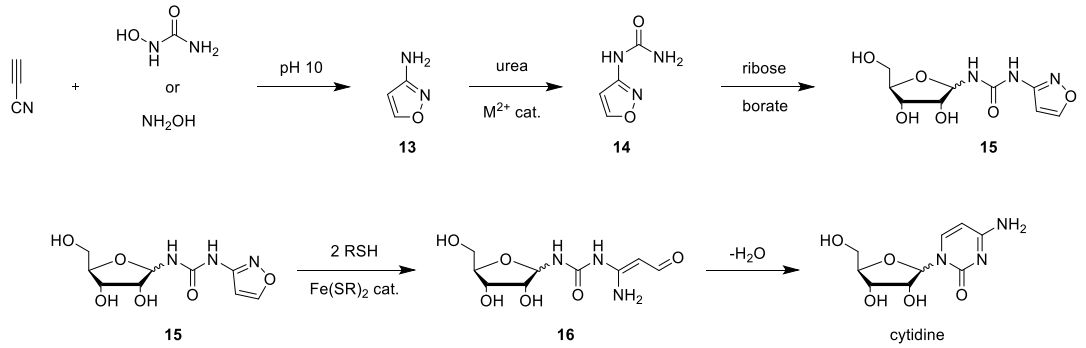
Scheme 2.4 Revisited FaPy route used an alternance of wet and dry conditions to promote chemical reactions and selectively crystalize key intermediates.

The authors also demonstrated in another study that a similar nitrite/urea chemistry afforded an efficient prebiotic route to the modification of already formed nucleosides (Scheme 2.5).[55] Indeed, nitrosylation of *N*-methylurea generated diazomethane in situ, that could methylate nucleobases. Similarly, ureas of amino acids were converted to the corresponding isocyanates that carbamoylated the nucleophilic amines of nucleobases to give products that are found in the current genomes of a wide diversity of species.



Scheme 2.5 Prebiotic methylation and aminoacyl carbamoylation of adenosine using nitrite-urea chemistry.

Remarkably, pyrimidine nucleosides were also accessible through a slightly different, yet compatible route[56] that involved a strategic isoxazole intermediate **(13**, Scheme 2.6**)**, prepared from either *N*-hydroxyurea or hydroxylamine. To this end, **(13)** reacted with urea in the presence of zinc (II) or cobalt (II) salts to give **(14)** that underwent ribosylation to **(15)**. The key step to the pyrimidine skeleton was then the reduction of the N-O bond, efficiently mediated by catalytic iron (II) sulfate with thiols as stoichiometric reductants. The linear intermediate **(16)** then cyclized to cytidine upon heating.



Scheme 2.6 Synthesis of pyrimidine nucleosides through a key isoxazole intermediate, under conditions compatible with the FaPy route to purine nucleosides.

Interestingly, the authors demonstrated that a phosphorylation step leading to CMP and UMP (through hydrolysis on the 5-position of the pyrimidine) could be integrated in a one-pot process. They used conditions already reported[57] that exploited the mineral lüneburgite ($Mg_3[(PO_4)_2 | B_2(OH)_6] \cdot 6H_2O$) as a natural source of phosphate in molten urea. The authors drew an interesting connection between their purine and pyrimidine nucleoside syntheses, that rely on a similar

set of reagents and catalysts. They were indeed able to synthesize the four Watson-Crick nucleosides in a single pot, thus demonstrating the high degree of selectivity of their reaction pathways.[56]

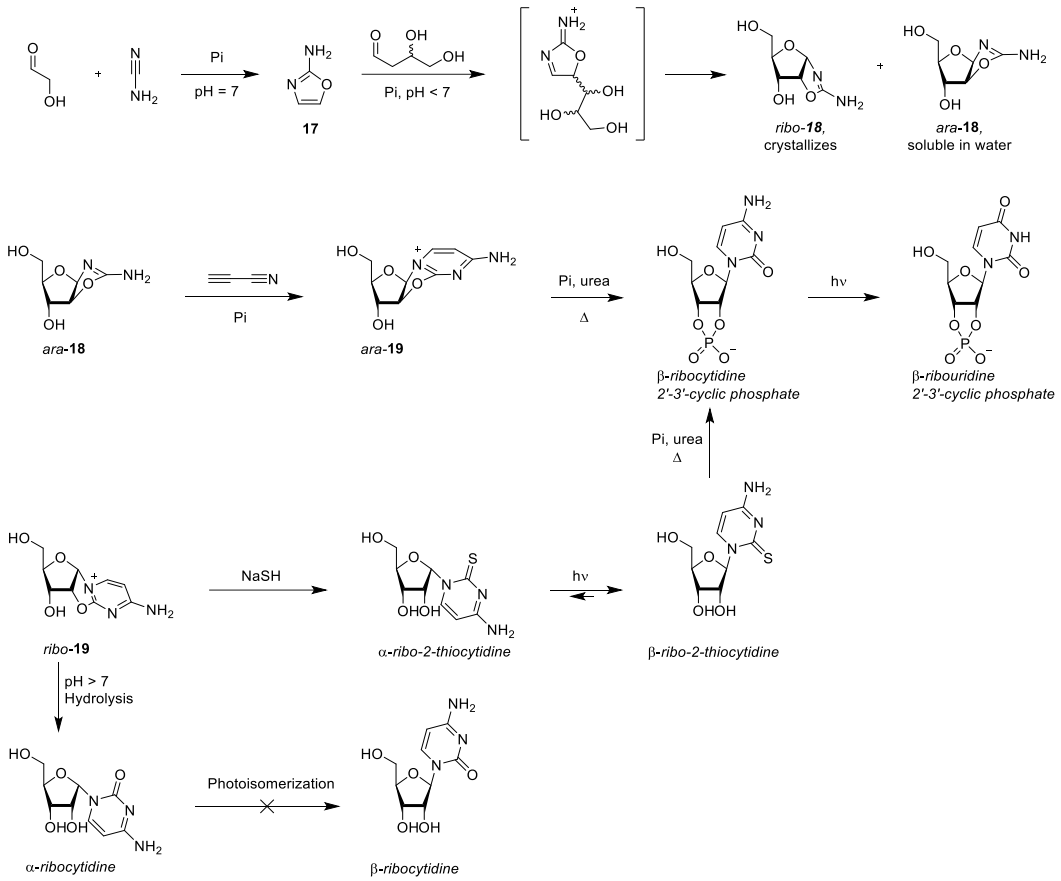
2.1.2.2. Synthesizing nucleotides without ribose

Carell's group strategy relies upon *"robust reactions that perform under "dirty" conditions"*.[54] One drawback of this paradigm is that the desired products are obtained in relatively small amounts among a complex mixture, with no particular selectivity. Indeed, the ribose required in their pathways is assumed to be produced through non-selective sugar syntheses. Even using pure ribose as a precursor, both pyranose and furanose ribonucleosides form, further complicating the outcome of the reaction.

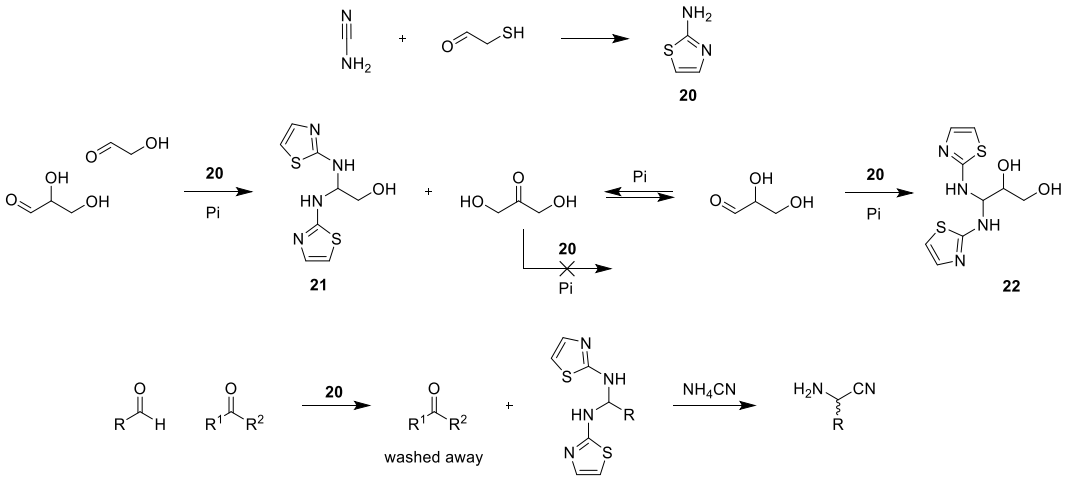
An alternative way to pyrimidine nucleotides that avoids the use of ribose was proposed in 1970 by Sanchez and Orgel but with very low yields.[58] It was revisited and significantly improved in 2009 by Powner and Sutherland (Scheme 2.7).[59] They first generated 2-aminooxazole **(17)** from cyanamide and glycolaldehyde, that then reacted with glyceraldehyde to give a bicyclic intermediate **(18)**. **(18)** combined elements from the forthcoming ribose and pyrimidine. Remarkably, the reaction yielded almost exclusively *ribo*-**18** and *arabino*-**18**, among all the other possible isomers (hexose-like isomers and other pentoses). *Ribo*-**18** and *ara*-**18** were further differentiated by the selective crystallization of the *ribo* isomer from water. Cyanoacetylene then completed the nucleobase scaffold to give anhydronucleosides **(19)**. *Ara*-**19** had the suitable stereochemistry to be opened under phosphorylating conditions (molten urea) to cytidine 2'-3' cyclic phosphate. The *ribo* isomer could not be anomerized directly but gave anomerizable 2-thiocytidine after thiolysis (Scheme 2.7, bottom). The sulfur atom was replaced with oxygen during the phosphorylation step.

This strategy only used glycolaldehyde and glyceraldehyde, the simplest carbohydrates, in lieu of pentoses. The latter was in interconversion with dihydroxyacetone, with an equilibrium strongly shifted toward the ketone form. The authors therefore needed an efficient separation process to successively release glycolaldehyde then glyceraldehyde, while avoiding dihydroxyacetone accumulation as much as possible. This was accomplished using the heterocycle 2-aminothiazole **(20)** as a selective precipitation agent that formed crystalline aminals with aldehydes but not with ketones (Scheme 2.8, top).[60] When glycolaldehyde and glyceraldehyde reacted with **(20)**, the aminal of glycolaldehyde **(21)** precipitated first within hours, followed by that of glyceraldehyde **(22)** several days later. Dihydroxyacetone was progressively consumed by the formation of **(22)**. This selective crystallization of aldehydes also provided a solution to a long-standing problem in the Strecker synthesis of amino acids: aldehydes and ketones both react with ammonium cyanide to yield α-aminonitriles, but only those originated from aldehydes give the natural α-monosubstituted amino acids. Starting from a mixture of carbonyl compounds,

the selective precipitation of aminals of aldehydes followed by washing away the remaining ketones resulted in an efficient and selective synthesis of proteogenic aminonitriles with excellent yields and no detectable formation of α,α-disubstituted aminonitriles (Scheme 2.8, bottom).

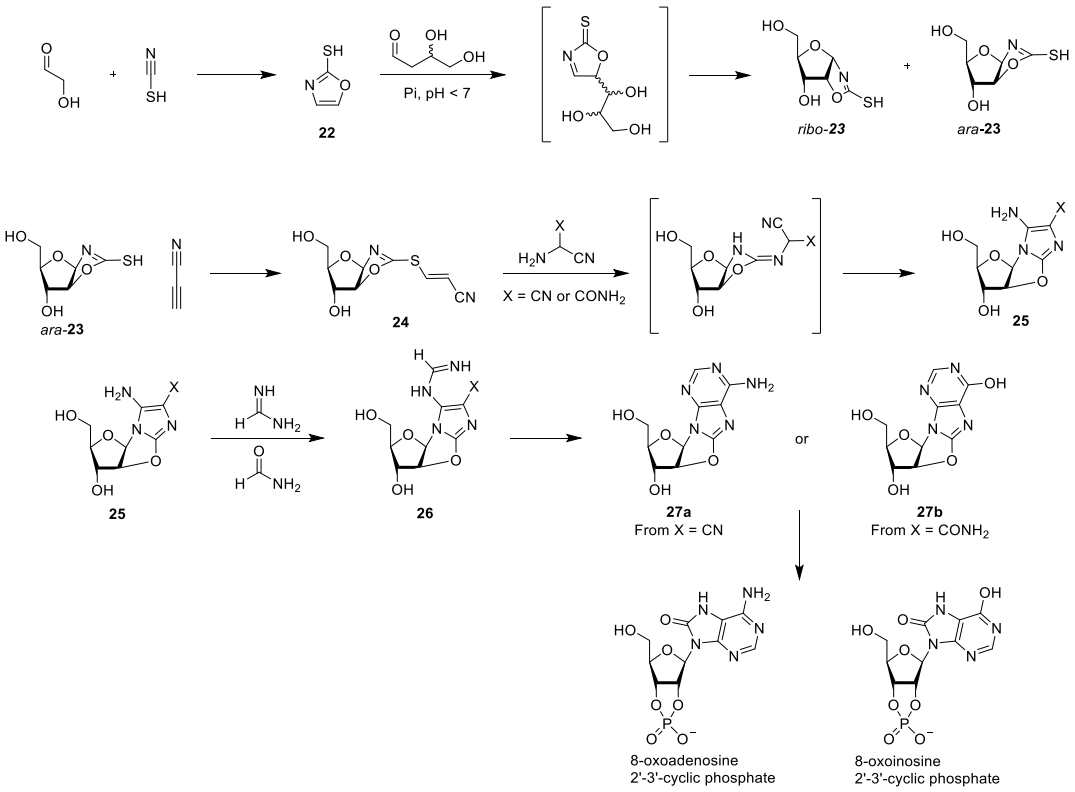


Scheme 2.7 Synthesis of pyrimidine nucleotides by Powner and Sutherland that did not require ribose



*Scheme 2.8 Selective precipitation of aldehydes as aminals of 2-aminothiazole **(20)**. Top: formation of **(20)**. Middle: stepwise precipitations of glycolaldehyde and glyceraldehyde depletes dihydroxyacetone. Bottom: selective formation of α-aminonitriles of aldehydes in the presence of ketones.*

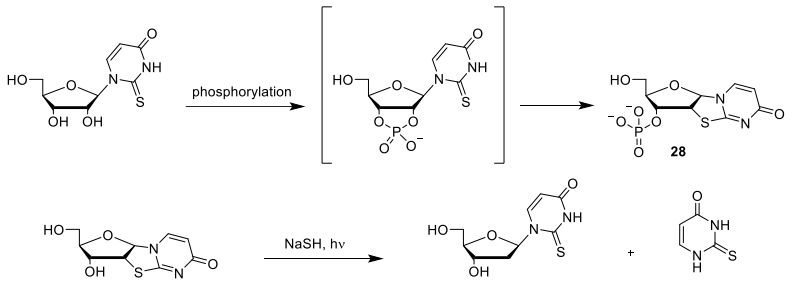
The strategy for the ribose-free synthesis of pyrimidine nucleotides was further extended to purines.[61] The overall route (Scheme 2.9) was very similar to the pyrimidine route, which suggested a high degree of compatibility between these. Indeed, several steps were successfully accomplished concomitantly by the authors. The route started with the preparation of 2-thiooxazole **(22)** that then reacted with glyceraldehyde to give *ribo*-**23** and *ara*-**23** similarly to the preparation of pyrimidine nucleotides. A divergence occurred as the nucleobase fragment had to be further elaborated. The sulfide of *ara*-**23** was cyanovinylated to give **(24)**, the cyanovinyl sulfide group was then displaced by a highly nucleophilic aminonitrile to give the intermediate **(25)**. The six-membered ring of the purine was then built using formamidine. Finally, simultaneous phosphorylation and opening of anhydro intermediates **(27)** yielded 8-oxopurines 2'-3' cyclic phosphates.



Scheme 2.9 Synthesis of 8-oxopurine nucleotides by Powner and coll. in a route closely related to their preparation of pyrimidine nucleotides.

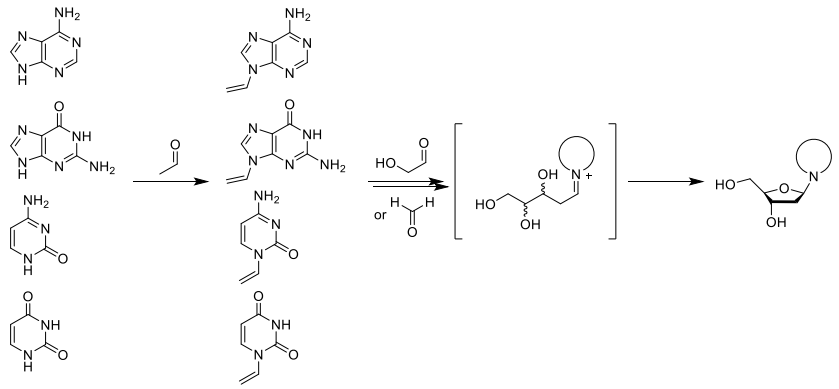
This synthesis was another demonstration that many constituents of living organisms could indeed be produced under a narrow set of conditions, starting from few precursors. It is interesting to note how this approach differed from the work of Carell's group. Here, every effort was made to use compounds that were presumably found in nearly pure form thanks to selective accumulation mechanisms. The reaction conditions were softer, and phosphate was used very frequently to buffer the pH around 7,[59] whereas Carell's scenario relied on pH cycling between values of 4 to 10, that they justified by various geological and weather scanarios.[56] The work of Powner and Sutherland therefore supports the idea that the bricks of life were largely selected by their chemical accessibility, before the advent of biology.[7]

Applying a systems approach to this synthesis of nucleotides also revealed unexpected pathways. Krishnamurthy and Sutherland[62] studied the phosphorylation of 2-thiouridine under three different conditions (Scheme 2.10): with inorganic phosphate either in hot formamide or in molten urea, or using the milder phosphorylating agent diamidophosphate (DAP). In this last condition, 2'-3' cyclic phosphorylation dominated but the product was further converted to the anhydronucleotide **(28)**. Upon photoreduction (see further discussion about prebiotic photoredox in section 2.1.2.4), a 2'-deoxynucleoside structure was obtained. The fact that some pyrimidine loss was observed during the process motivated the authors to study nucleobase exchange. Indeed, it was possible to obtain 2-deoxyadenosine by dry-heating 2-deoxy-2-thiouridine and adenine, albeit in very low yields and without clear α-β selectivity. Furthermore, 2-thiocytidine did not exhibit the same reactivity pattern and was only phosphorylated.



*Scheme 2.10 Top: 2-thiouridine gave the anhydronucleotide **(28)** through a cyclic phosphate intermediate. Bottom: photoreduction delivered 2'-deoxy-2-thiouridine alongside with base loss.*

Another access to 2'-deoxyribonucleosides without preformed ribose was proposed very recently by Trapp and co-workers.[63] They first converted preformed nucleobases (allegedly available through traditional prebiotic chemistry) to enamines upon reaction with acetaldehyde, then allowed the *N*-vinyl nucleobases to react with either formaldehyde or glycolaldehyde to generate a plethora of 2'-deoxyglyconucleosides (Scheme 2.11). Among the mixture, the desired 2'-deoxyribonucleosides were formed in yields from less than 0.5% to about 2.5% as determined by LC/MS. The regioselectivity of the vinylation was intrinsically controlled by the pK_as of the nucleobases and enabled canonical N1- (pyrimidines) and N9- (purines) glycosylation.



Scheme 2.11 Formation of 2'-deoxyribonucleotides directly from glycolaldehyde or formaldehyde, initiated by N-vinyl nucleobases. A large variety of glycosylated nucleobases were formed, of which only the desired ribo product is represented for clarity.

2.1.2.3. Could xenonucleic acids be ancestors of ribonucleic acids ?

Contemporary nucleic acids exploit a set of noncanonical bases to play various structural and functional roles.[64] It is not impossible that such noncanonical bases have arisen before the onset of life[54,55] and the variations offered by the synthetic paths described above illustrate this fact. Another hypothesis is that a completely different base-pairing polymer may have originated first and was only replaced later with canonical RNA.[65] Such structures include sugar backbone modifications such as pyranosyl RNA[66] (pRNA), threonucleic acid[67] (TNA), or the phosphate-free peptide-nucleic acid[68] (PNA, Figure 2.1).

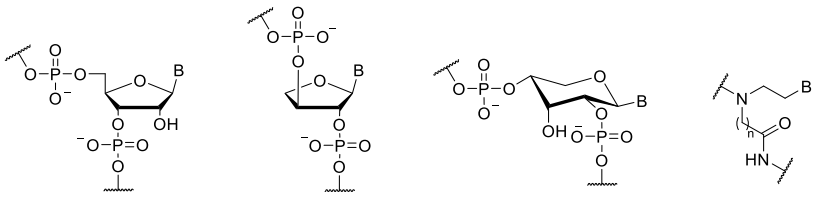


Figure 2.1 Nucleic acids with noncanonical backbone compared to canonical ribonucleotides. The name of the backbone is between parenthesis.

Some alternative nucleobases have also been envisioned, that sometimes solved the issues encountered in the glycosylation step, by displaying a greater reactivity and symmetry (and therefore obliviating the regioselectivity issues). For example, the reaction of 2-pyrimidone with ribose gave the nucleoside zebularine (Figure 2.2, left) under dehydrative conditions, under which the canonical nucleobases could not be efficiently ribosylated.[43] More recently, Hud and co-workers showed that barbituric acid and melamine, both symmetrical, underwent efficient ribosylation with ribose 5-phosphate under either mildly acidic (melamine) or basic (barbituric acid) aqueous conditions. The resulting nucleotides formed strong hydrogen-bonding networks and self-assembled into non-covalent hexameric macrocycles called rosettas. These rosettas could further stack into cylindrical fibers that preferentially incorporated β nucleotides (Figure 2.2, right).[42] In principle, such stacked material may be prone to easier oligomerization because of its preorganization, although this effect remains to be demonstrated (see in section 2.2.1.4 a discussion about the somewhat similar "midwives" approach). Comparable stacked structures are under intensive investigation in the Otto group using dynamic covalent libraries, with the purpose of producing fully synthetic systems that display activities characteristic of living organisms.[69]

Another example of nucleic acid with base pairs fairly different fron the canonical ones is the so-called "Wöhler RNA" that uses polyurea base pairs.[70] Notably, a triuret nucleotide was shown to base-pair with G residues following a hydrogen-bond pattern similar to the U-G wobble pair (Figure 2.3). Benner and co-workers also proposed an 8-letter nucleic acid system with two additional base pairs, that exclusively paired with one another and displayed essential features such as duplex stability.[71]

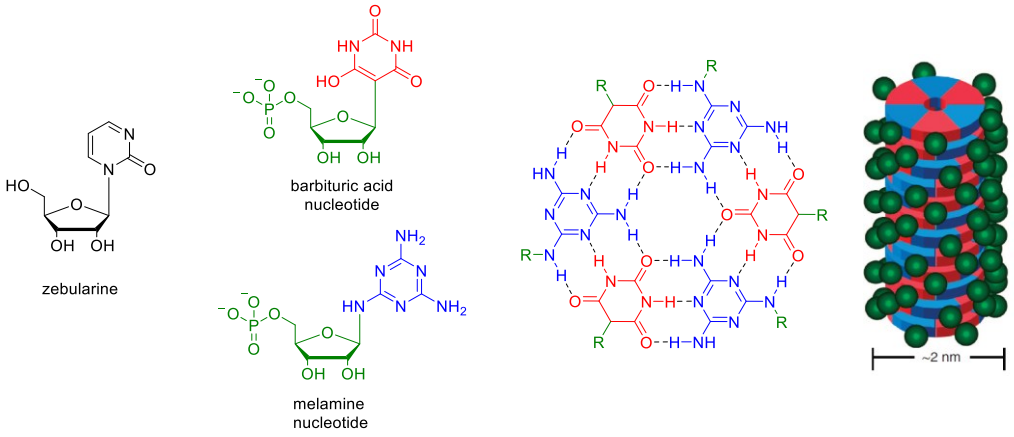


Figure 2.2 Left: the nucleoside zebularine. Center: nucleotides assembled from barbituric acid and melamine give rise (right) to hexameric structures through a dense hydrogen-bond network. Drawing under Creative Commons License reproduced from the original manuscript.[42]

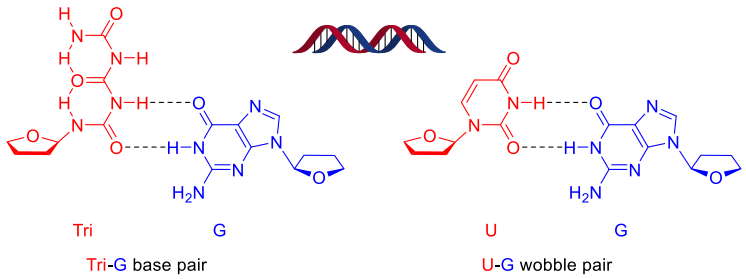


Figure 2.3 Triuret nucleotide pairs with guanosine residues (left) following a pattern similar to the U-G wobble pair (right).

Although these structures are not preserved today in living organisms, they may have been generated alongside with RNA in the prebiotic world, because of the lack of selectivity of certain reactions or the diversity of small molecules available through simple syntheses. It is still unclear how transitions could be accomplished between different genetic supports. Recent results suggest that mixtures of oligomers possessing the same base-pairing system but heterogenous backbones may spontaneously let homogenous sequences take over mixed ones through autocatalysis and self-amplification (see discussions about autocatalysis in the next section 2.2), providing a primitive selection mechanism for the transition to a homogenous nucleic acid system.[65]

2.1.2.4. Protometabolisms: the cyanosulfidic world

Cells build and maintain a dense and intricate network of reactions called metabolism, that is mostly driven by enzymes. The idea that such a network could emerge from chemistry before the onset of the first living organisms has received much attention and suggests that a common chemical origin of the building blocks of life is possible.

Sutherland and co-workers demonstrated that many intermediates in the synthesis of nucleosides and amino acids were available through a limited number of reactions in a cyanosulfidic environment, where hydrogen cyanide served as a

primary C1 building block, and hydrogen sulfide as a reducing agent.[9] The reactions in this protometabolism can be sorted in four categories (Figure 2.4):

a. homologation of a chain through cyanohydrin intermediates that are photochemically reduced by sulfides. This pathway gives access to the oxazole intermediates involved in nucleotide synthesis (section 2.1.2.2) as well as several amino acids.

b. reduction of oxygenated species promoted by sulfides, and coupled to the homologation procedure described in *a*. This gives glycerol, a precursor of lipids, and amino acids with branched saturated side chains.

c. acrylonitrile (from acetylene and hydrogen cyanide) and cyanamide react in a reductive homologation cascade leading to proline and arginine.

d. cyanoacetylene (from acetylene and hydrogen cyanide under oxidative conditions) and hydrogen cyanide give amino acids with carboxyl side chains.

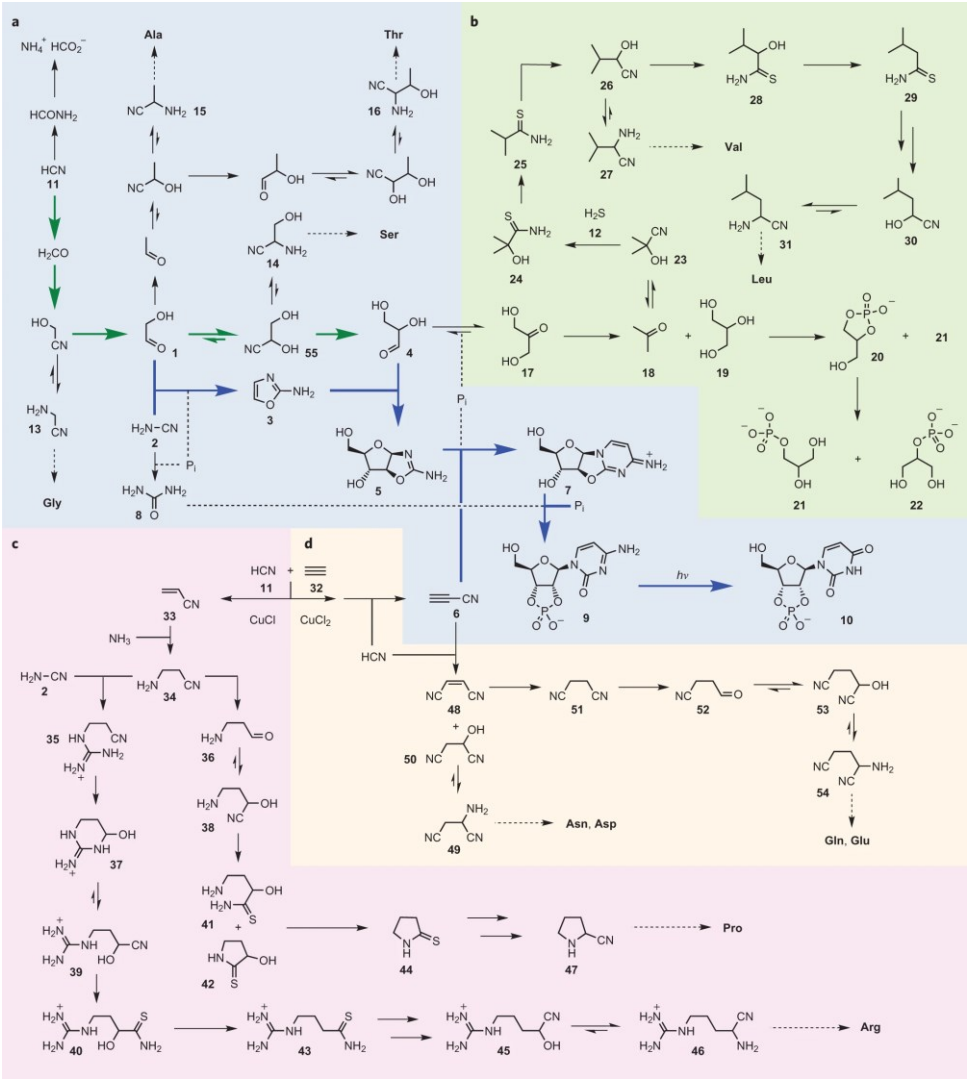
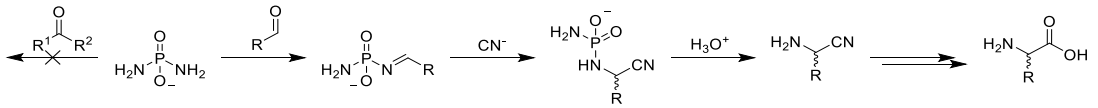


Figure 2.4 A cyanosulfidic protometabolism provides a large range of building blocks under a narrow range of self-compatible conditions. Figure reproduced from the author's manuscript[9] under license with the Editor.

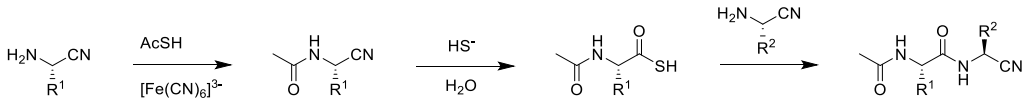
Through this narrow set of conditions (most reactions were reductive and used sulfide as the reductant) and using only three primary building blocks (HCN, acetylene and cyanamide), the authors obtained 11 amino acids and several critical building blocks for the synthesis of pyrimidine nucleotides (section 2.1.2.2) and phospholipids (section 2.1.2.6). Furthermore, many of the reaction conditions and intermediates involved in this protometabolism were also exploited in the synthesis of nucleotides without ribose

One conceptual flaw in this scheme was the Strecker reaction leading to amino acids. As already discussed (section 2.1.2.2), both aldehydes and ketones react with ammonium cyanide to give aminonitriles, providing undesired α,α-disubstituted aminonitriles among with the desired α-monosubstituted ones, and the cyanosulfidic protometabolism yields both aldehydes and ketones. This problem was circumvented by using DAP as a source of nitrogen instead of ammonia (Scheme 2.12). Indeed, an intermediate *N*-phosphorylated imine was formed selectively with aldehydes, leaving ketones unreacted. After cyanide addition, a *N*-phosphoryl amino acid was obtained, that was easily hydrolyzed.[50]



Scheme 2.12 Modified Strecker synthesis of amino acids using DAP as an aldehyde-selective source of nitrogen.

α-Aminonitriles were also exploited directly for peptide formation under cyanosulfidic conditions.[72] In this scenario, a starting aminonitrile, acting as the future *N*-terminus, was first acetylated with thioacetic acid under iron (III) catalysis ($[Fe(CN)_6]^{3-}$). Its nitrile was then hydrolyzed to the corresponding thioacid in aqueous hydrogen sulfide. Another aminonitrile then acted as a nucleophile to form the peptide bond (Scheme 2.13). Remarkably, the reaction proceeded with a retention of configuration when chiral aminonitriles were used and the selectivity against β-aminonitriles and functional groups on the side chains (including lysine) was excellent. This was due to the surprisingly low pK_a of the α-aminonitriles: around 5. At pH 7, the α-amine was therefore deprotonated and highly nucleophilic, to the contrary of side-chain amines and β-aminonitriles.



Scheme 2.13 Peptide bond formation between α-aminonitriles in a cyanosulfidic scenario.

2.1.2.5. Protometabolisms: metal catalysis without enzymes

Two very ancient and highly conserved metabolic pathways are the AcCoA reductive pathway[73] and the reductive tricarboxylic acid cycle (rTCA).[74,75] The

AcCoA pathway reduces acetate groups to form C-C bonds, notably the lipid tail of phospholipids. The rTCA cycle is more complex and involves 11 intermediates that are interconnected by more various reactions, among which carboxylations and reductions. Many of these reactions are promoted by metal-bearing enzymes and only a fraction depends on activation by ATP. In is then reasonable to think that analogues of the rTCA cycle were accessible under abiotic conditions through abundant metal catalysis, and that emerging organisms only integrated a useful, already existing network to their metabolism.[76]

Moran and co-workers demonstrated that native metals, in particular Fe, Ni and Co, reduced CO_2 to acetate in a hot, saline solution under CO_2 pressure. Pyruvate was also obtained to a smaller extent.[77] Methanol and formate, two C1 building block that are not involved in the rTCA cycle were also obtained. These molecules required to be desorbed from the metal surface to be recovered.

Reconstituting an enzyme-free analogue of the rTCA cycle required more elaboration. In a first attempt, Moran and coll. were able to mimic the three steps from oxaloacetate to succinate and three others from oxalosuccinate to citrate under a single set of conditions using a combination of hydrochloric acid, Fe^0 as the reductant and Zn^{2+} and Cr^{3+} as hydration/dehydration catalysts (Figure 2.5).[78] The adjunction of an artificial surfactant[79] generated micelles that enhanced the yield of some steps. Noteworthy, the very unfavorable dehydration of malate into fumarate was driven by the irreversible reduction of the latter, illustrating how Le Chatelier's principle may be exploited to promote difficult reaction when these are integrated into a deeper network. Alanine, an off-cycle product that results from the reductive amination of pyruvate, was also obtained when hydrazine was added as a source of nitrogen.

A further study by the same group showed that an even larger fraction of the cycle, embedding 7 steps over 11 and 9 intermediates over 11, was possible under metal catalysis, particularly iron (II) and without requiring acid, using pyruvate and glyoxylate as starting materials (Figure 2.5).[80] Four amino acids were obtained out of the cycle by reductive amination with hydroxylamine. As the rTCA is autocatalytic,[76] it may have emerged and been sustained by reagent feeding before the advent of living organisms. This work points toward a possible emergence of protometabolisms, although some concerns were raised about the plausibility of this hypothesis, both before and after the experimental contribution of Moran.[81,82]

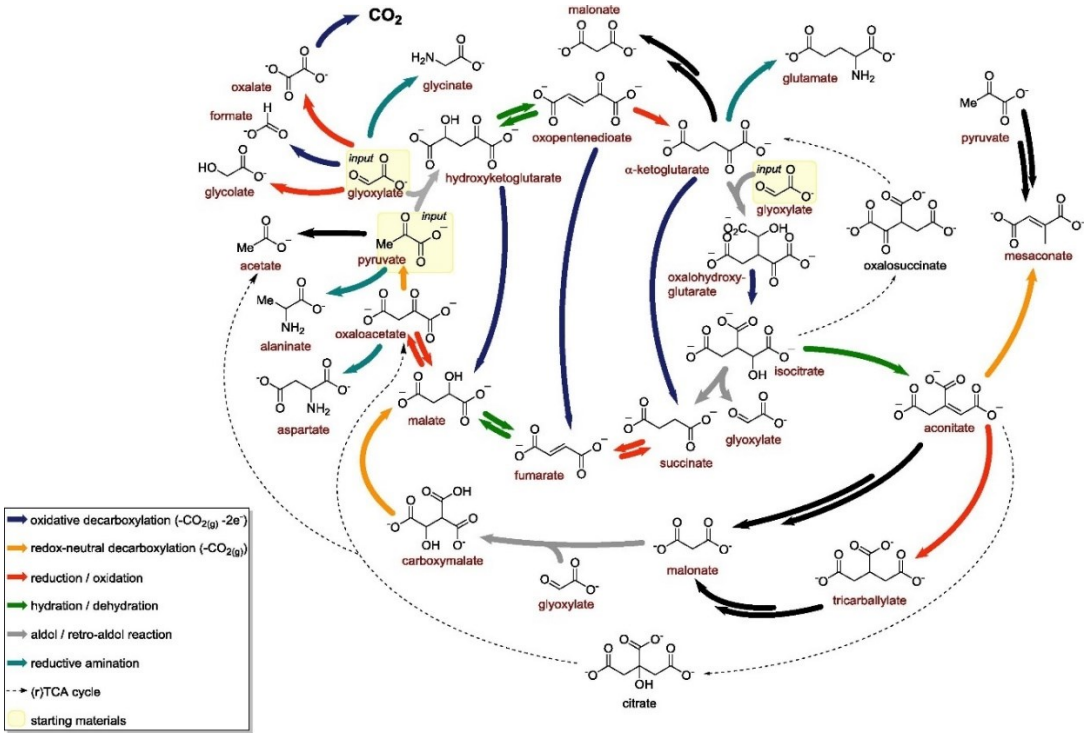


Figure 2.5 Prebiotic reconstitution of the rTCA cycle by Moran and co-workers. The thick colored arrows denote the reactions that were found to operate in their study while the thin dashed lines show reactions of the natural cycle that could not be reproduced. Figure reproduced from the manuscript[76] under license with the Editor.

2.1.2.6. The case of lipids and phosphorylation reactions

The phospholipids that constitute the majority of cell bilayers are rather complex molecules that are produced enzymatically in living organisms, and their prebiotic synthesis was probably not straightforward. It is also likely that intermediate lipids played a role in the formation of membranes (see section 2.3.1.1 for further discussion). On a chemical level, one may break down a phospholipid to three main parts: two hydrocarbon tails, a glycerol backbone and a phosphorylated head with a varying degree of elaboration (example of POPC in Figure 2.6).

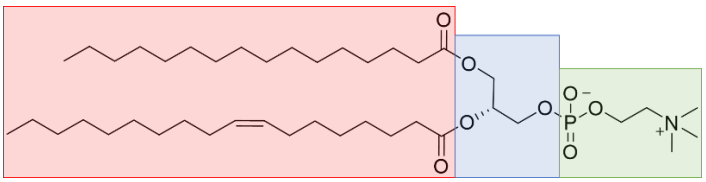
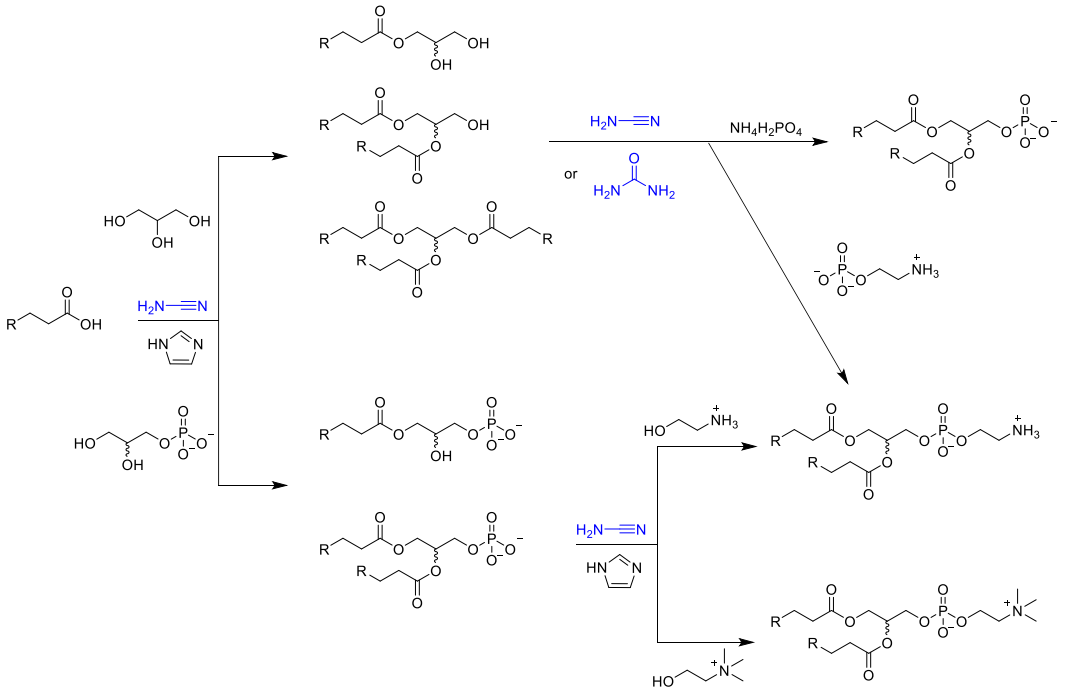


Figure 2.6 General structure of a phospholipid, here through the example of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC). Red: lipid tails; blue: glycerol backbone; green: phosphorylated head.

In living organisms, long saturated chains are built through either the AcCoA or the prenoid pathways.[83] Before the emergence of such sophisticated anabolisms, fatty materials may have been produced on a young, hot planet Earth under hydrothermal conditions,[84–87] or delivered from carbonaceous meteorites.[88–91] Glycerol is a potential product of the formose reaction and the cyanosulfidic

protometabolisms described above. It is interesting to note that all the reactions required to build a phospholipid from its elemental bricks are condensations, i.e. dehydrative reactions that need a water removal mechanism to be efficient. Purely thermal conditions were used,[92,93] but most pathways relied on chemical dehydrative agents, commonly cyanamide and urea. Cyanamide may have originated from geological sources and slowly hydrolyzes to urea. Urea itself can absorb a molecule of water to degrade into CO_2 and $2\times NH_3$. Both of these reactions were exploited to promote condensations.[21,94,95] Imidazole was often included in the coupling mixture as a promoter. Cyanamide, urea and imidazole have been used to promote the acylation of glycerol[96] and glycerophosphate[97] (obtained itself from glycerol and inorganic phosphate, Pi, under cyanamide conditions[98]) with fatty acids, yielding either mono-, di- and triglycerides or phosphatidic acids. Mono- and diglycerides could be further phosphorylated with either Pi or *O*-phosphoethanolamine using cyanamide or urea as a promoter.[99] Alternatively, phosphatidic acids were elaborated further to phosphoethanolamines[100] and phosphocholines,[101] again using cyanamide and imidazole (Scheme 2.14).

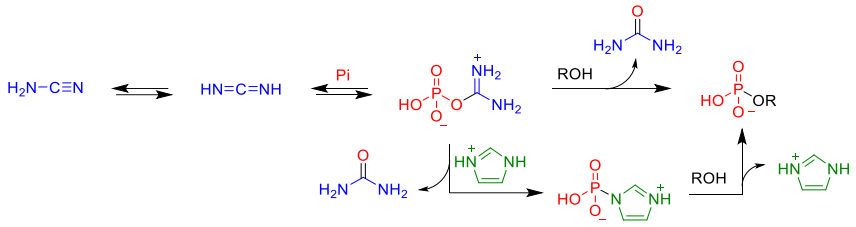


Scheme 2.14 Prebiotic lipid synthesis through condensations promoted by urea or cyanamide.

Other routes have been described to simpler amphiphiles. Monoalkyl phosphates were formed under various conditions,[102,103] and monoacylglycerols were phosphorylated with DAP to yield the corresponding cyclic phosphates.[104,105]

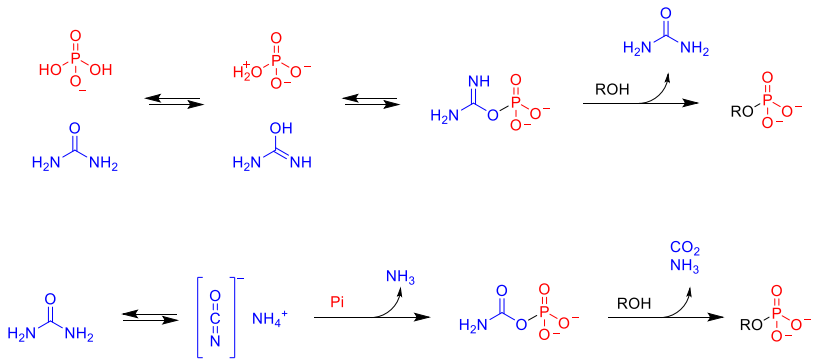
In a broader view, phosphorylation is a prerequisite of life. The synthesis of both nucleotides and phospholipids require *O*-phosphorylation, while some cofactors and energetic molecules such as phosphocreatine are also *N*-phosphorylated. A problematic point, known as the "phosphate problem", is that few forms of phosphate are soluble in water, and much of the natural phosphorus is trapped in minerals.[106,107] Many phosphorylation pathways were proposed[108] but two are of particular interest in a systems approach because they overlap with other reactions.

One is the cyanamide- or urea-promoted condensation. Many examples have been reported in this manuscript already, both for the synthesis of nucleotides and lipids. Cyanamide can tautomerize to the minor carbodiimide form that is electrophilic, and in the presence of Pi, be trapped as a phosphoryl isourea that is thought to be the active intermediate.[109] More elaborate mechanisms involving dicyanamide (the rapidly forming dimer of cyanamide) were also proposed.[110] The adjunction of imidazole, acting both as a base and a nucleophilic catalyst, was reported to enhance the reaction (Scheme 2.15).[96] This reaction occurs at relatively elevated temperature (usually 60-85°C) that may be necessary because of the involvement of a rare tautomer and overall modest kinetics. It is interesting to note that cyanamide was also widely employed to promote other condensation reactions, notably oligomerizations, very early in the history of prebiotic chemistry.[95] It is also a building block and a starting material in several of the syntheses described above.



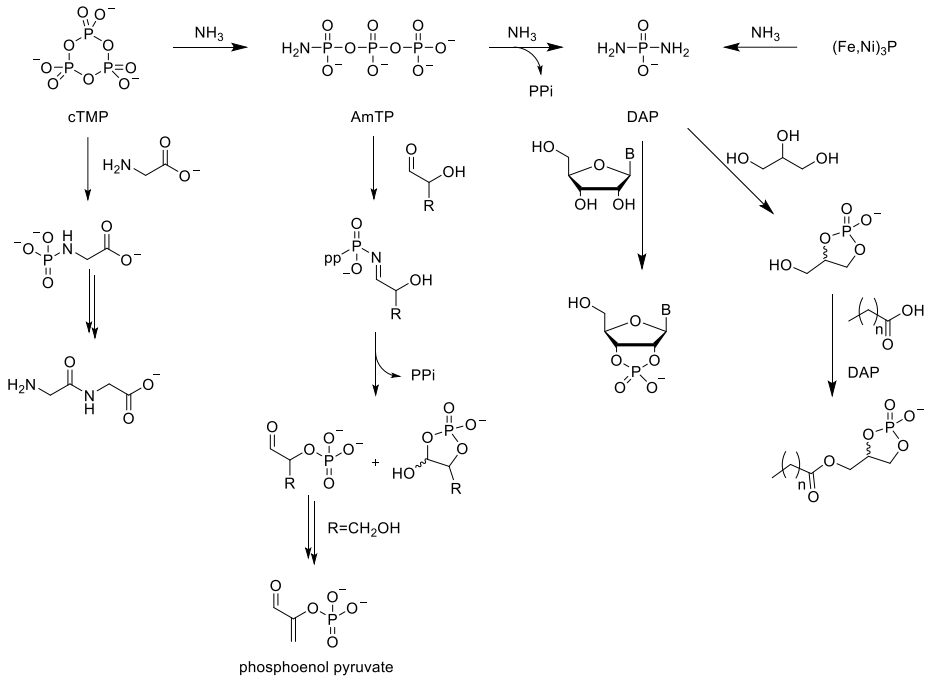
Scheme 2.15 Cyanamide-promoted phosphorylation of alcohols, with or without imidazole.

Urea is the product of hydrolysis of cyanamide and an important compound in organic and industrial chemistry. Upon hydration at elevated temperature, it releases CO_2 and ammonia, justifying its use as a dehydrative agent. As a prebiotic condensing agent, it is often used in the neat, molten or semi-molten form (commonly at temperatures above 100°C, while the melting point of urea is 133°C) and preferably phosphorylates 1,2-*cis*-diols to cyclic phosphates (see nucleotide syntheses above). The mechanism of urea-promoted phosphorylation is debated but it has been proposed to go through an isourea similarly to the cyanamide route.[111,112] However, this would involve a very unlikely phosphate tautomer. An alternative, yet undemonstrated route, involves the initial formation of ammonium cyanate, which results in a net consumption of urea instead of a catalytic action (Scheme 2.16). A similar activation of phosphate by isonitriles has been recently proposed.[113]



Scheme 2.16 Urea-promoted phosphorylation of alcohols. Top: through a rare phosphate tautomer with catalytic urea. Bottom: through an isocyanate with urea consumption.

The second source of activated phosphate that received much attention is polyphosphates (or metaphosphates). Most often, cyclotrimetaphosphate (cTMP, Scheme 2.17) was used as a model compound for these. cTMP is known to phosphorylate amines under basic conditions,[114] and even promoted the formation of peptide bonds between non-hindered amino acids to a limited degree.[115,116] The formation of peptide bonds went through a five-membered cyclic anhydride dubbed cyclic acylphosphoramidate (CAPA) that is a quite versatile electrophile.[114] Most often, the ammonolysis products of cTMP were used. When cTMP reacts with ammonia, it first generates amidotriphosphate (AmTP) in solution, then, at high ammonia concentrations and elevated temperatures, DAP (Scheme 2.17).[117] Corrosion of reduced phosphorus minerals such as schreibersite $(Fe,Ni)_3P$ also provided a source of mono- and diamidophosphate.[105] Eschenmoser and coll. demonstrated that aldoses were α-phosphorylated by AmTP through an imine intermediate,[118,119] which probably inspired Powner's aminonitrile synthesis with DAP (see section 2.1.2.4). DAP was also exploited in a potentially prebiotic synthesis of the high-energy phosphoenol pyruvate. cTMP also yielded amphiphilic cyclophosphoramidates[120] and DAP was used as a nucleoside phosphorylating agent[104] and to produce cycloglycerolipids (Scheme 2.17).[121]



Scheme 2.17 Prebiotic phosphorylations with polyphosphates and amidophosphates.

2.1.3. Overview

Early investigations of the abiotic synthesis of the key molecules of life has tackled multiple difficulties. The most intuitive routes to nucleosides and amino acids proved problematic because of a lack of selectivity toward canonical products and some inefficient key reactions such as nucleobase ribosylation. The disparate nature of the conditions leading to each product classes also raised the question

of the likeliness of a strict separation of synthetic routes. In particular, nitrogen-rich chemistry, mostly based on cyanides, was largely incompatible with the oxygen-rich chemistry of formaldehyde and carbohydrates.

Contemporary approaches tend to provide unified pathways to canonical and non-canonical but natural nucleotides, amino acids and lipids. The issues of compatibility as well as the selection of natural products among a soup of small molecules have been addressed by a systems approach in which reactions are interconnected in vast networks and share a limited number of starting materials and conditions. Current experimental investigations suggest that such protometabolic pathways may lead to the almost exclusive formation of biologically relevant matter in yields sufficient to overcome by-products and tar formation.

The conditions investigated as potentially plausible on a young Earth were also progressively revised. In the past, experiments were often performed in dilute aqueous solutions, maybe representative of seas, with deleterious effects on kinetics, and on the yields of dehydration reactions. Contemporary approaches favor concentrated solutions,[59] evaporative environments in which compounds are brought to dryness,[53] minimally hydrated pastes[104] and eutectics of water and salts (especially investigated in the context of oligomerization).[122]

2.2. Oligomerization reactions and self-assembly

Gathering a collection of molecular bricks in solution does not make a living organism quite yet. Although we still lack a universal definition of life, it appears that cells make up a large portion of autonomous living organisms and they are the traditional target of the studies of the emergence of life. The functions of cells are ensured by a complex chemical and spatial organization. Oligomers are critical to such functionalities because of their ability to fold, interact, store information and replicate it, which is probably not possible with small molecules, although fields such as dynamic covalent chemistry demonstrates how complex behaviors can emerge from all sorts of small molecules without irreversible bond formation.[3] In particular, the formation of peptides and oligonucleotides has attracted extensive research. The dehydrative reactions involved require a driving force as they are unfavorable in water (ca. 3 kcal/mol for peptide bonds[123] and 5 kcal/mol for DNA phosphodiesters[124]). Many conditions leading to small or medium-length oligomers were proposed, but not all offered the features required to obtain evolvable species, either because they focused on homosequences with little functional perspectives, or because the products were unable to break down and form again. Indeed, one-way syntheses do not give the flexibility required for evolution.[125] In the absence of the protective environments of cells, all oligomers are doomed to hydrolysis and must be built faster than they are destroyed to persist.[123] Despite the importance of these concerns, many oligomerization studies published in the last century, and some more recent ones, focused on "who got the longest" oligomer,[126] most commonly and readily a homosequence arising from the activation of a single residue.[127]

As for the synthesis of nucleotides and amino acids above, the chemist is faced with the practical constraints of time, analysis and experimental amenability and may therefore choose to diverge from geologically plausible scenarios. In particular, simple but slow or inefficient coupling agents may be replaced by more efficient ones, provided that they bring about the same reactivity patterns.[27] In this work, we consider that unraveling the basic principles of reaction networks and inscribing these in the perspective of more complex systems is more important than the quest for specifically prebiotic reagents, a topic that may be a source of never-ending bickering.

A systems approach also points toward the emergence of peptide-RNA conjugates. If amino acids and nucleotides were available together in the same environment, it seems almost impossible that they did not react together, and the evidences showing that mixtures of amino acids and nucleotides indeed produce mixed species upon chemical activation are accumulating.[27,29] Another hint in this direction is that cellular life is entirely dependent on peptide-nucleotide interactions, for processes as central as translation. Although it is not strictly impossible to envision living or life-like systems that use a completely different set of bricks and processes, the unity of the living organisms discovered so far

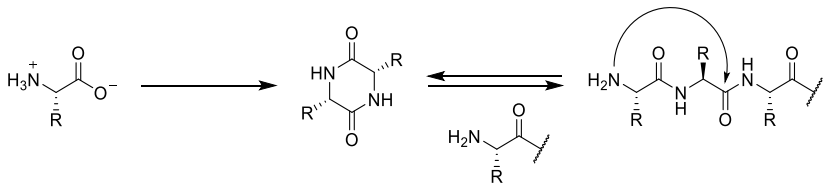
encourages the chemist to use this same (or a very similar) set of molecules and reactions to investigate the emergence of life.

2.2.1. Chemical activation and oligomerization reactions

2.2.1.1. Peptide elongation

Peptides and amino acids are kinetically relatively stable[128] and can withstand strong conditions to some extent. In neutral solution, amino acids are zwitterionic, which limits their spontaneous condensation. This is an important issue when considering amino acid oligomerization: at least one of their reactive ends (amine or carboxylate) is ionized at every pH. For this reason, esters[129] and even acyl chlorides[130] of amino acids can be prepared under acidic conditions and stored without observing amide formation. Another difficulty comes from the wide variety of their side chains, that are susceptible to compete with the backbone (amines in lysine, carboxylate in aspartate and glutamate…) or to degrade (oxidation of sulfur-containing side chains for example). Finally, peptides bearing active moieties on the *C*-terminus are inherently prone to racemization. While this is a longstanding problem in peptide synthesis,[131] it may be an interesting, perhaps necessary, feature in systems chemistry because it opens the way to chiral resolution through symmetry breaking and amplification.[125]

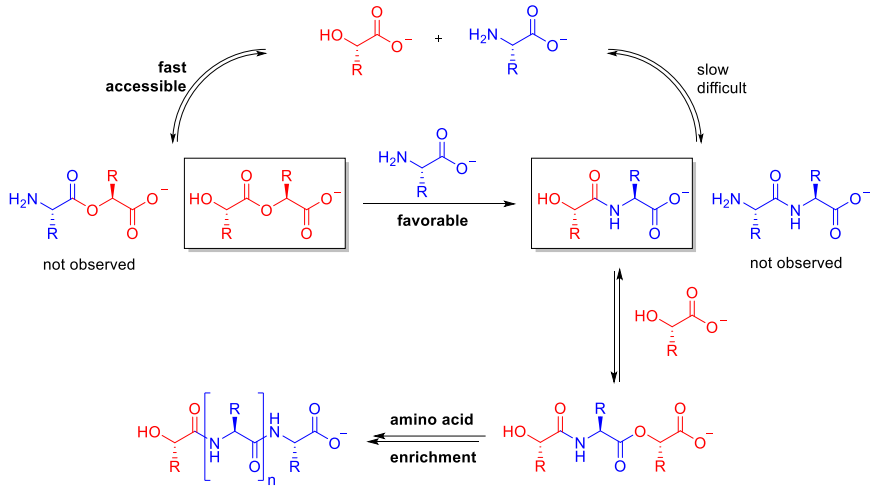
Short peptides can be produced by a variety of very simple physicochemical procedures. They were obtained by dry-heating mixtures of amino acids,[132] through successive hydrations and dehydrations of aqueous solutions over clay minerals,[133] in the presence of highly concentrated salts that acted as water scavengers (NaCl) and catalysts (Cu^{2+}),[134,135] under hydrothermal conditions at elevated temperatures and pressures[136] or using metal sulfides and carbon monoxide as promoters at high temperature.[137] In this last case, the authors noted that dipeptides were formed and cleaved under the same conditions, suggesting that the reaction would never go over the (unfavorable) thermodynamic equilibrium unless some fluctuations were introduced in the system. All of these studies gave relatively limited yields of small peptides, and some experienced the concurrency of the formation of diketopiperazines (DKP) (Scheme 2.18), a classical byproduct in peptide chemistry, that is also a key intermediate in peptide hydrolysis.[138,139] Interestingly, DKP can also be used as a building block for peptide elongation and may not strictly be considered a dead end.[140]



Scheme 2.18 Diketopiperazines (DKP) are side-products of peptide formation (left) and represent the main pathway for peptide degradation (right). They may also be opened to allow peptide elongation.

Thermal pathways are not the most attractive for systems chemistry: they require harsh conditions and do not offer much evolutive perspective, notably because of their reversible nature: the composition will equilibrate thermodynamically, preventing the formation of long sequences. Hot/cold and dry/wet cycles may be more promising,[141] as are interface-driven oligomerizations[142] (see section 2.3.2 for oligomerizations in lipid membranes).

An elegant way to overcome the difficulty of the direct formation of the amide function is to use esters as intermediate species. Methyl esters of amino acids were used as mildly activated substrates that react very slowly in the absence of catalytic additives[143] and thioesters may also be generated under prebiotic conditions.[144] Thioesters are the activated substrates used in the native chemical ligation (NCL), a very powerful tool for the selective ligation of peptide fragments.[145] A more integrative approach to the ester-promoted peptide growth is to form the esters in situ. Depsipeptides are peptide-like oligomers with a hybrid amide/ester backbone. Esters are kinetically more accessible than amides and are displaced by amines. Subjecting mixtures of α-amino acids and α-amino alcohols to dry/wet cycles produced depsipeptides that were progressively enriched in amide backbone.[146,147] Most species had a hydroxy acid at the *O*-terminus,[a] indicating that the first condensation yielded a α-hydroxy lactone rather than an α-amino lactone (Scheme 2.19).



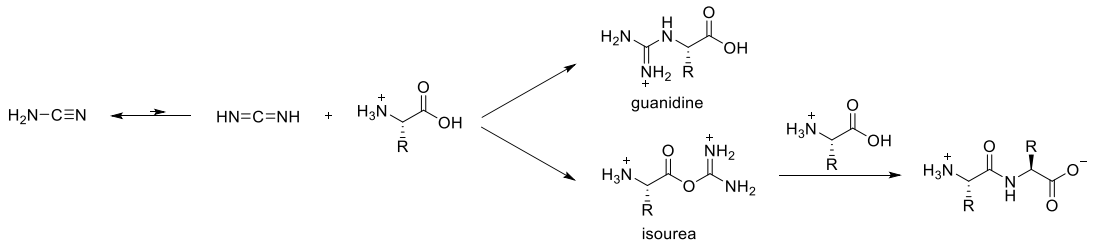
Scheme 2.19 Peptide elongation in dry-wet cycles through depsipeptide intermediates. Amide bond formation (top right) was difficult. Ester bonds were formed first (top left) then trans-amidated. Most sequences started with a hydroxy residue, indicating that the first condensation to occur was likely between two hydroxy acids.

Most oligomerization strategies capable of evolution are based on chemical activation. The core idea is that the systems relying on single inputs of energy, such as heating, are doomed to their thermodynamic equilibration, which permits very little selection and amplification mechanisms[3] and considerably limits the formation of long oligomers because of their equilibration to shorter ones.[148] At

[a] in a standard peptide this would be the *N*-terminus.

the opposite, constantly feeding a system with active species keeps it far from the thermodynamic equilibrium and enables kinetically controlled processes that are more relevant to systems chemistry.[13,149]

Early strategies for peptide activation relied on cyanamide under acidic conditions.[150,151] Modest yields of small peptides (up to 4 residues) were obtained. The reaction is believed to proceed through the carbodiimide tautomer of cyanamide that forms an isourea with the carboxylate of the amino acid. A major side-reaction is amino acid guanidylation, which hampers the synthesis considerably (Scheme 2.20). This is regulated by the protonation of the amino acids, justifying the use of acidic conditions. Interestingly, Hawker and Oró performed the reaction in the presence of nucleotides but without discussing the possible formation of phosphoramidates (see section 2.2.2).[150]

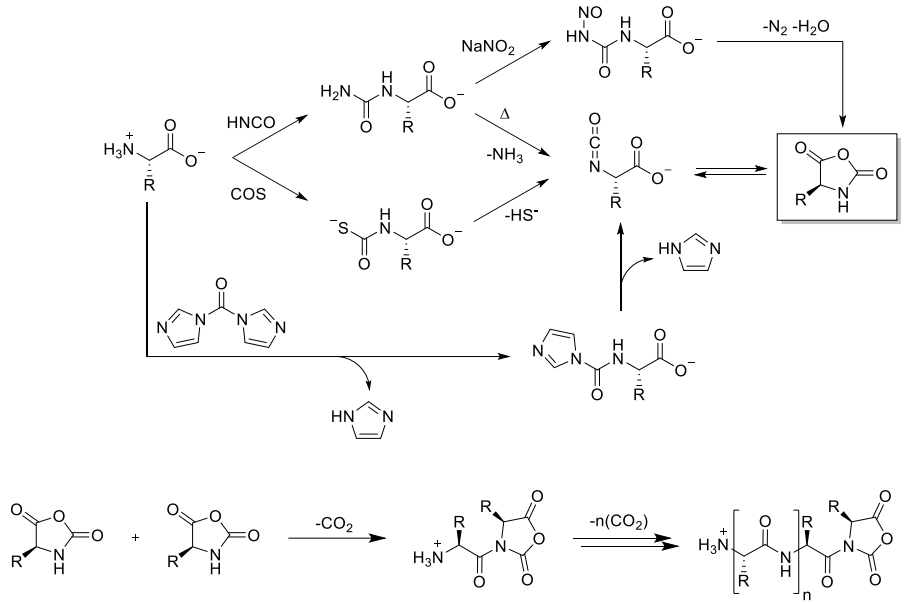


Scheme 2.20 Cyanamide-promoted peptide formation and the undesired formation of guanidines.

Most activated amino acids and peptide *C*-termini are readily converted into a small selection of metastable cyclic intermediates under a wide range of conditions.[125] This suggests that the precise nature of the chemical activating agents on the early Earth is probably of little importance. The most prominent of these intermediates is the *N*-carboxyanhydride (NCA), sometimes called Leuchs' anhydride (Scheme 2.21).[152] NCAs were formed from carbamoylated amino acids through nitrosylation[153] or thermal elimination,[154–156] from their thiocarbamates under reducing conditions[157] and were commonly generated in the laboratory using electrophilic reagents such as carbonyl diimidazole (CDI, Scheme 2.21).[158] NCAs then reacted with various nucleophiles with a loss of CO_2. They also oligomerized to give homo- or heteropeptides, when initiated by small amounts of either free amino acid[159] or a Lewis acid.[158] (Scheme 2.21, bottom). The α carbon does not epimerize in single amino acid NCAs.

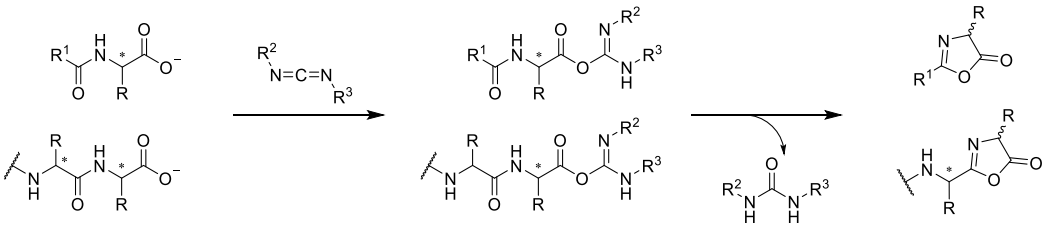
A common difficulty associated with amino acid activation is the wide variety of side chain functionalities, that may cross-react or be involved in the formation of more specific active intermediates. In the laboratory, the side functionalities are protected by late-stage protective groups.[160] This is unlikely to have occurred widely in a prebiotic context, although some examples may be postulated. For example, *O*-phosphoserine[161] may be considered a prebiotically plausible protected serine, and glutamic acid cyclizes upon activation to give pyroglutamate.[158] When amino acids are trapped as different active intermediates and oligomerize at very different rates, co-oligomerizations are problematic

because of the difficulty to incorporate the least reactive residues (unfortunately frequently the catalytically capable ones).[158]



Scheme 2.21 Top: formation of N-carboxyanhydrides of amino acids through multiple prebiotic and synthetic pathways. Bottom: oligomerization of NCAs to peptides.

When *N*-acylated amino acids (including the *C*-terminal residue of a peptide chain) are activated with carbodiimide reagents, intermediate 5(4*H*)-oxazolones are formed (Scheme 2.22)[162] and may lead to short (4-5 residues) peptides. Oxazolone intermediates are usually avoided in peptide synthesis because they are prone to racemization.[131] Although enantio-enriched amino acids have been detected in comets and meteorites,[163] amino acids were mostly expected to be available in their racemic form in the context of prebiotic chemistry. The possibility to epimerize a position was crucial to establish a feedback in potential chiral selection mechanisms (see below).[125] The degree of epimerization has also been used to estimate the extent of conversion of amino acids into oxazolones, and the efficiency of activation, with various carbodiimides and related agents, with the conclusion that most of these reagents did not deliver high yields of oxazolones in dilute prebiotic scenarios.[164]

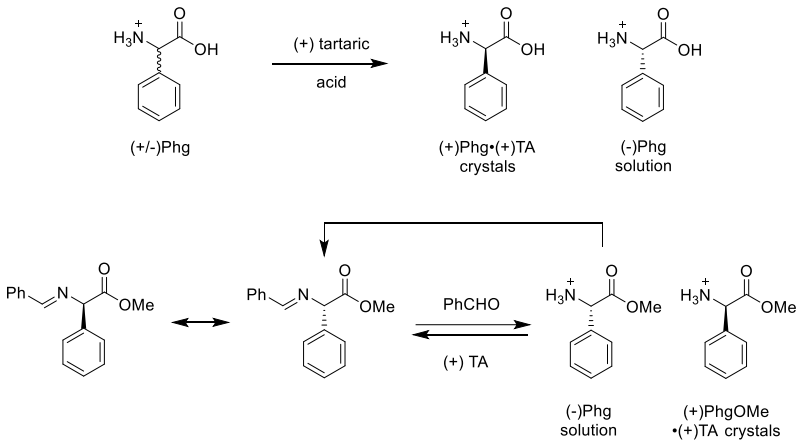


Scheme 2.22 Formation of 5(4H)oxazolones of N-acylated amino acids (top, R^1 = Me or Ph) or peptides (bottom) when activated with carbodiimides, accompanied by α-racemization of the residue.

2.2.1.2. The homochirality of amino acids

The homochirality of the molecules of life is a very vast topic that would deserve a full discussion on its own. Separative techniques have identified enantio-enriched molecules, in particular amino acids, in space, and the mechanisms leading to their formation are being discussed. This matter is out of our focus and has been reviewed elsewhere.[163] In this section, we briefly discuss some terrestrial mechanisms that could have led to an enrichment in amino acids of one configuration and are relevant to systems chemistry.

Enantioselective co-crystallization of amino acids, typically as tartrate salts, are long known but without the possibility of epimerization, a separation of the enantiomers occurs at best (Scheme 2.23, top), which results in a maximum theoretical recovery of 50%.[165] Esters or amides of amino acids are better suited, and reversible protection of the *N*-terminus as an imine further facilitates the epimerization and enrichment of one enantiomer at the expense of the other (Scheme 2.23, bottom). Numerous examples are known and have been reviewed elsewhere,[166] often exemplified with phenylglycine as a model substrate.

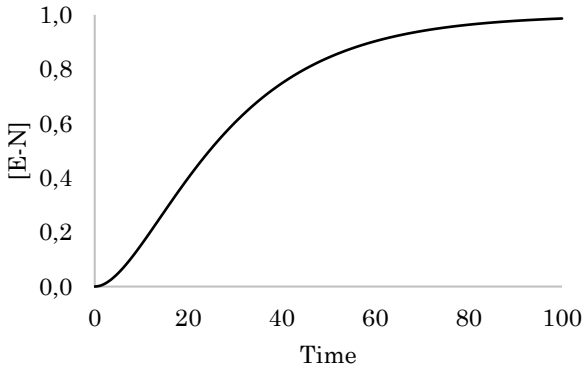


Scheme 2.23 Chiral resolution of phenylglycine. Top: co-precipitation with enantiopure tartaric acid. A maximum of 50% recovery is possible because the (-) isomer cannot be epimerized. Bottom: co-precipitation coupled with solution-phase racemization facilitated by the double derivatization (ester and imine). The desired enantiomer can theoretically be recovered quantitatively.

Scalemic mixtures with a very small enantiomeric excess ($ee = \frac{|R-S|}{R+S}$) can also be amplified to practically enantiopure crystals. This was demonstrated on phenylglycinamide[167] and phenylalanine methyl ester,[168] both reversibly protected as imines. The mechanism relied on the enantiomorphic formation of crystals coupled with a racemization occurring in a solvent in which the compounds were only sparingly soluble. Small crystals dissolved more readily than large ones, leading to an Ostwald ripening phenomenon that consumed one enantiomer at the profit of the other one.[169] The authors obtained a practically enantiopure phenylalanine derivative within days, starting with an *ee* of only 0.35%.[168]

2.2.1.3. Peptide self-assembly and autocatalysis

Although the roles of replication (possibly self-replication) and informational persistence are usually filled by nucleic acids, some peptides also have the ability to self-assemble and display autocatalytic behaviors. The self-recognition processes in folded peptides leads to the possibility of templated synthesis of a whole peptide from smaller fragments of its sequence. The first example was reported in 1996 in the Ghadiri group who identified a self-complementary α-helicoidal 32-mer that was autocatalytically reproduced from two of its fragments (an electrophile **E** and a nucleophile **N**) through a NCL at a judiciously chosen position.[170,171] As a characteristic of an autocatalytic process, the reaction started with a lag phase during which the concentration of the template **E-N** slowly built up then continued with a rapid catalyzed growth (by stacking of **E** and **N** over the preformed **E-N**) terminated with a slower phase in which aggregation dominated, leaving less free **E-N** behind to perform catalysis (Graph 2.1, Figure 2.7). Seeding the medium with **E-N** at the beginning of the reaction enhanced the initial rate.



Graph 2.1 Kinetic simulation (the author's own work based on an arbitrary kinetic modeling of the process described by Ghadiri and coll.[171]) of a sigmoidal pattern for the autocatalytic production of E-N using arbitrary values for equilibrium constants and initial concentrations (arbitrary time and concentration units). The curve displays a short lag phase at the beginning, corresponding to the buildup of the template, followed by a short exponential growth, terminated by a slow convergence due to the deactivation of most of the template through aggregation.

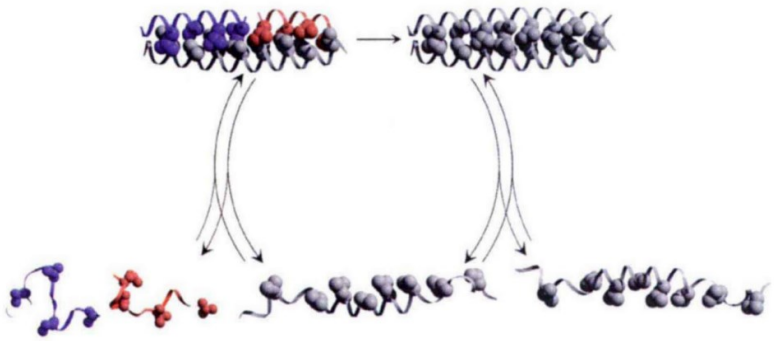
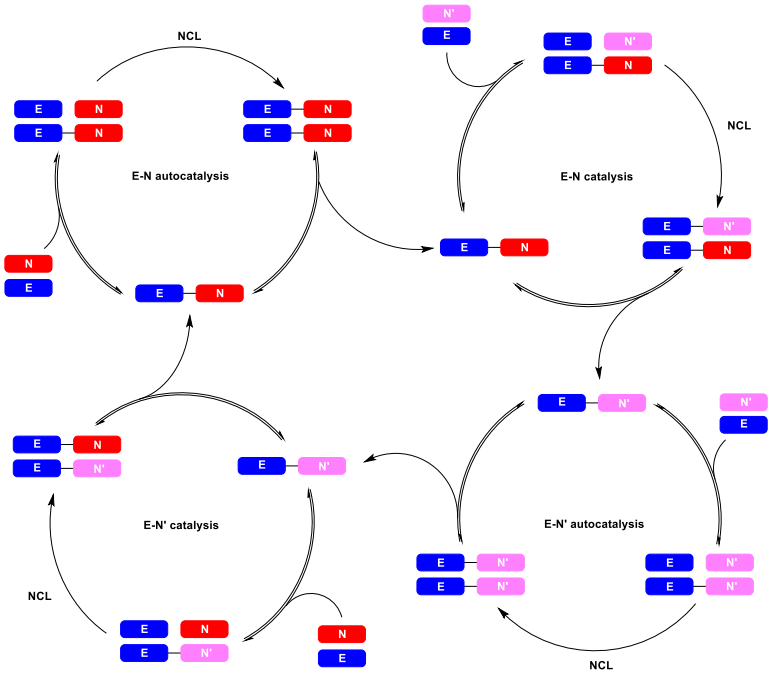


*Figure 2.7 Autocatalytic reproduction of an α-helicoidal peptide that displayed 1:1 coil-coil interaction. The fragments **E** (blue) and **N** (red) bound to the whole peptide then were coupled with ease to leave a double stranded complex that split to produce more template. Figure reproduced from the author's manuscript[170] under license with the editor.*

This system was robust against mutations: the introduction of mutant electrophilic ($\mathbf{E}_m$) or nucleophilic ($\mathbf{N}_m$) fragments produced a small amount of mutant peptides $\mathbf{E}_m$**-N** and **E-**$\mathbf{N}_m$. These peptides were able to catalyze the formation of the native peptide **E-N**, but not their own formation because of their inability to fold properly.[172] This was an example of error correction directed by aggregation properties, inherent to the subtle sequence-structure relationship encountered in α helix formation: changing one amino acid at a critical position may suppress entirely the aggregation phenomenon. Other peptides were selectively ligated using a similar system.[173]

Using two different nucleophiles **N** and **N'** such that **E-N** and **E-N'** were both autocatalytic in isolation but not complementary to each other gave rise to a chemical symbiosis mechanism (Scheme 2.24).[174] More specifically, **E-N** and **E-N'** did not form mixed aggregates but **N** and **N'** were still similar enough to perform an heterogenous catalytic cycle as described above with mutant residues. A four-cycle network then established spontaneously: **E-N** and **E-N'** both grew their own individual autocatalytic cycles (top left and bottom right), but the excess of **E-N** produced was exploited as a template for the formation of **E-N'** (top right) and reciprocally (bottom left). Practically speaking, this was achieved by mutating the hydrophobic regions of the nucleophiles: Val-Leu were switched to Leu-Ile fragments.



*Scheme 2.24 Autocatalytic production of **E-N** and **E-N'** were interconnected by the cross-catalytic production of **E-N** catalyzed by **E-N'** and reciprocally, in a proto-symbiotic network.*

One limitation of α-helix-based templating is that the recognition proceeds between two units that later need to be taken apart to act as a template, similarly to DNA or RNA replication. The need for strand separation inherently prevents the system from establishing an exponential replication rate. In contrast, amyloid

β peptides can stack into fibers, virtually without any limit, and do not require this separation step. Amyloid β-sheets are widely explored self-assembled peptide aggregates, with implications in numerous diseases, notably Alzheimer's disease.[175] They can be formed from very short peptides, down to the dipeptide Phe-Phe, which is compatible with the length of products generated in model prebiotic reactions.[176] The salt-induced formation of dipeptides was found to be more efficient with apolar amino acids,[148] and when mixtures of polar and non-polar amino acids were oligomerized by the CDI method, an insoluble peptide phase considerably enriched in lipophilic amino acids was obtained, that displayed amyloid-like X-ray patterns.[177] Alternating lipophilic and polar amino acids was another way to β-sheets, with the example of the $(Glu\text{-}Phe)_n$ sequence.[178] The implications for such peptide aggregates are numerous. Some of these are hollow, and allow objects to be molded and react in a space-constrained manner (Figure 2.8).[176] Dipeptides made of Phe and Phg residues were found to bind to RNA and prevented its hydrolysis (the "prebiotic sanctuary" in the word of the authors), opening a way to climb the thermodynamic hill of RNA elongation.[148]

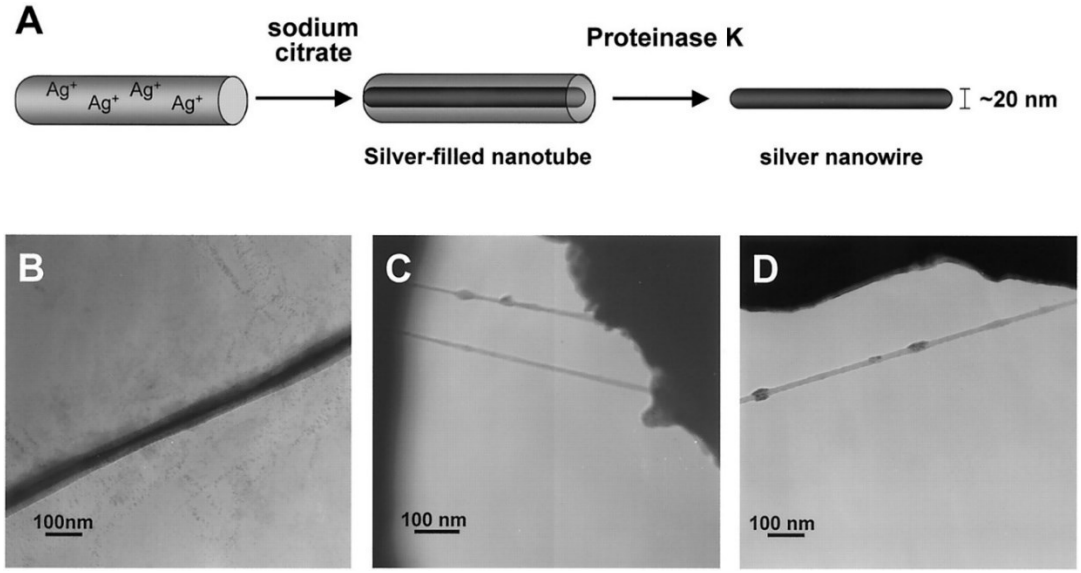
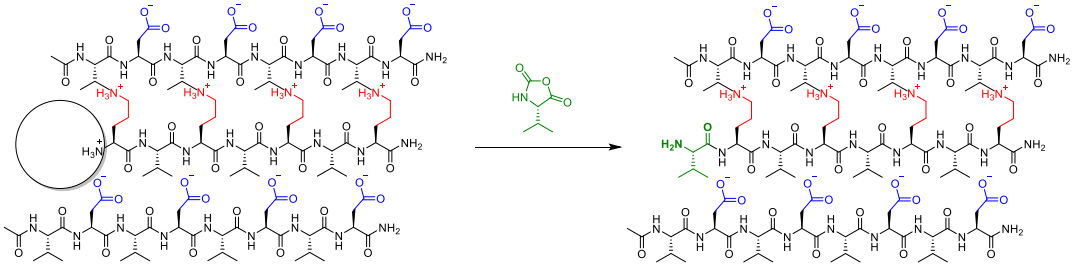


Figure 2.8 Nanotubes formed by the self-assembly of the Phe-Phe dipeptide hosted silver ions that were reduced to molded silver nanotubes. A: overview; B: electron micrographs of the silver-filled nanotubes; C and D: silver nanotubes after enzymatic digestion of the peptide shell. Figure reproduced from the manuscript[176] under license with the publisher.

Mihara and co-workers demonstrated that an α-helical peptide showing similarities to the **E-N** peptide discussed above could undergo a transition to β-sheets induced by hydrophobic defects.[179] This peptide could be generated autocatalytically from two fragments through NCL.[180] The reaction proceeded differently from the case of α-helices. Because β-sheet peptides form stacked aggregates instead of pair, there was no lag phase at the beginning of the reaction, and seeding the medium with the template generated no spectacular increase in rate, as the effective concentration of the aggregated template was relatively independent on the concentration of the peptide. Similar results were obtained

later in the group of Ashkenasy using a more simple peptide based on a repetitive $(Phe\text{-}Glu)_n$ motif that formed antiparallel β-sheets.[178]

Greenwald and coll. recently demonstrated that amyloid fibers could act as a template for peptide elongation through the addition of activated monomers rather than conjugation of specific peptide fragments. In their study, a complete template $(Phe\text{-}Glu)_4$ and an incomplete complementary sequence $Arg\text{-}(Phe\text{-}Arg)_3$ were both soluble in water when taken separately but assembled into fibers when mixed. The missing Phe residue at the *N*-terminus could be incorporated using Phe-NCA, with enhanced yields and selectivity when the template was present. The correct sequence completion with two missing fragments was also possible. The authors further extended the model to the more prebiotically relevant peptides $Val(Asp\text{-}Val)_4$ (template) and $(Orn\text{-}Val)_4$ (incomplete sequence). In this case, the completion properties depicted before were conserved but the extension also proceeded selectively on the backbone amine rather than on the terminal amines of the ornithines (Scheme 2.25).



Scheme 2.25 Selective incorporation of Val-NCA into β-sheet aggregated complementary peptides.

The formation of β-sheets may also have implications in the emergence of homochirality. Indeed, Lahav and coworkers demonstrated, using enantiomer deuteration, that isotactic (that is, homochiral) poly(Val) and poly(Leu) peptides formed to a much greater extent from a 1:1 mixture of D and L NCAs than what was expected from a statistical distribution.[181]

Prebiotically plausible oligomerizations mostly offer short peptides and, apart from Powner's aminonitrile coupling (described in section 2.1.2.4), suffer from the competition exerted by side chain functionalities. However, unlike nucleic acids, that need to be reasonably long to be able to carry significant amounts of information, very short peptides may play a defined role. Ser-His, a dyad involved in hydrolases, is known to be catalytically competent, although with limited efficiency.[182] Lipophilic peptides as small as Phe_2 can form defined structures which result in specific chemical and physical properties. It is assumed that aliphatic amino acids dominated on the early Earth, which implies that most prebiotic peptides were lipophilic and potentially insoluble. While this can be seen as a problem at first glance, it may have become a strength in a world where lipids were available. Lipophilic peptides without catalytic abilities may have played a crucial role as early anchors of nucleic acids to membranes.[20]

2.2.1.4. The polymerization of nucleotides

Nucleic acids are more delicate than peptides. When treated under excessively drastic conditions, they undergo various degradations such as hydrolysis, dehydration and depurination. Natural oligonucleotides possess a 3'-5' phosphodiester backbone and obtaining such regioselectivity from the 2-3' *cis*-diol pair represents an additional challenge, although some level of backbone heterogeneity may have been tolerated, if not beneficial, to the emergence of RNA biochemistry.[183] Forming phosphodiesters is also more difficult than forming amides, because of the higher thermodynamic hill in water and because alcohols are relatively poor nucleophiles. The need for sequestration mechanisms to allow long chain growth, as already discussed in the case of peptides, is still applicable to nucleic acids, that are even more sensitive to hydrolysis. This has led to the development of numerous surface-enhanced syntheses. Phosphates are readily trapped as stable active intermediates, and RNA growth studies may typically be split between those involving in situ activation and the oligomerization of pre-activated monomers. The latter naturally raise the question of the availability of such active species. The fact that the phosphorylation of nucleosides often yields 2',3'-cyclic phosphates[59,104] also questions the availability of 5'-phosphates, that are notwithstanding frequently used as starting materials for oligomerizations.

The attempts of thermally forming RNA phosphodiesters found little success, with only dimers and trimers detected,[184,185] leading very early to the conclusion that condensing agents were likely necessary to induce significant RNA growth. Carbodiimides and other condensing agents were able to promote phosphodiester formation,[186] but failed at delivering significant RNA chain lengths. Cyclic phosphates may be a source of self-activation. It was early demonstrated that 2'-3' cyclic phosphates could oligomerize in the dry state in the presence of catalytic amines, although a mixture of 2'-5' and 3'-5' backbones were obtained, with limited yields and chain length.[187,188] 3'-5' cyclic nucleotides gave long oligomers but in very small yields when an amine was added,[189] or in better yields upon dry-heating.[190,a] Highly concentrated salts in evaporative environments is another route to RNA that was recently suggested.[192] Similarly, Monnard and Deamer used an eutectic sodium chloride – water system at -18 °C to promote the condensation of pre-activated nucleotides (see below).[122] The hydrolysis was limited by the low temperature and the fact that water was mostly frozen, while monomers and salts were kept concentrated in the small cavities created between chunks of frozen water, leading to local concentrations much higher than the bulk concentrations introduced. Furthermore, their system was nearly equally effective for purine and pyrimidine nucleotides,[193] a feature that was difficult to achieve in water (purines were preferentially incorporated).[127] The benefits of

---

[a] These results[189] were partially contested, as they were found to be poorly reproducible by others,[190,191] a fact that the authors justified by the source of the starting material. However, the lack of reproducibility, added to the extremely low amounts of products, raise the question of both the purity of the starting material and the analytical methods employed, see further discussion in section 2.2.3.

eutectic environments were then further elaborated by the Holliger group in an RNA world perspective.[194,195]

Many model prebiotic reactions start with preactivated 5'-nucleotides (Figure 2.9). The most commonly employed leaving group is imidazole[196] or variants thereof,[197,198] but 1- and 3-methyladenine and other leaving groups[199,200] have also been used. The prebiotic pertinence of imidazole-activated nucleotides (ImpN) has long been criticized but the current lines of evidence are rather supportive. Lohrmann early suggested that ImpA could be produced by the thermal condensation of imidazole and adenosine polyphosphates,[201] the latter arising from the polyphosphorylation of adenosine with polyphosphates.[202] More recently, specific pathways to ImpN have been proposed through activation by isonitriles and simple aldehydes[113,a] or cyanogen chloride (generated from cyanide and bleach).[203] Furthermore, the carbodiimide activation of nucleotides in the presence of imidazole and other heterocycles intermediately yielded active nucleotides that accumulated to a significant extent.[32,204] All these points justify the use of imidazolides or other activated nucleotides as models in oligomerization studies.

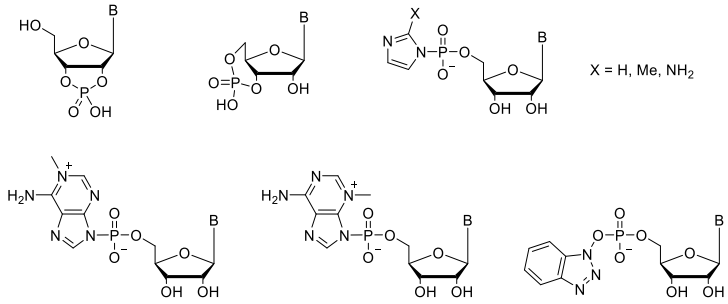


Figure 2.9 Selection of activated nucleotides used in model prebiotic oligomerization studies.

The difficulty to grow consequent RNA chains in solution or in the dry state prompted the use of supporting surfaces and templates. The self-pairing abilities of RNA naturally pointed toward a complementary oligomer as the template. This was experimentally demonstrated in 1968 by the synthesis of small oligomers of AMP using polyU as a template under EDC activation,[205] and then later for ImpG over a polyC template, where it was found that divalent cations were necessary and had a critical influence on the 2'/3' selectivity: a mixture of Mg^{2+} and Zn^{2+} gave 3'-5' phosphodiesters while Mg^{2+} and Pb^{2+} gave 2'-5'.[206] Substituting the imidazole leaving group for 2-methylimidazole afforded an efficient 3'-5' ligation in the sole presence of Mg^{2+}.[207] 2-methylimidazolides therefore became a "gold standard" in RNA elongation.

[a] At the time of publication of this manuscript, the Sutherland group had made two manuscripts available as preprints in ChemrXiv, that expand the use of methyl isocyanide and various imidazoles as an activation system. They were able to form peptide bonds, aminoacyl nucleotides and other related species.

Nucleotide analogues bearing an amine instead of a hydroxide on the 2' or 3' position were sometimes used to simulate RNA growth with better kinetics and efficiency.[111,208–210] In this case a phosphoramidate was formed in place of the phosphodiester. This approach was extensively used for templated RNA replication (see below).[209–215] The enhancement of the non-enzymatic formation of RNA in the presence of preformed strands suggested that the primary emergence of small sequences may then facilitate the offspring of RNA but did not solve the "chicken and egg" problem of synthesizing such first sequences.

The group of Ferris studied in depth the formation of oligonucleotides at the surface of minerals, in particular montmorillonite clays, using nucleotides activated as imidazolides or adenylates.[216,217] Montmorillonite is an aluminosilicate clay that presents a layered structure able to trap nucleotides between interlayers while leaving the reactive fragments (phosphate and 2'/3' hydroxides) available for reactions to occur.[218] Not all minerals capable of trapping nucleotides possess this crucial feature.[219] Montmorillonite-promoted elongation partially addressed the matters of regioselectivity[196,200,220] and stereoselectivity (incorporation of D- and L- ribonucleotides and the related emergence of homochiral oligonucleotides)[221] in RNA growth. Joshi and Ferris observed that monovalent and divalent cations were critical in the binding of the nucleotides at the surface of the clay.[222] The clay-promoted syntheses demonstrated that the simple immobilization of growing oligomers on a surface permitted the emergence of long sequences in addition to short ones (Figure 2.10). Further, strands obtained over clays may then act as templates for new sequences, without the intervention of the clay. This was demonstrated by the synthesis of polyG over an heterogenous polyC template prepared by clay-promoted synthesis.[223]

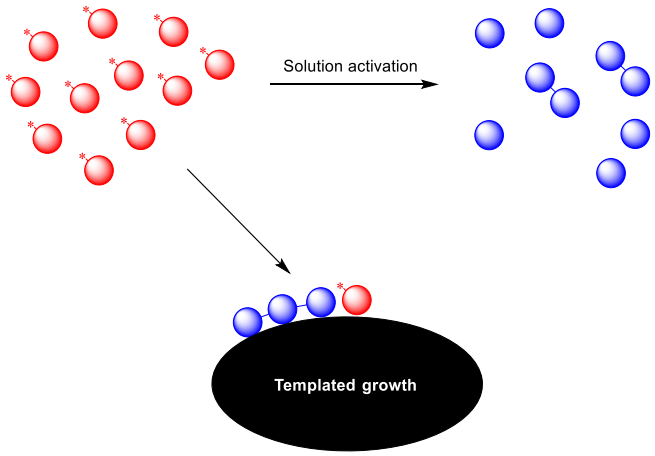


Figure 2.10 Solution-phase reaction of activated nucleotides (red spheres with asterisks) yields few short oligomers whereas hydrolysis (blue spheres) dominates. Templated growth allows the elongating oligomer to fight dilution and thermodynamic equilibration.

The growth of a linear RNA chain may be blocked easily by strand cyclization, in a way similar to the formation of DKP. A potential solution to this problem was found by Hud and coworkers using what they called "molecular midwives". These

are intercalating molecules that inserted between stacks of base pairs and prevented most cyclization by straightening the RNA strands during the coupling reactions.[224,225] This templating approach is innovative and attractive because it does not require preformed sequences.

2.2.1.5. The templated replication of oligonucleotides

Copying DNA and RNA sequences with high fidelity is a long-standing objective of molecular biology since the discovery of the structure of DNA. While the development of polymerase chain-reaction (PCR) technologies makes such copying readily available in any laboratory, non-enzymatic pathways to RNA replication remain elusive. The general idea of such templated replication is that a single strand of nucleic acid should receive activated nucleotides in the appropriate Watson-Crick positions to reconstitute the complementary sequence, possibly with the help of a primer sequence (Figure 2.11). Alternatively, non-activated monomers may be delivered in the presence of a suitable coupling agent. Since non-templated RNA growth is slow and inefficient, the formation of sequences independently from the template is not expected to be problematic.

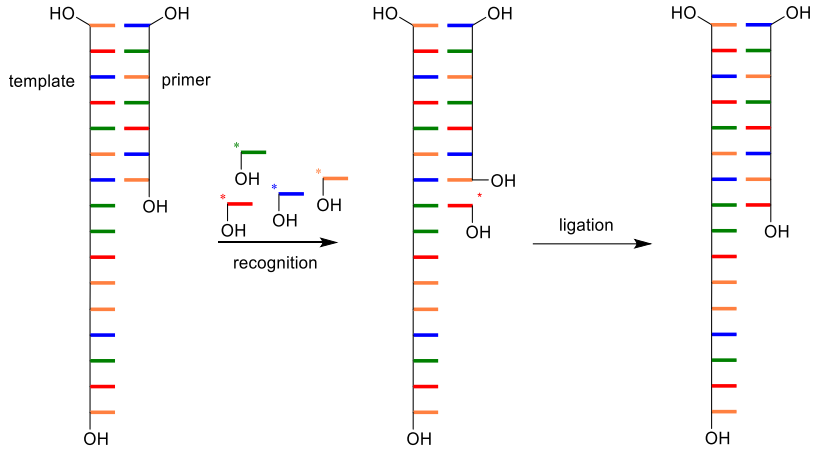


Figure 2.11 General principle of the templated elongation of RNA from a primer using pre-activated monomers.

An early demonstration of a nucleobase-specific template effect was the synthesis of T_{12} from $2\times T_6$ in the presence of dA_{12} using carbodiimide coupling agents.[226] Von Kiedrowski then applied this fragment coupling strategy to generate a self-complementary hexadeoxynucleotide in an autocatalytic manner from two protected trimers.[227] This approach has also been extensively developed outside of the prebiotic context for the synthetic modification of oligonucleotides.[228] The early attempts to elongate a primer with 2-methylimidazole-activated nucleotides have established a number of principles and have been extensively reviewed ([229,230] and references therein). First, the presence of Mg^{2+} is of critical importance, and high-salt buffers were typically used. Second, the replication proceeds with a satisfying 3'-5' selectivity. This is understandable from the conformation of the

ribose moiety in an A-type double helix: the 3'-*endo*-OH in an A-double-helix is favorably pointed toward the phosphate (Figure 2.12).

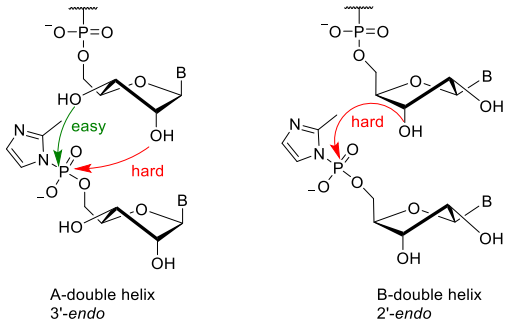


Figure 2.12 The pre-transition states in A-type RNA helices justifies the preferential formation of 3'-5' phosphodiesters. In comparison, putative B-type helices would not be copied efficiently at all.

Recently, leaving groups other than 2-methylimidazole were proposed. Richert and co-workers used OAt phosphoesters both with natural[199] and 3'-aminodeoxy RNA.[231] HOAt is not a prebiotic reagent, but the improved reactivity of OAt phosphoesters compared to 2-methylimidazolides is an interesting feature for the convenience of laboratory exploration. The Szostak group demonstrated that 2-aminoimidazole, a potentially prebiotic compound,[232] was a very efficient leaving group in primer elongation.[198]

The presence of a "helper" sequence, a complementary sequence placed downstream the elongation site, also facilitates copying.[233] This may be due not only to the expectedly better defined binding site, but also to the potential formation of highly reactive bridged imidazolides between the nucleotide being copied and the one downstream (Figure 2.13).[234–236]

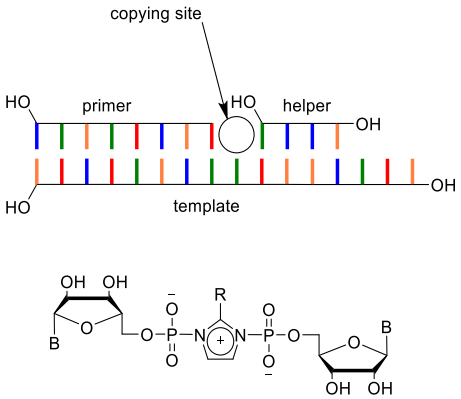


Figure 2.13 Helper-assisted templated replication. Top: general principle. Bottom: very active intermediates formed between the nucleotide in the copying site and the nucleotide downstream.

Mispairing is a common problem in templated replication. Poorly selective recognition of the base to be coupled leads to incorrect incorporation and the emergence of mutant sequences. Fidelity studies showed that the C/G pair is much more robust than the A/U, but the excessive representation of G residues generated aggregates (G-quadruplexes).[231,237] Richert and coll. further argued that in a primitive setting, copying from a two-base nucleic acid system would

likely be more faithful than from a four-base system.[231] Stronger binding may also be obtained when di- or trinucleotides are used as reagents instead of single nucleotides. In this case, the primer extension is in competition with cyclization, especially for trimers, but under some conditions the extension may still be quite efficient.[238]

The dependence of the copying efficiency to the activation form goes beyond basic kinetic preoccupations. Indeed, inactive monomers formed after hydrolysis of the active species tend to occupy the copying site, that is therefore blocked.[239,240] One way around this problem is to (re)generate the active species in situ, by the help of a coupling agent. Richert and coll. employed 1-ethylimidazole to produce the highly active 1-ethylimidazolium species in situ, fueled by EDC,[241] whereas forming and regenerating OAt phosphates in situ with EDC and HOAt was inefficient.[199] An alternative solution to the problem of deactivated monomers is to immobilize the template/primer strand and to provide it regularly with freshly activated material while washing out the hydrolyzed nucleotides. This may be achieved artificially using magnetostatic interactions[239] or in a more prebiotic simulation thanks to encapsulation in vesicles.[242] The detailed implications of activation sources and copying efficiency have been reviewed in greater detail elsewhere.[243]

It is unclear whether RNA could have been preceded by a proto-nucleic acid with a different backbone (see discussion in section 2.1.2.3). The transition between proto-RNA and RNA would then require a reliable copying as well as the tolerance for chimeric (mixed backbone) sequences.[230] Orgel and co-workers demonstrated that PNA, one of the most extensively studied potential RNA ancestors, could indeed serve as a template for RNA synthesis[244] and form information-carrying chimeric PNA-RNA sequences.[245] TNA may also possess this ability.[65,246] These lines of evidence suggest that transitions between different informative polymers possessing the same code (in other terms, the same set of bases) is conceptually possible, but the driving force for the rise of RNA over other polymers remains to be determined.

2.2.1.6. Overview

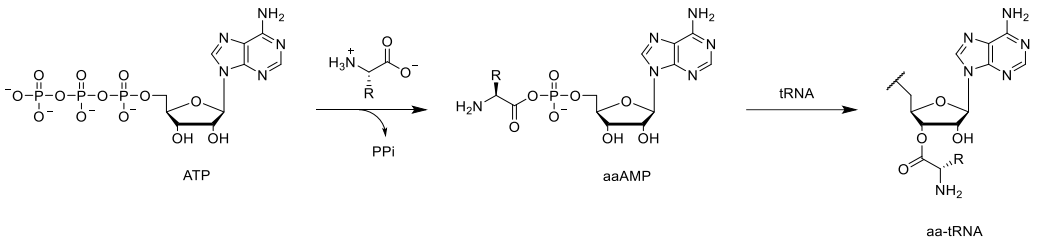
The replication of information-carrying polymers without evolved enzymatic tools is probably a requisite to any form of life. The formation of RNA without any kind of template is very inefficient under most conditions. Therefore, it is more likely that very primitive nucleic acid sequences may have arisen with the help of mineral surfaces or other confined environments. Templated replication mechanisms may then have provided an information flow. For now, practical challenges such as efficiency and regioselectivity of phosphodiester formation are still preventing the elaboration of functional systems without any modern construct (enzymes and ribosomes).

Peptides alone are unlikely to have carried genetic information but still display some level of dynamic and persistent behavior, such as self-reproduction. Peptides

as small as dipeptides may have been active catalysts,[182] and the simplicity of amino acid and peptide formation relative to nucleic acids prefigures that peptides may have played a very early metabolic role.

2.2.2. The emergence of co-oligomers

One of the central mechanisms of cellular life is the synthesis of proteins in the ribosome: the translation.[247] This operation requires a complex assembly of a ribozyme and a large collection of enzymes. Peptide chains grow by a *C*-terminus activation in the form of esters of nucleotides: aminoacyl tRNA (aa-tRNA), synthesized from a phosphoric anhydride intermediate (aaAMP) obtained from ATP (Scheme 2.26). aa-tRNA are unusually high-energy esters because of the vicinity of the ester with 2'-OH.[124,248] The synthesis of nonisomerizable 3'-fluoro-3'-deoxy analogues reinforces this hypothesis.[249] In the absence of the protective environment created by the active site of the ribosome, aa-tRNAs rapidly isomerize and hydrolyze,[250] and peptide bond formation from aminoacyl nucleotides in water is extremely slow compared to the rate in the ribosome,[251] a fact that must be kept in mind when proposing prebiotic scenarios for peptide synthesis.



Scheme 2.26 Aminoacyl-tRNA synthesis from ATP

It is difficult to envision a functional life form without a translation mechanism, and thus without a joint peptide-RNA chemistry.[252] This stands for many reasons. The first one is that we are naturally inclined to study forms of life that resemble the one that we know today. We know for a fact that a form of life relying on a peptide-RNA translation is possible, and it is actually the only form currently known. Furthermore, there is now strong evidence of the prebiotic formation of amino acids and small peptides, probably more easily than the formation of nucleotides and nucleic acids. As we will show in this section, the joint chemical activation of amino acids and nucleotides generates a covalent peptide-RNA chemistry. In this view, one may argue that a joint peptide-RNA world "could not *not* have happened"[a] very early in the prebiotic history, no matter how attractive other scenarios are. Finally, experiments aimed at developing a single-component world (usually an RNA world) always failed because of the difficulty of

[a] We borrowed the formula from a recent publication from Benner and colleagues, who employed it with the same meaning although on the topic of prebiotic nucleotide synthesis.[253]

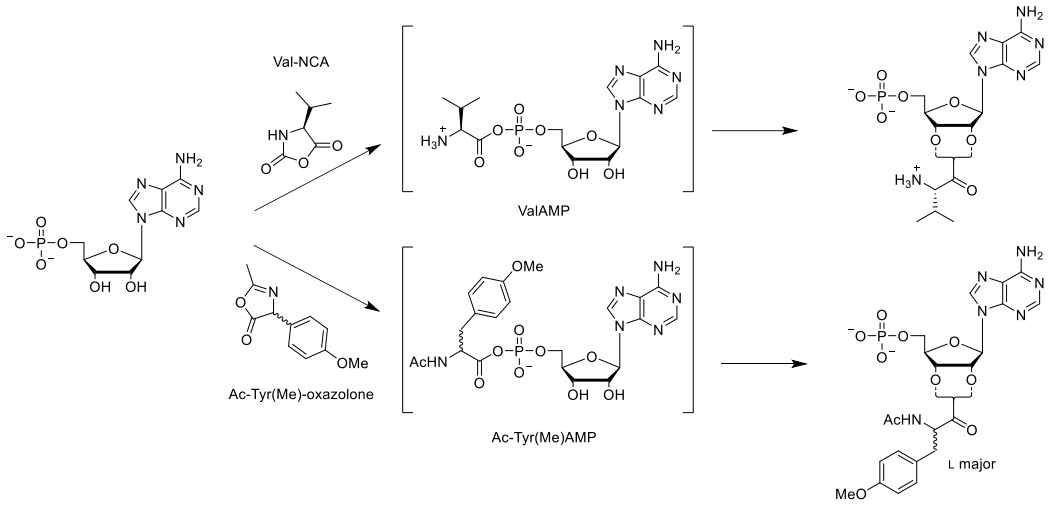
exponential self-replication without the participation of exogenous actors. The self-replication of palindromic RNA is inherently inhibited to rates below the exponential because the newly formed strand must be detached from their templates and because of self-hybridization.[213] Unless another molecule, typically an enzyme, can assist the strand separation, the accumulation of double-stranded sequences will quickly "kill" the self-reproduction process (see the discussion associated to Graph 2.1 in section 2.2.1.3). Although very preliminary to this point, observations such as Carny and Gazit's "prebiotic sanctuary"[148] described before let us envision how a joint peptide-RNA chemistry may overcome the inherent difficulties associated with a single-component world.

By surveying the affinity of short RNA sequences for amino acid-functionalized chromatography columns, Yarus proposed that the genetic code may have emerged from noncovalent recognition patterns between amino acids and their corresponding codons.[254] However, experimental lines of evidence for this principle are still limited and dive relatively deep into molecular biology rather than in prebiotically accessible chemistry.[255]

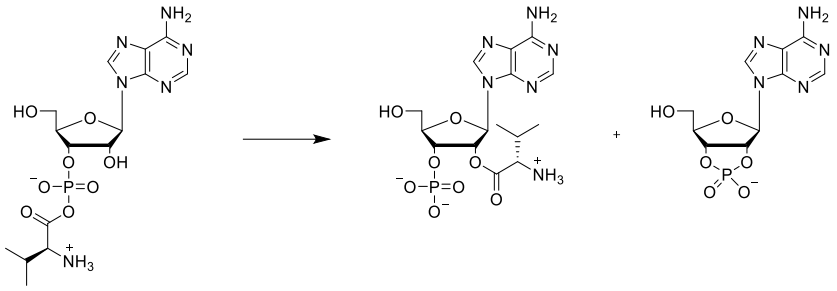
On the other hand, the molecular processes that could have yielded active peptide-RNA species are progressively unraveled. Pascal and coworkers demonstrated that NCAs[22] and oxazolones[256] could acylate AMP. In both cases, an intermediate mixed phosphoric anhydride (aaAMP with NCAs, its *N*-acetylated equivalent Ac-aaAMP with oxazolones) was formed and rapidly migrated to the 2' and 3' positions between which it equilibrated (Scheme 2.27). In the case of oxazolones, where racemization was possible, a preferential incorporation of the L enantiomer was noticed. The reaction did not proceed with 2'-deoxyAMP, which emphasizes the critical role of the 2'-3' pair in aminoacylation processes, and ultimately in translation. It must be noted that inorganic phosphate also reacted with NCAs to give a phosphoric anhydride,[257] instead of performing *N*-phosphorylation. The anhydride was then able to act as a phosphorylating agent for alcohols (demonstrated experimentally only in the case of methanol).

A similar pattern was observed with adenosine 3'-monophosphate, with the difference that the rearrangement of the anhydride was in concurrence with the formation of the 2'-3' cyclic phosphate (Scheme 2.28). A possible explanation for the efficiency of such processes is that the hydrolysis of aaAMP or Ac-aaAMP did not directly yield the inactive amino acids but rather the starting active form (NCA or oxazolone).[258] This was particularly visible in the case of NCAs, as aaAMPs were found to be quite stable in thoroughly degassed water but rapidly decayed to NCAs when CO_2 was added either as a gas or as bicarbonate ions. This explains the limited stability of aaAMPs in aqueous solution.

Oxazolones could acylate adenosine 5'-methylphosphate,[259] which demonstrated that the intermediate phosphoric anhydride may be bypassed. This result is interesting because it suggests that amino acids may be loaded to the 3' terminus of tRNA through oxazolone intermediates, where the RNA chain prevents the formation of 5' mixed phosphoric anhydrides. Again, 2'-deoxyadenosine 5'-phosphate did not react under these conditions, nor did NCAs.



Scheme 2.27 Aminoacylation of adenosine 5'-monophosphate with Leu-NCA (top) or Ac-Tyr(Me)-oxazolone (bottom) through mixed phosphoric anhydrides aaAMP and Ac-aaAMP. In both cases, a constantly equilibrating 2'-3' mixture was obtained.

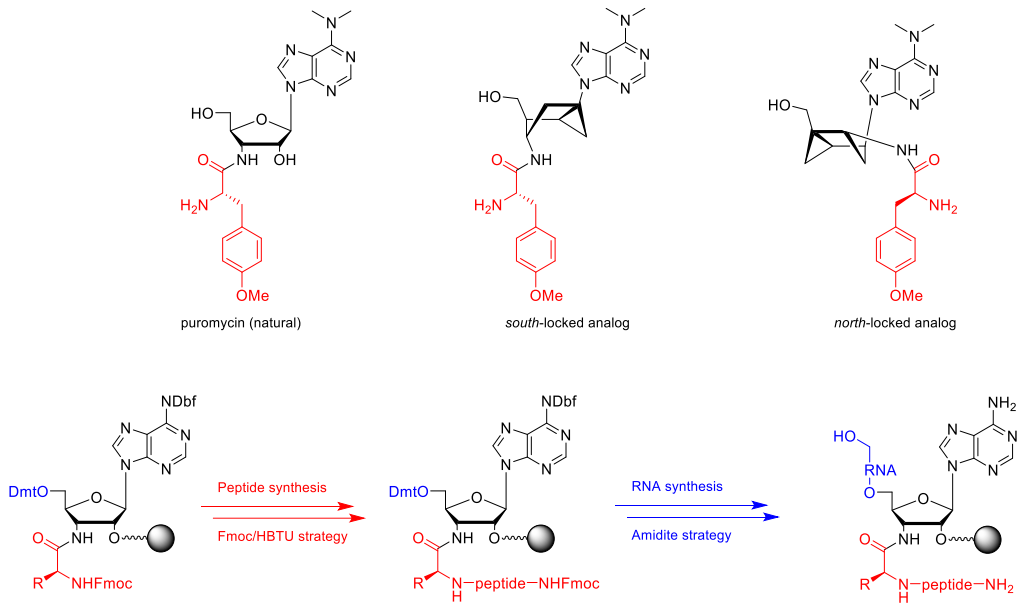


Scheme 2.28 Rearrangement of 3'-aaAMP to either 2'-valyl-adenosine-3'-monophosphate or adenosine 2'-3' cyclic phosphate.

A large body of research was dedicated to the synthesis of nonhydrolyzable aminoacyl-tRNA mimics and peptidyl RNA, both as an effort to elucidate the chemical mechanisms at work in the ribosome and with therapeutic motivations. Puromycin (Scheme 2.29, top) is a naturally occurring antibiotic that inhibits protein synthesis.[260] It resembles the aminoacylated end of tRNA but with a stable amide link instead of the expected electrophilic ester. Incorporation of puromycin therefore blocks protein synthesis. Strazewski investigated the synthesis of puromycin and its analogs (Scheme 2.29, top)[261–263] and chose a similar structure as a platform for the synthesis of stable peptidyl-RNA. The key fragment was attached to a resin through an appropriate linker on the unexploited 2' position, then the peptide synthesis was accomplished from the available *N*-terminus of the aminoacyl moiety, followed by conventional RNA synthesis from the 5' terminus of the building block (Scheme 2.29, bottom).

Comparable strategies were later exploited by the group of Micura to produce similar nonhydrolyzable amide-linked 3'-peptidyl-RNAs.[264–266] Producing full-length nonhydrolyzable tRNA by direct solid-supported synthesis would be very challenging. Indeed, tRNAs are approximately 80 nt-long and feature a large number of highly sensitive modified nucleobases that are not commercially available as amidite building blocks and sometimes incompatible with the

conditions employed in amidite chemistry. Micura and coll. circumvented this obstacle by enzymatically cleaving a small portion of the 3'-end side of natural tRNAs then replacing it with synthetically accessible material during a second enzymatic step (Figure 2.14).[267]



Scheme 2.29 Top: structure of puromycin and some conformationally locked analogs. Bottom: synthesis of peptidyl-RNA following a divergent solid-supported strategy by Strazewski and coll. The starting protected puromycin analog was synthesized separately and loaded onto a specific resin through a PEG-acid linker.

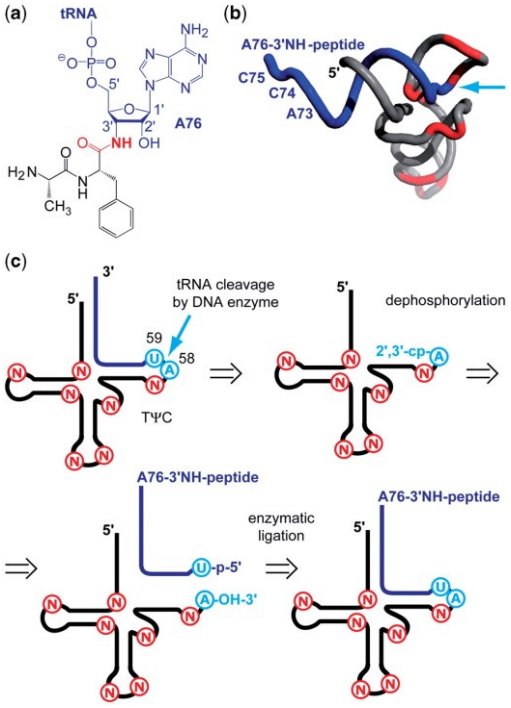
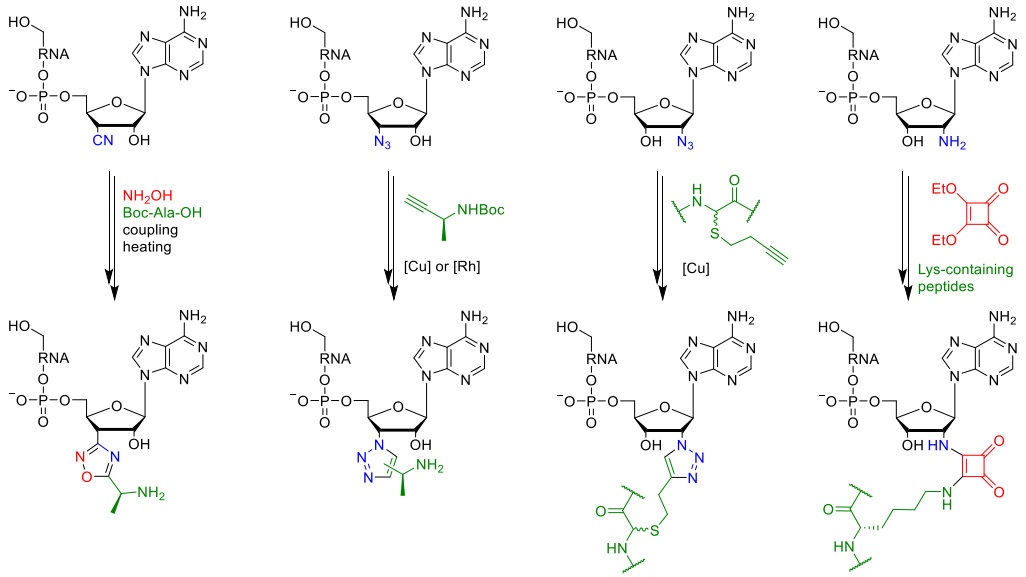


Figure 2.14 Overview of the semi-synthesis of nonhydrolyzable analogs of 3'-acyl tRNA. Modified nucleotides are marked with the letter N. Figure under Creative Commons license reproduced from the author's original publication.[267]

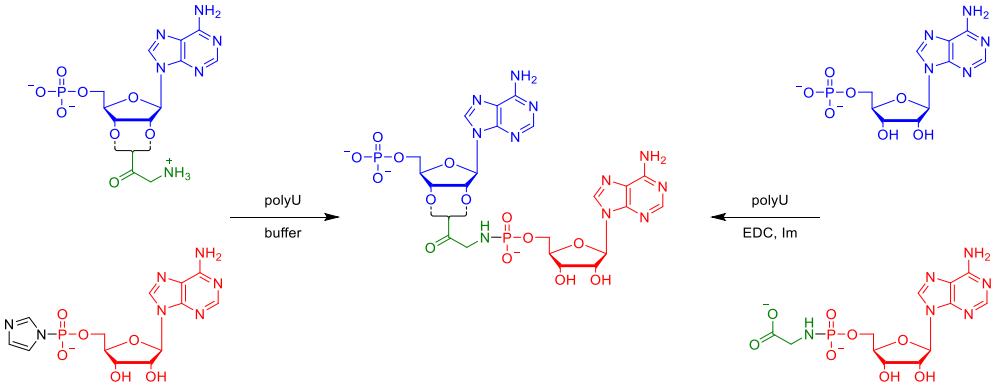
Linking fragments other than amides were proposed by Etheve-Quelquejeu and co-workers and were notably used to investigate processes related to the formation of peptidoglycans in cell membranes. Oxadiazoles were obtained from 3'-cyanodeoxy derivatives of tRNA,[268] while the classical 2'-azido-2'-deoxy and 3'-azido-3'-deoxy precursors, frequently used as masked amines in the syntheses of amide peptidyl-RNAs described above, were exploited to generate 1,2,3-triazoles through a Huysgen reaction with peptides containing *S*-homopropargylated cysteines or other alkyne-containing peptide-like compounds.[269–271] Squaramate was also introduced as a linker between 2'-amino-2'-deoxy tRNA and the side-chain of lysine-containing peptides (Scheme 2.30).[272] These fragments were typically synthesized on a small dinucleotide subunit then enzymatically linked to the end of longer strands.



Scheme 2.30 Generic view of several stable peptidyl-RNA analogues obtained by substituting the ester by a stable isostere.

The stable isosters described above are interesting synthetic models but are far from the considerations of primordial synthesis. From the perspective of systems chemistry, aminoacyl nucleosides and peptidyl-RNA were probably too short-lived to play a durable role, although they may have been constantly regenerated by wasting fuel.[13] However, it was soon discovered that capping the *N*-terminus of the aminoacyl moiety, and in particular phosphorylating it, increased the stability of the aminoacyl esters. Indeed, Orgel and coll. investigated the formation of alternating oligomers of AMP and glycine from ImpA and 2'/3' glycyl AMP. They obtained the same product starting from the phosphoramidate GlypA and AMP (Scheme 2.31).[273] The reaction was only efficient in the presence of a polyU template. Recently, Pascal and coll. further characterized the process by assessing the stability of the 2'/3' ester and providing a rationale for its formation.[29] They observed that lowering the temperature was critical to the stability of the ester, as well as neutralizing the positive charge of the amino acid by *N*-phosphorylating it, and proposed some implications of the reaction sequence for the emergence of translation. A similar reaction was also performed by Orgel and coll. with 3'-

amino-3'-deoxy GMP derivatives. The authors observed significant oligomerization to (GlypG-NH-)$_n$, in contrast to the adenosine study that only yielded one bond.[274] They were surprised to observe no enhancement of the oligomerization when a polyC template was included, an effect that they attributed to the formation of G-quadruplexes with or without the template.



Scheme 2.31 Formation of alternating Gly,A oligomers on a polyU template

Aminoacyl esters of nucleotides, even stabilized thanks to *N*-phosphorylation, remain highly sensitive to hydrolysis. A more versatile chemistry, relying only on the more stable phosphoramidate conjugation, recently received renewed interest, and is currently studied in detail by the Richert group. During their studies on templated replication without pre-activation, Richert and coll. developed a "general condensation buffer" composed of the coupling agent EDC·HCl and the catalyst 1-ethylimidazole (EtIm) in a HEPES buffer containing $MgCl_2$.[241] This buffer, at high concentrations and low temperatures that the authors deemed relevant in the context of an eutectic environment,[122] was found to promote a variety of coupling reactions. In particular, when amino acids and mononucleotides were mixed under these conditions, 5'-phosphoramidates were found.[27] Both peptide and RNA chains then grew efficiently to usual lengths (typically around 5 residues). The RNA growth produced both 2'-5' and 3'-5' phosphodiesters, that could be discriminated by NMR. Interestingly, omitting Mg^{2+} resulted in a favored formation of the canonical 3'-5' linkage.[33] The authors also observed the massive formation of pyrophosphates from AMP (AppA) and exploited this usually undesirable reactivity to produce cofactors featuring a pyrophosphate function, such as flavine adenine dinucleotide (FAD) from flavin mononucleotide (FMN) or nicotinamide adenine dinucleotide (NAD) from nicotinamide mononucleotide (NAM) (Figure 2.15).

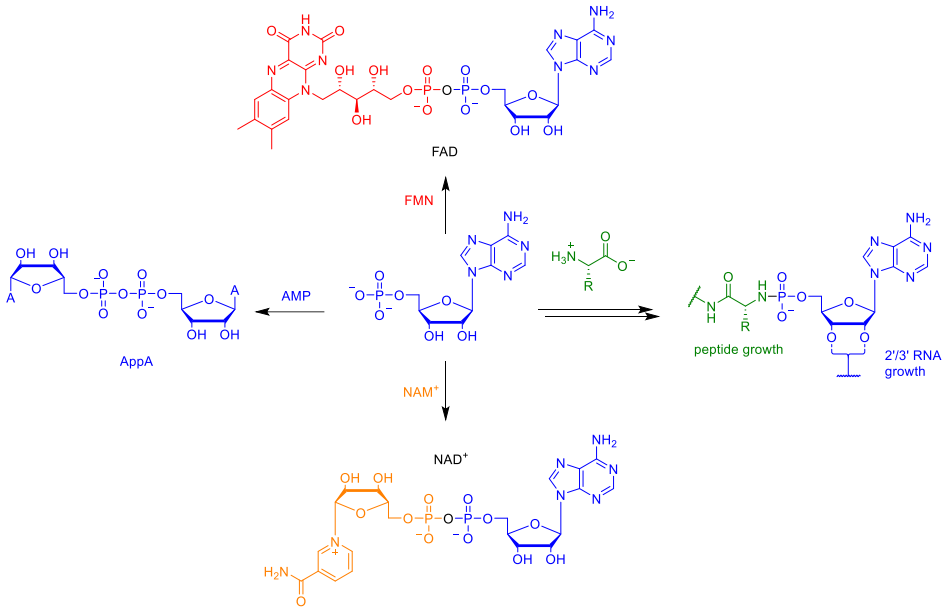
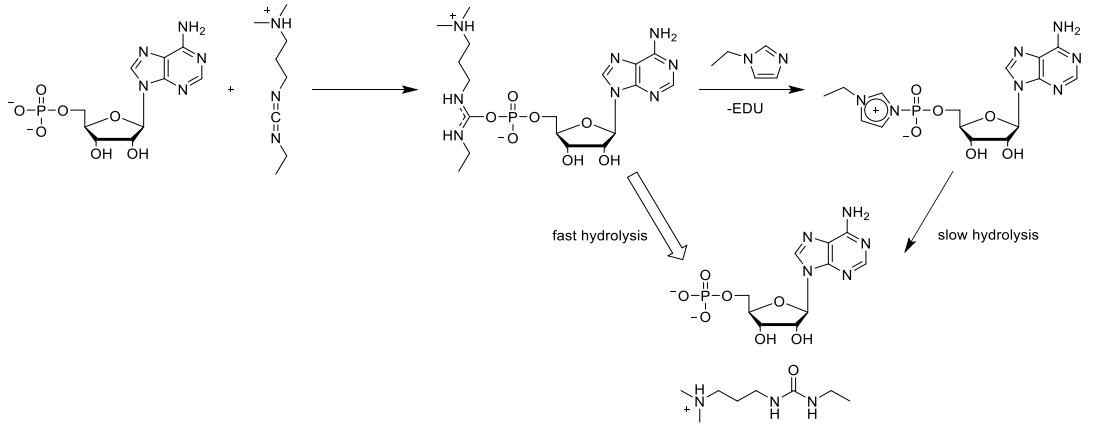


Figure 2.15 Products obtained from AMP in the general condensation buffer.

Although EDC is a well-known coupling agent in peptide synthesis, the activation of unprotected amino acids by EDC is inefficient, mostly due to the addition of the amine to the carbodiimide, forming a guanidine. This is avoided by blocking the nucleophilicity of the amine. In peptido-RNA, the initial, rapid formation of the phosphoramidate plays this role, and the elongation of the peptide chain is then allowed to proceed efficiently.[28] The carbodiimide-promoted formation of peptides in the presence of nucleotides was observed as soon as 1981 in the group of Oró,[150] but the authors did not discuss the formation of phosphoramidates, nor gave any rationalization for adding a nucleotide to their reaction mixture. When screening the twenty most common proteogenic amino acids with the four canonical ribonucleotides for peptido-RNA formation,[31] Richert and coll. noticed a significant disparity. As it could be expected, amino acids with functionalized side chains produced non-canonical products (branched, connected through another function than the α-amine…). Histidine mainly yielded AMP imidazolides, whereas cysteine, aspartic acid, tyrosine and glutamic acid were poorly coupled and mainly provided NppN in the addition of the starting nucleotides. Proline, which differs from all other amino acids by its cyclic nature and its secondary amine, was particularly reactive and furnished the best yield of peptido-RNA.

The role of EtIm in the reaction was studied in detail[32] and revealed to be deeper than the loose "general nucleophilic catalysis" that is usually invoked. A kinetic study, in which experimental rate constants obtained through NMR experiments were injected, revealed that the intermediate 1-ethyl phosphoroimidazolium accumulated in place of the less stable EDC-activated isourea, and prevented the rapid, phosphate-catalyzed hydrolysis of EDC (Scheme 2.32). This is an interesting bridge between the dogmas of pre- and in situ activation, that meets the results of McGown and coll. with imidazole.[204]



Scheme 2.32 EDC-activated AMP was quickly hydrolyzed but could be preserved as a kinetically more stable ethyl phosphoroimidazolium.

2.2.3. A note on the analytical methods used in prebiotic oligomerization studies

Studying complex mixtures of compounds is sometimes considered "the chemist's nightmare" but unfortunately, this is often what systems chemistry is about. The appearance of new analytical methods added to the spectacular development of others offers the contemporary chemist a very wide range of techniques to meet his needs. This was not always the case, and prebiotic and systems chemistry long suffered from the lack of appropriate analytical tools, especially in the field of separations, which is critical when working with mixtures. Moreover, inappropriate procedures and data curation may lead to erroneous conclusions. Some reactions seemed to give very different outcomes depending on the method used to analyze the samples, especially when it came to quantitation. This is in part understandable given the amount of work required to set up a proper quantification, which is usually out of reach of a study focused on something else than the analysis itself. Satisfyingly, several recent studies tend to find a compromise between robust but reasonably tedious analytical campaigns.

2.2.3.1. Separative techniques

Separation has early been performed by either paper chromatography or electrophoresis (for larger molecules) using radioactive (^{14}C or ^{32}P typically) substrates. Radioactive counting enabled a convenient quantification of very small samples but intrinsically limited the substrates to those available in the labeled form. Paper chromatography and thin-layer chromatography (TLC) are considered low-resolution techniques to current standards. Another difficulty with these techniques was the identification of the products, because no hyphenation of paper chromatograms or TLC to an analytical tool is usually available.[a] Comparing with standards or inferring from other experiments is

[a] Now automated TLC-MS is available but is not widely used.

generally the only option, and the risk of misidentifying compounds or missing unexpected products is high.

The development of high-performance liquid chromatography (HPLC) enabled much more performant separations. While reverse phase elution remains the most common technique, the very polar nucleotides are sometimes better separated by ion exchange or ion pairing procedures. The latter offers the considerable advantage of using much less salt and can therefore be hyphenated to mass spectrometry (MS). Moreover, hydrophobic interactions, which are quite weak in ion exchange chromatography, can be exploited efficiently in ion pairing methods. For example, siRNA and lipids in a therapeutic formulation were separated and detected in a single HPLC run using ion-pairing on a low-retention reverse-phase column.[275]

Nucleotide-derived compounds and most peptides are conveniently detected by ultraviolet absorbance. For compounds without a chromophore or with insufficient absorbance, evaporative light-scattering detection (ELSD) is an interesting alternative.[276,277] HPLC/MS is a valuable choice because it enables the rapid identification of most products. Fernandez and co-workers used a complex hyphenation of UHPLC, ion mobility (IM) and MS/MS sequencing to study the evolution of depsipeptide mixtures.[147] Despite this arsenal, they still found that some sequences were difficult to tell apart, which suggests how much information can be routinely lost from more simple analytical setups. The value of LC/MS over direct mass spectrometry is twofold: it provides a form of quantitation (usually by UV) and prevents several false positives and adduct formation. For example, in their first study about peptido-RNA, Richert and co-workers reported the formation of mixed sequence hexa- and heptanucleotides by MALDI-MS analysis of ion exchange HPLC fractions.[27] While the identity of the compounds by mass spectrometry raised no question,[a] the HPLC trace showed, after a 50- to 300-fold magnification depending on the runs, a barely discernable rise over the baseline, demonstrating that such long products were extremely minor before shorter ones. Direct MALDI-MS would not have provided such information. Another example is the montmorillonite-promoted oligomerization of nucleotide imidazolides. In a comparative study, Ferris and coll. found that oligomers over 30 nt length could be detected by MALDI-MS but not by HPLC (products up to 12 nt gave discernable peaks).[278] The question was addressed in a dedicated experimental study[279] that showed that, although MALDI-MS may be useful to detect minute amounts of products that are otherwise difficult to identify, it was also prone to false positives due to aggregation of small products that gave a high apparent mass. The authors demonstrated, both on oligonucleotide standards and on crude reaction mixtures, that strictly excluding dust and desalting the products before spotting was crucial to get reliable MALDI results.

[a] The authors used low-resolution mass spectrometry but given the focus of the study (formation of oligomers) there is relatively low risk to misidentify compounds.

A similar case occurred with gel electrophoresis. In 2009, Costanzo and coll. reported the formation, in extremely low amounts, of unusually long polyG by simply incubating 3'-5'-cyclic GMP (cGMP) in Tris buffers.[280] When receiving critiques from researchers who were unable to reproduce their results, they published a note[191] stating that in order to perform well, the reaction required the free acid form of the cyclic nucleotide, and not the usual sodium form. This is debatable at best, as it was never mentioned in the original paper, and because the polymerizations were performed in Tris buffer at quite elevated pH. They also mentioned that only radioactive counting was sufficiently sensitive to read the electrophoretograms. In a more recent publication,[281] the group repeated a similar experiment using cAMP that they checked for the absence of sodium salts. They also verified that no oligomers could be present in the sample before the experiment. They used Tris buffers at high pH again and found, under these properly controlled conditions, products in the usual range of oligomerization studies (less than 10 nt long).

2.2.3.2. Structure determination

The matter of structural information has been largely addressed by modern NMR techniques. The Richert group showed that a system as complex as their "general condensation reaction" could be tracked in detail by combining ^{31}P (nucleotide-derived products), ^{13}C (peptide growth) and ^{1}H (coupling agents) NMR.[32] They also extensively used ^{1}H-^{31}P HMBC sequences (Figure 2.16) to facilitate their attributions,[31] a technique that knows a growing use in the non-specialized community.[61,282]

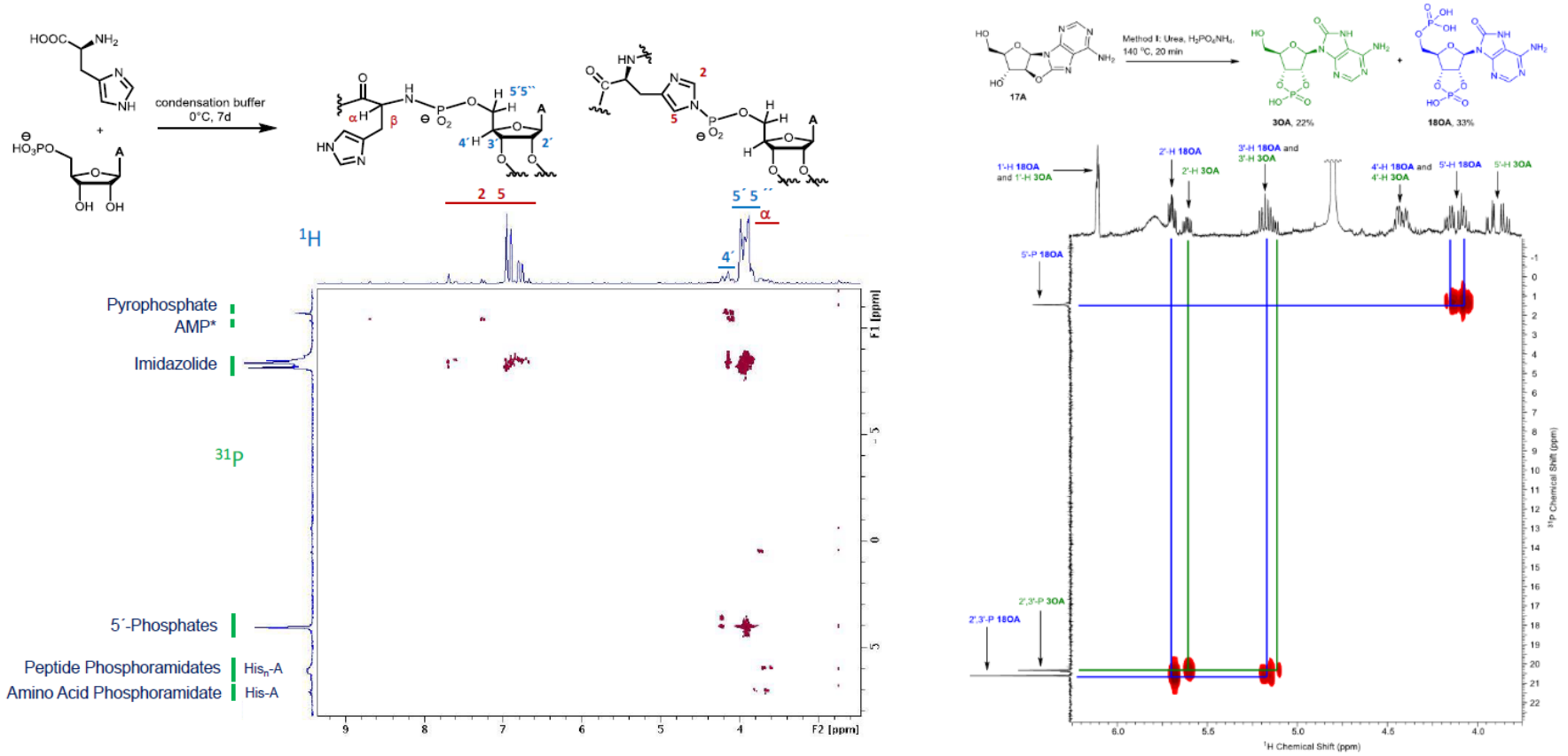


Figure 2.16 The growing use of ^{1}H-^{31}P HMBC to attribute ^{31}P peaks in complex mixtures. Left: attribution of many product classes in a general condensation reaction. Reproduced from the SI of the original manuscript[31] under Creative Commons License. Right: determination of the regio- and chemoselectivity of a phosphorylation reaction. Reproduced from the SI of the original manuscript[61] under Creative Commons License.

For very complicated or crowded spectra in one dimension, selective excitation methods[283] are a useful tool and increasingly performing methods are being developed.[284] Diffusion-ordered spectroscopy (DOSY), a specific subset of diffusion NMR in which the diffusion data is used to generate a pseudo-second dimension that indexes diffusion constants,[285,286] is also useful but suffers from peak overlapping. Islam and coll. circumvented this problem by fitting a pure-shift pulse program into a DOSY experiment to study a very complex mixture of TNA and RNA that included completely unknown compounds (Figure 2.17).[287] Thanks to the pure-shift method, each spin system was resolved to a singlet, greatly facilitating the attribution of severely overlapping broad signals.

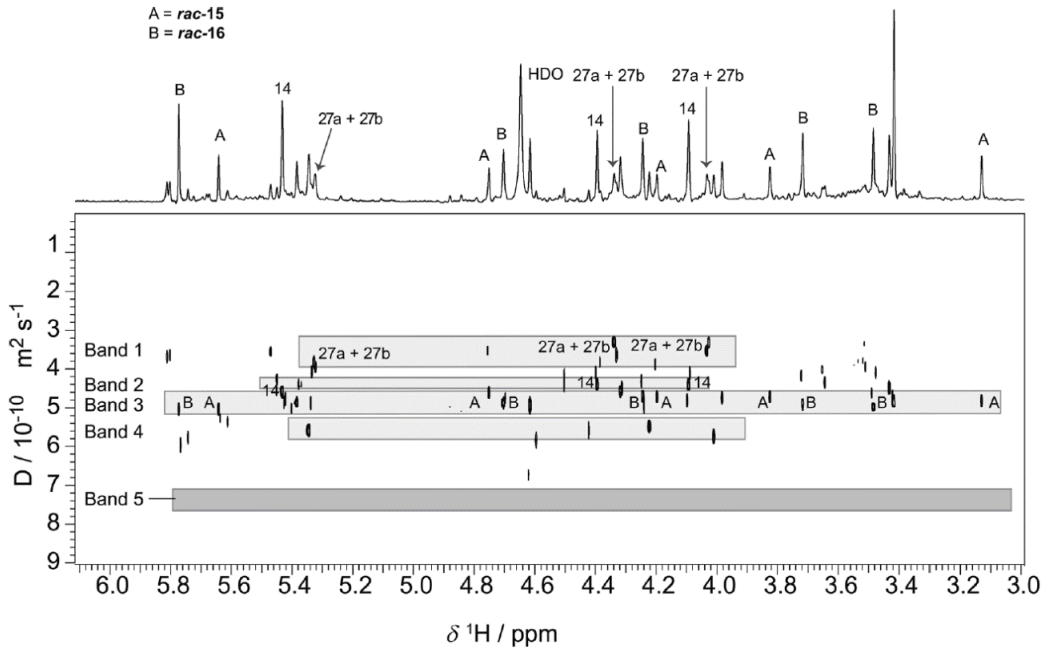


Figure 2.17 Pseudo-2-dimensional DOSY spectrum obtained on a complex aldol mixture using a combination of pure-shift ^{1}H NMR and diffusion NMR (One-Shot DOSY sequence). Figure reproduced from the original manuscript[287] under license with the Editor.

A common issue in RNA elongation and replication studies is the 2'/3' regioselectivity, as discussed before. The proportion of 2'-5' and 3'-5' phosphodiesters is often determined indirectly by selective enzymatic cleavage of one bond type followed by analysis of the fragments, with addition of other enzymes to remove 5'-phosphates and replace them by ^{32}P-labeled ones (see for example the representative study of Monnard and coll,[193] who used a set of 6 enzymes). These methods suffer from the lack of specificity of the enzymes and yield statistical results only. For this reason, the Richert group developed a method based on multidimensional NMR that allows to determine the 2'/3' repartition in small oligonucleotides, including non-canonical species (capped by a 5'-phyrophosphate for example), that they applied to RNA growth[33] and templated elongation.[288]

More advanced analytical tools are made available at an unprecedented rhythm. Although their misuse by non-specialists has sometimes been a source of errors

and overstated results, they also enabled studies that were impossible in the past and improved workflow efficiencies. One may also wonder how existing and, more importantly, widely known and affordable tools, affect the way research is carried out. It is indeed natural to first investigate what is readily accessible.

2.2.4. Overview

Forming and maintaining oligomers of sufficient length is a prerequisite of life. Without the action of enzymes and the protective environments found in cells, pathways to such long oligomers remain elusive. Although generally treated separately, peptide and oligonucleotide elongation face similar challenges: regioselectivity, growth of long chains, preservation against hydrolysis. Mechanisms that enhance the production of specific peptides or nucleic acids once a templating sequence is present are progressively unraveled.

The crucial importance, both conceptual and practical, of translation in cells motivated the search for a nucleic acid-assisted peptide synthesis. Mechanisms leading to peptidyl and peptido-RNA are being discovered, some of which provided enhanced peptide synthesis. The establishment of a code, where amino acids are selectively activated by given nucleic acids, is still far from reach.

2.3. Implications of lipid membranes in systems chemistry

Cells are surrounded by a membrane that is essential to their function and survival. Contemporary cell membranes are composed of a majority of phospholipids (phosphatidic acids, phosphoethanolamines, phosphocholines and more exotic lipids) and reinforced by a cell wall. De novo biosynthesis of amphiphiles is a relatively complex task, involving C-C bond forming pathways (Ac-CoA in eukaryotes and prokaryotes, isoprenoid pathway in archaea)[83] and have been used as a marker of biological evolution.[289] Prebiotic lipid synthesis, in other terms the geological origin of long chain material and its abiotic conversion to higher lipids, has been reviewed in section 2.1.2.6.

Although amphiphiles spontaneously self-assemble when dispersed in water, obtaining stable vesicles with the desirable physical properties to enable the formation of organized systems is not straightforward. Here we review some aspects of the implications of vesicles in prebiotic studies, and particularly their contribution to the formation and encapsulation of oligomers.

2.3.1. Lipid vesicles and the importance of compartmentalization

2.3.1.1. The formation of lipid vesicles from prebiotic amphiphiles

Vesicles of contemporary phospholipids are routinely produced in the laboratory.[290] However, complete phospholipids were, at best, a late product of the chemical evolution,[21] and maybe even of the beginning of biological evolution.[289] Early protocells must then have accommodated with more simple lipids. Fatty acids self-assemble into vesicles in water when added above a concentration known as the critical vesiculation concentration (CVC).[291] This concentration is highly dependent on the conditions, notably pH,[292] and fatty acid vesicles easily disrupt when exposed to high salt concentrations.[293] Indeed, the stability of fatty acid membranes is largely due to the formation of a dense hydrogen bond network between carboxylic and carboxylate heads, which requires the pH to be around the apparent pK_a of the acid in water (from 6.5 for octanoic acid, the shortest membrane-forming amphiphile, to 8.5 for oleic acid).[294] Moreover, even low concentrations of Mg^{2+} and Ca^{2+} complex fatty acid, causing membrane disruption.

The stability of fatty acid vesicles against salts may be improved by including relatively small amounts of co-surfactants. Nonanoic acid/nonanoate vesicles display a CVC of 85 mM at pH 7 in pure water, but this value was lowered to 20 mM by the addition of only 2 mM nonanol, with net improvement in pH tolerance.[292] Monoacylglycerols (Figure 2.18), that are intermediates in prebiotic phospholipid synthesis, also displayed similar stabilizing effects.[295,296] The phosphorylation of monoacylglycerols with DAP gave single-chain

cyclophospholipids (Figure 2.18).[104] When mixed with fatty alcohols, cyclophospholipids formed vesicles that were much more robust against mono- and divalent cations and pH changes.[121] We also demonstrated that crude mixtures of amphiphiles obtained by urea- or cyanamide-promoted phosphorylations of long-chain material could give vesicles upon hydration.[99,103] Similarly, carbonaceous meteorite extracts were shown to form membranes upon hydration.[89] More surprisingly, small molecules such as nucleobases[297] and amino acids[298] may also stabilize fatty acid vesicles upon adhering to them. Mixtures of fatty acid and the corresponding phospholipids also yielded stable membranes that were conveniently prepared in the lab, which is an interesting exploration of the evolutionary span of lipids (from primordial, prebiotically generated fatty acids and late phospholipids, maybe chemically synthesized, maybe produced by a primitive catabolism).[299]

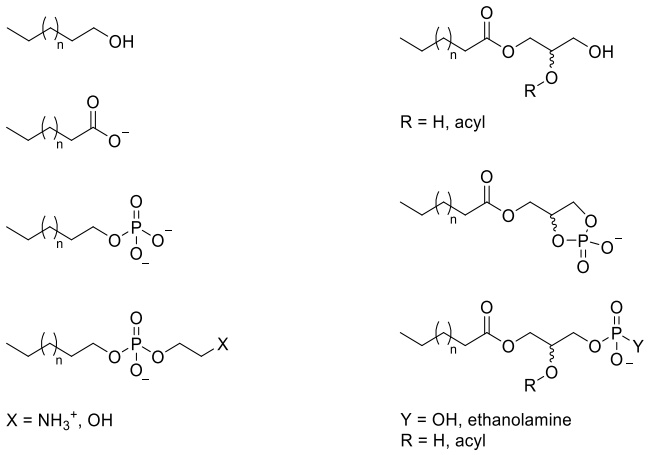


Figure 2.18 Possibly prebiotic amphiphiles. Left: single chain amphiphiles (from top to bottom fatty alcohols (alkanols), fatty (alkanoic) acids, alkyl phosphates, alkyl phosphoethanolamines and hydroxyethyl phosphates); right: glycerolipids (from top to bottom mono- and diacyl glycerols, cyclic acylglycerophosphates, mono- and diacyl phosphatidates and phosphoethanolamines.

Another important phenomenon, that may be a difficulty as well as a strength, is the phase transition temperature of the lipids. Lipids must be fluid in order to form membranes,[300] but short chain fatty acids typically remain solid at elevated temperatures. In biology, the contribution of unsaturated lipid chains bearing (*Z*) double bonds helps increasing the fluidity of the membranes at room temperature (for example T_m = 13 °C for oleic acid and 69 °C for stearic acid, its related saturated acid). However, it seems relatively unlikely that unsaturated chains, especially with the labile (*Z*) configuration, were formed by chemical synthesis and preserved durably, although oleic lipids are commonly used for convenience in prebiotic vesiculation studies.[99,299] Here again, mixed compositions provide a potential answer.[292,295]

2.3.1.2. Laboratory methods to prepare and study giant vesicles

The early descriptions of phospholipid vesicle preparation techniques are half-century old.[301] The so-called "natural swelling" or "gentle hydration" involves spreading a thin phospholipid film on a glass surface by evaporation of a stock solution (typically in chloroform or chloroform/methanol) then delicately hydrate it with an appropriate aqueous solution without mechanical disturbance (Figure 2.19, A). The original description included a pre-hydration step with wet gas that is not systematically applied today but remains helpful when performing difficult hydrations, for example at physiological salt concentration.[302] An alternative option is to prepare the film over a thin layer of agarose gel, although the agarose tends to be partially incorporated, resulting in vesicles with altered properties.[303] The natural swelling yields giant unilamellar vesicles (GUVs, diameter 1-50 µm), alongside with smaller multilamellar vesicles (MLVs) that result from hydration of excessively thick films or ripping of portions of the film, and small and large unilamellar vesicles (SUVs and LUVs). This very rustic method has a high relevance to prebiotic processes as it involves no sophisticated intervention. In the laboratory, the resulting heterogenous vesicle suspension may be further processed to produce more regular SUVs (<100 nm) and LUVs (>100 nm).[a]

The injection method is another simple process leading to unilamellar vesicles.[304] It involves rapidly mixing (usually injecting with a syringe) a concentrated solution of lipids in ethanol or methanol in an aqueous buffer. If the conditions are well chosen, quite regular liposomes are obtained, with an average radius typically around 100 nm.[305] The method was used for the encapsulation of drug molecules originally dissolved in the alcohol phase[306] and is relevant to growth and division studies (see below).

The growing application of vesicles in both fundamental sciences and biotechnology motivated the development of methods to produce vesicles with regular sizes, lamellarity, or efficiently encapsulating solutes in their lumen.[290] They include the use of oil/water emulsions with a lipid layer sitting at the oil/water interface (Figure 2.19)[307] and electroformation at the surface of electrodes.[308] Electroformation is only effective for some lipids. Oil emulsion methods are useful to encapsulate small molecules. In this technique (Figure 2.19, B), a water-in-oil emulsion is first prepared, in which fine droplets of an aqueous solution are dispersed in mineral oil containing small amounts of lipids, and get coated by a single layer. These droplets are then added to an oil-water biphasic system between which a monolayer lies. Gravity (or centrifugation) brings the droplets down through the interface, where a bilayer is formed, without allowing

[a] The exact threshold between « small » and « large » vesicles is not well-defined. In particular, vesicles of 100 nm in diameter may lie in each category depending on the authors. Here we chose to call LUVs the vesicles with a standard diameter (either expected or determined by measurements) of 100 nm or more, and SUVs vesicles with a standard diameter strictly below 100 nm.

the solutes in the droplets to equilibrate with the bulk. Unfortunately, some oil may accumulate in the bilayer, yielding altered vesicles.

Regularly sized, unilamellar vesicles of phospholipids (SUVs and LUVs) may be prepared from a crude suspension of giant vesicles obtained by natural swelling. Repeatedly freezing the vesicle suspension in liquid nitrogen then thawing it in warm water reduces drastically the multilamellarity (unilamellar vesicles can be obtained after 10 cycles) and breaks down the biggest vesicles to LUVs, as observed by microscopy and quantitatively studied by ^{31}P NMR.[309] Narrower size distributions are then obtained by extrusion techniques, in which the vesicles are forced through polycarbonate membranes with specific monodisperse pore sizes (usually 100-400 nm).[310] The solutes encapsulated in the crude vesicles before extrusion tend to leak out during the extrusion step, especially small molecules.[311] Extruded vesicles are well-suited for spectroscopic studies because of their regular size and thickness.

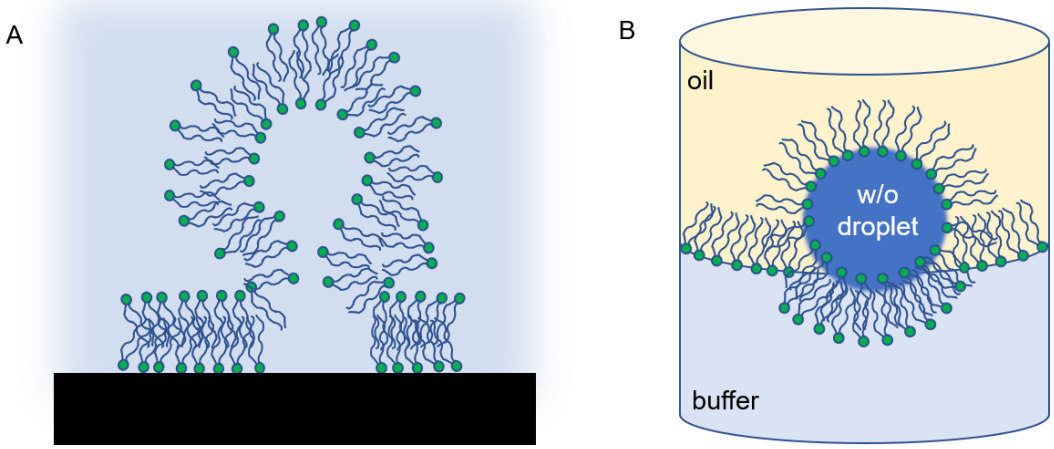


Figure 2.19 Formation of vesicles. A: hydration of a thin lipid film deposited on a surface or an electrode. B: water-in-oil method. A single layer is formed at the interface between oil and buffer, then a water-in-oil emulsion containing lipid-coated water droplets is added and falls through the interface by gravity, forming a bilayer while retaining its encapsulated substrates.

Fatty acid-based vesicles are obtained by methods simpler than phospholipid vesicles. Fatty acids usually do not yield a suitable film when evaporated, as phospholipids do. Therefore, they are commonly mixed in the desired buffer (or in water with an appropriate stoichiometry of NaOH) and gently heated until a translucent phase is obtained.[294] A more elaborate method involves first dispersing the fatty acids at a pH 0.5-1 unit more elevated than their apparent pK_a to generate a limpid micelle suspension (if needed, co-surfactants are added at this stage), then adjusting the pH back to the optimal value (pH = pK_a^{app}).[292] Moreover, fatty acid vesicles are typically produced at high amphiphile concentration because their CVC is elevated.[294]

Several methods have been developed to separate vesicles from external, non-encapsulated solutes, or to achieve size discrimination. SUVs and LUVs are typically purified by size-exclusion chromatography over gel columns.[312] In this

technique, the sample is applied to a gel and elutes down by gravitation. Small samples are more conveniently handled in spin-tube gel columns. The larger objects (here the liposomes) do not interact with the gel and elute first, followed later by small molecules and salts that must diffuse through the gel particles. Unfortunately, GVs do not resist this technique. They can be dialyzed or retained by tangential filtration, a technique in which defined pore size filters are fitted to the horizontal walls of a spin-tube.[298] The advantage of tangential devices when compared to classical ultrafiltration is that the vesicles are not pressed against the bottom of the filtration device, reducing the risk of rupture associated with the mechanical stress. Another method is to hydrate the lipids with a dense buffer, then to dilute them with an isotonic but less dense buffer.[311] Typically, sucrose (dense) and glucose (less dense) are used, at relatively high concentrations (200 mM). This allows to spin the dense vesicles down to a pellet that can be washed repeatedly until the level of outside contamination is suitably low. Methods based on size-specific filtrations are also available. Using wide-pore (typically 10 µm) ultrafiltration membranes, only GVs are retained, while smaller liposomes are washed out at the same time as non-entrapped solutes (Figure 2.20).[313,314] We recently proposed a variation of this method in which much narrower (0.02 µm) nylon membranes were used.[315] Vesicles encapsulating calcein were repeatedly diluted in calcein-free buffer and concentrated by filtration up to three times to yield samples with very low background fluorescence (Figure 2.21). Unfortunately, a large amount of lipids was lost due to compression of the vesicles against the filter, that induced vesicle breakdown.

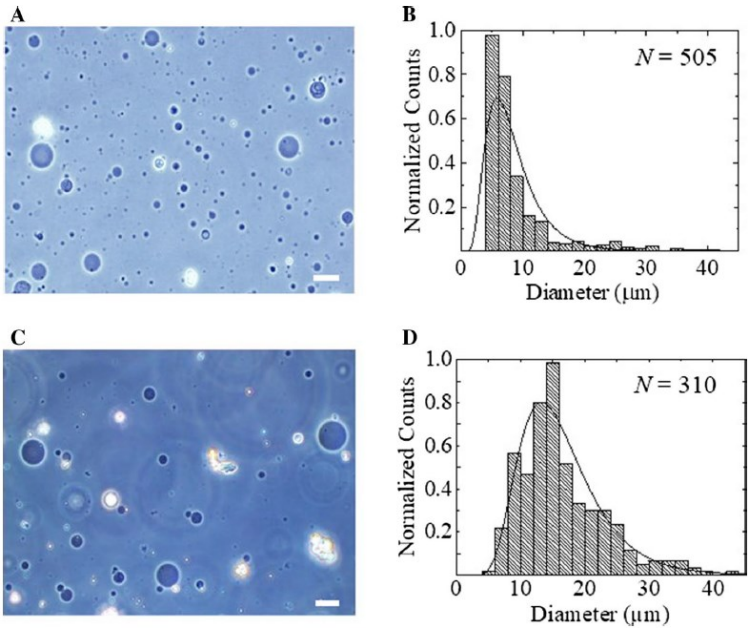


Figure 2.20 Phase-contrast micrographs (left) and size distribution (right) of GVs before (A, B) and after (C, D) filtration over 10 µm pore filters. The size distribution highlights the removal of liposomes with a diameter below 10 µm. Figure reproduced from the publication of Karal and coll.[314] under license with the publisher.

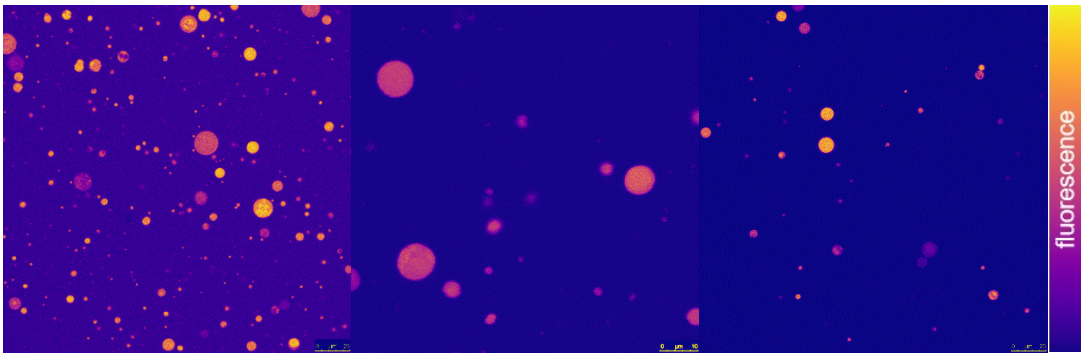


Figure 2.21 Confocal micrographs (the author, personal collection) of giant vesicles prepared by gentle hydration with a calcein-containing buffer, from left to right: before purification by filtration, after one filtration, after two filtrations.[315] A highly contrasted look-up table (LUT, depicted in the right bar) is applied to evidence the decrease in background fluorescence on low-contrast printouts and screens.

Analytical and imaging techniques for vesicles are numerous. GVs are large enough to be observed by optical microscopy, either directly or through the use of fluorescent probes. Calcein,[315] pyranine and carboxyfluorescein[294] are among the most popular water-soluble probes to reveal the lumen, as they do not cross bilayers easily unless a stress is applied. The membranes may incorporate a small amount of a fluorescent lipids (added to the composition of the thin film to be hydrated) or be stained externally with a lipophilic dye, most commonly Nile Red™ (Figure 2.22).

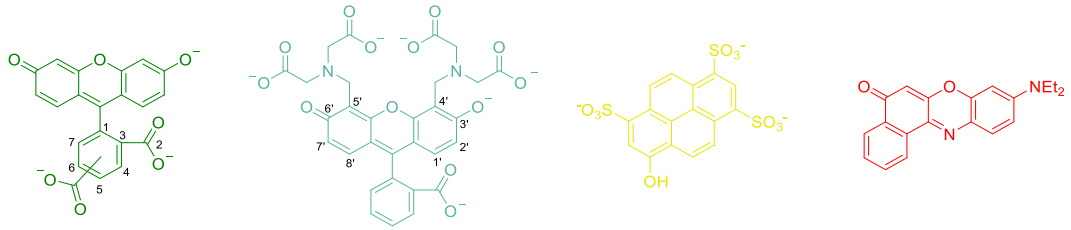


Figure 2.22 Structure (from left to right) of (5/6)carboxyfluorescein, calcein (sold as a mixture of 2'-7' and 4'-5' isomers), pyranine and Nile Red™.

When available, confocal microscopy with laser excitation fluorescence (example in Figure 2.21 above) is best suited for quantitative and time-resolved studies of giant vesicles[26,315] as it allows for a very narrow depth-of-field and a great spatial resolution, in addition to the possibility of particle tracking, tomography and other features. Phase-contrast microscopy (example in Figure 2.20 above)[314] is also used for vesicles encapsulating solutes, as well as freeze-fracture microscopy[90] for general purposes, but these techniques are spatially less resolved. A range of techniques usually called "super-resolution microscopy" or "subwavelength imaging" are under constant development and deliver images with a resolution beyond the limits imposed by optical diffraction.[316]

SUVs and LUVs are too small for optical microscopy and are usually observed by electron microscopy (example in Figure 2.23 in the next section), after flash-freezing and slicing (cryo-TEM technique).[317] A metallic stain may be needed. Flow cytometry, a tool developed to analyze and sort cell populations, can also be

used with giant vesicles[99] and is convenient for quantitative encapsulation assay by fluorescence, or size distribution calculation by light scattering, two tasks that are tedious by microscopy imaging and image treatment. Finally, the lipid concentration and composition may be determined by specific phospholipid assays,[318] ^{31}P NMR for phospholipids,[319] ^{1}H NMR for non-phosphorylated lipids[99] or chromatography (LC/MS, LC/ELSD,[276] GC[320]).

2.3.1.3. Encapsulation of biomolecules toward protocells

A common scenario for the emergence of protocellular organisms is that biomolecules were entrapped in vesicles or were able to permeate their membranes selectively, without the assistance of a biological machinery, to give "encapsulated" systems.[13,14] Such systems are shielded against the outer environment because of the selective permeability of the membranes. For example, catalase trapped in fatty acid/alcohol vesicles was protected against the action of a protease that would have degraded it in solution, and retained its activity inside of the vesicles thanks to the high permeability of its substrate, hydrogen peroxide, through such membranes.[292] Contemporary cells are also highly dependent on crowding mechanisms for maintaining their structure and functions.[321,322]

Numerous tools are used to estimate the accumulation of molecules in the lumen of vesicles. The direct method, separating vesicles from the bulk solution and directly assaying the desired compounds in the denatured vesicle sample, is a valid option for small vesicles, but handling larger ones may be problematic (see discussion in section 2.3.1.2). Labeled biomolecules were sometimes used, including fluorescent-tagged nucleic acids[299] or DNA revealed by intercalating dyes.[292] Ferritin was used as a protein model because its high iron content allowed to discretely detect it through transmission electron microscopy (TEM).[323] Mg^{2+} could be revealed by specific fluorophores,[299] and changes in the refractive index was used for sugars.[291] General leakage and disruption were monitored by the release of encapsulated membranophobic dyes.[295]

The Szostak group observed that RNA generated on the surface of montmorillonite clays (as described in section 2.2.1.4) could be trapped directly inside vesicles.[324] They adsorbed a synthetic A_{15} oligomer tagged with cyanine 3 (Cy3) to the clay, then fed it with fatty acid micelles (myristoleic acid, an unsaturated fatty acid) that produced small vesicles, some of which encapsulated RNA. The Luisi group studied the spontaneous encapsulation of dilute macromolecules inside vesicles of varying sizes.[323] For this purpose they hydrated lipid films with appropriate aqueous solutions and studied the repartition of macromolecules by microscopic techniques. Under these conditions, it was naively expected that a statistical repartition of the solutes would occur. In relatively small vesicles (typically 50-200 nm in diameter), encapsulating a discrete number of molecules, that can be counted with appropriate methods, gives rise to concentrations in the μM-range.[290] When a solution of ferritin, a spherical, iron-

rich protein, was used to hydrate POPC[a] films, vesicles were formed with encapsulated ferritin molecules. The vesicles were separated from the bulk, then the proteins in individual vesicles were counted by cryo-TEM. Contrarily to the statistical expectations, most ferritin was trapped in a small number of vesicles (0.1-0.5 % of the vesicle population), the other ones being empty (85 % of the population) or filled with only a few protein units (15 % of the population, Figure 2.23). Even more surprisingly, the highly loaded vesicles were among the smallest ones. No protein aggregation or membrane binding was observed in the process that would be sufficiently significant to explain this behavior.[317]

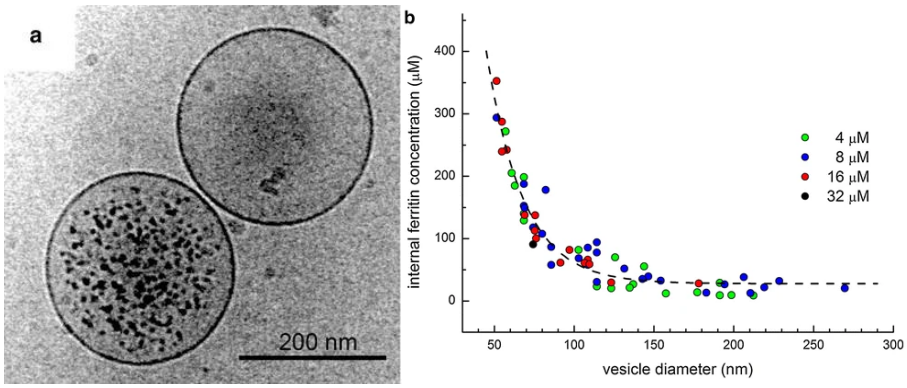


Figure 2.23 Crowding of ferritin in LUVs. a) cryo-TEM micrograph showing two LUVs, one filled with 150 ferritin molecules (fine, regular black dots), the other empty. If ferritin were statistically spread, such coexistence would be too improbable to be practically observed. b) plot of ferritin content vs vesicle diameter revealed that the highly filled vesicles were the smallest, which is opposed to a statistical prediction. The filled vesicle contained 150 ferritin units, giving a local concentration around 6 mM, which was three orders of magnitude above the bulk concentration in the supernatant. Figures reproduced from the original manuscript[323] under copyright with the Editor.

This crowding mechanism could be applied to metabolic systems. The PURE system is a synthetic, purified kit of biomolecules (ribosomes, enzymes, DNA etc) and cofactors required to express a protein autonomously, used as a convenient minimal system in model organisms.[325] Luisi and coll. encapsulated the PURE mixture in micron-sized vesicles by spontaneous entrapment and observed that the components of the system spontaneously crowded in a small selection of vesicles, as revealed by the expression of the enhanced green fluorescent protein (eGFP) in these vesicles but not in the background (Figure 2.24).[326] Indeed, the PURE system is composed of more than 80 substances and a slight dilution dramatically impairs its function, which reflects the critical requirement of crowding mechanisms for the emergence of protocells. This spontaneous crowding behavior is not well-characterized yet,[327] but was observed by others with immobilized vesicles hydrated with a fluorophore (CoroNa) solution[328] and by us using flow cytometry with calcein-filled vesicles made of prebiotic mixtures of lipids, although to a much less spectacular extent.[99]

[a] A table of the name, abbreviation and structure of the lipids mentioned in this work is given in the appendices

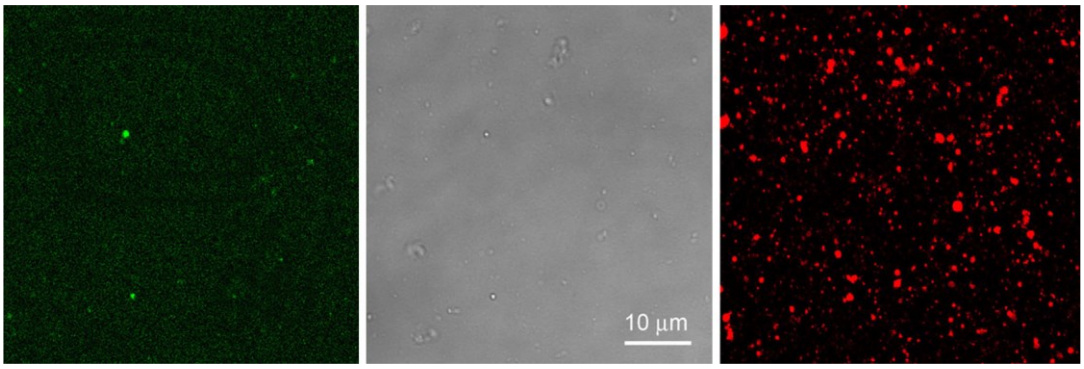


Figure 2.24 Confocal micrograph of vesicles prepared by ethanol injection in 0.65× PURE. Left: GFP channel revealed very bright eGFP-producing vesicles against a dark background. Middle: bright field. Right: staining with the membranophilic dye Nile Red™ revealed that many vesicles were formed, while only a few expressed eGFP. Figure reproduced from the original manuscript[323] under license with the Editor.

Although some membranophobic or large molecules may have been encapsulated spontaneously during vesicle formation, it was necessary that some substances could get it and out of protocells. Indeed, metabolic pathways must be fed by nutrients and eliminate some forms of waste. This led to considering membrane permeation as a critical parameter in protocell design. Fragile membranes such as fatty acid vesicles are usually highly permeable to a large range of substrates, including small organic molecules, ions, and molecules such as hydrogen peroxide. At the opposite, phospholipid membranes are rather robust and impermeable, and transfers across the membrane in contemporary cells is essentially dependent on enzyme-mediated transport mechanisms. Such a strong barrier could be a problem in the establishment of protocells, suggesting again that intermediate lipid membranes less fragile than pure fatty acid but more permeable than pure phospholipid could have played a critical role.[299,329]

Deamer and coll. measured the permeability of phospholipid membranes to polar substrates such as Pi and amino acids and found it to be several orders of magnitudes below the rates observed in cells, suggesting that phospholipid vesicles were essentially impervious to polar substrates in the absence of transportation mechanisms.[330,331] Yet, lipid membranes displayed enhanced permeability a temperatures close to their phase transition (liquid-disordered/gel) temperature T_m. ADP could permeate dimyristoyl phosphocholine (DMPC) vesicles (T_m = 23 °C) at a rate sufficient to allow a polymerase-promoted reaction to take place.[332] The same year, the Luisi group performed a similar experiment in oleic acid vesicles.[333] Both these examples were coupled with growth and division processes (see section 2.2.1.5). In 2008, the Szostak group studied a prebiotic variation of this experiment, encapsulating a template DNA with a long dC region and its complementary primer (leaving the dC region to be extended), and feeding the system with imidazole-activated 2′-amino-2′,3′-dideoxyguanosine. They observed template-directed elongation of the DNA strand in vesicles made of non-phospholipid amphiphiles (using myristoleate chains to gain fluidity) but not in phospholipid vesicles (Figure 2.25).[334]

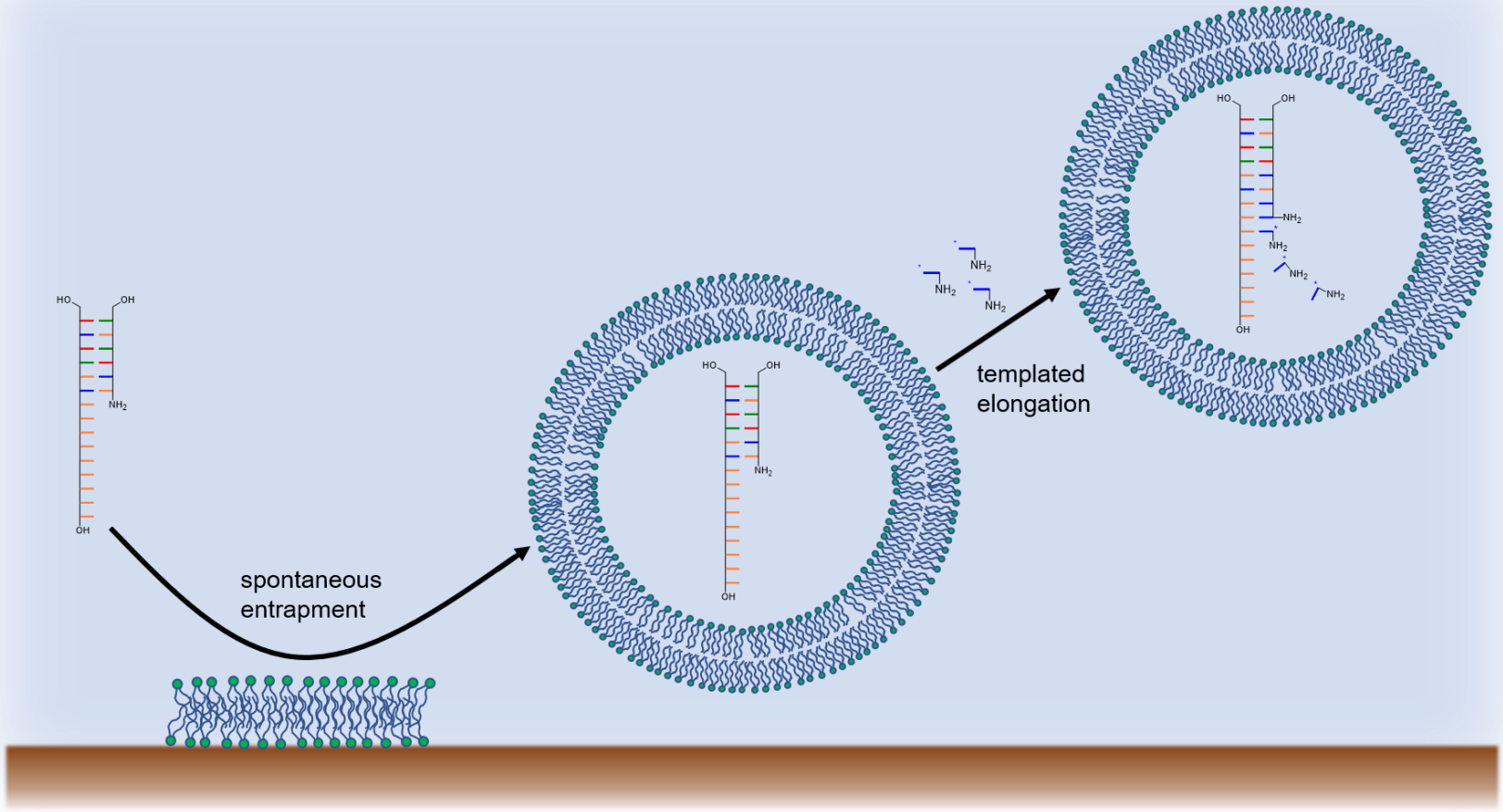


Figure 2.25 Templated replication of DNA inside giant vesicles as described by Szostak and coll. A template-primer duplex is first spontaneously entrapped in vesicles by dispersion of the lipids in a solution of the duplex. Activated monomers are then fed, cross the membrane, and allow templated elongation.

Several mechanisms may explain the permeation of charged molecules across bilayers. The lipid tail region itself resembles a hydrocarbon phase and the partition of most solutes in it is too low to explain any significant crossing rate.[91,335] Instead, the membranes may present local defects allowing the formation hydrated channels that are exploited by solutes to cross the membranes (Figure 2.26, A).[336] Another mechanism is the so-called "flip-flop" exchange, by which lipids are exchanged between the inner and the outer leaflet of a bilayer (Figure 2.26, B).[337] Without the intervention of enzymes called flippases, this exchange is very slow for phospholipids[338] but fatty acids, especially in their protonated form, are much more prone to flipping.[339]

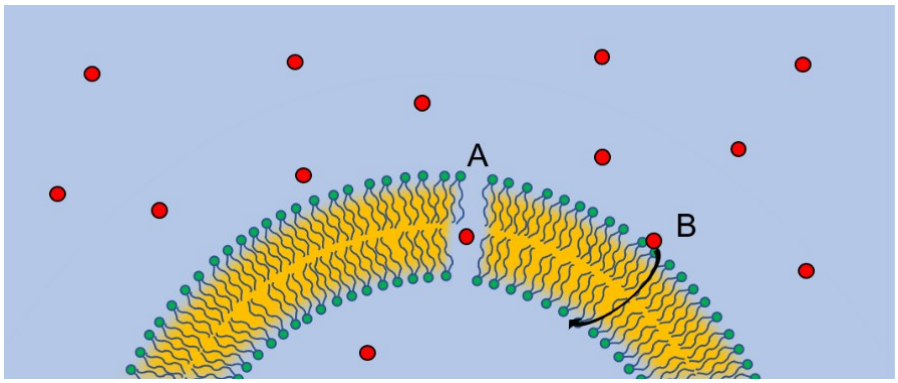


Figure 2.26 Penetration of solutes (red spheres) through bilayers, exploiting hydrated defects in the membrane structure (A) or the flip/flop of the lipids (B).

2.3.1.4. Spontaneous growth and division of vesicles

Living species persist through long geological times thanks to the self-replication of organisms. For cells, this implies the existence of a mechanism of reproduction of the cell membrane. A minimal protocell embedded in a vesicle would then require a primitive mechanism triggering the proliferation of its vesicle boundary using locally available resources. A good level of fidelity in the replication of the membrane's chemical composition and physical properties is required, as is the transfer of at least some of the components originally encapsulated in the lumen of the "mother" vesicle to the lumen of the "daughter" vesicles.

Early demonstrations of this "autopoietic reproduction" were produced by the Luisi group using reverse micelles of cationic surfactants.[340,341] Four years later, they were able to carry out the first practical demonstration of fatty acid vesicle growth and division, feeding oleate vesicles with oleic anhydride.[342] Because oleic anhydride is practically insoluble in water, it did not hydrolyze readily, like usual anhydrides would. Its insertion into pre-established oleate membranes enhanced its hydrolysis considerably, generating an autocatalytic system because the excess of fatty acid released in the medium then formed new vesicles (Figure 2.27).The process was characterized kinetically with SUVs then observed by optical microscopy with GVs.[343]

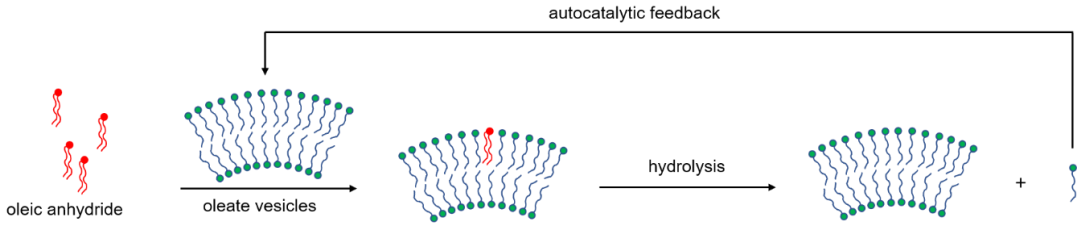


Figure 2.27 Autopoietic growth of oleate vesicles fed with oleic anhydride that hydrolyzed in the membranes.

The practical difficulties associated with the immiscibility of oleic anhydride could be overcome by using aqueous micelles or concentrated alcoholic solutions of oleate as the feeding material.[344] With this method, Luisi and coll. established that oleate vesicles followed a templated behavior: the initial size distribution was conserved throughout the feeding. This was not the case when POPC was used as the feeding material, because of the very rapid de novo formation of POPC vesicles, independently of the preformed oleate vesicles. Reversely, preformed POPC vesicles acted as a template when fed with oleate,[345] although this approach was questionable in the light of the prebiotic chemistry of phospholipids (phospholipids were produced much later than fatty acids). More recently, the Devaraj group proposed synthetic models for autocatalytic phospholipid formation and vesicle growth, exploiting efficient reactions such as copper-promoted cycloadditions or the histidine ligation.[346–348] These reactions were very far from prebiotic considerations as judged by the reagents involved and their overall complexity, and were proposed as models for the biotic synthesis of lipids from its precursors. Indeed, in living cells, the last steps of lipid synthesis are carried out inside the membrane, using transmembrane enzymes, inducing the feeding and growth of the cell wall.

Szostak and coworkers proposed a model to simulate the successive growth and division of fatty acid vesicles fed with micelles.[324] In their system, extruded (100 nm) vesicles were first prepared and then fed with equimolar micelles. The growth of the vesicles was followed by DLS measurement of the vesicle radius, which expectedly increased by a factor of 1.4 ($\sqrt{2}$, as expected when doubling the surface of the vesicles). A subsequent extrusion step brought the vesicles back to their original size distribution, but in increased number. To monitor the inclusion of a new population of lipids into the original vesicles, they used a well-known fluorescence method based on the Förstner resonance energy transfer (FRET) effect between NBD and rhodamine-labeled lipids (Figure 2.28).[349] When the micelles fused with the pre-existing vesicles, the membrane budded and the fluorophores were scattered away, causing fluorescence enhancement by decreasing the FRET effect, which is extremely strongly dependent on distance.

A notorious problem in growth and division studies is that successive generations of vesicles are mixed together and cannot be distinguished or separated. Luisi and coll. studied the growth of POPC/oleate vesicles using free-flow electrophoresis (FFE, Figure 2.29, left), a technique that is able to separate non-denatured supramolecular objects based on their isoelectric point or their overall charge.[350]

This enabled the separation of the vesicles based on their oleate content and provided a first approach to the study of the chemical composition throughout growth and division. Recently, Monnard and Fiore proposed a system in which the first generation of vesicles was built around glass microspheres, then fed with ethanol solutions of fatty acids. The successive generations of lipids could therefore be collected separately after centrifugation (Figure 2.29, right).[320] Using this system, they observed by optical microscopy the major phenomena involved in the growth and division of giant vesicles and characterized the incorporation and exchange of lipids between generations of vesicles. Interestingly, this tool was compatible with the use of giant vesicles (the microspheres had a diameter of 5 µm), whereas electrophoretic studies used LUVs.

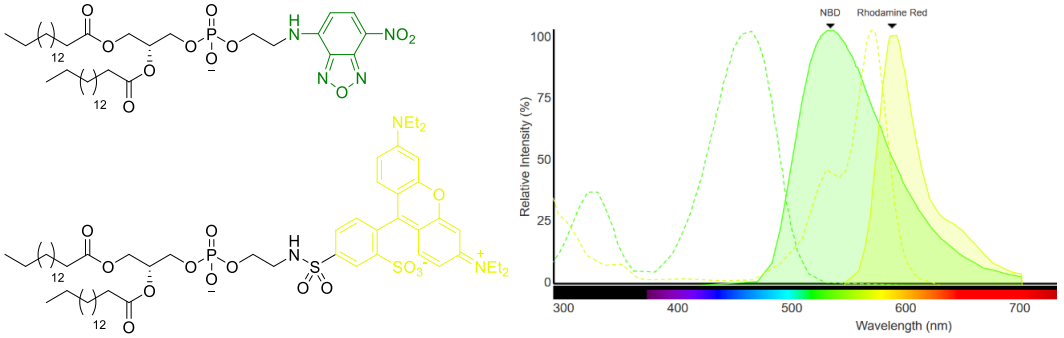


Figure 2.28 Structure of a FRET pair of lipids frequently used in membrane budding and fusion. Left: structures of NBD-DPPE (top, green) and Rhod-DPPE (bottom, yellow). Right: fluorescence spectra showing the overlap between the emission (filled curves) of NBD and the excitation (dashed curves) of rhodamine. Curves generated with Thermo Fisher's online tool Fluorescence SpectraViewer.

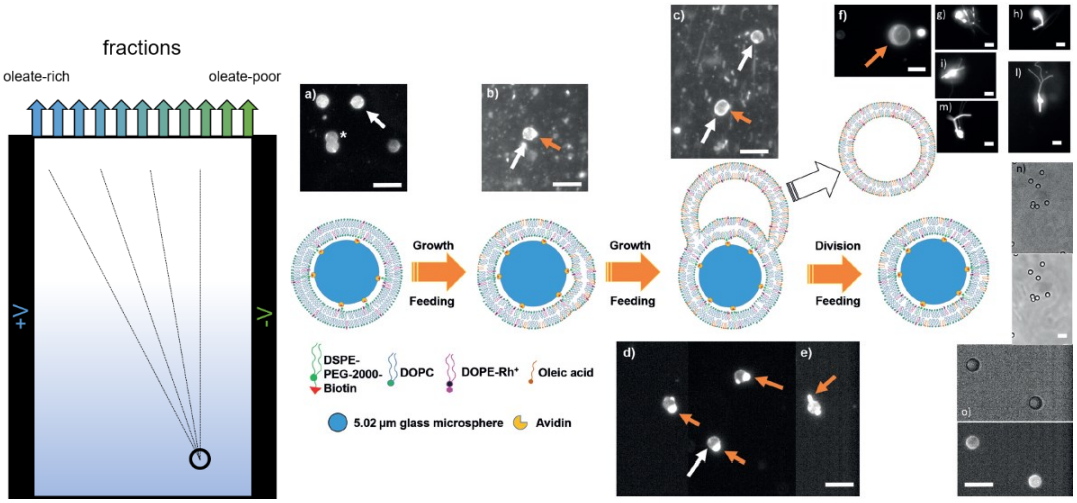


Figure 2.29 Techniques used to monitor the evolution of the chemical composition in growth and division studies. Left: free-flow electrophoresis. Non-denatured LUVs were spotted (black circle) and eluted vertically by buffer flow, while a transversal electrical field was applied. Negatively charged oleate-rich vesicles were biased to the left (+V) while near-neutral POPC vesicles (bearing a small apparent charge due to polarizability) eluted straight. Right: glass microsphere-supported vesicles were separated by centrifugation between each feeding run. Figure reproduced from the author's manuscript[320] under license with the editor.

In order to be applicable to protocell reproduction, a growth and division mechanism must be compatible with the distribution of the content encapsulated in the mother vesicles to the daughter vesicles. Indeed, if the compartments were to leak out most of their content, or non-selectively exchange it, there would be no proper "reproduction" taking place. Several studies then tackled the possibility to divide vesicles while retaining their content. Luisi and coll. were able to encapsulate a replicase and its substrates and cofactors in oleate vesicles and to grow the system by addition of oleic anhydride while preserving some catalytic activity.[351] They inhibited the action of the enzyme outside of the liposomes to ensure that the encapsulated system was indeed the source of the replicating activity. Similarly, a self-cleaving hammerhead ribosome was encapsulated in myristoleate/1-myristoleoyl glycerol (MMG) vesicles and maintained its activity through a growth process.[296] Since the activity of most ribosomes is dependent on Mg^{2+}, both the robustness and the permeability of MMG-containing vesicles to divalent cations was a key to the success of the experiment.

The Szostak group carried out a series of studies highlighting the potential synergetic interactions between growth processes and metabolic reactions. They first observed that osmotic pressure was a suitable lever to induce vesicle growth.[352] Indeed, vesicles filled with high concentrations of ions were able to grow by collecting lipids from isotonically filled vesicles, until their salinity reached an equilibrium. This discovery is relevant because many biotically relevant reactions are expected to be accompanied by changes in osmolarity,[321] and therefore could induce growth mechanisms. In some cases however, durably maintaining non-isotonic condition may be required. For example, pH gradients between the lumen and the external medium are necessary for certain metabolisms, and could provide a mechanism for the accumulation of amino acids in their ester form: under relatively basic conditions, the neutral form of the amino acid ester would be lipophilic enough to cross the membrane, but would be protonated once in the lumen, preventing it from crossing back.[331] Fatty acid vesicles are capable of very rapid ($t_{1/2}$ in the order of seconds) equilibration of their internal pH with the external medium because of the ease with which they flip-flop, carrying protons in and out. However, coating the surface of oleate vesicles with arginine prevented this mechanism and enabled the durable sustainment of a pH gradient throughout the course of a growth experiment.[353] Very recently, it was also demonstrated that the degradation of urea, promoted by encapsulated urease, provided a suitable pH shift to induce the division of POPC/oleic acid GUVs in unbuffered media.[354]

Another study by the Szostak group showed the possibility of altering membrane properties by the accumulation of lipophilic products other than new lipids. The dipeptide H-Ser-His-OH has been reported to catalyze the formation of peptide bonds between protected amino acids. The model product dipeptide used, Ac-Phe-Leu-NH_2 (formed from Ac-Phe-OMe and H-Leu-NH_2) is extremely lipophilic.[143] When Ser-His was encapsulated in fatty acid vesicles, it catalyzed the formation of the lipophilic protected product dipeptide that accumulated in the membrane and positively biased the filled vesicles toward growth.[355] A similar effect of

enhanced membrane growth induced by membranophilic peptide was observed when polyarginine was added to phosphocholine vesicles.[356]

2.3.2. Lipid vesicles as reaction promoters

Contemporary chemists are now facing the multiple downsides of their very frequent use of solvents: elevated costs, recycling or disposal processes, petroleum shortage, acute and long-term harm for the populations and the biosphere… Few industrial processes are compatible with neat conditions and the typical "green" solvent, water, often does not meet the requirements of solubility and reactivity necessary to chemical reactions. Chemists are searching for a way around this problem by using micelles, emulsions or other formulations of amphiphilic compounds in water that may provide sufficient solubility and reactive lipophilic "nests" for organic processes to perform, while being relatively harmless and, in principle, recyclable.[357]

The drastic conditions under which lipids were formed on the early Earth (or delivered to it), coupled to their somewhat simple structure relative to nucleic acids and peptides, strongly suggests that amphiphilic matter may have been present early in the first chemical steps toward life.[358] This directly raises the question of the influence that lipid matter may have on chemical reactions. We have discussed above the relevance of the encapsulation of relatively complex systems in the lumen of vesicles, and the related reactions. It is however unclear whether earlier and more simple chemical reactions have drawn a significant benefit from lumen confinement. One study tackled the production of glycine peptides from encapsulated glycine in a microfluidic simulation of hydrothermal vents (pure thermal conditions), and found some degree of reaction promotion, albeit the overall yields are extremely low.[359] The permeation of monomers through the membranes, the elimination of by-products and the stability of vesicles under the conditions required to perform efficient prebiotic chemistry are so many obstacles in the way of a primitive encapsulated chemistry.

The availability of lipid membranes is a more attractive option. It provides a lipophilic environment, with properties very different from the aqueous bulk, and the charged headgroups may bind some substrates by dipole interaction. Most studies on prebiotic chemical reactions in the presence of lipids (with the exclusion of the enzymatic or synthetic biological systems described in section 2.3.1.3) exploited the membrane itself and not the lumen.

2.3.2.1. Mechanisms for membrane binding and substrate localization

Analytical evidence supports that lipid membranes are able to organize small molecules in a way that may be favorable to their condensations. Black, Deamer and co-workers showed that nucleobases[297] and amino acids,[298] especially the lipophilic ones, were retained at the surface of fatty acid vesicles, most likely by

ion-pairing mechanisms. Among common sugars, ribose also showed the best retention in membranes in their studies. Such binding also had a positive effect on vesicle stability against salts, which joins the conclusions, previously exposed, on the addition of co-surfactants[295] or small molecules[353] to fatty acid vesicles. Nucleotides are polar, negatively charged molecules and have therefore very little chance to spontaneously bind lipophilic, anionic membranes, but stochastic entrapment during multilamellar structure formation may lead to the formation of organized layers of nucleotides as demonstrated by X-ray crystallography.[360] Nucleic acids stochastically entrapped during drying and hydration cycles displayed enhanced production of complementary sequences.[361] RNA may also be localized at the proximity of membranes either by covalent attachment to lipophilic peptides[26] or by electrostatic interactions with membrane-bound cations.[362] Small peptides were also capable of binding membranes by lipophilic interaction and formation of secondary structures (see below).

Localizing substrates to a membrane is usually believed to increase local concentrations, as can be illustrated by simple geometrical considerations. Let a vesicle be modeled as a sphere of radius r, with a wall of thickness Δr and an external surface S. In a first approximation, if the wall thickness is negligible before the diameter of the vesicle, the number of lipids forming the inside and the outside walls are considered equal. Given that Δr is in the order of 4 nm for PC vesicles,[290] this hypothesis is reasonable for LUVs and GUVs. A lipid head has a surface of S_{lipid} that will be averaged to 0.60 nm² for PC vesicles using literature data.[363] The local concentration of N_{local} molecules of substrate in a membrane of volume V is $C_{local} = \frac{N_{local}}{V}$. Let $f_{bound} = \frac{N_{local}}{N_{lipids}}$ be the molar ratio between the substrate bound to the membrane and the lipid. This term is convenient because it encompasses both the relative concentrations of substrate and lipids and the extent of binding in a given experiment. The volume of the membrane is:

$$V = \frac{4}{3}\pi(r^3 - (r - \Delta r)^3)$$

$$\Leftrightarrow V = 4\pi r^2 \Delta r \left(1 - \frac{\Delta r}{r}\right)$$

We can further write:

$$N_{local} = f_{bound} \times N_{lipid}$$

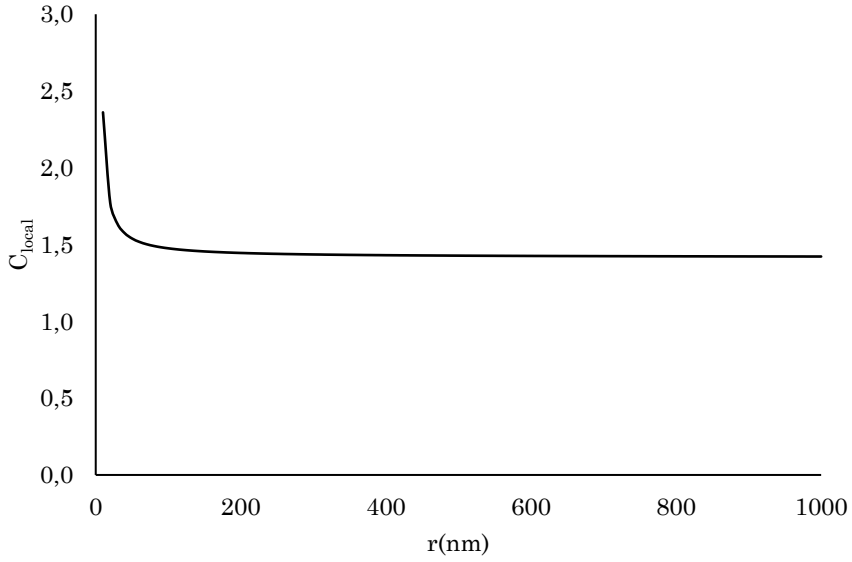
$$\Leftrightarrow N_{local} = f_{bound} \times 2\,\frac{S}{S_{lipid}}$$

$$\Leftrightarrow N_{local} = f_{bound}\,\frac{8\,\pi r^2}{S_{lipid}}$$

Combining these expressions in C_{local} gives:

$$C_{local} = \frac{2 f_{bound}}{S_{lipid}\,\Delta r \left(1 - \frac{\Delta r}{r}\right)}$$

We see that, when r increases, the concentration rapidly converges to $\frac{2\,f_{bound}}{S_{lipid}\Delta r}$, which is consistent, as $\frac{1}{2}\,S_{lipid}\Delta r$ is representative of the volume of one lipid in the simplified geometrical representation used here. C_{local} is expectedly proportional to f_{bound}. A graphical representation of $C_{local}(r)$ is given in Graph 2.2 with f_{bound} fixed to 0.001, which represents, for example, the case in which 10 µM of a substrate is concentrated in the bilayer of 10 mM vesicles.



Graph 2.2 Calculated local concentration (mmol.L^{-1}) of a substrate embedded in the membrane of a vesicle of radius r (in nanometers), for a value of f_{bound} fixed to 0.001.

With giant vesicles, the effective concentration of the substrate in this example would therefore be around 150 times the concentration initially introduced in water. For lipophilic peptides (see below), f_{bound} may be as high as 0.05 in dilute experiments, leading to spectacular local concentrations, especially if small liposomes are involved. Note that the geometrical simplifications involved in this model directly suggest that the bilayer is practically flat and made of an equal number of lipids on the inside and outside layers, which is not verified for small radii. Accordingly, a physically non-meaningful divergence is registered when $r \rightarrow \Delta r$.

Most biologically relevant substrates are however unlikely to make their way to the inner volume of the bilayer, which can be assimilated as a hydrocarbon phase. It is much more likely that polar substrates bind the surface of the vesicle, either by electrostatic pairing or by interaction with the shallow layer constituted by the polar head and the glyceric backbone, but rarely going deeper. The crowding associated with surface binding is most conveniently appreciated by considering the average distance between molecules, a data that is, in addition, meaningful to reactivity in the usual "efficient collision" model that is commonly presented to students. In an ideal solution of concentration $\mu = \frac{N}{V}$ (note that this concentration is expressed in molecules per cubic meter, and must be further converted to moles

per liter to get practical values), this average intermolecular distance is given by the Wigner-Seitz radius:

$$\langle r_s \rangle = \left(\frac{3}{4\,\pi\mu}\right)^{\frac{1}{3}}.$$

For the same 10 µM homogenous solution described above, this radius is around 34 nm. At the surface of a bilayer, a similar formula, but expressed in two dimensions, applies:

$$\langle r_s \rangle = \left(\frac{1}{4\,\pi\rho}\right)^{\frac{1}{2}}$$

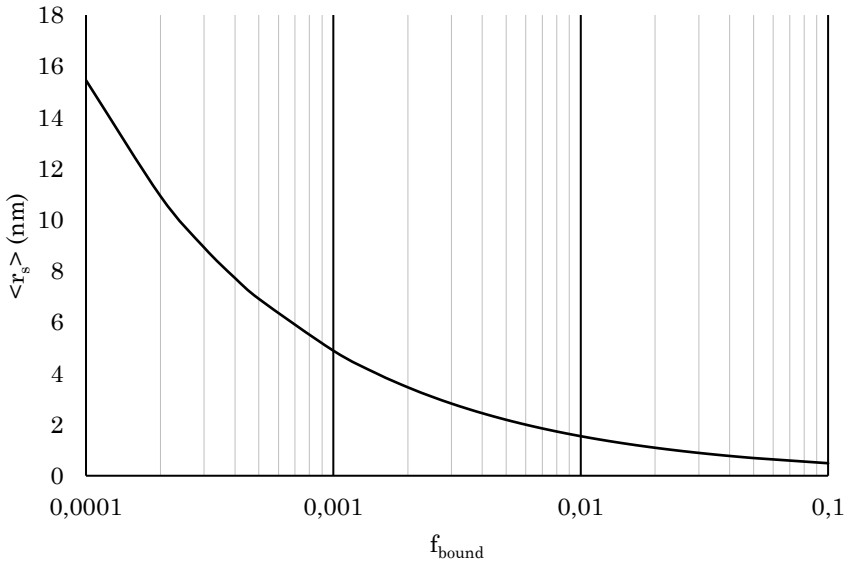
with this time $\rho = \frac{N_{local}}{S}$. Using the expressions for N_{local} already exposed above, it comes that:

$$\langle r_s \rangle = \left(\frac{S_{lipid}}{8\,\pi f_{bound}}\right)^{\frac{1}{2}}$$

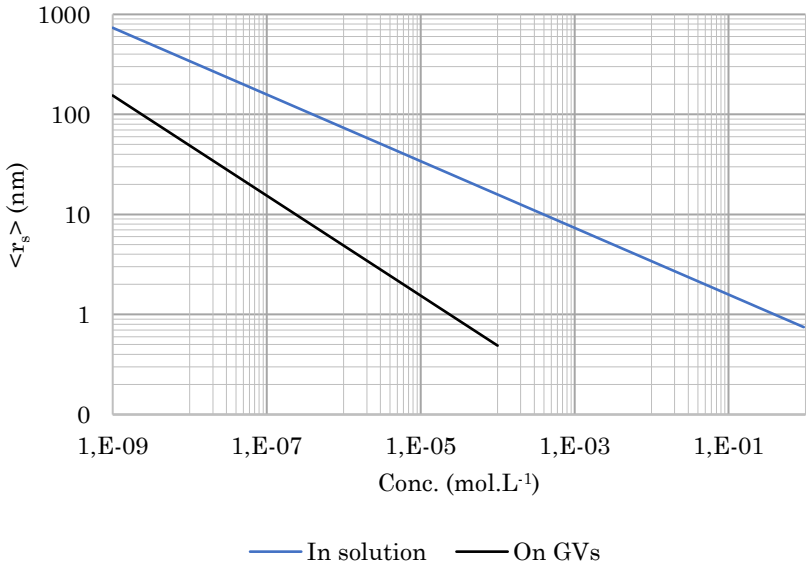
Interestingly, the size of the liposomes does not play a role anymore. This is understandable as the surface/inner volume ratio of the membranes, which is radius-dependent, is irrelevant when considering surface binding. However, these calculations assume that all the membranes are unilamellar, and therefore fully accessible (for their outer leaflet) to a solute. Calculating r_s for practical values of f_{bound} (Graph 2.3) yields a closer proximity than in solution. For example, a relatively low bound fraction of $f_{bound} = 10^{-3}$, that could be obtained with a 10 µM solution of a moderately binding substrate (only 10% bound) in 1 mM vesicles, gives an average distance between particles around 4.9 nm, which in solution would be achieved at a much higher concentration of 3.3 mM. With a strong binder at the same concentrations, one would get $f_{bound} = 0.01$ and $r_s = 1.55\,nm$, which is to be compared to a 108 mM solution. If the molecule considered is a strong membrane binder, that partitions entirely to the membrane, the local proximities obtained will almost never be outcompeted by solutions with realistic concentrations (Graph 2.4).

An increase in effective (or local) concentrations is crucial to the outcome of bimolecular reactions. Oligomerization and conjugation processes are typically well-described as bimolecular reactions between an active electrophile (potentially formed in a preliminary step) and a nucleophile. They compete with deleterious paths such as hydrolysis of the electrophile. Since water is available in virtually unlimited excess at usual working concentrations, the Ostwald principle states that hydrolysis reactions will exhibit an apparent first order, with no concentration dependence. In other terms, in excessively dilute systems, the hydrolysis will always take over, no matter how fast the activation processes are. This has two practical consequences, that can both be, in principle, prevented by membrane binding:

- in oligomerization, the longest oligomer will always be the least concentrated species and therefore small oligomers will tend to accumulate, preventing the growth of long chains (see other discussions of this idea in section 2.2)
- poorly soluble reagents are unlikely to react efficiently because of their low effective concentration



Graph 2.3 Average distance between two molecules (two-dimension Wigner-Seitz radius in nanometers) adsorbed at the surface of unilamellar liposome of any radius, in function of f_{bound} (log_{10} scale).



Graph 2.4 Wigner-Seitz radius (in nanometers) calculated as a function of substrate concentration (in $mol.L^{-1}$) in 1 mM vesicles for a free solute (blue) and a solute entirely bound to the membranes (black). Solutes with intermediate membranophilicity would result in a curve with an intermediate slope. The values are stopped at 0.1 mM for membrane-embedded substrate because it seems unlikely that more than 10% of the membrane may be loaded with a substrate without rupture or at least a drastic change in properties. A double log_{10} scale is applied.

We note that the group of Tomas proposed in 2015 an original interpretation of membrane-promoted peptide bond formation.[364] In their study, a protected lipophilic peptide (Boc-Trp-Trp-NHBzl) was formed from tryptophan benzylamide (H-Trp-NHBzl) as the nucleophile and Boc-tryptophan pyranine ester (Boc-Trp-OPyr) as the electrophile. The bimolecular reaction was in competition with the first-order hydrolysis of the electrophile. In dilute aqueous solution the hydrolysis dominated, and no coupling was detectable. When adding reagents to a vesicle suspension, the hydrolysis was reduced, and the dipeptide was formed. Pyranine was initially chosen by the authors as a convenient spectroscopic probe, but also revealed a very specific behavior: it was highly lipophilic (and therefore adsorbed to membranes) but its negatively charged sulfonates prevented it from crossing membranes, resulting in durable encapsulation in the lumen of vesicles. When the reaction was performed in the encapsulated state, the hydrolysis rate became extremely slow and the reaction extremely efficient. The authors explained this effect by the much higher lipid/water ratio at the inner layer of vesicles, especially small ones. This observation advocated that encapsulation may strengthen the effects of membrane-promoted synthesis.

The benefits of membrane embedding are not limited to the increase in local concentration. Conformations are more rigidly fixed in membranes than in solution. Ohkubo and coll. studied the hydrolysis of peptide esters that were modified to be embedded in cationic membranes made of the cationic surfactant didodecyl dimethylammonium bromide (DDAB) under catalysis by histidine-containing lipophilic peptides.[365] They found that the steric constrains imposed by the environment allowed the reaction to proceed with a stereoselectivity that was not observed in solution (Figure 2.30). Indeed, the histidine residue and the ester to be cleaved were coplanar in the *match* case, and on opposite planes in the *mismatch* case. A NOESY study confirmed the 3-dimensional arrangement.

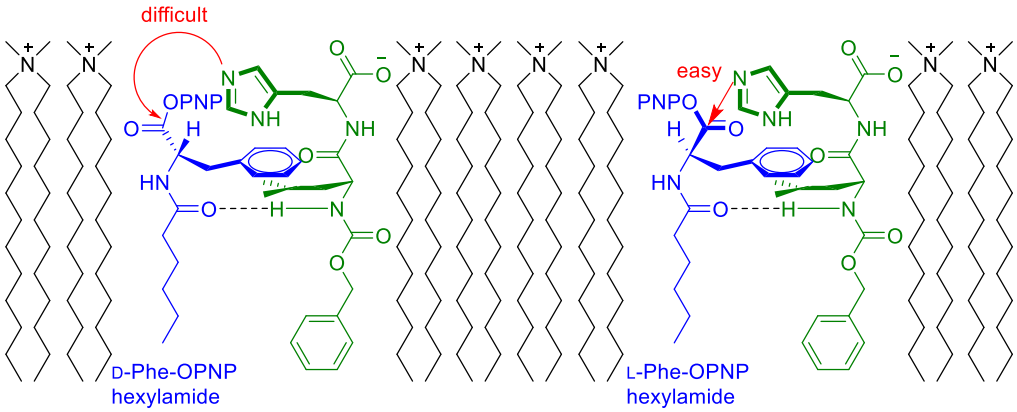


Figure 2.30 Hydrolysis of N-hexanoyl phenylalanine 4-nitrophenyl ester catalyzed by Z-Leu-His-OH in DDAB membranes was stereoselective because the ester to be cleaved in the L enantiomer pointed up, to the same side as the L-His residue of the catalyst. Figure redrawn with adaptation from the original manuscript.[365] PNP = p-nitrophenyl.

Chiral phospholipids enantioselectively adsorb amino acids. Namely, L-amino acids are selectively adsorbed at the surface of membranes of phosphocholines of the natural *sn* configuration, by a combination of hydrophobic interactions (for

hydrophobic amino acids), hydrogen bonding and dipole interactions (Figure 2.31).[366] This adsorption, sometimes up to $f_{bound} = 0.1$, alters the membrane properties. Adsorption of the hydrophobic Trp rigidifies the membrane by repelling surface hydration, while the hydrophilic His has the opposed effect.[366] L-proline is commonly employed in asymmetric aldol reactions. In a recent study, membrane-embedded L-Pro catalyzed the Michael addition of acetone to *trans* - β -nitrostyrene in a vesicle suspension with good enantioselectivity, while the reaction did not perform in pure water and displayed limited selectivity in DMSO.[367]

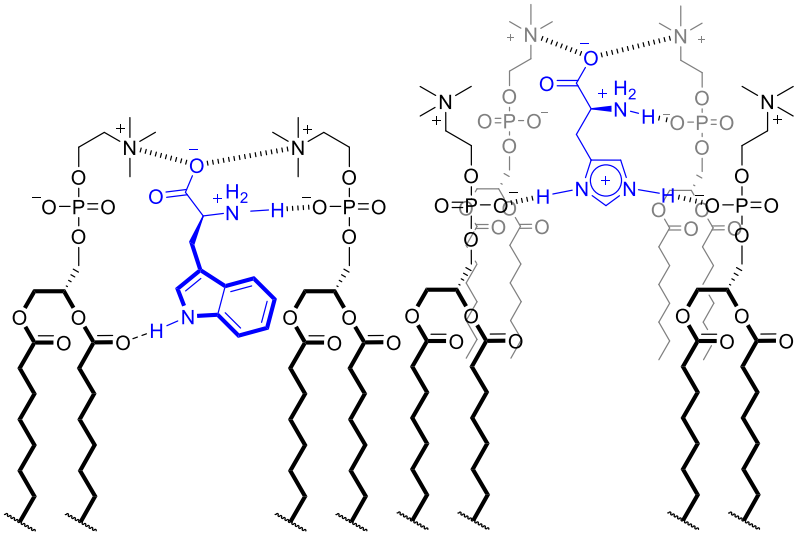


Figure 2.31 Adsorption of L-Trp (left) and L-His (right) at the surface of sn-glycero-3-phosphocholine membranes through the coordination with two of four choline heads. Adapted from the figure in the original manuscript.[366]

2.3.2.2. Analytical tools to assess membrane binding

The membranophilicity of small molecules may be assessed by two types of methods. The first is to measure the partition of the substance between the lipid phase and the aqueous phase after a separation, which is only possible when the membrane embedding is kinetically durable. The second is to observe and quantify the bound fraction in situ by spectroscopic means. This approach also delivers interesting structural information such as folding or aggregation state, or the nature of the lipid-substrate interaction but may be complicated by the heterogenous nature of the sample, which is often detrimental to spectroscopic techniques. Fortunately, monodisperse LUVs usually behave conveniently in these studies.

The methods described above (section 2.3.1.2) to separate vesicles from bulk solutes are applicable to the determination of the partition of a substance that durably resides in the membrane. Umakoshi and co-workers separated 100 nm phospholipid LUVs containing proline by ultrafiltration over membranes with a 50 kDa cutoff (equivalent to around 5 nm in pore size as calculated with

equivalence curves[368]) after 48 hours of equilibration and quantified proline in the effluent by the ninhydrin test.[367] Wimley and White used equilibrium dialysis (cutoff 12-14 kDa, 3 nm) and HPLC quantification to assess the binding of peptides to the same type of LUVs.[369] Black and coll. processed fatty acid vesicles using 3 kDa (approximately 2 nm) tangential filters in spin-tubes, and performed a fluorescamine assay on the effluent to quantify amino acids.[298] Fluorescent-labeled oligomers may be quantified without prior separation based on their fluorescence intensity inside and outside membranes. This was accomplished by Strazewski and co-workers by confocal imaging of vesicles bearing a fluorescein-labeled peptidyl-RNA.[26] Size-exclusion chromatography of LUVs was usually applied to compounds encapsulated in the lumen, but not bound to the membrane, although this is in principle feasible. Tomas and coworkers noted in their study involving pyranine peptides that membrane binding that lipophilic unencapsulated substrates tended to be carried through the gel by the vesicles and, after leaking back to the bulk, resulted on a false positive background signal.[364]

The case of peptides has been particularly well-studied, notably because of the importance of transmembrane enzymes and cell-penetrating peptides in biology and biotechnical applications. Wimley and White dedicated a lot of work to establishing meaningful empirical membranophilicity scales for amino acid residues within short peptide sequences.[335] Older methods relied on indirect determinations, such as the partition of amino acid side chains between water and an organic solvent (octanol or cyclohexane), a measurement that was poorly pertinent to the reality of membranophilicity. All peptides possess an amide backbone that is extremely polar and prevents virtually any penetration of the peptide in the carbon chain part of the lipid bilayer.[370] The charges on the carboxylate and the ammonium termini further reinforce this effect. Therefore, peptides are only able to enter the headgroup region. Further, random-coiled peptides must be very lipophilic, and hence rather long, to display favorable partitioning in membranes. On the other hand, Wimley and White found that simple short peptides were able to form β-sheets in membranes to a much greater extent than in buffer. The protected hexapeptide Ac-Trp-Leu-Leu-X-Leu-Leu-OH (where X was the residue to be studied) formed β-barrels in POPC membranes at pH 7 when X was a nonpolar residue, with the exception of glycine because of its specific flexibility and proline because of its kink (Table 2.1).[371] At pH 2.5, the protonation of the carboxylic terminus considerably facilitated the membrane penetration and the formation of membrane-embedded β-barrels. The aggregation state was assessed by circular dichroism using the tryptophan residue as a spectroscopic probe.

Residue	$\Delta G_{H_2O \rightarrow POPC}$ (kcal.mol^{-1}) at pH 7	f_{bound} at pH 7	f_{bound} at pH 2.5
Arg	-4.5	0.03	0.59
Lys	-4.3	0.03	0.52
Pro	-4.4	0.03	0.73
Ala	-4.6	0.04	0.85
His	-4.6	0.04	0.53
Ser	-4.7	0.05	0.85
Val	-4.7	0.05	0.84
Gly	-4.8	0.06	0.79
Cys	-5.1	0.09	0.9
Leu	-5.4	0.14	0.93
Phe	-5.9	0.29	0.98
Trp	-6.7	0.58	0.99

Table 2.1 Membranophilicity (POPC LUVs against buffer) of hexapeptides Ac-Trp-Leu-Leu-X-Leu-Leu-OH for a selection of X residues as determined by Wimley and White. The color of the cell reflects the membranophilicity (from blue, membranophobic, to green, membranophilic).

NMR spectroscopy is now among the most frequently used techniques to access structural and conformational information. An efficient way to detect non-covalent interactions between two structures is to measure diffusion coefficients. The diffusion coefficient of a small molecule in solution is the reflect of the quadratic average if its free path due to Brownian movement (see a more complete discussion in section 3.2.4). Diffusion NMR is now well-established[285] thanks to the routine availability of probes with gradient units and high-sensitivity spectrometers, and has become a standard technique in supramolecular chemistry.[286] Small molecules typically diffuse quickly in water but are considerably slowed down by adherence to lipid membranes. These two regimes can be detected by diffusion measurements and the binding extent of a substrate extracted from it. More subtle changes in folding or conformation can also be detected by smaller modifications of the diffusion rate.[372] This procedure is applicable to a wide range of analytes from amino acids[298] to macromolecules such as polymers.[373] Multidimensional NMR has also been exploited to give additional information on the binding mechanisms of peptides into membranes. The insertion of the protected tripeptide H-Ala-Phe-Ala-OtBu in sonicated DMPC vesicles was determined by nuclear Overhauser effect NMR spectroscopy (NOESY) and the data obtained injected into a balls-and-stick model (Figure 2.32). The NOE revealed interactions between the acyl chains and the hydrophobic phenyl ring of Phe, as well as several headgroup interactions, and a degree of conformational rigidity of the peptide that was absent in solution.[374]

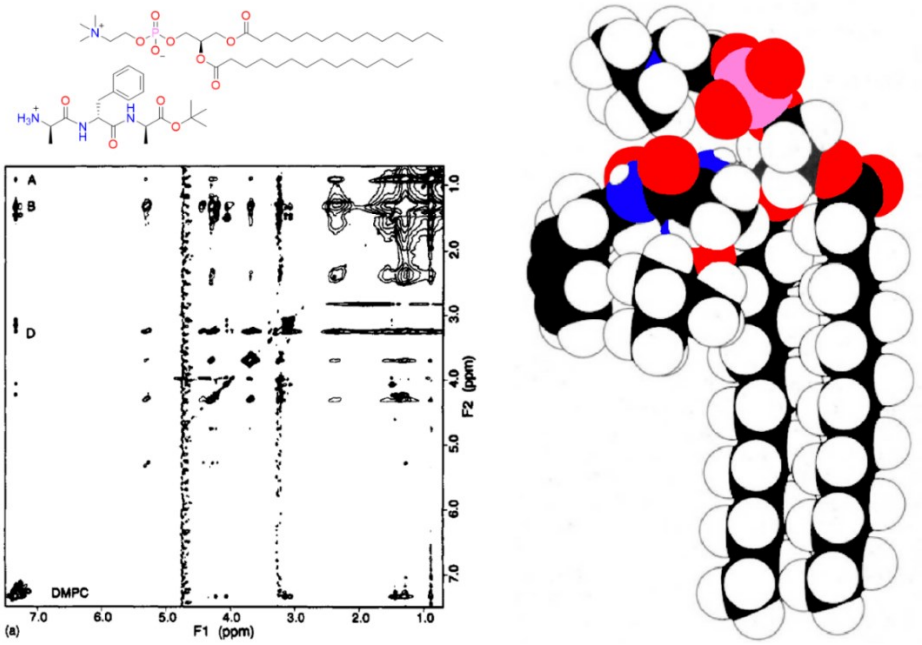


Figure 2.32 Structure of the complex between the protected peptide H-Ala-Phe-Ala-OtBu and DMPC membrane as elucidated by NOESY NMR studies. Top left: molecular structure. Bottom left: 2D-NOESY spectrum. Right: ball structure obtained from a molecular dynamics simulation constrained with the NOE data. Spectrum and ball model reproduced from the original manuscript[374] (colored from the original black and white model) under license with the Editor.

Similar methods were applied to the membranophilic nonapeptide bradykinin (H-Arg-Pro-Pro-Gly-Phe-Ser-Pro-Phe-Arg-OH) in various phospholipid membranes.[375] Line broadening revealed that bradykinin preferentially bound negatively charged membranes containing DOPA, which was expected from the presence of two cationic Arg residues in the peptide sequence.

The addition of paramagnetic salts ($MnSO_4$ or today more commonly Eu^{3+} complexes) to liposome suspensions induces a faster T_2 relaxation of spin systems outside of the vesicles that may be either measured, typically by spin-echo,[376] or assessed indirectly by its effect on peak broadening.[377] This technique notably allows for distinguishing the inner layer from the outer one in PC membranes thanks to the splitting of the NMe_3 signal of the choline headgroup, as well as the determination of liposome binding and encapsulation. Lipid aggregates such as bicelles (micelles prepared from a specific mixture of long- and short-chain lipids) were specifically designed to be easily implemented in liquid-state NMR but their structure differs a lot from native membranes.[378]

Rotational echo double resonance (REDOR), a technique issued from solid-state NMR, also enables the determination of the spatial proximity between heteronuclei. It is applicable to frozen vesicle samples. Peptide-membrane interactions are typically probed between ^{13}CO-labeled peptides, and either the ^{31}P nuclei of phospholipids, or ^{2}H-labeled positions. Labeling specific sites on the lipid yields spectroscopic "handles" at various membrane depths (Figure 2.33), allowing to finely study the position of the peptide within the bilayer.[379] While diffusion and relaxation studies are now relatively amenable in conventional NMR facilities, REDOR requires a high-resolution solid-state spectrometer with

an appropriate triple-channel probe able to keep the sample frozen, and the acquisitions are typically long, which still limits its use to specialists.

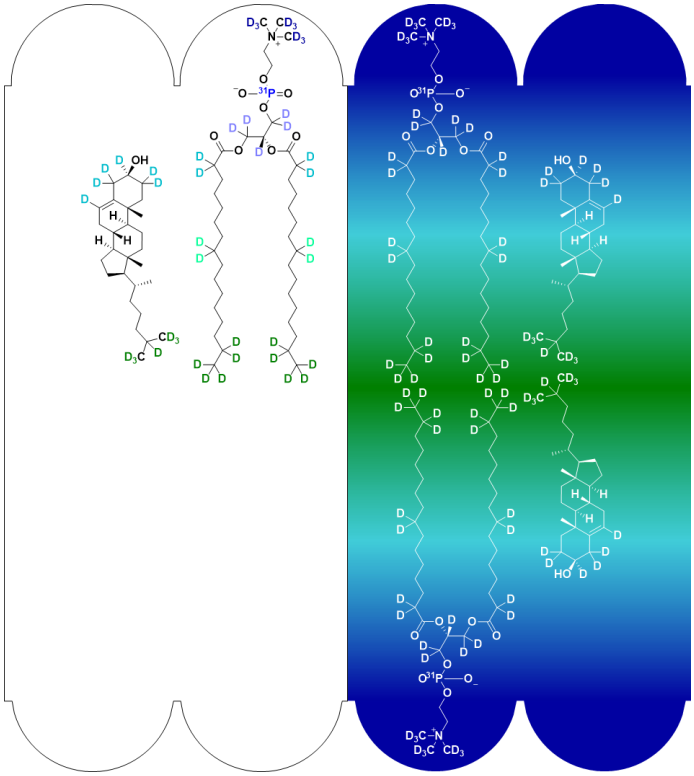


Figure 2.33 Typical deuterated phospholipids and cholesterol used in localization studies by REDOR NMR. The availability of various deuterated positions allows to probe the whole membrane depth (here represented by a color gradient from green core to blue head zone).

2.3.2.1. Oligomerization reactions promoted by lipid membranes

Chemical reactions promoted by lipid membranes have been the subject of a vast body of literature, recently reviewed by others.[380] Among these studies, only a fraction is relevant to the processes of early life. In particular, many experiments involving cationic surfactants are in contradiction with the common assumption that most, if not all, lipids of life were neutral, zwitterionic or anionic. Moreover, the reaction substrate(s) must necessarily interact with lipid membranes in order to see the course of the chemical reactions influenced by it. While this is usually not a problem in organic synthesis because most substrates are highly lipophilic and poorly soluble in water, it drastically narrows down the range of biotically relevant reactions potentially affected by lipids. Studies involving chiral catalysts or lipids interestingly sit in the middle of the line: prebiotic chemicals must have first occurred in a racemic form but the pathways to their stereochemical enrichment are highly relevant.

Nucleotides are not membranophilic. The only reports of lipid-enhanced production of oligonucleotides therefore relied on their entrapment between lipid layers in the course of hydration/dehydration cycles in an acidic medium. Under

these conditions, small oligonucleotides were obtained (significant yields up to 4-mers as quantified by SAX HPLC and minute amounts of longer sequences only detectable by very sensitive methods), accompanied with degradation reactions.[381,382]

Peptides were more likely to have benefited from membrane-promoted oligomerizations. Luisi, Walde and coll. studied the formation of the very hydrophobic oligotryptophanes in the presence of POPC vesicles, either starting from Trp-NCA or Trp_2 with the hydrophobic condensing agent EEDQ.[383] In water, the oligomerization was rapidly inhibited by the precipitation of the growing oligomers. POPC vesicles were able to host the growth of chains up to 23 units as determined by HPLC. As discussed in greater detail in the dedicated section 2.2.3, HPLC does not usually yield false positive signals and only reveals products that are present in a meaningful concentration. Detecting chains longer than 20 amino acids by HPLC is a highly significant result. However, the overall increase in yield, including all peptide lengths, was not as spectacular as one could have expected. Non-lipophilic amino acids were also eligible to this membrane-promoted elongation when ion-pairing interactions would be established. This was demonstrated by the authors in the case of glutamic acid with cationic liposomes containing DDAB, and arginine and histidine with anionic liposomes containing DOPA.[159] The stereochemistry of the reaction was also investigated by LC/MS studies of reactions involving one deuterated enantiomer, but no significant chiral induction by the chirality of the phospholipids could be noted.[384] This was an echo to a previous publication of the group that showed that chiral fatty acids ((*R*)- or (*S*)-2-methyl dodecanoic acid) did not yield enantio-enriched vesicles at room temperature (they did at 10 °C however).[385]

Pascal and Ruiz-Mirazo recently described the formation of a lipophilic dipeptide from protected amino acids through oxazolone chemistry in fatty acid membranes.[386] In their study, the oxazolone of *N*-acetylated *O*-methyltyrosine reacted with leucinamide while adsorbed on the membrane to give Ac-Tyr(Me)-Leu-NH_2, a product that was almost insoluble in water. The reaction was not promoted by phospholipids or long-chain ammonium or sulfate surfactants, but general acid catalysis in organic solvents with acetic or decanoic acid was efficient. This suggested that the carboxylic group of the fatty acid membranes played a catalytic role in the reaction, in addition to the colocalization.

2.3.3. Overview

The case of lipids illustrates well how prebiotic systems chemistry goes beyond the geologically plausible chemical synthesis of molecular bricks. Self-assembly, and early pathways to autonomous behaviors may be more important than the precise nature of the components in a self-replicating evolving system. Lipid boundaries, in particular giant vesicles, may play the multiple roles required to unify building blocks into a system. Such vesicles are capable of gathering macromolecules into a specific environment, prevent them from virtually

unlimited dilution, protect them against degradative pathways and provide them with a micro-environment that differs from the external medium. They can grow and self-replicate. The transition between early and more elaborate amphiphiles reflects well the physico-chemical challenges that such systems must have faced.

Lipid boundaries may have acted as a lipophilic environment in a world dominated by water and the dry state. Reactions that do not perform efficiently in water because of the poor solubility of the substrates or because of dilution may benefit from membrane binding, although the experimental demonstrations directly related to the chemistry of oligomers are still limited in scope and less efficient than what may be expected from geometrical predictions. Finally, studying systems embedded in lipid membranes, that are by nature heterogenous, is an analytical challenge that is progressively taken over by the refinement of tools, some of which were known for a long time but are just becoming accessible to a larger public.

2.4. Nucleotide phosphoramidates: synthesis and applications

Studies in basic and applied sciences generate a constant demand for chemicals, materials, and molecules with specific structures and features. When natural sources are unable to fulfill the requirements of the research in question, scientists turn themselves to chemical synthesis.[387] Peptide-RNA interactions are well-known, notably in the case of translation and strategies to covalently bind peptides and oligonucleotides were developed accordingly. Here we will only review the formation of (oligo)nucleotide phosphoramidates that are central to the investigations of this thesis work.

Phosphoramide is a relatively delicate chemical function characterized by a phosphorus (V) center featuring one P-N bond involved in conjugation similarly to amides. Compounds with two/three P-N bonds are called phosphoro(di/tri)amides. It is found in high-energy natural compounds, notably phosphocreatine. Phosphoramides are sensitive to acid hydrolysis and the amine moiety can slowly act as a leaving group for oxygenous nucleophiles. This is expected by comparison with their well-known P(III) relatives, phosphoramidites, that serve as electrophilic phosphorus centers under weak acid catalysis, especially in oligonucleotide synthesis (Figure 2.34).

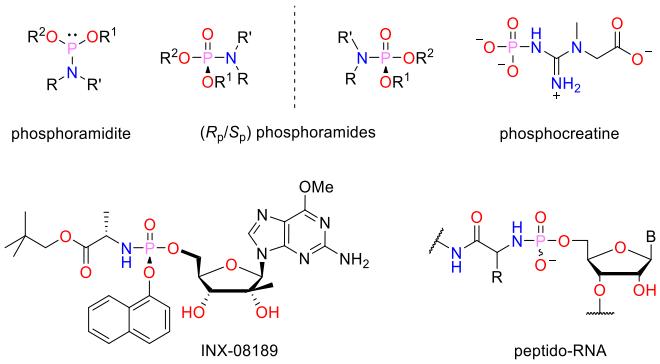


Figure 2.34 Structures of phosphoramidites, phosphoramides (two enantiomers), phosphocreatine, INX-08189 (an example of nucleotide phosphoramide antiviral prodrug) and peptido-RNA (generic form).

Phosphoramides resist basic conditions relatively well. They are conveniently detected by ^{31}P NMR (typically between 6 and 8 ppm for alkyl species), sometimes as two distinguishable rotamers when originated from tertiary amines (see many examples in the results of this work). When not uniformly substituted, phosphoramides are chiral, and the S_p and R_p isomers may be isolated when the rotations around the phosphorus atom are sufficiently slow. In phosphoramidates, that feature an unsubstituted hydroxy residue, the exchange is too rapid to discriminate the enantiomers. In mass spectrometry, the low mass difference between O and NH (only 1 Da) groups calls for high-resolution techniques and careful calibration.

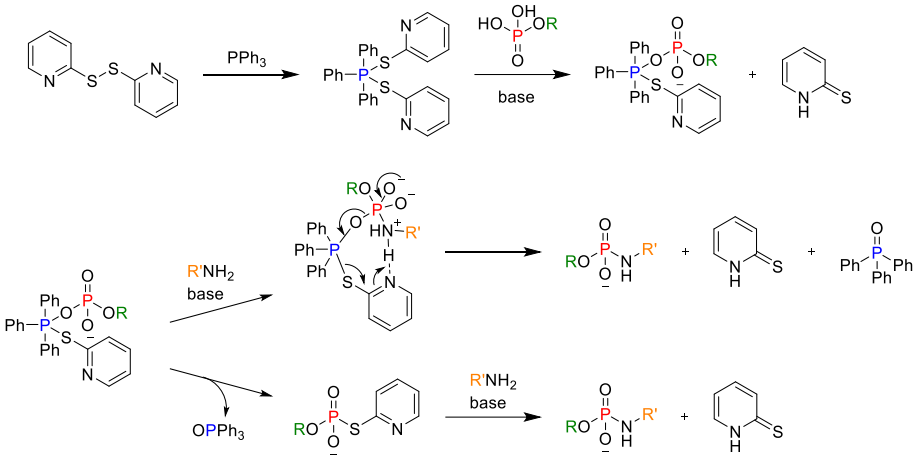
Peptide-RNA conjugates connected through a phosphoramidate have first been dubbed "peptidyl RNA" by analogy with aminoacyl RNA chemistry. The recent intensive investigations by the Richert group raised the concern of a more appropriate designation, and "peptido-RNA" was chosen to account for the phosphoramido group, in contrast with the acyl connection that exists in peptidyl RNA.[31] No standard sequence nomenclature has been officially established to date. We used Richert's notation as a basis to develop a system as clear and convenient as we could, that is fully explicated in the appendices. Note that because of the *N*→5' connection, either the RNA or the peptide strand must be enumerated in an order opposite to the usual convention. We chose to read the sequence from the peptide moiety to the nucleotide moiety, in adequation with the usual representation of nucleotides with the base on the right and the 5' terminus on the left, and therefore enumerating the peptide from the *C* to the *N* terminus.

The synthesis of peptido-RNA has been typically carried out on small scales, notably because oligonucleotides are expensive and are not conveniently prepared in large scale in the laboratory. In most cases, a 5'-phosphorylated oligomer was synthesized first then coupled with a peptide, either directly or by isolating an active phosphoester intermediate. Peptido-nucleotides and compounds alike, some of which are of pharmaceutical interest, can be produced on larger scales by a wider range of techniques.

2.4.1. Synthesis of peptido-RNA by direct coupling

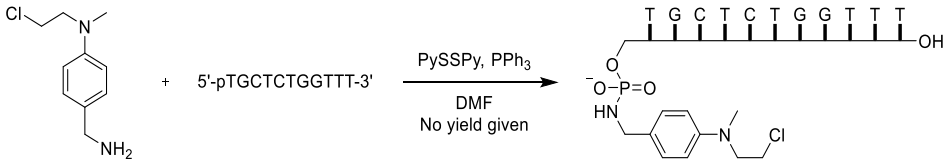
The literature on peptide-RNA conjugation through phosphoramidates is relatively scarce and its study is further complicated by archiving issues. In particular, a significant body of publications, including several contributions from the Zarytova group at Novosibirsk University, was published in USSR journals before 1990 that were never translated in English (or only in private collections). Phosphoramidates have attracted much less interest than conventional conjugation methods such as phosphodiesters or nucleobase decoration through linkers, possibly because of their limited stability.

Many conventional amide coupling mixtures (including carbodiimides with imidazole, HOBt and related) lead to more or less stable intermediates when applied to phosphates (see next section). To our knowledge, all the direct couplings between peptides and 5'-phosphorylated oligonucleotides were performed using a redox coupling system initially developed by Mukaiyama for peptide coupling, a domain in which it remained scarcely used.[388] The reaction relies on mixing an oxidant and a reducing agent that cannot accomplish their redox reaction without the presence of water (Scheme 2.33). Among the many reagent pairs proposed in the original paper, many of which involved toxic mercury salts, 2,2'-pyridyl disulfide and triphenylphosphine have been universally adopted. Originally, a concerted mechanism was postulated for the carboxylate activation (transcribed to phosphates in Scheme 2.33). In the case of phosphate activation, an active thiophosphate intermediate has been observed by ^{31}P NMR (Scheme 2.33).



Scheme 2.33 Mechanistic propositions for the PySSPy/PPh₃ coupling of amines and phosphates: concerted addition or thiophosphate intermediate.

This method was applied in the Zarytova group to the decoration of DNA strands with an alkylating group (Scheme 2.34). Sequence complementarity allowed to alkylate specific positions of a complementary strand.[389] Sumbatyan and co-workers compared the efficiency of direct coupling against active intermediate isolation in the case of antisense RNA-lipophilic peptide conjugation.[390] They found that, under the conditions tested, direct coupling promoted by PySSPy and PPh_3 in DMF or DMF/DMSO was superior. They enhanced the solubility of the oligonucleotide by using cetyl trimethylammonium bromide (CTAB) as a counterion and used a large excess of peptide (at least 10 equivalents). The products were purified by denaturing gel electrophoresis, probably because of the aggregating nature of the peptides used.

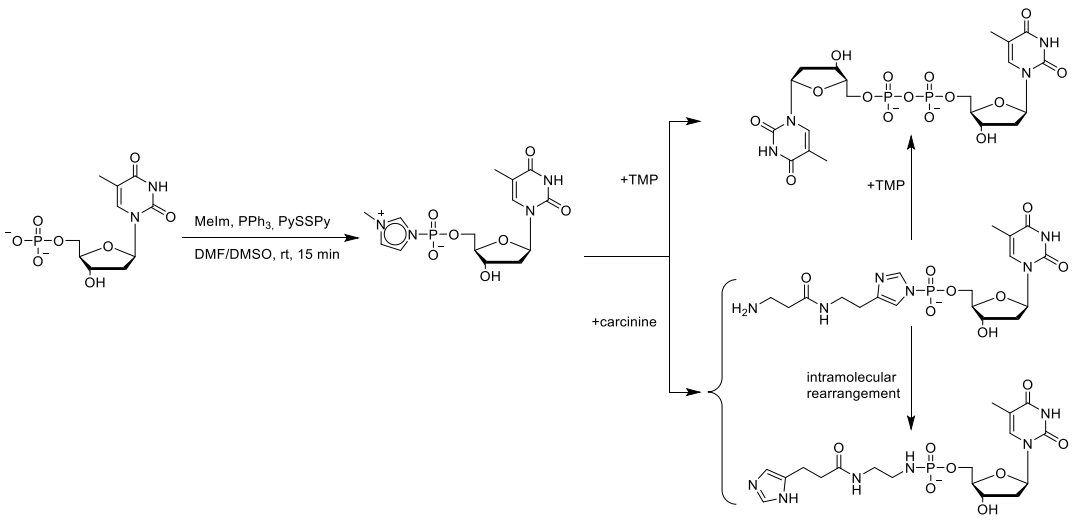


Scheme 2.34 Coupling of a DNA 11-mer to an alkylating group featuring a 2-chloroethylamine function.

2.4.2. Synthesis of peptido-RNA with a pre-activation step

More commonly, 5'-phosphorylated oligomers are pre-activated, isolated, and coupled in a subsequent step with the peptide or the amine. The leaving groups investigated for phosphate activation include HOBt, 1-methylimidazole and 4-dimethylaminopyridine-1-oxide (DMAPO). In the study mentioned just before, Sumbatyan and co-workers activated 5'-phosphorylated oligonucleotides with HOBt and EDC in water/DMF. They then isolated the intermediate OBt phosphoester by precipitation and coupled it to various peptides in aqueous 1-methylimidazole buffers with varying proportions of DMF as a co-solvent. For medium-length peptides (10-15 mers), the yields were modest when compared to the one-pot coupling.[390]

Garipova and Silnikov studied in detail the reaction of thymidine 5'-monophosphate (TMP, in the form of a bis(triethylammonium) salt) with carcinine through an in-situ activation step.[391] They used an excess of 1-methylimidazole in DMF/DMSO and PySSPy/PPh_3 to form the methylimidazolium of TMP then added a solution of carcinine. They found that observing specific reaction times was important. In particular, a lengthly activation of TMP led to the massive formation of TppT. Carcinine also possesses an imidazole ring that can serve as a platform for further transformation (Scheme 2.35).



Scheme 2.35 Synthesis of TMP-carcinine conjugates and multiple rearrangement paths.

Zarytova and co-workers coupled oligonucleotides to the copper(II) complex of bleomycin A5 (Figure 2.35), a highly elaborate molecule known for its antibiotic properties, and that naturally features an alkylamino linker.[392] They pre-activated the oligonucleotide 5'-phosphate (CTAB salt) with DMAPO using the PySSPy/PPh_3 couple at high dilution (2.5 μM) in DMSO and precipitated it as a lithium salt by employing a solution of lithium perchlorate in acetone. The conjugation step was then carried out in a sodium bicarbonate buffer at pH 8.5 with 10 equivalents of bleomycin. They obtained conjugates with diverse DNA sequences in 60-80% yields after HPLC purification.

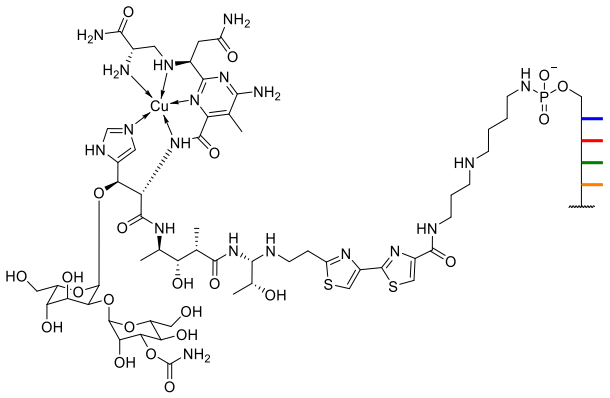
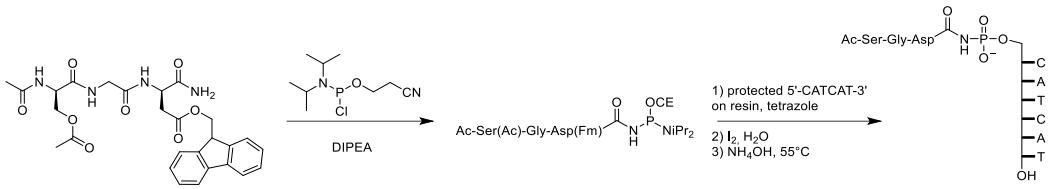


Figure 2.35 Structure of conjugates between oligonucleotides and the Cu^{2+} complex of bleomycin A5.

Grandas and co-workers provided an original approach by preparing *N*-acylphosphoramidates, connected to the amidated *C*-terminus of the peptide, rather than the conventional phosphoramidates.[393] They synthesized protected peptides in the amide form, then phosphitylated the amide terminus with a conventional chlorophosphoramidite reagent and thus obtained an active acylphosphate species. The coupling to a solid support-bound oligonucleotide 5'-end was then carried out under conditions similar to RNA synthesis: tetrazole-promoted coupling followed by oxidation. The conjugate was finally cleaved and deprotected under typical RNA conditions and purified by HPLC with yields between 22 and 26 %, which was very good considering that it took the RNA synthesis into account (Scheme 2.36).



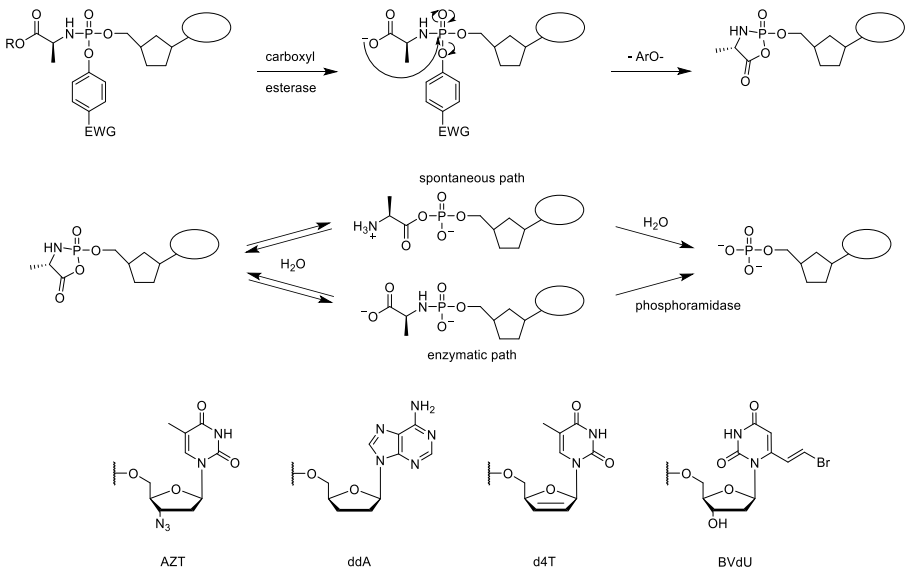
Scheme 2.36 Synthesis of acylphosphoramidate conjugates through peptide amide N-phosphitylation and coupling with resin-bound RNA.

2.4.3. Synthesis of small phosphoramidates

Peptido-nucleotides, composed of a single nucleotide residue and small peptides or even single amino acids, can be prepared on a larger scale because the expensive RNA synthesis is not necessary. Strategies similar to those depicted above have been used by Richert and co-workers to prepare small peptido-nucleotides, either by displacement of an OAt-activated nucleotide or by EDC-promoted coupling of an amino acid methyl ester with the nucleotide, followed by saponification.[28] These strategies were attractive because of their straightforward aspect and the avoidance of protective groups. In both cases however, the purification proved difficult. We will detail this issue in the results and discussion section.

The discovery of the ProTide pro-drug strategy by the McGuigan group in the 1990s gave a significant kickstart to the synthesis of numerous phosphoramides of nucleotide analogs that found many therapeutic applications, particularly against HIV.[394,395] Nucleoside drugs such as AZT, d4T and many others (Scheme 2.37) are not directly active in cellulo in the nucleoside form and must be converted to triphosphates. The phosphorylation steps required to form the triphosphates, especially the first one, are responsible for the relatively poor efficiency of the drugs in vivo. Administering monophosphates may in principle solve this problem, but these do not permeate membranes.[395] After early attempts to use phosphotriesters as prodrugs, trisubstituted phosphoramides bearing an esterified amino acid were found to be ideal candidates. Among the many possible variations, it was found that alanine esters (preferentially with bulky chains such as isopropyl or neopentyl, but not tert-butyl) were by far the most active, that β-

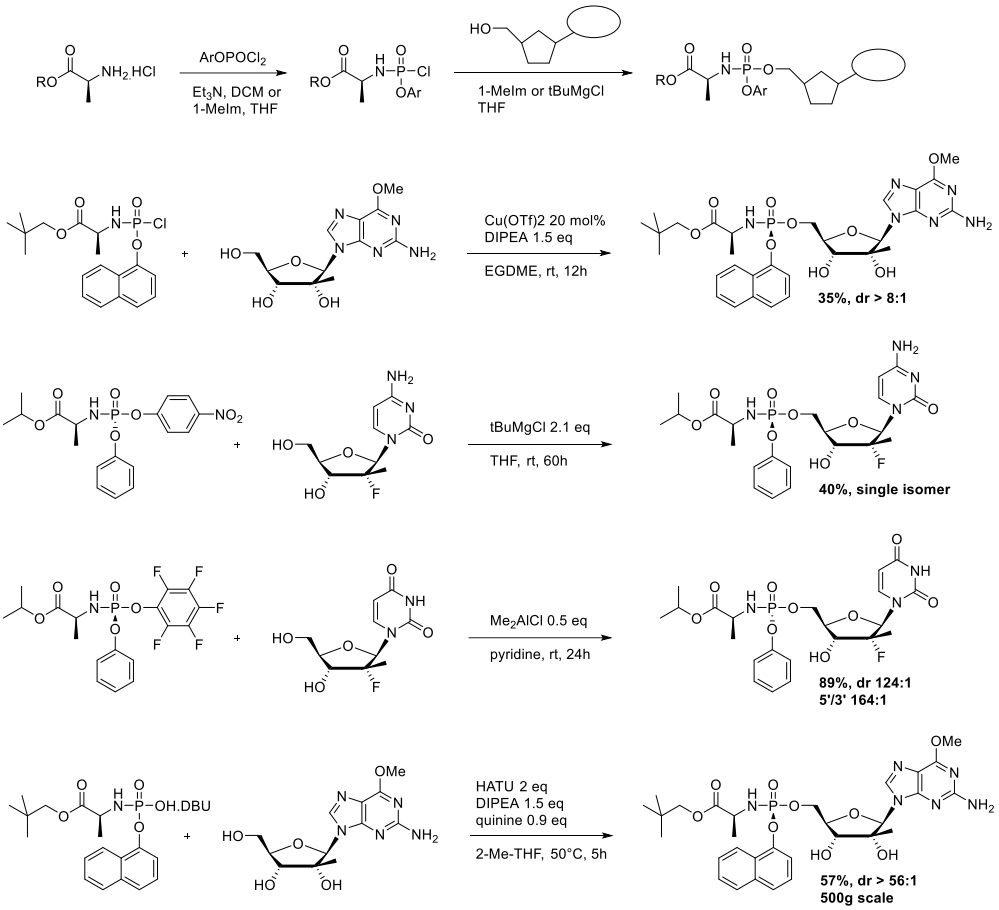
and γ-amino acids were poorly active and that electron-poor aryloxy groups were the most suitable.[394] All of these points are well-understandable when considering the mechanisms involved in the release of the nucleotide.[396] Indeed, the presence of an α-carboxylic acid has a dramatic influence on the hydrolysis of the phosphoramide, that proceeds through a CAPA (Scheme 2.37, top). The hydrolysis may be spontaneous or, in cells, be promoted by phosphoramidate-specific hydrolases. A very large range of nucleoside analogs were then conjugated to alanine esters following this ProTide approach (a few examples are represented in Scheme 2.37, bottom). Other amino acids (especially with branched side chains such as Val, Leu or Ile) were less desirable because of their lack of sensitivity to enzymatic ester hydrolysis.[395]



Scheme 2.37 Top: chemical and enzymatic degradation of ProTide into active nucleotide drugs through a CAPA intermediate. Bottom: a few examples of nucleotide analogs that are in current use or were tested for therapeutic applications. AZT: 3'-azido-3'-deoxythymidine; ddA: 2'-3'-dideoxyadenosine; d4T: 2'-3'-dideoxydyhydrothymidine; BVdU: 5-(E)-bromovinyl deoxyuridine.

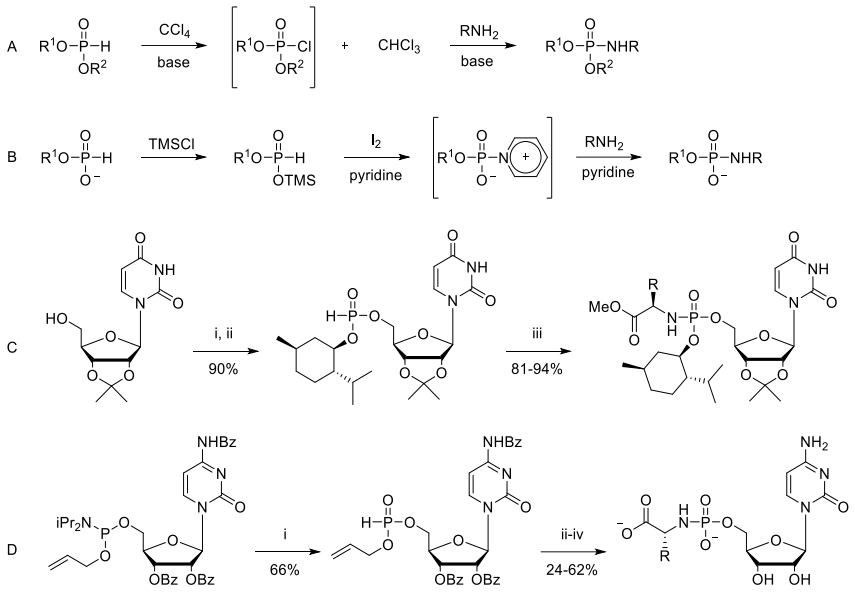
The early syntheses of ProTides followed a straightforward and flexible strategy relying on chlorophosphates in organic solvents (Scheme 2.38).[397–399] However, more recent clinical trials revealed that only the R_p diastereoisomers were active in vivo, motivating the search for stereoselective syntheses, and particularly phosphorylation conditions able to preserve the chirality of an active phosphoester. The McGuigan group proposed a modification of the classical conditions employing copper (II) salts as catalysts in THF or glyme.[400] In most cases however, phosphorylating agents milder than chlorophosphates were used. Ross and co-workers used 4-nitrophenoxy phosphoramidates that were first prepared in enantiopure form by either seeded crystallization or chiral chromatography with supercritical CO_2.[401] Another similar approach was taken recently with pentafluorophenol as the leaving group and Lewis acid activation of the enantiopure phosphate.[402] To avoid the need for an enantiopure phosphorylating agent, a research group from Brystol-Myers Squibb used dynamic kinetic resolution.[403] The prepared in situ the OAt ester of the

aryolxyphosphoramidate using HATU, that was able to interconvert easily. Addition of quinine as a chiral ancillary led to the major formation of the desired isomer (Scheme 2.38).



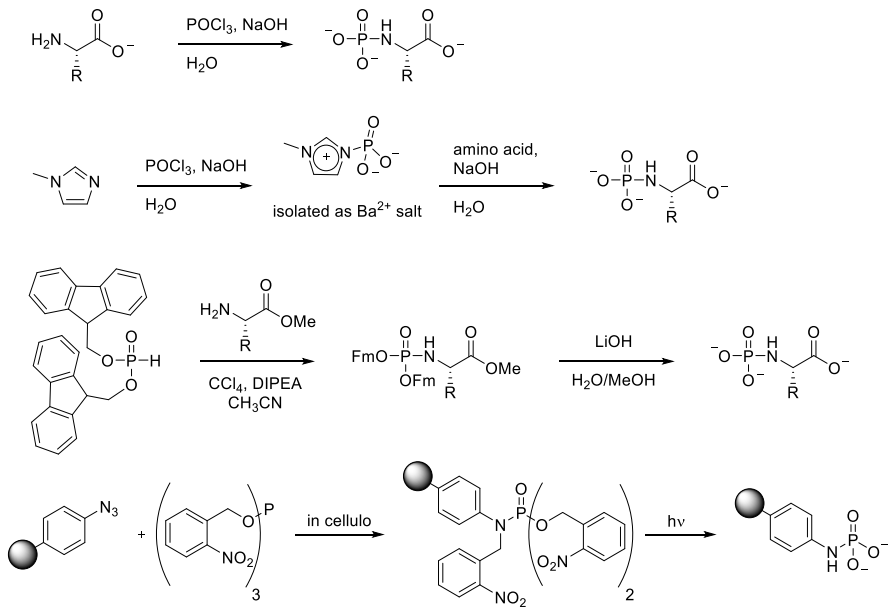
Scheme 2.38 Synthetic strategies to obtain phosphoramide prodrugs. Top: general non-diastereoselective strategy. Bottom: various diastereoselective strategies.

Relatively few studies have tackled the preparation of phosphoramides and phosphoramidates outside of the ProTide-related studies, maybe because the function is uncommon in natural products. The first historical example is now known as the Atherton-Todd reaction, an oxidative coupling between a *H*-phosphonate and an amine, usually mediated by carbon tetrachloride.[404,405] This strategy has been applied to the synthesis of phosphoramidate-linked analogues of oligodeoxynucleotides.[406,407] A variant, using iodine as the oxidant and proceeding through a pyridinium intermediate, was also used for the same purpose.[408] The intermediate phosphonate may be prepared by various strategies, using PCl_3[409] or amidites[410] as intermediates (Scheme 2.39).



Scheme 2.39 Atherton-Todd inspired strategies to prepare phosphoramides and phosphoramidates. A: the classical Atherton-Todd reaction. B: a variation used in the synthesis of phosphoramidate analogues of nucleotides. C: Synthesis of uridine phosphoramidates through an Atherton-Todd coupling. i: PCl_3 10 eq, Et_3N, CH_2Cl_2, 0°C, 3h. ii: menthol 1 eq, tBuOH 1 eq, Et_3N 2 eq, CH_2Cl_2, 10 min. iii: amino acid methyl ester 1.3 eq, Et_3N, CCl_4, water, CH_3CN, rt, 10 min. D: synthesis of cytidine phosphoramidates through an Atherton-Todd coupling. i: water, tetrazole, CH_3CN, rt. ii: amino acid methyl ester, CCl_4, Et_3N, CH_2Cl_2, rt. iii: $Pd(PPh_3)_4$, PPh_3, morpholine, CH_2Cl_2, rt. iv: MeONa, MeOH, NaOH, rt.

Finally, let us briefly discuss the *N*-phosphorylation of amines. Free phosphoramidates are hydrolytically stable only under basic conditions and degrade quickly in acidic or neutral aqueous solution. They are typically prepared by reacting aqueous amines with electrophilic phosphates, while keeping the pH high. This includes the slow addition of $POCl_3$, along with NaOH to compensate the release of HCl,[411] and the preparation of 1-methyl-3-phosphorylimidazolium chlorides, from $POCl_3$ and 1-methylimidazole, that can be stored as various salt forms.[412] The Atherton-Todd method described above has also been used with *H*-phosphonates that could be easily be deprotected (bis(fluorenemethyl)phosphite for example).[413] An elegant green strategy was recently described,[114] inspired by prebiotic phosphorylation studies.[115] It used cTMP in warm water that was maintained at pH 11 by automated titration. Crystallization was used to separate the main product from inorganic pyrophosphate (PPi) that was generated as a side-product. PPi could be reconverted to cTMP in very good yields by dry-heating, thus compensating the poor atom-economy of cTMP. All of these methodologies have been mainly exemplified with amino acids (Scheme 2.40). In a now seminal paper, Bertozzi and co-workers reported the chemoselective in cellulo modification of proteins bearing azide groups by Staudinger ligation.[414] Hackenberger and coll. adapted this reaction to photosensitive phosphites, that operated a similar click-like *N*-phosphorylation of azide-bearing proteins (Scheme 2.40, bottom).[415] The strategy was also applicable to synthesis.[416]



Scheme 2.40 Synthetic access to O,O' unsubstituted phosphoramidates. From top to bottom: direct $POCl_3$ phosphorylation. Use of imidazolium phosphate intermediates. Atherton-Todd reaction with protected H-phosphonates. Staudinger-phosphite reaction that could be performed in cellulo.

2.4.4. Overview

The synthesis of phosphoramides and phosphoramidates is essentially articulated around a small selection of relatively robust reactions. Some are relatively specific (Atherton-Todd, Staudinger-type ligations), while more general strategies for coupling and leaving group displacement are efficient and selective thanks to the high nucleophilicity of the amines. When applied to small molecules, conventional techniques involving protective groups, strong activation and anhydrous conditions are most common. In contrast, the formation of peptido-RNA is preferentially conducted under mild aqueous conditions, although strategies involving protected or supported oligomers are also in use. A bridge between these two paradigms is essentially missing: how to prepare small conjugates in efficient ways under mild activation and without (or with a minimum of) protective groups?

2.5. Objectives of this work

This thesis work takes part in a five-year collaborative research project with the Richert group (Universität Stuttgart, Germany) and the Egli group (Vanderbilt University, Nashville, TN, USA). The project, funded by the Volkswagen foundation under the code name "Molecular life", aims at better understanding how small molecules can make an efficient use of chemical fuel to produce growing oligomers, and how these integrate in self-assembled systems. The expected contribution of our group in this collaboration lies in the use of lipid assemblies (vesicles) to influence reaction networks and produce encapsulated systems.

The Richert group developed a set of "general condensation conditions" under which RNA and peptides grow simultaneously, essentially as phosphoramidate-linked co-oligomers: peptido-RNA. These conditions involve high concentrations of monomers and a carbodiimide coupling agent in a near-neutral buffer, a system that they named the "general condensation buffer" to reflect its ability to produce several different useful chemical bonds. A first step in the transition of this peptide-RNA world into a peptide-RNA-lipid world is cross-compatibility: do peptido-RNA efficiently form in the presence of lipids? Can amphiphilic conjugates be synthesized under similar activation conditions, and do lipids assemble into stable membranes in the general condensation buffer? Meeting these conditions would establish the plausibility of a highly complete peptide-RNA-lipid world relying on a single source of chemical activation, given the wide acceptation that most lipid synthesis steps can be performed under carbodiimide activation.

One attractive prospect of peptido-RNA is the formation of amphiphilic compounds, in which a sufficiently lipophilic peptide chain would be able to interact with lipid membranes. For this reason, it is suspected that the presence of lipid membranes can influence the course of the general condensation reaction when lipophilic amino acids are involved. To investigate this hypothesis, we propose to study the formation of the lipophilic yet soluble conjugates of the general type Val_nA_p under the general condensation conditions in the presence of mixtures of lipids. The studies that tackled oligomer formation in the presence of lipids are often unsatisfactory because they are limited to products that are insoluble in water, and therefore easily accumulate in membranes, or to processes that are very inefficient in the background. Others only afford modest improvement, such as increasing the maximum length but in low yields only. In this work, we will focus on soluble products that are amphiphilic but not hydrophobic, and will evaluate whether membrane-promoted reactions are performing well enough to be of any interest when compared to background reactions in water.

In addition to quantitative ^{31}P NMR spectroscopy, already well described by the Richert group, we will use UHPLC/MS to gain a detailed insight into the variety

of the products and by-products formed during the general condensation reaction. The use of chromatography is of critical importance both to apprehend the complexity of the samples and to reduce the risks of false positives and misidentification that direct mass spectrometry poses. Because the general condensation reaction is complex, analyzing the samples is expected to be time-consuming and drawing meaningful conclusions could be difficult. To develop systematic studies and investigate detailed parameters, we will also investigate more simple conjugation reactions, while conserving the core aspect of amphiphilic phosphoramidate formation. These results should allow us to screen the parameter space more efficiently. The general conclusions drawn from such model reactions could guide us into more elaborate experiments exploiting the general condensation reaction.

Another important aspect of the project is to probe the interaction of amphiphilic peptido-RNA with lipid membranes, on the way to peptide-promoted encapsulation of nucleotides. This task will first require producing synthetic samples of pure peptido-nucleotides and peptido-RNA. The synthetic routes to peptido-nucleotides described so far suffer from either poor yields or the need for protection-deprotection steps, and several of these are described on the micro- or nanomole scale. It will therefore be necessary to develop new general, low-effort routes to a library of peptido-nucleotides. The interaction of these small amphiphilic conjugates with lipid membranes will then be probed in situ using diffusion NMR spectroscopy. The objective is to identify simple, short peptide sequences that would be able to attach peptido-nucleotides to membranes, with the perspective of delivering the nucleotide to the lumen of the vesicle through membrane permeation. Indeed, most prebiotic studies found that nucleotides had utmost difficulties to cross evolved membranes. With a more contemporary and applied perspective, parallels can be drawn with the ProTide strategy, but with the important difference that mild syntheses without protective groups are possible. A collaboration is planned with the Richert group for the evaluation of the binding of a longer peptido-RNA to giant vesicles using fluorescence microscopy. A peptido-RNA featuring a membranophilic peptide and a fluorescent-tagged RNA chain will be produced in Stuttgart and its interaction with phospholipid membranes studied in our group by confocal microscopy.

Taken together, these axes should provide elements of answer toward the goal of the “Molecular life” project: how to efficiently harness chemical energy to drive the evolution of a peptide-RNA system? How can such system be influenced by lipids and, ultimately, evolve toward an encapsulated proto-organism?

3. Results and discussion

3.1. The synthesis and the purification of nucleotide phosphoramidates

Synthesizing peptido-nucleotides and peptido-RNA is not the primary objective of this thesis. However, reference compounds were necessary, as well as pure samples for several studies. Furthermore, a lot of information about reactivity is available through synthesizing small scopes of molecules under different conditions. Peptido-nucleotides derived from small peptides (1-5 residues typically) and mononucleotides are small molecules. They can be produced on reasonable scales from affordable starting materials, quantified by weight and analyzed by NMR easily. Peptido-RNA, even from dinucleotides, are very different from a practical point of view, because RNA synthesis is an expensive and time-consuming activity. Amounts inferior to the micromole are typically handled, which changes most practical aspects: all the substrates must be transferred as stock solutions and quantified by concentration measurement (typically through UV absorbance), analyses are carried out by HPLC or mass spectrometry, NMR spectroscopy is rarely an affordable option. The implications of small and larger conjugates for systems chemistry are also different.

Purification techniques also differ between scales. For such complex mixtures of sensitive molecules, chromatography is preferred. The choice between HPLC and flash chromatography is a matter of scale, difficulty, availability of specific stationary phases, but also purity requirements and time-efficiency. The purification methods and the degree of purity that must and can be achieved is sometimes overlooked by the literature in synthetic organic chemistry. The phosphoramidate function is relatively labile and hydrolyzes very easily under slightly acidic conditions. Many peptido-nucleotides, in particular the small ones, are extremely polar species that represent a chromatographic challenge beyond what was expected at the beginning of this research. The overall effort toward efficient purification methods in this work is probably not lower than the synthetic effort itself.

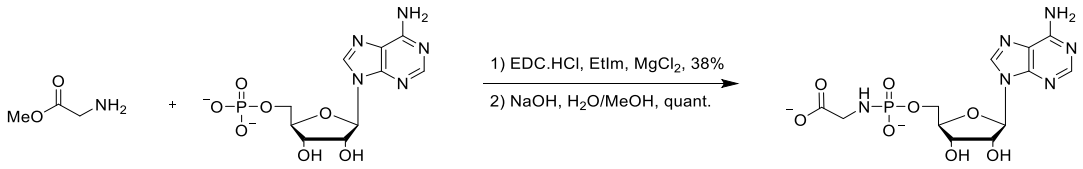
We mainly developed the synthesis of small, amphiphilic species, attempting to conjugate short peptides, in particular lipophilic ones, with mononucleotides in the most efficient ways. The extensive use of protecting groups was avoided, and aqueous media were often our first choice. This was a consequence of the solubility of our substrates and the focus of our project. In a context of an urgent need for cleaner chemistry and withdrawal from petroleum dependence, simple synthetic ways to conjugates that find growing therapeutic uses are also welcome, although this is not our primary goal here. Perspectives on the transfer of these methodologies to solid-supported synthesis are briefly discussed, as are reflections on the reactivity of active nucleotide phosphates in water.

3.1.1. The preparation of peptido-nucleotides

3.1.1.1. Direct coupling methods

An intuitive retrosynthesis of peptido-nucleotides involves the direct coupling between the nucleotide and the amino acid or short peptide. As further developed in section 3.2.1, coupling conditions applied to a solution of unprotected nucleotides and amino acids afford a complex mixture of products including short peptides, peptido-nucleotides, pyrophosphates and other by-products. Isolating a single product, even a major one, from such sample would be, at best, tedious, and likely low-yielding. Protecting the *C*-terminus of the amino acid or peptide would greatly reduce the number of products by blocking peptide growth. Phosphodiesters of the nucleotides are only slowly formed in competition with phosphoramidate coupling, and the protection of the ribose 2' and 3' OH groups is therefore optional. Pyrophosphates and by-products inherent to the specific chemistry of each coupling agent cannot be avoided by the use of protective groups but their formation may be limited by improving coupling conditions. Furthermore, protection strategies always present the drawback of adding at least two steps and are only efficient if the protection and deprotection reactions proceed in good yields with high selectivity.

Richert and coll. prepared the phosphoramidate GlyA[a] (Scheme 3.1) in moderate yield from glycine methyl ester and adenosine monophosphate through a coupling mediated by EDC·HCl and 1-ethylimidazole in water, followed by a saponification of the ester.[28] They purified both the intermediate and the final product by reverse-phase cartridge chromatography. This method suffers from the potentially harsh saponification step, the limited yield and the difficulty to purify highly polar species (see section 3.1.2 for further discussion).



Scheme 3.1 Synthesis of GlyA by the coupling/deprotection synthesis.

We proposed to overcome these difficulties by using a benzyl ester instead of the methyl ester. Not only are benzyl groups orthogonally deprotected by hydrogenation, but they also add significant lipophilicity to the protected intermediates, resulting in easier purifications. We used AMP and commercially available L-valine benzyl ester as model substrates to develop efficient coupling conditions. A mixture of pyridine and water was chosen as the solvent for its ability to dissolve a large variety of hydrophobic peptides, often difficult to solubilize properly, as well as the very hydrophilic nucleotides. When both substrates were mixed in pyridine-d_5/D_2O (75:25) in the presence of EDC·HCl and

[a] In the absence of a standard nomenclature for peptido-RNA, we developed a short and a detailed notation, both based on Richert's first proposals.[27] These notations are explicated in detail in the appendices, section 2 on page 163.

DIPEA, the desired phosphoramidate BnOValpA quickly started to form as evidenced by ^{31}P NMR. Due to the presence of water, several equivalents of EDC were necessary. The reaction was rather slow and with time, the 5'-diadenosyl pyrophosphate AppA, *N*-phosphorylureas and other by-products were formed (see section 3.2.1.3 for further discussions about by-products arising from EDC chemistry). We found that repeatedly adding smaller amounts of EDC while monitoring the reaction by ^{31}P NMR was more efficient and generated less by-products than providing an excess of EDC in a single addition. Three successive additions of one equivalent of EDC were found to be optimal and the desired product was formed with 83% conversion after 18 hours. Small amounts of AppA (<3%) was the only by-product. Adding more EDC did not improve the conversion (Figure 3.1).

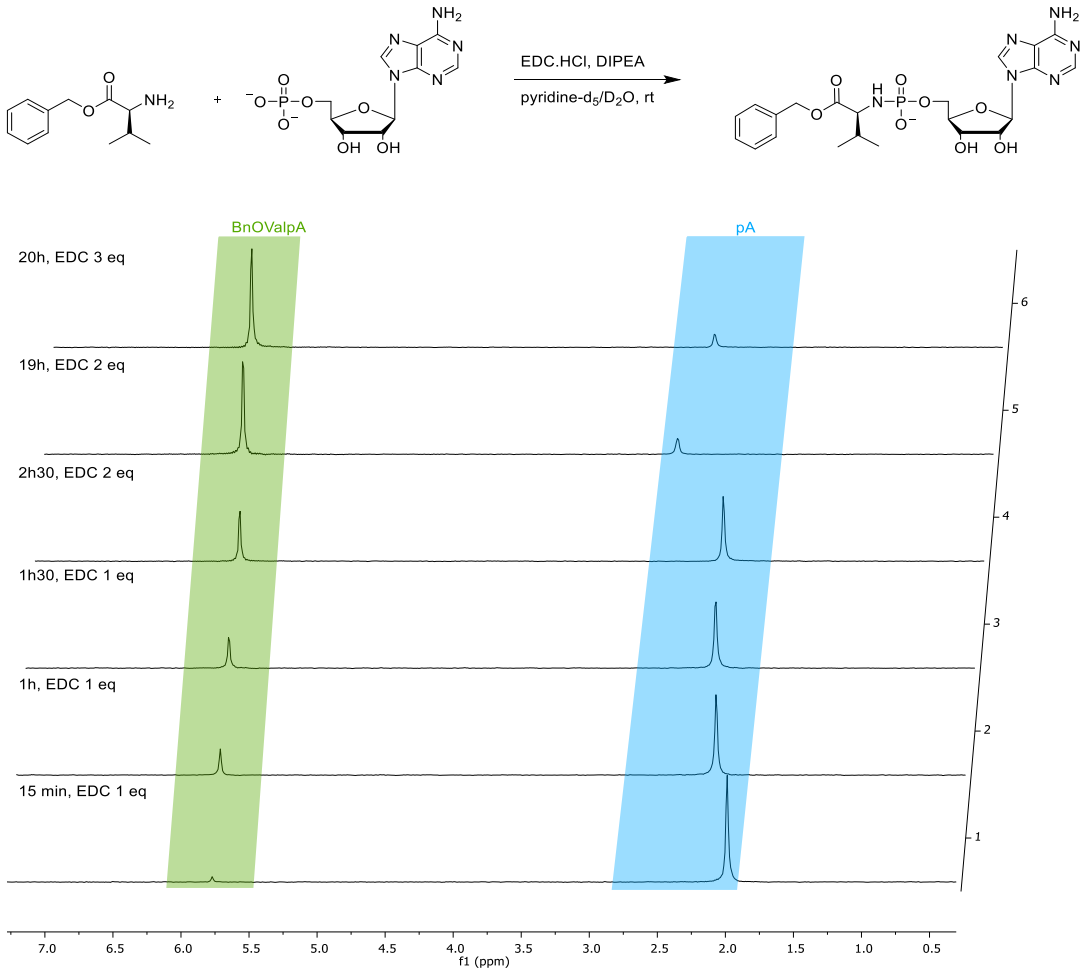
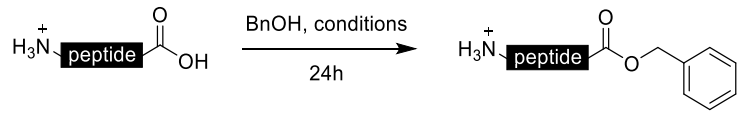


Figure 3.1 Crude ^{31}P NMR (pyridine-d_5/D_2O, 120 MHz) of a coupling reaction between H-Val-OBzl and AMP (2-2.5 ppm, sensitive to the pH and salinity changes throughout the additions) yielding the phosphoramidate BnOValpA (5.8 ppm).

The reaction was successfully repeated on a 0.5 mmol scale and the crude product was isolated by precipitation from acetone/sodium perchlorate in 60% yield. The benzyl group was then cleaved by hydrogenation over Pd/C. The deprotection was monitored by HPLC and found to be successful. However, the title compound could not be efficiently separated from AMP by reverse-phase chromatography. An alternative approach would be to purify the protected intermediate and to use

optimized cleavage conditions yielding the product in good purity without the need for purification.

Most amino acids are commercially available as benzyl esters, but peptides rarely are. To make our route applicable to peptides, it was therefore necessary to develop a general and efficient benzylation method. The protection of carboxylic acids as esters is widely applied but most conditions rely on dissolving the substrate in the alcohol of interest.[417] Our peptides were essentially insoluble in benzyl alcohol and did not react when thionyl chloride was added to the suspension. We tried several different conditions on three model substrates (L-valine and the dipeptides Val_2 and Gly_2) with little success: only Dean-Stark azeotropic water removal with TsOH as the catalyst gave positive results, but under poorly reproducible conditions, especially on a small scale, which is more representative of typical peptide chemistry, where substrates are expensive and scarce. Val_2, the most representative substrate for lipophilic peptides, gave very low yields (Table 3.1). Alternative esters that may be cleaved under mild conditions include 2-cyanoethyl (removed by sulfides[418] or TBAF[418,419]) and cyanomethyl (removed by sulfides[420]), but they were not investigated experimentally.



Entry	Substrate	Conditions	Yield (%) [1]
1	Val	BnOH 5 eq., TsOH 1.2 eq, toluene, Dean-Stark, reflux	0-66 [2]
2	Gly_2	BnOH 5 eq., H_2SO_4 conc. cat., toluene, reflux	nd
3		BnOH 5 eq., H_2SO_4 fum. cat., 4Å MS, dioxane, reflux	nd
4		BnOH 5 eq., TsOH 1.2 eq, toluene, Dean-Stark, reflux	65
5	Val_2	BnOH (solvent), $SOCl_2$, reflux	nd
6		Dry HCl, BnOH 5 eq., toluene, co-evaporations	nd
7		BnOH 5 eq., TsOH 1.2 eq, toluene, Dean-Stark, reflux	0-15 [3]

Table 3.1 Attempts to prepare the benzyl esters of amino acids and small peptides. [1] the compounds were precipitated as p-toluenesulfonate salts by addition of diethyl ether and the yield determined by the weight of pure product when obtained. [2] the reaction performed well on the gram-scale but was inefficient on a smaller scale. [3] some product was observed by TLC and NMR, but the precipitation failed to provide a pure compound. nd: product not detected.

As described in section 2.4, several authors discussed the conjugation of small oligonucleotides to peptides[421] and other kinds of amines.[391,392] Many strategies employed couplings promoted by the 2,2'-dipyridyl disulfide/triphenylphosphine couple, as first reported by Mukaiyama's group for peptide couplings.[388] Usually, DMAP or 1-methylimidazole were added to form an activated intermediate that reacted rapidly with the peptide. This reaction required anhydrous conditions and was therefore more appropriate when the substrates were soluble in organic solvents, often DMF or DMSO. These conditions were investigated with oligonucleotides as substrates, on very small scales (tens to hundreds of nanomoles), using considerable excesses of peptides and coupling agents. We tried

to adapt these conditions to a larger scale and at higher nucleotide concentrations to prepare small peptido-nucleotides. However, only a limited choice of lipophilic peptides were soluble in DMF, and in all of our assays, AppA was the only product to be formed, likely because of the important concentrations in use.

It must be noted that coupling agents such as HBTU, or coupling conditions employing imidazole, HOBt, morpholine or similar compounds will quickly lead to the formation of an active phosphate derivatives that needs to be displaced to yield the desired phosphoramidate, a matter that is discussed in the following section. For this reason, HBTU, HATU and similar reagents will not act with phosphates as coupling agents in the classical meaning, but as pre-activators that yield isolable, and sometimes storable, active phosphates.

3.1.1.2. Pre-activation of the phosphate

The selectivity issues encountered when using coupling reactions directly between nucleotides and peptides may be overcome by an activation-displacement strategy, taking advantage of active phosphates that are stable enough to be isolated and even stored, but reactive enough toward amines. This strategy is widely used in automated peptide synthesis, where *N*-protected amino acids are activated at the *C*-terminus, usually as an OBt- ester. Phosphates tend to form active esters that are more stable than these from carboxylates, which leads to a decreased reactivity but reduces drastically the extent of side reactions and makes aqueous chemistry possible.

We investigated a range of leaving groups for the phosphate function of nucleotides and displaced them with peptides to yield peptido-nucleotides. Among known leaving groups, imidazolides[422] were by far the most common and well-documented. Other examples included (aza)benzotriazolides,[199] morpholidates,[200] pentafluorophenoxides[401] and other electron-poor phenoxides, and several charged groups such as 1-methylimidazoliums[391] or pyridiniums[392] that were usually generated in situ only (Figure 3.2). Ferris and coll. made a systematic evaluation of 8 leaving groups for RNA elongation reactions.[200] Strongly activated derivatives such as chlorophosphates were poorly compatible with unprotected substrates and preclude aqueous chemistry. *N*-hydroxysuccinimide (NHS), a common leaving group in carboxylate chemistry, was only reported once, in a patented procedure with few details, for nucleotide activation.[423]

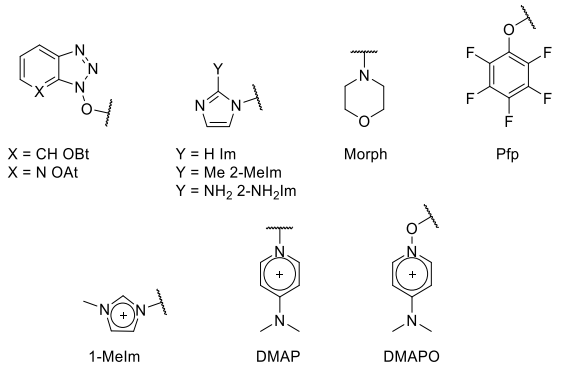
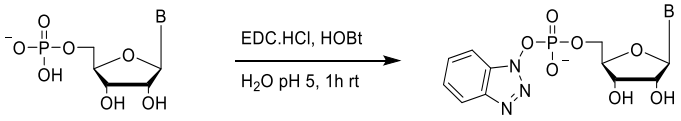


Figure 3.2 Selection of leaving groups used to activate 5'-phosphates of nucleotides. Top: leaving groups that will bear no charge once coupled, usually resulting in a stable intermediate; bottom: leaving groups that usually result in positively charged, less stable intermediates that must be generated in situ just before use.

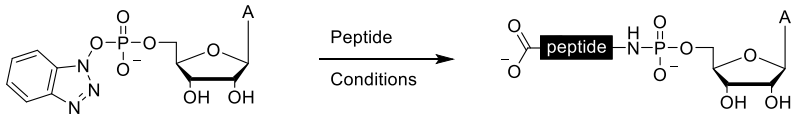
Richert and coll. prepared azabenzotriazolides of nucleotides either from from HOAt using EDC·HCl as the coupling agent in aqueous solution,[199] or in lower yields from HATU/HOAt in DMF.[424] At the time of this study, we were only able to purchase dilute DMF solutions of HOAt and we therefore prepared the similar benzotriazolides of nucleotides using HOBt and EDC·HCl in water (Scheme 3.2). A precipitation from acetone/sodium perchlorate afforded a crude mixture that mainly contained the desired active nucleotide (AMP-OBt) along with the excess of HOBt and residues derived from EDC. The product was partially hydrolyzed when reverse-phase chromatography was attempted, and it was therefore used without further purification. In our hands, using the HBTU/HOBt couple in anhydrous acetonitrile afforded comparable crude mixtures but required more effort.



Scheme 3.2 Preparation of benzotriazolides of nucleotides in water.

When AMP-OBt reacted with a 5-fold excess of the hydrophilic peptide Gly_2 in water at pH 8-9 (adjusted with NaOH; Gly_2 played the role of the buffer), a phosphoramidate was quickly formed as evidenced by a new peak at 7.55 ppm in the ^{31}P NMR spectrum. After 24 hours, 80% of AMP-OBt was converted to the desired phosphoramidate, the rest being hydrolyzed back to AMP. Similar results were obtained when UMP-OBt (82% conversion) and GMP-OBt (68% conversion) were used. We repeated the same experiment with L-valine as the nucleophile, but solubility limitations prevented us to use more than one equivalent of valine. As a consequence, the formation of the phosphoramidate was considerably slower, and hydrolysis dominated, followed by the formation of small amounts of AppA, resulting from the condensation of AMP with AMP-OBt. No suitable co-solvent was found to improve the solubility of valine. We turned ourselves to its benzyl ester, that we had already used in direct coupling experiments, and found that aqueous methanol was a suitable solvent for the reaction. The addition of small

amounts of pyridine accelerated the process but using it as a co-solvent resulted in very quick hydrolysis and lower yields (Table 3.2). The overall yield was lower than when using the direct coupling procedure as previously described, but this alternative pathway might be preferable for other substrates, in particular peptides with functionalized sidechains that would produce by-products under simple coupling conditions. The benzyl group was removed as described in the precedent chapter. The dipeptide Val_2, representative of the lipophilic peptides of our interest, reacted to a very limited extent under the same set of conditions.



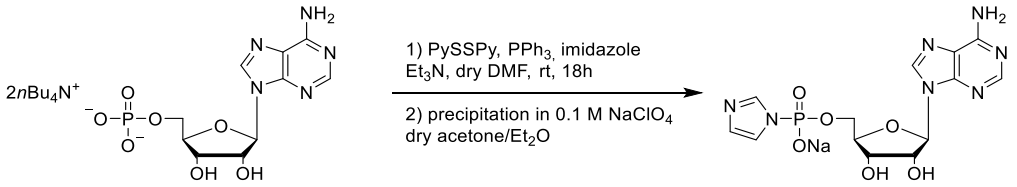
Entry	Peptide	Conditions	Conversion (%) [1]
1	Gly_2	5 eq peptide, water pH 8-9, 24h	80
2	Val	1 eq valine, water, pH 8-9, 1 week	<10
3	ValOBzl	1 eq valine ester, 4 eq DIPEA, pyridine/water 75:25, 24h	20
4		1 eq valine ester, 4 eq DIPEA, methanol/water 8:2, 4 days	40
5		1 eq valine ester, 4 eq DIPEA, 1 eq pyridine, methanol/water 8:2, 48 hours	45 (38) [2]
6	Val_2	1 eq peptide, 4 eq DIPEA, methanol/water 8:2, 8 days	<10

Table 3.2 Formation of phosphoramidates by nucleophilic displacement of a benzotriazolide. [1] All the reactions were performed on a 0.1 mmol scale at room temperature in deuterated solvents and followed by ^{31}P NMR. [2] The reaction was repeated on a 1 mmol scale and the product was obtained in 38% isolated yield.

As benzotriazolides offered an overall disappointing reactivity, especially with lipophilic peptides, we turned ourselves to imidazolides, one of the most common strategies to pre-activate the phosphate of nucleotides. Three main strategies were described for their preparation:

- reaction with an excess of carbonyldiimidazole (CDI) in an organic solvent, typically DMF or acetonitrile.[425] While this strategy was straightforward, we found in our hands that several by-products were formed, in particular through purine acylation and cyclic carbonate formation at the 2'-3' cis-diol.
- aqueous coupling with EDC and an excess of imidazole, by analogy with the preparation of the OAt esters.[199] This method provided a crude mixture after precipitation from sodium perchlorate in acetone, similarly to the OBt strategy. While imidazole was efficiently washed off, EDC and EDU residues considerably polluted the samples.
- coupling in anhydrous organic solvents, mediated by 2-pyridyl disulfide and triphenylphosphine, followed by a precipitation.[426] Despite the additional care required, this method provided very clean samples of imidazolide in almost quantitative yields and was adopted as our reference procedure.

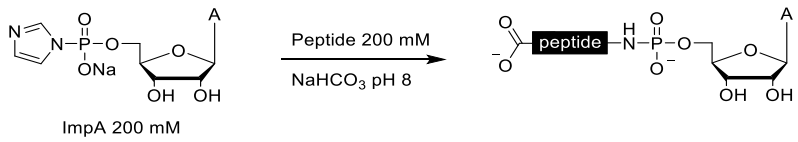
Nucleotides must be converted to lipophilic salts to be soluble in organic solvents. While bis(triethylammonium) (TEA) salts were often reported in literature, we found that bis(tetrabutylammonium) (TBA) salts were superior, leading to a 10-fold increase in the solubility of AMP. This was particularly important as the purification was operated by precipitation under anhydrous conditions and required at least 10 times the volume of the reaction solvent. For millimole-scale couplings using TEA salts, very large volumes of anhydrous precipitation solvent would be required. Starting from AMP as its TBA salt, we could routinely prepare the imidazolide ImpA, on a 2 mmol-scale with high purity and in quantitative yield, using only 12 mL of DMF as the reaction solvent and 150 mL of sodium perchlorate solution in anhydrous acetone/diethyl ether for the precipitation (Scheme 3.3). ImpA did not hydrolyze readily under aerobic conditions but was highly hygroscopic and had to be precipitated, handled, and stored under anhydrous conditions to avoid the irreversible formation of a sticky oil. Sodium perchlorate served as a counter ion exchanger, allowing the desired sodium salt to precipitate while triethylammonium or tetrabutylammonium perchlorates remained soluble. This exchange was efficient as assessed by ^{1}H NMR, where residual tetrabutylammonium signals were usually below 5 mol% (appendices, Figure S2).



Scheme 3.3 Preparation of ImpA on a 2 mmol-scale using optimized Mukaiyama conditions. The desired compound was obtained quantitatively as a pure sodium salt.

ImpA reacted with a variety of water-soluble peptides to yield the corresponding peptido-nucleotides. The reaction occurred under neutral conditions but was most effective under slightly basic conditions, typically at pH 8.5. Organic buffers such as MOPS were convenient for kinetics and reactivity studies, but not appropriate for preparative applications because of the very large amounts that needed to be removed by chromatography. Furthermore, many organic buffers possess primary or secondary amines that may compete with the peptide. Among the inorganic buffers, we found sodium borate and sodium bicarbonate, adjusted to a pH of 8 to 9, to be suitable. Ammonium bicarbonate, a lyophilizable buffer, was not used because of the potential reactivity of the ammonium ion. It must be noted that borate ions form stable complexes with the 2'-3' cis-diol of nucleotides. While this was never a problem in terms of reactivity or chromatography, ^{1}H and ^{31}P NMR spectra were much affected by this complexation (appendices, Figure S3). Sodium bicarbonate, 0.1 M, pH 8 (prepared in either H_2O or D_2O) was therefore used to dilute, analyze, and store pure samples. Val_2 and Val_3 reacted efficiently with ImpA to provide the corresponding phosphoramidates in excellent conversion and good isolated yields. Valine afforded inferior conversions. The more lipophilic peptide Leu_3 was poorly soluble in water, forming thick gels at 200 mM. When it

was first converted to the hydrochloride salt, it could be dissolved at a maximal concentration of 50 mM at pH 8, and Leu_3A formed much more slowly than in the other experiments, due to the dilution (Table 3.3). When organic (co)-solvents were employed, the reactivity of the imidazolide was drastically changed and no significant amounts of the desired product were obtained. This effect was already noticed in benzotriazolide chemistry, but no clear rationalization was found to this day. Among other activating groups, we briefly investigated pentafluorophenol (Pfp), but both procedures developed for OBt and imidazolides failed at providing Pfp-activated nucleotides in good yields.



Entry	Peptide	Conversion (%) [1]	Yield (%) [2]	Remarks
1	Val	44	17	Difficult purification
2	Val_2	71	52	
3	Val_3	70	40	
4	Ala_3	63	35	
5	Leu_3	50 [3]	45	Low solubility [3]

Table 3.3 Synthesis of peptido-nucleotides by the imidazolide method in water. [1] The conversion was determined by analytical HPLC after 7 days of reaction, except for entry 1 and 5 which were followed by ^{31}P NMR. [2] The yield was determined after HPLC or column chromatography purification by dissolving the sample in a known volume of buffer and measuring the absorbance at 260 nm. [3] due to the low solubility of Leu_3 in water, the reaction was performed at 50 mM and was therefore much slower. It was stopped after 4 months, at which point less than 10% of the starting imidazolide remained.

3.1.1.3. On the reactivity of activated nucleotides

Throughout the synthesis of peptido-nucleotides, we observed several intriguing behaviors that contrasted with the straightforward aspect that coupling and pre-activation-displacement reactions may have at a first glance. Although the topic of this thesis was not to provide mechanistical or even systematic insight into aqueous phosphoramidate synthesis and reactivity, we will share here certain observations and reflections that arise from our experimental data and from discussions with the Richert research group, that pursued much deeper reactivity studies.

When glycylglycine was coupled with AMP using EDC in the presence of pyridine, Gly_2A quickly formed and the starting nucleotide was consumed in less than 24 hours. However, if the crude mixture was allowed to stand for a longer period, the reaction slowly reverted to AMP. It was unlikely that direct hydrolysis of the phosphoramidate occurred under such basic conditions, and we have stored similar conjugates for more than a week in aqueous solution at pH 8 without signs of degradation by ^1H NMR and HPLC. A nucleophilic displacement by pyridine was more plausible, leading to an intermediate phosphopyridinium that was

easily hydrolyzed. Richert and co-workers have recently described the crucial role of 1-ethylimidazole in the EDC-promoted formation of phosphoramidates,[32] using GlyA as a model. Without EtIm, the active intermediate AMP-EDC was hydrolyzed faster than it reacted with glycine. When EtIm was present, a phosphoroimidazolium formed. It was kinetically more stable than AMP-EDC and served as a reservoir of active AMP (Figure 3.3). 1- and 3-methyladenine had the same effect, and other nucleophilic heterocycles such as pyridine or DMAP are expected to behave similarly. The authors termed this effect "organocapture" as a contrast to the common use of the term "nucleophilic catalysis" that should refer to the acceleration of a specific, desired reaction, rather than the inhibition of an undesired reaction.

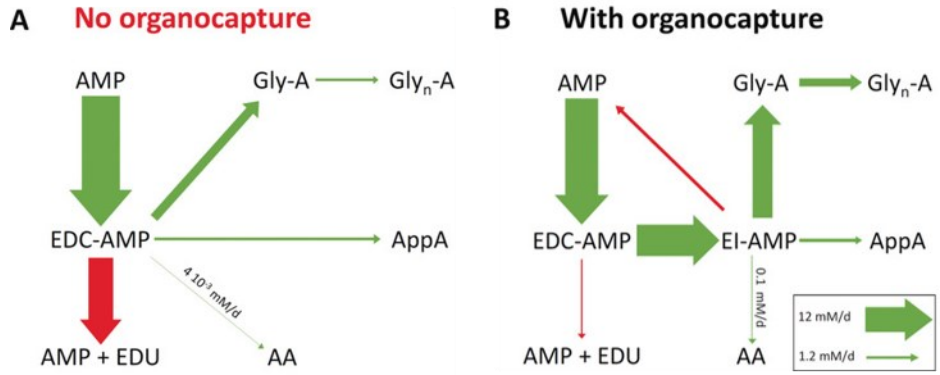


Figure 3.3 Organocapture by 1-ethylimidazole (EI) as described by Richert et coll.[32] for the EDC-promoted formation of GlyA and other related species. The width of arrows represents the chemical flux (in mM per day). Quantitative kinetic data obtained from ^{1}H, ^{13}C and ^{31}P NMR experiments was injected in kinetic computations to provide a quantitative description of the network of reactions taking place. Figure under Creative Commons License reproduced from the original manuscript.[32]

An indirect conclusion of this work is that the efficiency of a phosphoramidate coupling merely comes down to how much the peptide displaces the leaving group. Owing to the principle of micro-reversibility, this step must be reversible and the "organocapture agent" can in principle displace the peptide back from the phosphoramidate. When charged, strongly activated intermediates such as imidazoliums are used, it is quite clear that a highly nucleophilic substrate should, under basic conditions, be strongly favored. The situation is less evident when stable species such as AMP-OBt or ImpA are used.

AMP-OBt was isolated as a crude mixture containing, among others, a considerable excess of HOBt. When well-soluble peptides such as Gly_2 were used in a 5-fold excess, the reaction was efficient. When only one equivalent of peptide was employed, for solubility reasons for example, the reaction was slower and the yield very low. HOBt has a pK_a (HOBt/BtO$^-$) of 5.5. It was therefore deprotonated under the basic conditions employed, quite nucleophilic, and present in excess. It would not be surprising to see this excess of good nucleophile displace the small amounts of peptido-nucleotides that have progressively formed. Furthermore, Gly_2 is a good buffer at pH values around 8, when used at elevated concentrations (1 M in our study). When lower concentrations of peptides were used, the release

of the acidic HOBt upon substitution of AMP-OBt lowered the pH beyond the buffering capacity of the peptides and favored phosphoramidate hydrolysis.

ImpA was isolated in pure form. Only a stoichiometric amount of peptide was required to achieve good yields of phosphoramidates from it. With a pK_a (ImH^+/Im) of 6.95, the release of imidazole during the reaction had no drastic influence on the pH. The nucleophilicity scale of Mayr's group[427] classifies imidazole as a relatively weak nucleophile, more than three orders of magnitude below pyridine. Unfortunately, no direct comparison with HOBt was possible since this compound was included in no similar nucleophilicity tables. Furthermore, Ferris and coll. found in systematic studies that the correlations between basicities and leaving abilities were only valid in series of tightly related leaving groups. Indeed, they noticed that effects such as charge delocalization and formation of stable functions such as pyridones and guanidines played a crucial role.[200]

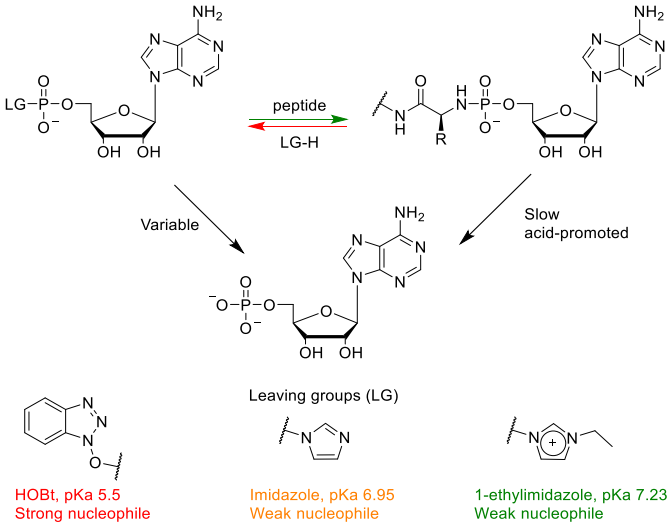
The case of the intermediate 1-ethylimidazolium in the work of Richert and coll. combines several advantages: 1-ethylimidazole is slightly basic with a pK_a ($EtImH^+$/EtIm) of 7.23[428] and may be used in catalytic amounts. Its nucleophilicity on Mayr's scale in water is similar to that of imidazole.[a] However, the fact that it is used only as an intermediate "organocapture agent" makes a strong difference, as the hydrolysis of EtImpA back to AMP may be durably compensated through consumption of the feedstock of EDC. Using preformed EtImpA as a starting material would be very different.

From our data and literature information, we reason that the displacement of the leaving group in the pre-activation strategy is likely reversible, not only theoretically owing to the micro-reversibility principle but to practical kinetic extents. The conversion is then governed in part by the nucleophilic competition between the peptide and the leaving group (LG, Scheme 3.4), that is influenced not only by their chemical structure and nature but also by their stoichiometry.

It is then profitable to use an excess of peptide when possible and avoid the presence of competing nucleophiles. Furthermore, the hydrolysis rates of both starting material and product must be considered. This point is very important regarding strongly activated intermediates, that may suffer from major hydrolysis before the coupling reaction occurs. The case of Richert's imidazolium differs by the presence of a durable energy source that compensates for the hydrolysis. Finally, the pH value plays an important role in all the aforementioned points. Phosphoramidates are poorly stable under acidic conditions and nucleophilicities may vary considerably with pH changes. Neutral to basic leaving groups are preferable, especially at high reaction concentrations where buffering is impractical. These considerations are summarized in Scheme

[a] Mayr and coll. only reported 1-methylimidazole, a much more common compound than 1-ethylimidazole. However, the slight change in chain length is not expected to have a significant effect here, as 1-methyl and 1-phenylimidazole displayed very similar data despite a greater chemical difference.

3.4. It must be noted that no dedicated experiments were carried out to confirms our views, but they may be considered as a plausible rationale in the expectation of additional investigations.



Scheme 3.4 Chemical pathways proposed to rationalize the differences in reactivity between benzotriazolides and imidazolides in the displacement strategy.

3.1.1.4. Perspective: solid-supported peptido-RNA synthesis

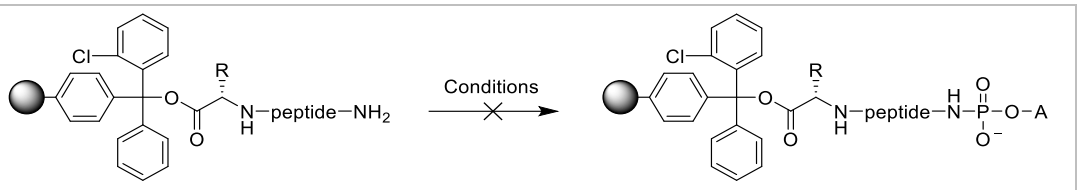
Peptides and oligonucleotides are essentially synthesized by solid-supported iterative techniques, in which the first residue is covalently linked to a functionalized material (resin or glass), and iterative deprotection and coupling steps are performed successively with intermediate solvent washes. This technology, combined with coupling methods of very high efficiency, allow small and medium peptides and oligonucleotides to be prepared in unprecedented yields and scales.[429] A single workup procedure is necessary to cleave the oligomer from the solid support and remove late-stage protecting groups, followed by a single purification, a considerable advantage over classical multistep synthesis, where nearly every step is followed by an individual workup and purification routine.

Co-oligomers of peptides and RNA, namely peptidyl-RNA analogs, were prepared in our group in the past by a fully solid-supported strategy (see a description in section 2.2.2), using puromycin analogs as both the key bridging residue and the anchoring point to the support.[23] This strategy may not be applicable directly to peptido-RNA. Indeed, the phosphoramidate connection is considerably weaker than the amide found in puromycins, and it is located between the 5'-end and the *N*-terminus, both of which are the iterative nucleophiles pointing away from the resin in conventional solid supported synthetic strategies. A fully iterative strategy, i.e. starting from one end of the conjugate to get to the opposite one solves only one half of this problem. If the peptide was built first, the *N*-terminus may be coupled to a nucleotide, but the RNA synthesis should next be performed in the unusual, but not unknown,[430] synthetic direction 5' to 3'. Among the potentially harmful steps for the phosphoramidate, the acidic detritylation, the capping step (delivering acid anhydrides and the nucleophilic catalyst 1-

methylimidazole) and the iodine oxidation in pyridine must be investigated. Alternatively, building the RNA first will pose the similar problem of synthesizing the peptide moiety from the *C*-terminus to the *N*-terminus, a strategy known[431] but not routinely operated yet. The Fmoc deprotection step by piperidine may have to be replaced by a less harmful procedure.

A complete investigation of the feasibility of such solid-supported strategy was beyond the objectives of this thesis work and probably deserves a dedicated project. However, we envisioned that solid support would be a valuable tool for the synthesis of peptido-nucleotides, where no further RNA synthesis was required. A peptide may be synthesized in the conventional *C*→*N* direction, then its free *N*-terminus coupled to a single nucleotide. Peptide solubility issues would be reduced, as would be the need for their purification. For short peptides sequences, where the overall yield is usually excellent and very few *n-1* sequences are generated, one may even hope to recover a sample of sufficient purity after the resin cleavage and the workup. With this perspective in mind, we briefly explored the solid-supported synthesis of peptido-nucleotides and report here our first conclusions.

The Wang-type resin typically used in peptide synthesis was not appropriate since the peptide must be cleaved from it under strongly acidic conditions (50-100% trifluoroacetic acid).[160] We used instead 2-chlorotrityl resins, that may be cleaved very mildly by treatment with 20% 1,1,1,3,3,3-hexafluoroisopropanol (HFIP) in dichloromethane for 30 minutes according to manufacturer's procedures, or other mild conditions.[432] Short peptides were first synthesized by conventional Fmoc/HBTU chemistry, with the final Fmoc off, on commercially available 2-chlorotrityl resins loaded with a Fmoc-amino acid. The resin was then recovered, washed, dried and swollen in anhydrous dichloromethane before the beginning of the coupling with AMP under various conditions (Table 3.4). We attempted to follow the consumption of the nucleotide by HPLC analysis of supernatant samples, but the large excess of it, combined with the presence of many organic solvents and reagents, resulted in inconsistent measurements. We therefore stopped the coupling after arbitrary times, flushed out the reagents, extensively washed the resin and proceeded to the cleavage with 20% HFIP in dichloromethane. The product was then precipitated from the crude mixture using diethyl ether and analyzed. Unfortunately, none of the coupling conditions explored yielded the desired conjugates as judged by NMR and MS analyses, and the cleaved residues contained only the chosen peptide. The absence of unbound nucleotide in the sample demonstrated that hydrolysis during the cleavage was not the problem, but rather an absence of reactivity.



Entry	Peptide sequence	Conditions
1	2-ClTrtO-VV-H	AMP, DCC, DMAP, DIPEA, DMF, 6h
2	2-ClTrtO-VV-H	AMP, PySSPy, PPh_3, Et_3N, DMAP, DMF/DMSO, 24h
3	2-ClTrtO-GVV-H	AMP, DCC, DMAP, DIPEA, DMF, 24h
4	2-ClTrtO-GVV-H	ImpA, DIPEA, DMF, 24h

Table 3.4 Coupling conditions that failed to provide resin-bound peptido-nucleotides. All reactions were performed in anhydrous solvents under argon, at room temperature and with mechanical shaking, as magnetic stirring tends to mechanically damage the resins.

All of these trials used DMF, a solvent that did not swell the resins very efficiently. The loading (0.5-0.8 mmol/g) was typical from peptide chemistry but much higher than typical loadings of commercial CPG supports used in RNA synthesis (usually less than 0.05 mmol/g) and the pores were also narrower. A more appropriate set of resin, solvent, loading and maybe linker may solve the reactivity problems. Furthermore, highly reactive 5'-phosphoramidites are commercially available for reverse-direction oligonucleotide synthesis and may represent a more efficient, albeit sensitive, active nucleotide, that could be used in acetonitrile or dichloromethane.

3.1.2. The purification of nucleotide phosphoramidates

Phosphoramidates are relatively sensitive chemical species. They display a good tolerance for basic media but are quickly hydrolyzed under acidic conditions. Their rate of hydrolysis is also very much influenced by their structure (see the discussion in 2.4.3). In synthetic studies, peptido-RNA was typically cleaved under mildly acidic conditions (15% acetic acid, 50°C, 40 min[390] or sodium formate buffer at pH 3.5 overnight at 37°C[392]). In our hands, we found that exposing small peptido-nucleotides to dilute formic acid (0.1%, ca. 25 mM, pH 2.5) or even ammonium formate buffers at pH 4 during purification procedures, for periods of less than 2 hours, was sufficient to induce significant degradation. This precluded the use of any acidic conditions in workup and purification procedures. Samples of peptido-nucleotides were stored at pH 8 (sodium bicarbonate buffer) at room temperature for more than a week and in a freezer for months without detectable degradation. The gentle evaporation (40°C) and lyophilization of fractions from neutral to mildly basic buffers, typically ammonium acetate, was also suitable.

3.1.2.1. Reverse-phase chromatography

Besides stability issues, the challenge of purifying small peptido-nucleotides resides in their extreme polarity. Silica gel chromatography was immediately excluded for this reason. Typical reverse-phase chromatography procedures used

for small molecules failed at providing decent retention and chromatographic resolution. We turned ourselves to specific chromatographic columns, known under various names including “polar C18”, “aqueous C18” or “C18 aq” depending on the manufacturer, that combine the lipophilic behavior of a C18-silica column with polar endcapping groups. These columns offered better retention for our polar compounds in addition to be compatible with 100% aqueous eluents. They were available in all formats, including analytical HPLC and UHPLC columns, preparative HPLC, preparative flash chromatography and SPE cartridges. This point was important, as we were able to keep consistent chromatographic behaviors from analytical studies to various preparative scales.

In analytical setups, we used a 3 µm, 3×150 mm polar C18 column (Phenomenex Luna Omega Polar C18) to develop our chromatographic conditions. An acidic eluent system, consisting of 0.1% formic acid in a water/acetonitrile gradient provided the best retention and resolution that we could obtain. Formic acid was preferred over the usual trifluoroacetic acid for its better compatibility with electrospray mass spectrometry. A fully aqueous plateau of 3 minutes (analytical conditions) or 5 minutes (preparative conditions) afforded a much better retention than starting a gradient from the beginning of the run and was useful to remove the salts and buffers. Figure 3.4 (top) shows a representative analytical injection was made from a crude aliquot of a Val_2A synthesis. The compounds to separate included the starting ImpA, the desired Val_2A, the excess of peptide, AMP arising from the hydrolysis of ImpA and minor residues. All of these compounds were sufficiently retained and separated from each other with a good resolution. Neutral eluents consisting of aqueous ammonium acetate (pH 7.0, 10-100 mM) and acetonitrile (Figure 3.4, bottom) provided less retention and limited resolution between ImpA and Val_2A. Furthermore, the resolution between different phosphoramidates was better under acidic conditions than under neutral conditions (data not shown).

The chromatographic qualities of the acidic conditions transposed well to the preparative scale. However, when pure fractions obtained under acidic conditions were quenched to a pH of 7-8 with ammonia or NaOH immediately after collection, then immediately lyophilized, the resulting sample contained up to 30% of AMP arising from a post-purification hydrolysis. Ammonium acetate did not present this issue and was chosen as a preparative buffer for reverse-phase chromatography despite a less appealing chromatographic performance in our case. Species such as Val_2A and Val_3A were successfully purified by preparative HPLC with purities over 99.5% (analytical HPLC) using these conditions on a 10 mm diameter column. For larger sample sizes or increased time-efficiency, flash chromatography on a coarser and larger column (Interchim C18 Aq, 30µm, 40g) gave satisfying results with HPLC purities over 98%. Less hydrophobic peptido-nucleotides were more challenging. Gly_2A co-eluted partially with AMP, and despite drastic fraction cuts, the final samples presented large amounts of residual AMP. Gly_2U was even more problematic, as it co-eluted with both UMP and excess Gly_2 and could not be isolated in satisfying purities. For such small

species, the direct coupling of esters (section 3.1.1.1) may be an interesting alternative. It must be further noted that the considerable excess of HOBt and residues from the activation of nucleotides as OBt esters typically tailed on reverse-phase columns. While this was not problematic in the case of Gly_2A and Gly_2U, that eluted earlier, it may be very difficult to separate more lipophilic peptido-nucleotides from these residues. Imidazole eluted as a sharp, poorly retained peak which is another argument in the favor of imidazolides over benzotriazolides.

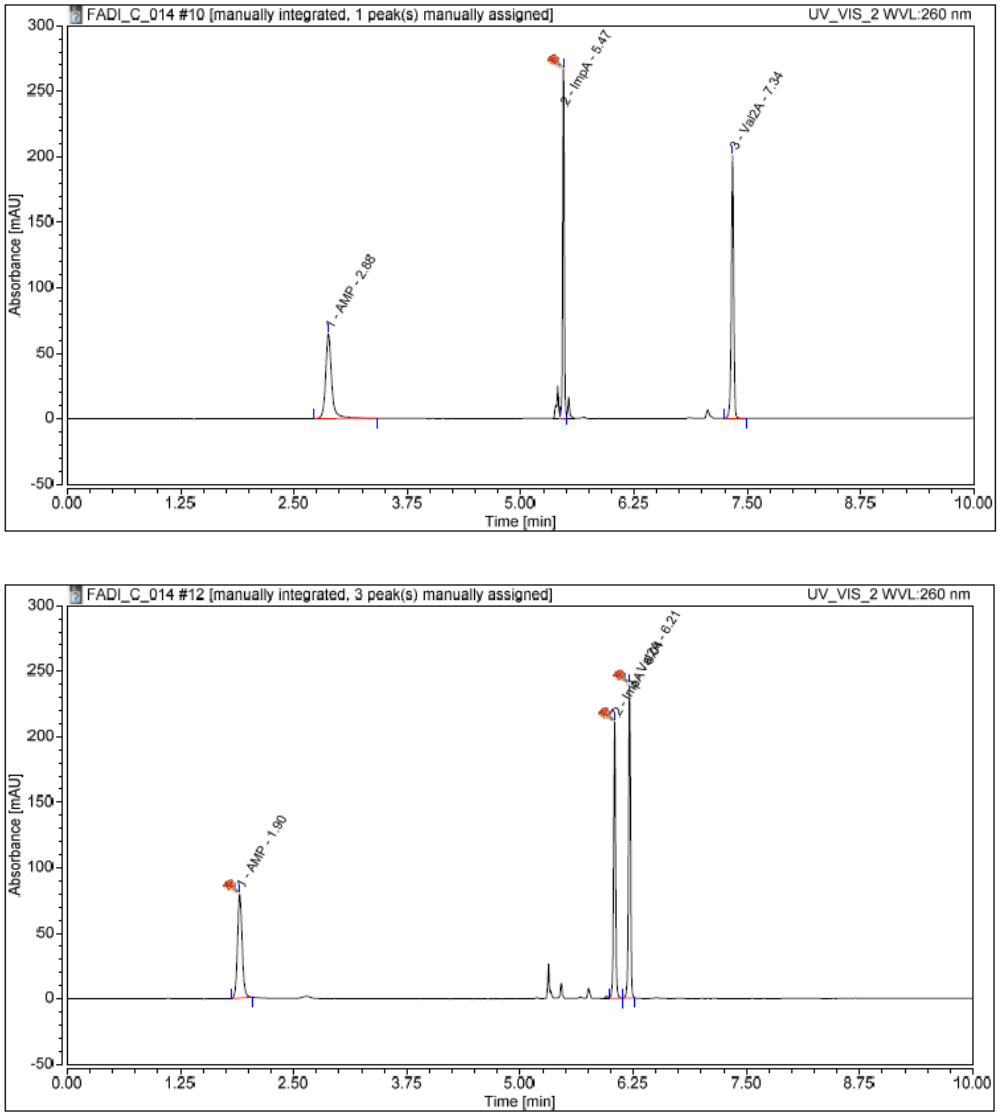


Figure 3.4 Analysis of a crude aliquot from a Val_2A synthesis by UHPLC on a Phenomenex Luna Omega Polar C18 3µm 3×150 mm column, 0.65 mL/min, detection at 260 nm. Top: using 0.1% formic acid in water/acetonitrile; bottom: using 10 mM ammonium acetate pH 7.0/acetonitrile. The lower resolution and retention times under neutral conditions were clearly visible.

3.1.2.2. Anion exchange chromatography

Strong anion-exchange chromatography (SAX) is a routine analytical and purification technique for relatively short oligonucleotides, and a good alternative

to gel electrophoresis, especially when large amounts must be purified. Typical SAX columns are polymer-based materials functionalized with quaternary ammonium groups that retain anions, with little influence of other retention mechanisms. We attempted to purify Gly_2A using a typical oligonucleotide purification column (Dinoex DNAPac 100, see general methods in the appendices). We reasoned that, at a pH around 5, the excess of AMP would be present as a single anion, whereas Gly_2A would be a dianion. While the excess of peptide and coupling ancillaries (notably HOBt) were efficiently washed out at the beginning of the chromatography, the resolution between AMP and Gly_2A was less satisfying. Moreover, AppA, a dianion, co-eluted with the desired product (Figure 3.5), albeit its limited formation (typically around 3%) may be tolerable in some cases. Another drawback of SAX was the use of concentrated salts as eluents, typically sodium chloride or sodium perchlorate. These salts must later be separated from the desired products, which represents additional work and expenses, and a potential source of difficulties when working with highly polar species that cannot be easily desalted by reverse-phase procedures. Volatile salts, usually ammonium formate, were sometimes used to avoid a desalting step, but the resolution suffered a lot from this substitution.[221]

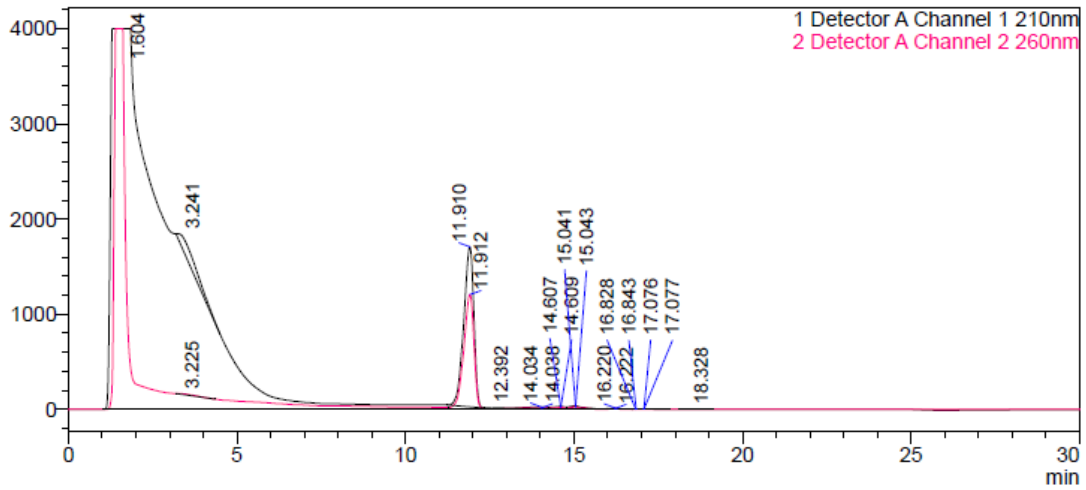


Figure 3.5 Preparative SAX chromatography of a crude sample of Gly_2A, eluting with 0-100 mM sodium perchlorate in 10 mM ammonium formate adjusted to pH 5. The desired product co-eluted with AppA at 12 minutes.

It must be noted that the Dionex DNAPac columns are primarily marketed for the purification of medium oligonucleotides, with enough negative charges to induce an efficient electrostatic retention.[433] The columns developed for single-charge interactions are typically dedicated to inorganic ion analysis in water and may not be appropriate for organic species. We could not test these in this work. Columns that combine ion-exchange and lipophilic retention on the same support are available under the generic name "mixed-mode" and have found niche applications in drug analysis. However, these columns are still uncommon. Most manufacturers only provide them in analytical sizes and as SPE cartridges for sample preparation.

3.1.2.3. Ion-pairing chromatography

An alternative to ion-exchange chromatography is ion-pairing. This strategy became popular for the analytical separation of highly charged species such as oligonucleotides in LC/MS setups, where the salt concentration must be kept to a minimum to preserve the mass spectrometer's performance and integrity. It involves the use of reverse-phase columns (usually C18) and eluents that contain an ion-pair that strongly binds to the stationary phase, enabling temporary ionic interactions between the analytes and the column. This retention mode does not require strongly saline eluents. In a sense, this technique exploits a concept similar to mixed-mode columns: combining a hydrophobic retention mode to an ionic interaction. The application field of this type of chromatography is wide, as both stationary and mobile phases can be varied. It may also help when samples contain species with very different properties. Indeed, a formulation of double-stranded oligonucleotides (small interfering RNA) in phospholipid liposomes was successfully analyzed in a single run using a phenylhexyl HPLC column eluted with dibutylammonium acetate as the ion pairing agent.[275] The weak retention of the phenylhexyl phase was suitable for phospholipid separation under 100% acetonitrile elution, while the ion pair provided sufficient retention for the siRNA.

For oligonucleotides, triethylammonium salts (usually acetates or bicarbonates) are frequently used for both LC/MS analysis and preparative chromatography. We have found that the smallest peptido-nucleotides could be purified by this technique to a better extent than simple reverse-phase chromatography. ValA was purified manually on a hand-packed, coarse (50 μm) C18 silica column, eluting with 50 mM triethylammonium acetate and acetonitrile. The desired product was recovered with a purity greater than 90%, the contaminants being AMP and AppA. In contrast, classical reverse-phase HPLC typically returned ValA polluted with 30% AMP under the same conditions. A desalting step was required to exchange the triethylammonium counterions. A precipitation from 0.1 M sodium perchlorate in acetone has been found to be efficient for this purpose. Applying ion pairing to preparative HPLC may lead to even more satisfying levels of purity, but a dedicated HPLC column would be needed, as triethylamine tends to bind irreversibly to the stationary phase. This could not be tested in this work.

3.1.3. Conclusions

To prepare our library of small peptido-nucleotides, we chose to avoid protecting groups and to work under aqueous conditions as often as possible. This led to efficient synthetic routes involving the mild pre-activation of nucleotides with water-compatible leaving groups, followed by a substitution in water. This strategy was very efficient when soluble peptides were employed. Less soluble peptides still afforded decent yields, but the reaction times became impractically long when high dilutions could not be avoided. Single amino acids afforded inferior yields.

Several chromatographic techniques were tested for the purification step. The challenge of separating very polar species could be partially overcome by using ion pairing eluents or dedicated polar stationary phases. However, efficient methods to obtain high-purity samples of the smallest, most polar peptido-nucleotides still lack.

3.2. On the role of membranes in oligomerization and conjugation of oligomers

The abiotic emergence of peptides and short oligonucleotides have often been treated separately. Abiotic pathways to both amino acids and ribonucleic acids or alternative nucleic acids have been proposed, and their homo-oligomerization has attracted considerable efforts. However, it is unlikely that a living world may have emerged based on the sole intervention of one of these classes. It has been proposed that primitive cell-like systems have arisen from the separate formation of lipid membranes and oligomers, later joined through encapsulation. In such scenario, primitive lipids would self-organize to give stable vesicles in water, and solutes including biomolecules would get entrapped inside the lumen of such vesicles. This entrapment may be spontaneous or occur thanks to membrane permeation. In this work, we reason that lipids may have played a role earlier than during the encapsulation step, by providing a privileged environment to certain substrates and reactions. We investigated the possibility of performing oligomerization and conjugation reactions in the presence of lipid membranes, with the goal of bringing more efficiency and some selectivity to the otherwise random formation of oligomers.

For this purpose, two main chemical reactions were studied: the first is a complex model in which nucleotides and amino acids were activated together in solution, leading to mixtures of (co)-oligomers. This reaction network was rather difficult to track properly because of its complexity, the variety of products and by-products generated in a single run. We later turned ourselves to simpler conjugation reactions, in which a nucleotide was coupled to a small peptide to form a phosphoramidate, following the pre-activation strategy described in section 3.1.1.2. This reaction could be tracked more efficiently, and we varied the conditions to assess whether some reactions would be sensitive to the presence of lipid membranes. The main novelties of this work compared to the published examples of membrane-promoted reactions are that we deliberately avoided the cases of reactions that did not perform at all in water, and substrates with very low water solubility. These cases were already documented to some extent. In contrast, we were interested in general effects applicable to a broader range of chemistry relevant to the emergence of cellular life.

3.2.1. Model oligomerization and conjugation reactions

3.2.1.1. Peptido-RNA arises from random co-oligomerization reactions

The group of Richert reported the formation of a large variety of oligomers and small molecules when a solution of nucleotides and amino acids was activated by a carbodiimide coupling agent.[27] They proposed a “general condensation buffer”, able to promote these reactions in an efficient way, consisting of EDC·HCl as the

activator, 1-ethylimidazole as a crucial partner, in a variety of saline buffered solutions at near-neutral pH. Under these conditions, peptido-nucleotides were formed, accompanied by pyrophosphates, small amounts of phosphodiesters and free peptides, and many by-products arising from the side-reactivity of EDC, or from couplings occurring with organic buffers.[31] Cyanamide provided a similar reactivity pattern but with a much weaker activation. The reaction times were therefore extended to months. When 1-ethylimidazole was omitted, the overall efficiency of the reaction network went down because of the fast hydrolysis of the EDC-activated intermediates.[32]

We used this "general condensation reaction" network starting with AMP and L-valine as model substrates. We chose valine as a relatively lipophilic amino acid, with no side chain functions and a sufficient solubility to maintain homogenous conditions throughout the reaction. Furthermore, the screening studies of the Richert group showed that valine exhibited a decent reactivity in the general condensation buffer.[31] We hoped that the poly(valine) peptides and peptido-nucleotides that would be generated in the course of the reaction could to interact with lipid membranes and that selectivity patterns would be observed.

3.2.1.2. *N*-phosphorylation and the reactivity of amino acids

A key element in the formation of peptides by the carbodiimide chemistry is that the activation of the carboxylate is in competition with the high nucleophilicity of the amine, that forms guanidines with EDC or cyanamide. As an example, 1-^{13}C glycine was activated using the general condensation buffer (EDC·HCl, 1-ethyimidadole, $MgCl_2$ in MOPS buffer at pH 7.5, see below for further discussion) and its oligomerization was followed by ^{13}C NMR (Figure 3.6). After 24h of reaction, the integration of the labeled carboxyl peaks revealed that glycine was essentially converted to N-acylureas (38 %) and guanidine (42 %). Oligomerization represented less than 20% of the conversion.[a]

Locking this nucleophilicity or taking advantage of it is therefore essential to the formation of long peptides. Pascal and coll. demonstrated that *N*-acetylated or *N*-benzoylated amino acids were converted to 5-(*4H*)-oxazolone intermediates that further reacted to form peptide bonds, with the drawback that racemization would occur in this intermediate form.[162] The *N*-phosphorylation of amino acids to give phosphoramidates is an alternative pathway to solve the *N*-terminus problem, with the double advantage that the intermediate has a low susceptibility to racemization and that the peptide may be cleaved from the phosphate source

[a] Here we assume that the integration of ^{13}C NMR peaks gives a quantitative overview of the distribution of species. In principle, such assumption is only verified if a full relaxation of every spin system is reached, either by increasing the relaxation delay, which lengthens the analyses considerably, or by adding paramagnetic relaxation promoters. Richert and coll. used this last approach.[32] In the specific case of this example, we assume that the relaxation of all the chemically-similar carbon atoms that we considered was equivalent, although we could not find exhaustive data on the topic.

under mild conditions,[28] whereas the oxazolone pathway provides acylated peptides (Scheme 3.5).

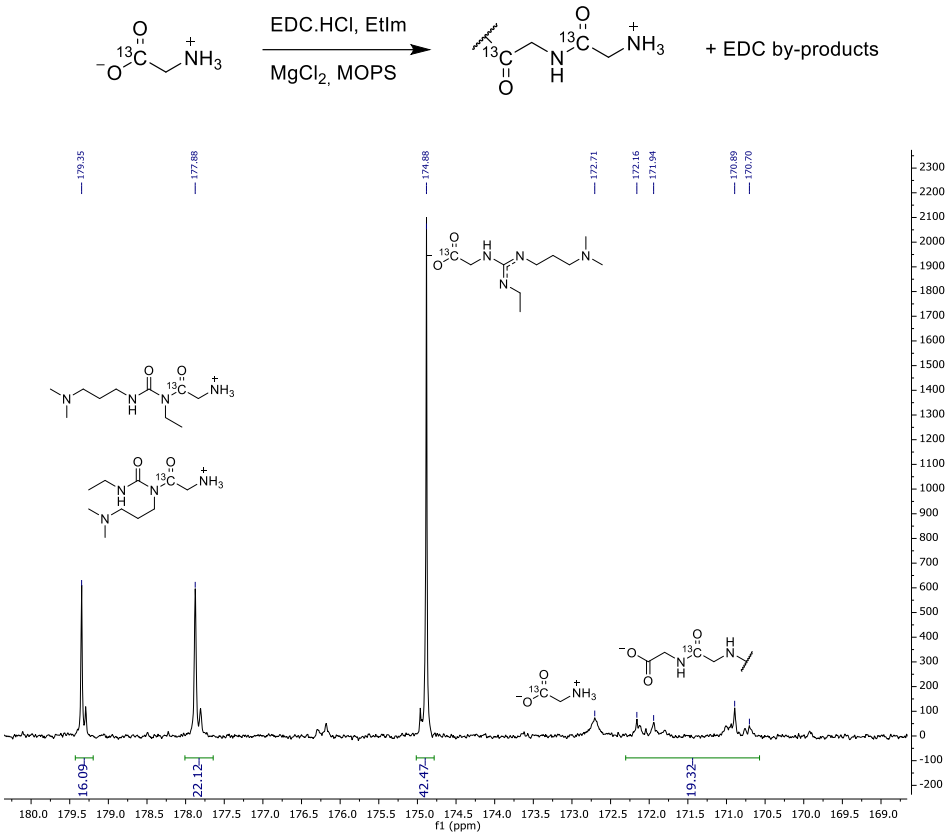
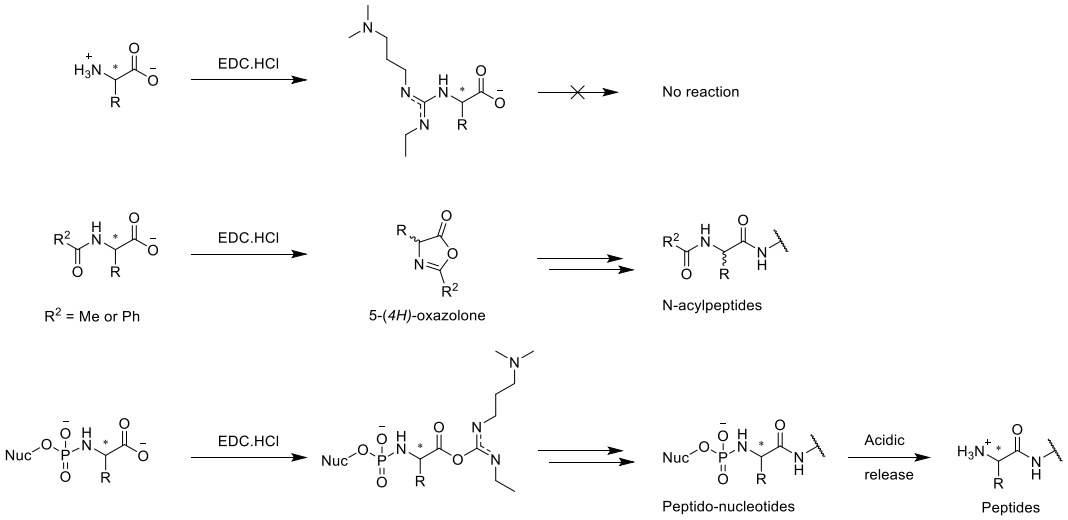


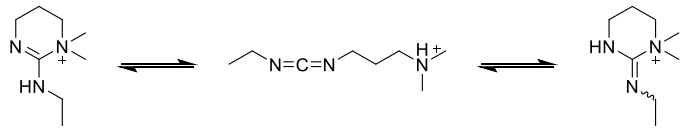
Figure 3.6 ^{13}C NMR spectrum (H_2O, 100 MHz, 32 scans) of 1-^{13}C glycine in the general condensation buffer after 24 hours. Less than 10% glycine (172.7 ppm) remained. The integration of the other signals was summed up to 100.



Scheme 3.5. Top: amino acids give unreactive guanidines with EDC. Middle: N-acetylated or N-benzoylated amino acids give racemic N-acylated peptides through 5-(4H)-oxazolone intermediates. Bottom: N-phosphorylated amino acids yield peptido-nucleotides with preserved stereochemistry. The weak phosphoramidate bond can be cleaved under slightly acidic conditions to release the peptide. Nuc = nucleoside

3.2.1.3. Aqueous carbodiimide chemistry: by-products and efficiency

EDC is a complex activating agent and its reactivity has been studied in detail. Its superior stability in neutral aqueous solutions ($t_{1/2}$ = 37 h at pH 7 and 25°C[434]) may be tracked down to the existence of cyclic guanidinium tautomers (Scheme 3.6).[435] However, it is much less stable under slightly acidic conditions ($t_{1/2}$ = 3.9 h at pH 5 and 25°C) and some salts are known to accelerate its hydrolysis, in particular inorganic phosphate ($t_{1/2}$ = 3.7 h at pH 7 and 25°C in the presence of 10 mM Pi, this value drops down to 0.4 hours when 100 mM Pi are present).[434] This result also suggests that EDC benefits from a privileged reactivity toward phosphates, that partially explains the very favorable formation of phosphoramidates in the general condensation buffer.



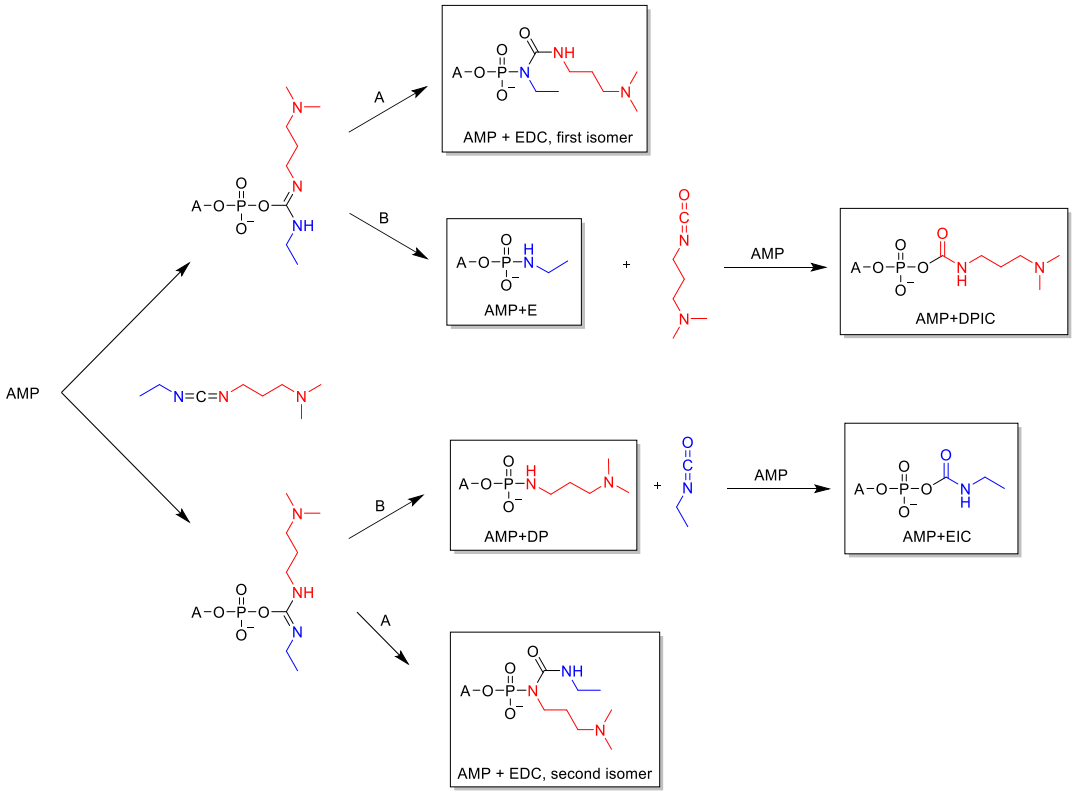
Scheme 3.6 EDC exists in equilibrium between the linear form and several cyclic tautomers that provide additional stability in neutral aqueous solutions.

Besides the formation of guanidines, a common side-reactivity of EDC is the rearrangement of the activated intermediate *O*-acylurea in the very stable and undesirable *N*-acylureas. This happens to larger extents when the coupling reaction that must be performed is slow, and that the intermediate is accumulated. The addition of a nucleophilic catalyst is often required to circumvent this problem. Despite the use of EtIm, the general condensation reaction still yields important amounts of *N*-acyl and *N*-phosphorylureas arising from activated carboxylates and phosphates. The accumulation of these products represents an additional analytical burden.

To obtain an overview of the diversity of by-products that are formed in a typical condensation reaction, we ran a coupling involving valine and AMP in the general condensation buffer (200 mM AMP, 200 mM valine, 800 mM EDC·HCl, 150 mM 1-ethylimidazole, 80 mM $MgCl_2$, 500 mM HEPES adjusted to pH 7.5, kept at 0°C for 7 days) and we analyzed the outcome by LC/MS/MS (detailed data in appendices). In addition to the known by-products described above, we discovered that other non-canonical species were present in the sample. The mechanism of their formation is not yet solved but hypotheses can be advanced (Scheme 3.7). *N*-phosphorylureas formed from the rearrangement of EDC-activated AMP were found with an increase in mass of 155.14 Da that corresponded to one EDC unit. Two isomers of this compound were found (Scheme 3.7, pathways A).

We further identified increases in mass of 128.10 Da and 71.03 Da that corresponded to fragments of EDC with the respective formulas $C_6H_{12}N_2O$ and C_3H_5NO. These fragments were troublesome at first, because their formulas perfectly matched lysine and alanine residues, two amino acids that were not used in this assay. Each of these products was accompanied by a formally

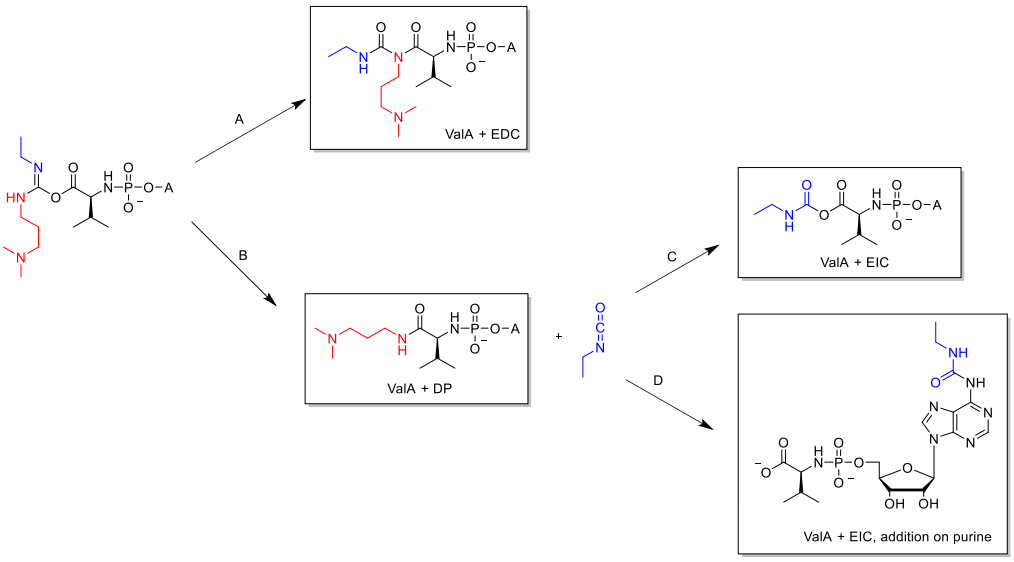
decarboxylated equivalent. As an EDC unit must be broken in halves to form such fragments, we hypothesize that they were formed simultaneously from a common intermediate. Seel and Klein reported the formation of carbamoylphosphates from the reaction of inorganic phosphate with isocyanates in water and observed that these species were relatively stable.[436] The direct splitting of a unit of EDC or EDU into an amine and an isocyanate is unlikely but the rearrangement of *O*-phosphorylureas may lead to a phosphoramidate and an isocyanate (Scheme 3.7, pathways B) instead of the more conventional *N*-phosphorylurea. This would explain both the formally decarboxylated compounds (AMP + $C_6H_{12}N_2O$ – CO_2, dubbed AMP + DP or dimethylaminopropylamido, and AMP + C_3H_5NO – CO_2, dubbed AMP + E or ethylamido) and the simultaneous release of ethyl or dimethylaminopropyl isocyanate may yield the compounds of formula AMP + $C_6H_{12}N_2O$ (AMP + DPIC, dimethylaminopropyl isocyanate) and AMP + C_3H_5NO (AMP + EIC, ethyl isocyanate).



Scheme 3.7 Mechanistic proposition for the formation of by-products from EDC-activated nucleotides. Pathway A represents the usual rearrangement leading to stable N-phosphorylureas. Pathway B was postulated to yield stable phosphoramidates and isocyanates that further reacted to give carbamoyl phosphates. Each pathway could apply to the two isomers of EDC-activated AMP, leading to a total of 6 by-products. All the corresponding species have been observed in LC/MS analysis.

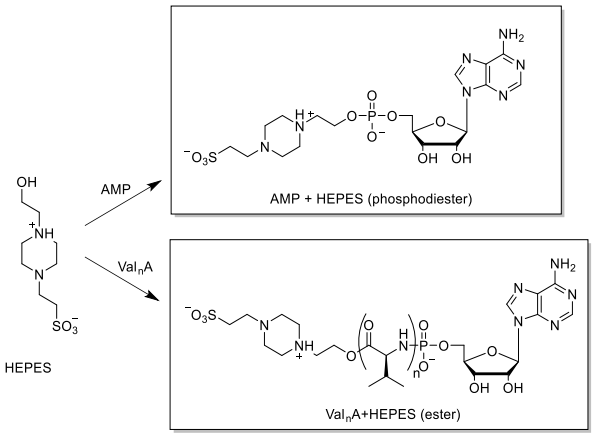
Similar mass increases were detected on products that contained one or more valine residues. The formation of *N*-acylureas can be explained by a usual rearrangement similarly to pathway A. The types of addition products arising from the pathway B are more difficult to rationalize in carboxylate chemistry. Indeed, an ethyl- or dimethylaminopropylamide of valine would be a stable species, giving a consistent explanation of products of the type ValA + DP and

ValA + E, but the addition of a carboxylate on an isocyanate would generate a relatively unstable anhydride. It is then possible that the species of the type ValA + DPIC and ValA + EIC underwent an isocyanate addition to the purine, forming a urea, rather than on the carboxylate (Scheme 3.8). This was further supported by MS/MS data, from which the loss of $C_6H_{12}N_2O$ or C_3H_5NO was drawn either from the molecular peak or from the purine fragment.



Scheme 3.8 Mechanistic proposition for the formation of by-products from EDC-activated ValA (only one isomer is drawn for clarity, the same pathways apply to the other isomer). Pathway A is the known N-acylurea rearrangement. Pathway B may lead to a stable amide and to an isocyanate that would be trapped either by the carboxylate end (pathway C) or by the purine (pathway D).

The general condensation buffer proposed by Richert and coll. originally used HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonate) to regulate the pH around 7.5. This organic buffer possesses an alcohol function that reacted under the coupling conditions to yield phosphodiesters with AMP and esters on the valine residues (Scheme 3.9). In addition to represent another set of by-products, these reactions terminated the oligomerization process and may have led to an underestimation of the growth capacities permitted by these conditions.



Scheme 3.9 The reaction of HEPES with AMP generates stable phosphodiesters. The reaction of HEPES with species of the type Val$_n$A terminates the peptide growth by forming stable esters.

We counted the number of species displaying canonical structures, derived from EDC rearrangements, HEPES couplings and other side-reactions (Table 3.1). As quantitative information was difficult to extract from crowded chromatograms, this analysis was only qualitative and provided insight on the diversity of the products formed during the coupling, rather than on the molar fraction that they represented.

Type of species	Occurrences	Frequency (%)
Canonical	22	42
+ EDC	8	15
+ EDC fragments	7	13
+ HEPES	7	13
Others	8	15
Total	52	100

Table 3.5 Repartition of the products obtained from the reaction of AMP and valine in the general condensation buffer. Canonical products had the expected general formula Val_nA_p, "+ EDC" products arose from the rearrangement of EDC-activated intermediates as N-acyl or N-phosphorylureas, "+EDC fragments" products were described in the paragraph above, "+ HEPES" products contained a HEPES unit in their structure. Other products contained non-listed modifications, such as the loss of an extra molecule of water, or could not be elucidated from their MS/MS data.

It was interesting to note the very large number of non-canonical species obtained in this assay. HEPES by-products may be avoided easily by the use of another buffering system but the products inherent to EDC were more problematic, and the use of 1-ethylimidazole only partially prevented their formation. The qualitative information acquired on each compound by LC/MS was coupled with quantitative data obtained from ^{31}P NMR (Figure 3.7). As most of the chemistry of peptido-RNA is centered around one or more phosphorus atoms, ^{31}P NMR offered an efficient and reliable way to quantify the outcome of general condensation reactions by sorting the products by general structure. After having measured the relaxation of the species of interest, we routinely set the relaxation delay `d1`[a] to 10 seconds. Proton decoupling was always applied using the inverse gate technique (`zgig`). Correlations with ^{1}H NMR using a HMBC sequence facilitated the attribution of the peaks. It must be noted, however, that the ^{1}H spectra of such mixtures were very difficult to read because the signals of interest were often very weak in comparison to those arising from EDU and HEPES (representative ^{1}H-^{31}P HMBC in appendices, Figure S5). Furthermore, the high concentration of ionic compounds in the sample resulted in a viscous, highly conductive analyte that had a deleterious effect on pulse calibration and shimming procedures.

[a] NMR spectroscopy was performed on various Bruker systems using the Topspin 3.5 software line as detailed in the appendices. Since most laboratories use Bruker NMR system, we found useful to include the specific name of delays, pulse programs and other essential technical information, as referred to in Topspin. To distinguish these manufacturer-specific names from general NMR jargon, they are written in `console typeset`.

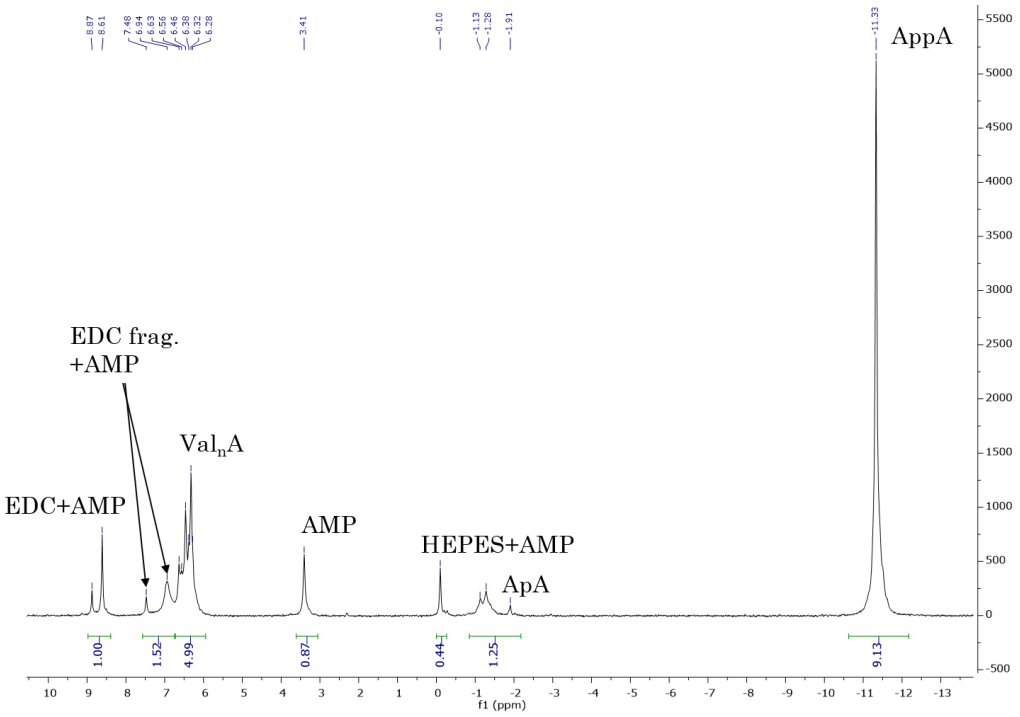


Figure 3.7 ^{31}P NMR spectrum (H_2O, 200 MHz) of a crude general condensation reaction involving valine and AMP after 7 days. Each peak represented a variety of species of similar structure, rather than one species. The attributions were based on literature data and ^{1}H-^{31}P HMBC acquisitions.

The quantitative NMR analysis (Table 3.6) highlighted the formation of large amounts of pyrophosphate, a product that could not be quantified directly, nor distinguished from regular 2'-5' and 3'-5' dinucleotides by LC/MS. It also revealed that the diversity of by-products observed by LC/MS actually did not represent an overwhelming mole fraction, as more than 75 mol% of phosphorylated products had a canonical structure.[a] It must be noted, however, that side reactions happening at the carboxylate end of peptido-nucleotides were not covered by ^{31}P NMR. These products may be detected by ^{13}C NMR, but we did not engage in this task. The Richert group used this technique to quantitatively follow the formation of peptido-nucleotides and *N*-acylureas.[32]

Overall, the general condensation reaction appeared to be relatively complex considering the apparently simple reaction conditions. Its intrinsic qualities include a new approach to the carbodiimide-promoted condensation of amino acids, the formation of a variety of stable or labile chemical bonds in one pot and the spontaneous emergence of a joint peptide-RNA chemistry that knew relatively little detailed precedents. The extent of the formation of by-products represented

[a] When AppA and unreacted AMP were counted as canonical products. When considering only the fraction of new products that were not pyrophosphates, the ratio dropped to 68% canonical products. In their quantitative evaluation of the efficiency of the general condensation network, Tremmel and coll. counted pyrophosphates as "useful bonds",[32] echoing their original report that pyrophosphate-containing cofactors such as NAD were formed under the general condensation conditions.[27] However, one may argue about the pertinence of the domination that pyrophosphates exert on other species in some contexts, since the "usefulness" of peptido-RNA seems generally higher than that of pyrophosphate cofactors.

an analytical challenge but waste generation and side-reactivity are inherent to systems chemistry, and having these reactions directly integrated in the network may be an interesting model of complex chemical systems in which the energy must be guided toward desired, rather than parasite, reactions.

Type of species	Mole fraction (%)
AMP (unreacted)	6
Peptido-nucleotides	34
Phosphodiesters	8
AppA	31
Total canonical products	79
HEPES phosphodiester	3
N-phosphorylureas	7
Phosphoramidates from EDC	10
Total side-products	20

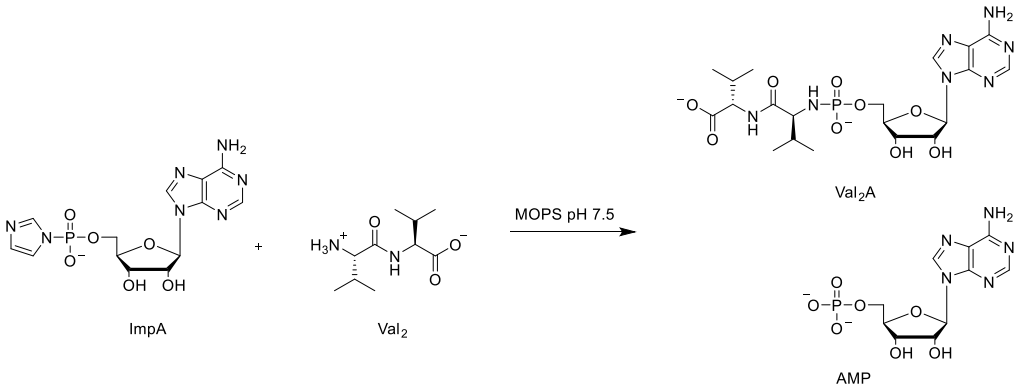
Table 3.6 Quantitative data extracted by relative integration from ^{31}P NMR analysis of a crude general condensation reaction involving valine and AMP after 7 days.

3.2.1.4. Pre-activated nucleotides and displacement reactions

A simpler model, involving fewer chemical reactions than the general condensation reaction, was also desirable to investigate the effect of lipids on the formation of amphiphilic peptido-nucleotides. We chose to focus on the formation of the phosphoramidate, because separate peptide and RNA growth were already studied in-depth by others, and because phosphoramidates form more readily than many other condensation products. Our model reaction (Scheme 3.10) involved the substitution of imidazolide-activated AMP (ImpA) with the dipeptide Val_2, that was a compromise between lipophilicity and aqueous solubility. The reaction was carried out at a pH of 7.5 using MOPS (3-morpholinepropane-1-sulfonate) as a buffer instead of HEPES, to avoid the formation of HEPES phosphodiesters. Reasonable concentrations of MOPS (typically twice the concentration of each reagent) were sufficient to avoid any pH drift during the reaction. ImpA was produced with satisfying purity as described in section 3.1.2.2. The simple nature of the process allowed us to monitor the reactions by HPLC, with a UV detector set up at 260 nm. It was assumed that all the species of interest (ImpA, AMP and Val_2A) had identical absorbance coefficients and the peak areas were used to establish the relative quantitation of each species. AppA, that only formed in small amounts under these conditions, was excluded from this quantification because of its higher absorbance coefficient.[437] No racemization on the phosphorylated Val residue of Val_2A was observed throughout the reaction.

A reference reaction, involving 50 mM ImpA and 50 mM Val_2 in 100 mM MOPS at pH 7.5 was allowed to proceed for 2 weeks at 0°C and followed by HPLC. After this time, 22% of the ImpA was converted to Val_2A, while 59% was hydrolyzed to AMP. These values were confirmed by ^{31}P NMR measurements (appendices, Table

S6, Figure S9, Figure S10). When 100 mM imidazole was included in the buffer, the same amount of Val_2A was obtained but only 18% AMP was formed by hydrolysis. After 28 days, the conversion increased to 37% of Val_2A, with 31% of hydrolysis to AMP and 32% remaining ImpA. This illustrated the preservative effect of imidazole on ImpA, an effect that may be echo Richert's imidazolium intermediates.[32]



Scheme 3.10 Model reaction between the active nucleotide ImpA and Val_2 that yielded the desired conjugate Val_2A in competition with hydrolysis to AMP.

3.2.2. The formation of peptido-RNA in the presence of lipid membranes

3.2.2.1. Simplified condensation conditions

We began our investigations on the formation of amphiphilic peptido-RNA in the presence of membranes using the general condensation reaction model. Our two objectives were:

- assessing the possibility of forming stable membranes in reaction mixtures containing concentrated reagents and salts
- evaluating the impact of the presence of lipid aggregates on the efficiency, selectivity, and diversity of the outcome of general condensation reaction.

Because of the numerous by-products and the high amounts of pyrophosphate formed in the general condensation buffer, we simplified the reaction conditions. When $MgCl_2$ was omitted, the pyrophosphate formation was drastically limited. HEPES was also omitted to avoid the formation of the related esters. The reaction mixture was adjusted to pH 7.5 just before adding the coupling agent. Throughout the progress of the reaction, the pH shifted to 8.5. This had no negative impact on the coupling efficiency. In the original report, the reaction was performed at 0°C. Since the formation of lipid membranes by the hydration of a lipid film proceeds only above the transition temperature of the lipid mixture, obtaining membranes at low temperature may be more difficult than in warm environments. We ran the simplified condensation reaction (without $MgCl_2$ and HEPES) at 0°C and 25°C and followed their evolution by ^{31}P NMR. As it could be expected, the reaction running at 25°C was faster and a stable composition was obtained after 5 days,

at which point more than 80% of AMP was consumed to give other phosphorylated species. The reaction at 0°C was almost equilibrated after 12 days and gave a lower overall conversion, with only 55% of consumed AMP. The repartition of the product classes, based on the amount of AMP converted, were measured for each sample. In both cases, pyrophosphate formation was minimal. The formation of phosphodiesters was also very limited, but this point was not crucial for our goal of studying amphiphilic molecules. The reaction at 0°C formed slightly less EDC-derived by-products. At this temperature, ValA was the main species of the peptido-nucleotide population, with limited growth to longer peptides. At room temperature, the overall yield in peptido-nucleotides was slightly lower, with more AMP converted to EDC-derived by-products, but the growth of the peptide chains was slightly more consequent (Table 3.7). With this comparison in hands, we decided to run the assays involving lipid membranes at room temperature.

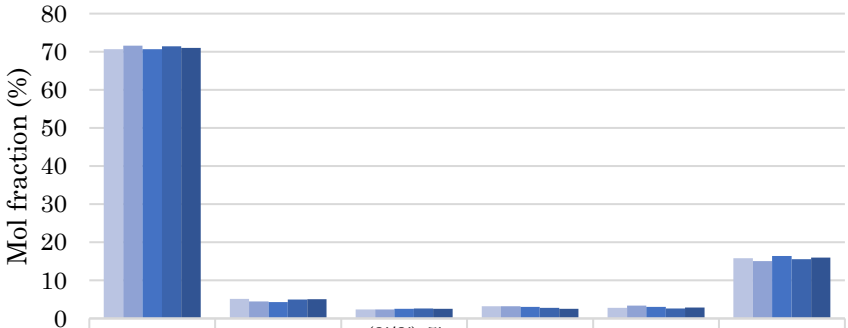
	0°C, 12 days	**25°C, 5 days**
AMP consumed (%)	55	82
Product repartition		
ValA	46	29
Peptido-nucleotides	35	44
Phosphodiesters	<1	2
AppA	<2	<2
Total side-products	16	23

Table 3.7 Quantitative description of a simplified condensation reaction involving valine and AMP at 0°C and 25°C, as determined by ^{31}P NMR integrals (data in appendices). The product repartition was calculated on the reacted fraction only, by summing the integrals of all the peaks with the exception of AMP to 100.

3.2.2.2. Simplified condensation reaction in the presence of a gradient of lipid concentration

The simplified condensation reaction between AMP and valine was carried out in the presence of an increasing concentration of lipids. DOPA (1,2-dioleoyl-*sn*-glycero-3-phosphate) was chosen as a model lipid. The reagents were dissolved in water (200 mM valine, 200 mM AMP, 800 mM EDC·HCl, 150 mM 1-ethylimidazole) then added to a thin film of lipid in a glass tube. Up to 5 mM DOPA was used, as higher concentrations resulted in thick films that did not hydrate properly, and therefore did not come in contact with the reagents. The reaction was allowed to proceed without stirring, and the quantitative outcome was determined by ^{31}P NMR after 7 days. As the samples containing lipids were inhomogeneous and therefore difficult to analyze by NMR, MeOD was added until a clear solution was obtained. A control reaction that involved no lipid was prepared under the same conditions and analyzed by the same means. The oligomerization performed well under these conditions, and high yields of peptido-nucleotides were obtained. However, the presence of the lipid film had no significant influence on the outcome of the reactions. All the samples, including

the control experiment, gave very similar results (Graph 3.1), suggesting that the DOPA film had no impact on the outcome of the reaction.



	ValnAp	Val1Ap	(2'/3')-5' phosphodiester	Pyrophosphate	AMP	By-products
0	70,7	5,17	2,36	3,22	2,78	15,77
0.5	71,6	4,44	2,37	3,2	3,35	15,04
1	70,70	4,27	2,5	3,01	3,08	16,44
2	71,4	4,98	2,63	2,76	2,66	15,57
5	71	5,09	2,55	2,54	2,86	15,96

Graph 3.1 Quantitative distribution of species obtained from ^{31}P NMR integrals for a simplified condensation reaction performed in the presence of 0 to 5 mM DOPA for 7 days at 25°C.

3.2.2.3. Simplified condensation reaction in the presence of evolutionary lipid mixtures

To determine whether the absence of effect of the lipid films was specific to DOPA, we repeated a similar experiment using a scope of lipid compositions. In a recent publication, our group demonstrated that 1,2-dioleoylglycerol could be phosphorylated under simple thermal conditions to yield racemic phospholipids, including DOPA (using Pi as the phosphate source) and DOPE (using *O*-phosphoethanolamine), with cyanamide as a promoter. Several mixtures of lipids, representative of increasing evolutionary stages from fatty acids from phospholipids, gave vesicles upon hydration with aqueous buffers (Table 3.8, top and Figure 3.8).[99] The fact that both lipid synthesis and the general condensation may be promoted by cyanamide (to a lesser extent for the second one) prefigured a scenario in which many different classes of molecules may have evolved in parallel using similar resources, to eventually constitute complex systems. We decided to investigate the general condensation reaction over lipid films with compositions ranging from oleic-rich mixtures to phospholipid-rich mixtures (Table 3.8, bottom). An overall concentration of 2 mM of lipids, that was typical for vesicle formation, was used.

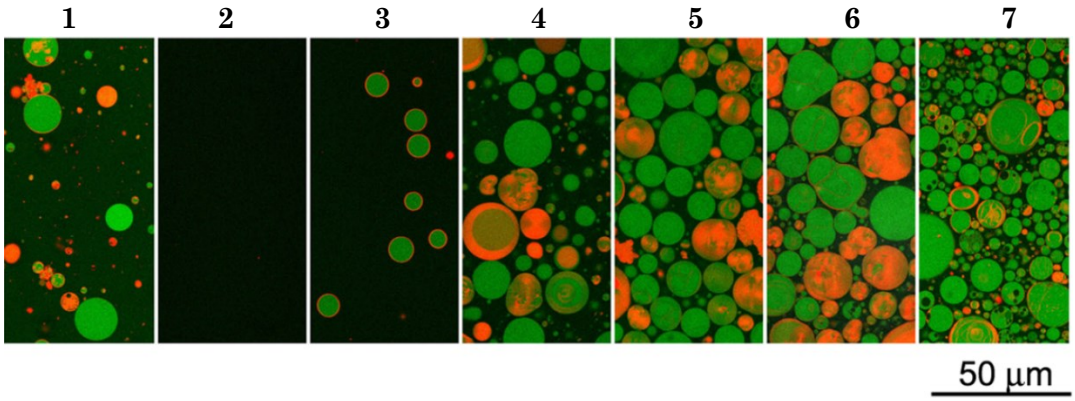
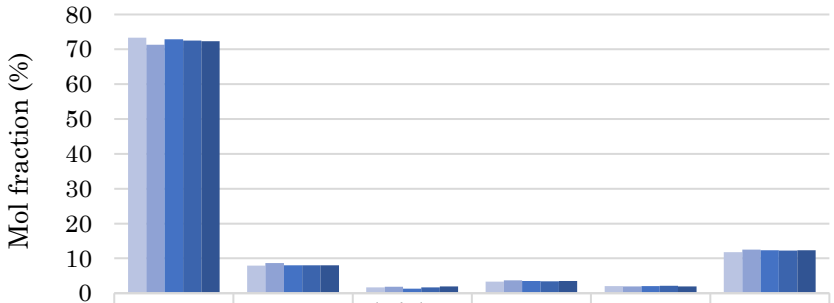


Figure 3.8 Vesicles and other aggregates obtained from the hydration of thin films of potentially prebiotic mixtures of lipids in the report of Fayolle and coll, as observed by fluorescence confocal microscopy.[99] The compositions are given in the table below the picture. The lipid mixtures were supplemented with 0.15 mol% rhodamine-labeled DOPE (orange channel) then hydrated with buffers incorporating 2 µM calcein (green channel) and separated from the fluorescent background by the sucrose-glucose gradient method as detailed in the publication. Figure under Creative Commons License.

Mixture	Oleic acid	MOG	DOG	DOPA	DOPC
Lipid mixtures in Fayolle et al.					
1	8	2	-	-	-
2	6	2	2	-	-
3	4	2	2	2	-
4	2	2	2	4	-
5	-	2	2	6	-
6	-	-	2	8	-
7	-	-	-	10	-
Lipid mixtures used in this work					
A	8	2	-	-	-
B	1	1	2	6	-
C	-	-	-	10	-
D	-	-	-	3	7

Table 3.8 Lipid mixtures (molar fractions) used to prepare the vesicles depicted above by Fayolle et al (top); and in the simplified condensation reaction in this work (bottom). MOG: 1-oleoylglycerol; DOG: 1,2-dioleoylglycerol; DOPA: 1,2-dioleoyl-glycero-1-phosphate; DOPC: 1,2-dioleoyl-glycero-1-phosphatidylcholine. The total amount of lipids was 2 mM in each experiment.

Following the same protocol as for DOPA, simplified condensation reactions were allowed to proceed over the lipid films for 7 days at 25°C. After this time, 25 µL aliquots of each sample were kept for microscope observation and the rest was homogenized with MeOD for quantitative NMR analysis. The quantitative outcome of each reaction was very similar to what was observed with DOPA. Peptido-nucleotides were formed in good yields, but the presence of the lipids induced no visible impact on the reaction efficiency, nor selectivity among the products (Graph 3.2).



	ValnAp	Val1Ap	(2'/3')-5' phospho diester	Pyropho sphate	AMP	By-products
Control	73,3	7,9	1,66	3,35	2,02	11,77
A	71,3	8,67	1,86	3,67	1,96	12,54
B	72,90	7,97	1,32	3,5	2,02	12,29
C	72,5	8,06	1,67	3,44	2,1	12,23
D	72,35	7,98	1,95	3,5	1,93	12,29

Graph 3.2 Quantitative distribution of species obtained from ^{31}P NMR integrals for a simplified condensation reaction performed in the presence of lipid mixtures A, B, C, D and control (no lipid) for 7 days at 25°C.

Samples from the crude reaction mixtures were observed by optical microscopy after staining with the membranophilic dye Nile Red™. All the lipid mixtures presented fatty clumps of various sizes, but mixtures A and D also displayed giant vesicles with diameters above 10 μm and thick, massively multilamellar membranes (Figure 3.9). This was especially unexpected in the case of the "early" mixture A, that contains a large fraction of oleic acid. This observation confirms recent observation that "intermediate" and "incomplete" lipids may increase drastically the stability of vesicles against salts and high concentrations of solutes. The reaction performed over DOPA did not give vesicles but long, fiber-like structures that decanted at the bottom of the sample.

The absence of chemical influence of the lipid films may be explained by the overwhelming excess of reagents over the lipid aggregates. Because of this, most reactions would occur in the bulk aqueous phase without interfering with the lipid areas. The lipid concentration could only be increased to a certain point, after which the lipids were observed to agglomerate and stick to the walls of the vessel, and therefore became inaccessible from the aqueous phase. The surface area and the concentration of lipids may be increased by preparing unilamellar vesicles, but this "artificialized" approach would bring us away from our original goal. Another problem was that the simplified general condensation reaction is an efficient networks that relies on concentrated reagents and high-yielding chemical processes. Improvements are more likely to be visible from a more modest system than in such an optimized process. In comparison, most systems in the literature used reactions that did not perform, or not well, in water. Finally, the complexity of the general condensation reaction may be a fundamental strength, but it hampered the readability of the outcome and quantitative evaluation of simple effects may be, at first, difficult.

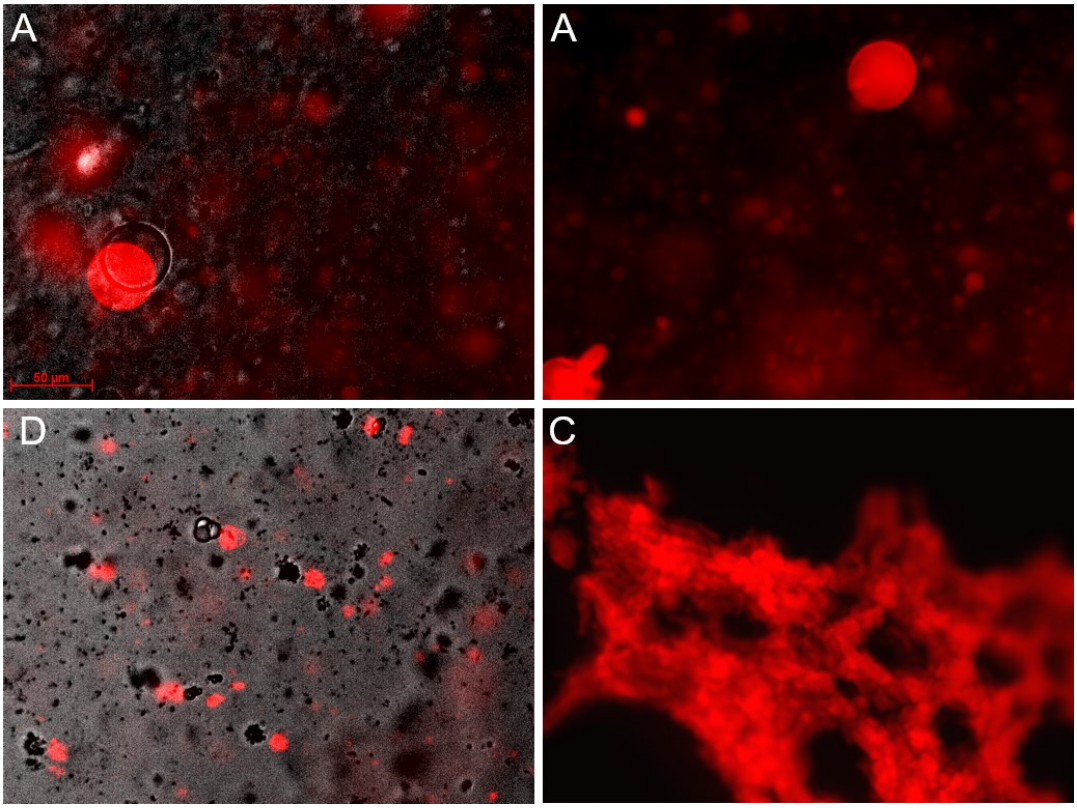


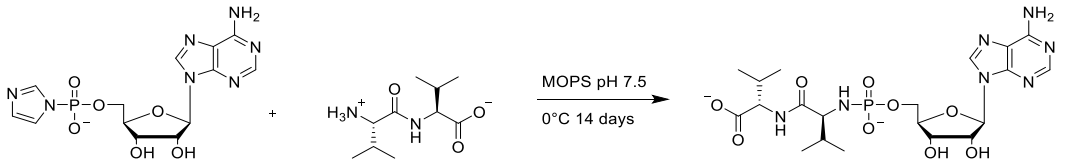
Figure 3.9 Micrographs of aggregates found in samples A, C and D by microscope observation after staining with the membranophilic fluorescent dye Nile Red™. Grayscale channel: optical, red channel: fluorescence. The scale bar applies to all pictures. The lack of superimposition between the channels on the pictures on the left is due to the high mobility of the samples combined with long exposition times.

For these reasons, we reasoned that a simpler model would provide more insight in a first approach. The discoveries made under simple conditions may be later transposed to the general condensation reaction or any complex chemical network. The imidazolide displacement reaction was promising for this task: it involved only two substrates of very distinct lipophilicity; the main process was the formation of an amphiphilic phosphoramidate; one hydrolysis reaction modeled efficiency and competitivity considerations; no significant amounts of by-products were formed.

3.2.3. Peptide-nucleotide conjugation in the presence of lipid membranes

We performed the displacement reaction described in 3.2.1.4 in the presence of various lipid membranes with the hope to unveil quantitative effects of the presence of the latter on the outcome of the reaction (Scheme 3.11). The reaction was performed at 0°C to minimize the rate of hydrolysis of the imidazolide and avoid any potential reactive rush in the first hours of the reaction. In a fast process, the majority of the conversion may indeed be reached before the lipids were properly swollen. As detailed above, we also expected to detect significant differences more easily from a mild, relatively slow reaction than from a fast and very efficient process. We used the modern phospholipid POPC as a model

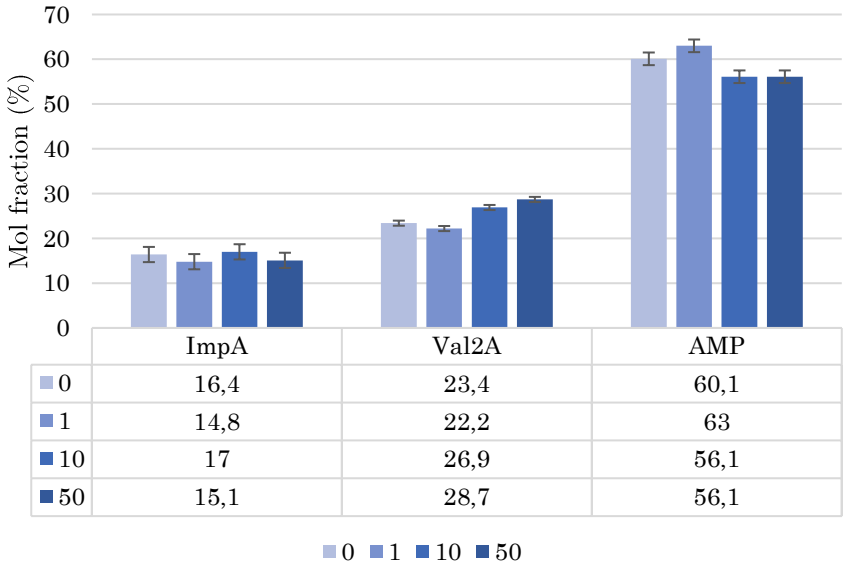
membranogenic compound, as its T_m was compatible with the formation of vesicles at 0°C. As for the general condensation reaction, a thin film of POPC was prepared then hydrated with a freshly prepared solution of reagents and allowed to stand without stirring at first. After 24 hours, the vesicles were formed, and a gentle stirring was applied to prevent the vesicles from floating or sedimenting. The progress of the reaction was monitored regularly by HPLC analysis of dilute samples. The control reactions without lipids and the reactions involving less than 1 mM lipids gave homogenous samples after a 100-fold dilution in water and were injected directly. When more lipids were present, they were removed prior to the injection by adsorption onto a dedicated material.



Scheme 3.11 The model displacement reaction used for quantitative assays in the presence of lipids.

In a first set of experiments, the concentration of both reaction partners was set to 50 mM, in the presence of POPC concentrations of 0 (control), 1, 10 and 50 mM. 1 mM was a typical concentration for the formation of phospholipid vesicles by the natural swelling method. At 10 mM, vesicles were expected to be present as well as larger clumps of lipid material. 50 mM was typically too concentrated for proper vesicle preparation without additional processing and represented a saturation situation. With this experimental setup, no clear effect of the of vesicles on the outcome of the reaction was identified. Under all the conditions, the reaction reached a similar composition after 15 days, with only small differences that may be attributed to experimental factors such as the exact stoichiometry or the integration of the HPLC traces (Graph 3.3). A comparison of the HPLC quantification with the previously used ^{31}P NMR method confirmed that HPLC was reliable for Val_2A and AMP, which were the two main compounds of interest (appendices, Table S6).

To explain the absence of influence of the presence of lipids, we propose that the concentration of the reagents may still be disproportionate for the amount of lipids available. In the most favorable case, one molecule of lipid was present for each molecule of peptide and oligonucleotide. The actual ratio may even be lower, as many lipids were probably aggregated into clumps, thus lowering the amount of lipids actually accessible to the solutes. In such scenario, the formation of Val_2A would have essentially occurred in water, with little regard for the presence of lipids, as this was already suggested from the general condensation reaction.

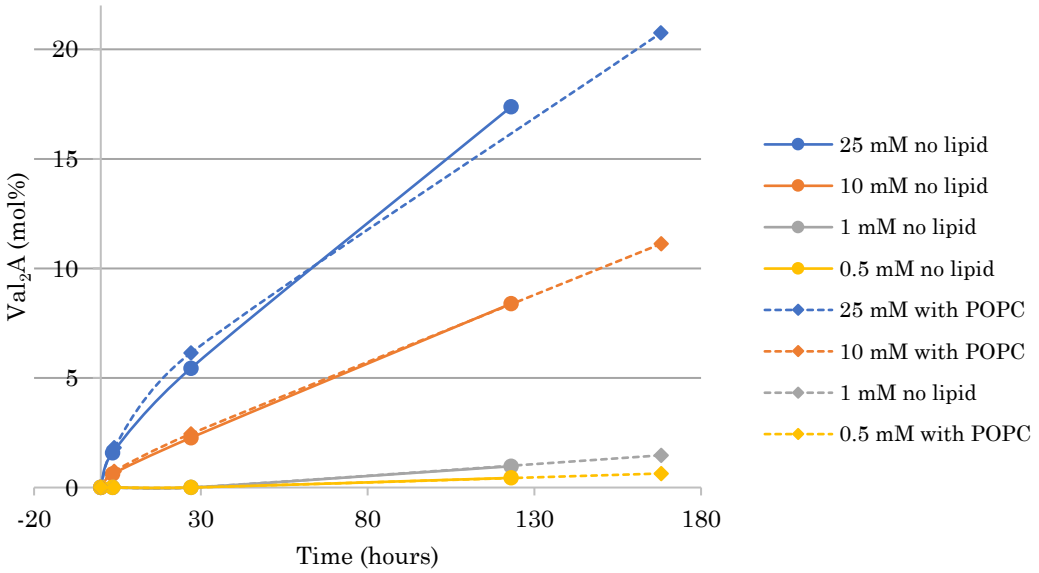


	ImpA	Val2A	AMP
0	16,4	23,4	60,1
1	14,8	22,2	63
10	17	26,9	56,1
50	15,1	28,7	56,1

Graph 3.3 Quantitative outcome of the formation of Val_2A from Val_2 and ImpA in the presence of 0-50 mM POPC for 15 days at 0°C, as determined by HPLC with UV detection at 260 nm. Error bars based on the agreement between UV HPLC and ^{31}P NMR measurements. Higher errors due to a lack of repeatability cannot be excluded.

To access more favorable lipid-to-reagent ratios, we envisioned to dilute the reagents while the concentration of lipids remained fixed. The formation of Val_2A from ImpA and Val_2 was expected to be assimilable to a bimolecular process, and thus the speed of the reaction would be dependent on the initial concentration of the reagents. This effect was confirmed during the synthesis of peptido-nucleotides by the displacement method, where the concentration of the substrates played an important role on the efficiency and the speed of the synthesis (section 3.1.1.2). At the contrary, the hydrolysis of ImpA back to AMP was expected to resemble a monomolecular process and to be independent on the concentration of ImpA. In such setting, the kinetics and efficiency of the reaction would degrade upon successive dilutions. Localization of the substrates to the bilayers may preserve ImpA from hydrolysis and enhance the formation of Val_2A in dilute setups.

We performed another set of displacement reactions in the presence of POPC at the fixed concentration of 50 mM, while the concentration ImpA and Val_2 (equimolar in all cases) varied between 25 mM and 0.5 mM each. A set of control experiments without POPC was prepared similarly. The reactions were followed by HPLC during one week and their progress was plotted against time. As expected, the formation of Val_2A was drastically slower in the more dilute samples (Graph 3.4, solid lines). The hydrolysis of ImpA to AMP proceeded instead quickly at every concentration (appendices, Graph S2). At every concentration, the presence of lipids (Graph 3.4, dashed lines) did not induce significant kinetic differences with the reference samples (solid lines). At the lowest concentration of the study, the theoretical reagent/lipid ratio was 1:100 with the hypothesis of entirely available lipid bilayers. Lower concentrations of substrates could not be used because of the sensitivity of the UV detector of the HPLC.



Graph 3.4 Formation of Val_2A from Val_2 and ImpA at different concentrations of reagents, in the presence of 50 mM POPC (dashed lines) or without lipid (solid lines).

The most likely explanation for this absence of effect is the poor binding of the substrates to the membranes. Indeed, the most dilute cases offered a consequent amount of lipidic matter to bind to, and the background reaction (occurring in water) was very slow and did not represent a serious competitor.

The results that we obtained so far are in general accordance with the literature precedents. It seems that only a restricted selection of reactions is actually promoted by the opportunity to occur in or at the surface of lipid membranes. In the context of prebiotic studies, this immediately excludes a large range of polar molecules, such as nucleotides and many peptide sequences. Furthermore, the benefit seems to be restricted to reactions that perform poorly in the background, often because their substrates are poorly soluble in water. In this work, we chose a model reaction involving a competition between two fundamental chemical processes, conjugation and hydrolysis. This differs from other studies tackling only the maximum length of oligomer (most often homo-oligomers) detectable by a given analytical method. In addition to the risks of false negatives[279] inherent to every technique and the problems of reproducibility, questioning the quantitative efficiency of a reaction, in other terms how many bonds are created from one mole of activator (or activated substrate) under given conditions, is crucial when assessing the viability of a chemical process in a systems chemistry perspective. Furthermore, short sequences may be ligated together by several processes, leading to much longer and more varied ones without requiring a long initial polymerization extent. To our knowledge, phosphoramidate formation was never attempted in the presence of vesicles before, and we sought to demonstrate, as a proof of concept, the enhanced phosphoramidate formation between membranophilic substrates.

3.2.4. The search for membranophilic peptides

The membranophilicity of peptides and amino acid residues has been extensively studied. Many membranophilic peptides, including cell-penetrating peptides, are relatively long, exploit residues that are not usually considered prebiotic, or sequence-specific aggregation patterns. In this work, we were interested in small species, with typically less than 6 amino acids. Such short peptides suffer from the high membranophobicity of the amide backbone, to which must be added the weights of the charged carboxylate and ammonium residues at neutral pH. Even the most lipophilic side chains do not compete with this polarity and efficient membranophilicity of short peptides is only expected to arise from aggregation.[370] Indeed, Wimley and White used a model hexapeptide and POPC vesicles to provide a membranophilicity scale of amino acid residues.[371] In their study, the valine residue that we used extensively in our previous study only sparingly partitioned into POPC membranes. In contrast, leucine was more promising. The most lipophilic acids in this study were phenylalanine and tryptophan, but these often form insoluble species[143,383] that we specifically wanted to avoid in this work, to keep a "fair" competition with background aqueous chemistry.

To evaluate the extent of membrane binding of our model poly(Val) peptides and other peptides of interest, we measured their diffusion coefficients in vesicle suspensions using diffusion NMR, a technique of growing interest to assess and quantify binding phenomena.[286] The rate at which a molecule diffuses in solution is determined its apparent hydrodynamic radius, the characteristics of the solvent and the experimental conditions, notably temperature. A diffusion coefficient D is associated to it and reflects the quadratic average of the free path:[a]

$$\langle r^2 \rangle = 6Dt$$

where $\vec{r}$ is the displacement vector and t is the diffusion time. This leaves D in $m^2 s^{-1}$. Under given conditions, changes in the conformation or folding of a given molecule may be detected by a subtle change in its diffusion rate. Similarly, a change in the environment of the molecule modifies its diffusion rate. For an ideal spherical liposome of fixed radius R, the diffusion coefficient D is given by the Stokes-Einstein equation:

$$D = \frac{k_b T}{6\pi\eta R}$$

where k_b is the Boltzmann constant (in $J.K^{-1}$), T is the temperature (in kelvins) and η the dynamic viscosity of the surrounding fluid (in Pa.s). From this equation, we understand that a liposome with a diameter of 100 nm will diffuse much more slowly than a small molecule, with an apparent diameter below 1 nm, so that vesicles will be virtually immobile at the molecular scale. A substrate bound to a vesicle membrane would be diffusing along with it, at a very slow rate. This

[a] The average displacement of molecules diffusing in an NMR tube is $\vec{0}$ since there is no overall movement of the liquid inside the tube in the absence of shaking and convection.

substrate-lipid interaction can therefore be detected by the appearance of a new, slow diffusion regime in the presence of liposomes. The same approach is applicable to the efficient screening of protein-ligand interactions and is in growing use.[438]

Model membranes consisting of monodisperse large unilamellar vesicles (LUVs) are used routinely in diffusion experiments. In contrast with crude giant vesicles obtained by gentle hydration, LUVs are regular in size and thickness, are made of a thin bilayer, and their relatively small size (usually 100 nm in diameter) affords quasi-homogenous samples that are compatible with several spectroscopic techniques, including NMR.

The diffusion rate of a molecule in a liquid sample may be assessed by NMR using different techniques that rely on magnetic field gradients. In usual static single-dimension NMR, a homogenous vertical magnetic field $\vec{B}_0$ is applied along the $\vec{z}$ axis of the sample. Every effort is made so that this field remains homogenous throughout the tube, so that the spatial position of the analyte inside the tube has no influence on its spectral signature. In diffusion NMR, this homogeneity is broken in a controlled manner by adding a linear gradient of magnetic field $\vec{B}(z)$ along the $\vec{z}$ axis of the sample. The Larmor frequency ω at which a spin is observed therefore becomes dependent on the spatial position of the spin itself:

$$\omega = \gamma B_z \; and \; B_z = B_0 + B(z)$$

$$\Leftrightarrow \omega(z) = \omega_0 + \gamma B(z)$$

where γ is the gyromagnetic ratio of the nuclide. During the spin precession, the phase of each signal is influenced by the spatial position of the spin because of this. In a typical diffusion NMR sequence, a gradient is first applied in one direction to induce diffusion-dependent dephasing of the signals. In a second step, an inverse gradient is applied that refocuses the phase of any spin that would have stayed at an identical spatial position.[a] An immobile chemical species will therefore remain unaffected by this combination of gradient pulses. At the contrary, any species that moved from its original position because of diffusion will experience a decreased signal due to imperfect refocusing.[439] The extent of this decrease in intensity is directly linked to the strength of the gradient (stronger gradients induce higher signal decay) and to the mobility of the species (higher diffusion rates induce higher signal decay). The delay between the application of each gradient conditions the time allowed to the molecules to diffuse before the refocusing occurs. When this procedure is repeated several times with gradients of increasing strength, a stack of spectra is obtained, in which peak intensities are given by the equation:

[a] Most modern NMR sequences flip the spin to 180° and apply the gradients in the same direction twice, which has exactly the same effect. Many other improvements, such as the use of bipolar or shaped gradients to reduce eddy currents and other artefacts were developed and are now part of the routine pulse programs of NMR workstations.

$$I(\delta) = I_0 e^{-(\gamma g \delta)^2 D\left(\Delta - \frac{\delta}{3}\right)}$$

where g is the strength of the magnetic field, δ the duration of the gradient pulse and Δ the diffusion time between the two gradients. An exponential fit of $I(\delta)$ allows the determination of the diffusion coefficient D and the pre-exponential term I_0.

If a molecule is in slow equilibrium[a] between several environments that induce different diffusion behaviors, then $I(\delta)$ becomes a multi-exponential function, in which each exponential term represents one diffusion mode. In the case of a species bound to LUVs in water, assuming that the exchange between the aqueous phase and the bound phase is slow, separate I_0^{bound} and I_0^{aq} parameters will be found, associated to different diffusion coefficients D^{bound} and D^{aq}. D^{bound} is expected to be equal to the diffusion coefficient of the LUVs and therefore will be much smaller than the diffusion coefficient in water D^{aq}. The mole fraction of bound species is determined according to the expression:

$$f_{bound} = \frac{I_0^{bound}}{I_0^{bound} + I_0^{aq}}$$

An intrinsic limitation of this method is that the lipid phase has a finite (and rather small) hosting capacity and may therefore be saturated. In this case, using high concentrations of analyte would result in a very small f_{bound} even with molecules of high membranophilicity. NMR is not a highly sensitive technique in itself and a compromise must be found between achieving a signal-to-noise ratio (SNR) sufficient to obtain clean signals and keeping f_{bound} to practical values. The SNR can be improved by increasing the number of scans, which also increases the total duration of each increment and leads to sample and environment evolution (precise internal temperature, room temperature, subtle variations of the baseline and the phase…) throughout the measurement and hence produces artefacts. Additional features of the pulse program, such as water presaturation, also lengthen the increments and tend to build up heat that considerably degrades the quality of the acquisition because of convection.

We began by evaluating the partition of Val_3 between POPC membranes and water. 10 mM POPC was first hydrated in a sodium bicarbonate buffer (pH 8, 100 mM) to form crude vesicles, which were then frozen and thawed repeatedly to eliminate multilamellar structures. The resulting GVs were then extruded at 1 µm then 100 nm by repeatedly forcing the sample through polycarbonate membranes. This gave a LUV sample that was slightly viscous, homogenous, and almost transparent. In contrast, crude GVs at the same concentration are very turbid and almost opaque (Figure 3.10).

[a] At the timescale of NMR, that is, seconds. A fast equilibrium will yield an average diffusion which requires additional calibration experiments.

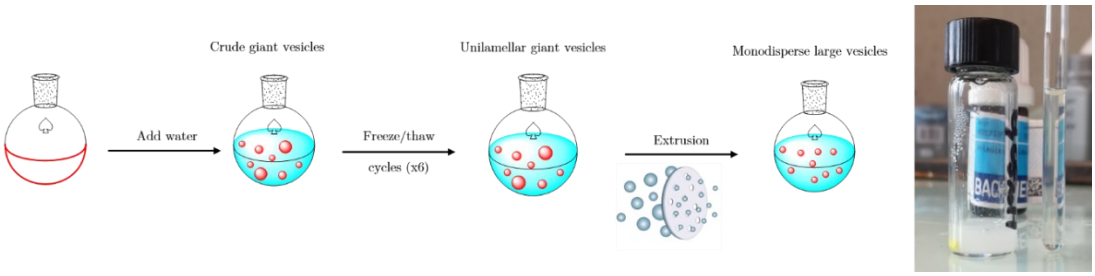


Figure 3.10 Left: preparation of LUVs by hydration of a thin film of lipids, freeze/thaw cycles and final extrusion(s). Right: crude vesicles obtained after hydration (large vial) were turbid and opaque, while LUVs (NMR tube) were clear and almost transparent.

5 mM Val_3 was then dissolved in and a 500 µL sample was prepared for NMR. In our first attempts, we used a buffer in D_2O/H_2O 1:9 to prepare the vesicles, which resulted in a tall water peak. We tried different suppression techniques to remove it. Among them, excitation sculpting[440] was the only satisfying one, as other techniques (including watergate and continuous wave presaturation) caused baseline distortion or phase inconsistencies. The high dilution of the analyte forced us to use strong saturation pulses, and while this was acceptable for 1D ^1H NMR, phase issues were encountered in diffusion NMR. Indeed, because of the residual artefacts and baseline distortion around the suppressed water resonance, a consistent phase throughout the full spectral width could not be obtained for several increments, which then had to be discarded. Changing the solvent to 99% D_2O resulted in a much smaller water peak and efficient suppression was obtained with the excitation sculpting pulse without baseline distortion. After testing different diffusion programs, we chose a stimulated spin echo (STE) sequence with bipolar gradients. Before the acquisition, the pulse power was adjusted to yield a precise 90° flip. This was repeated for every new sample, as pulse values tend to vary greatly under saline conditions. The diffusion delay Δ was set to 100 ms, while the pulse duration δ was adjusted between 1500 and 2000 µs for each analyte.[a] A maximum of 64 scans per increment could be accumulated before baseline distortion occurred, likely caused by fluctuations during the lengthy acquisition, and convection in the tube. After the acquisition, the peaks were integrated manually and attemptedly fitted to a biexponential decay using the software Bruker Dynamics Center.

For Val_3, the α proton at 3.80 ppm provided a clean, well-resolved peak and was used as a reference. The fitting only returned a single exponential profile, confirming our hypothesis that short valine peptides were poor membrane binders. When analyzing the more lipophilic Leu_3, a biexponential profile was obtained from the CH and CH_2 groups, confirming an ability to bind POPC membranes (Figure 3.11). The terminal methyl groups could not be used because

[a] As D is obtained through a regression of $I(\delta)$ and not by direct reading, it is not mandatory to keep identical gradient strengths between analyses. All the other parameters, in particular the temperature and those related to sample preparation, must be strictly identical. The pulse length entered in the sequence corresponds to a single gradient. Because the gradients are bipolar, the actual pulse length is twice the entered value.

they overlapped too severely with lipid signals. The partition could not be quantified with precision, because of the high uncertainties of the fitting: $I_0^{bound} = 7.1 \pm 0.8$ for the methyl peaks and $I_0^{bound} = 2.6 \pm 0.7$ for the gathering of the CH_2 and CH peaks.[a] The uncertainty was even higher for D^{bound} (appendices). This was a double consequence of the low SNR of the last increments and the limited contribution of the bound regime because of a saturation of the membrane with the analyte (appendices, Figure S29). The diffusion coefficient of the LUVs was roughly equal to D^{bound} as expected. Surprisingly, it was about one order of magnitude higher than what was calculated by the Stokes-Einstein formula for a sphere of 100 nm diameter in water ($\eta = 1.00\ mPa.s$ at 20°C, from CRC Handbook online). This suggested that the viscosity of the LUV suspension was about 10 times higher than pure water, which was believable from the physical aspect of the liquid. However, such a difference was not observed for small molecules, for which D^{aq} only slightly differed between aqueous solutions and LUV suspensions (calculated for Val_3A and Leu_3A in the appendices). This suggested that more elaborate mechanisms were at work.[441]

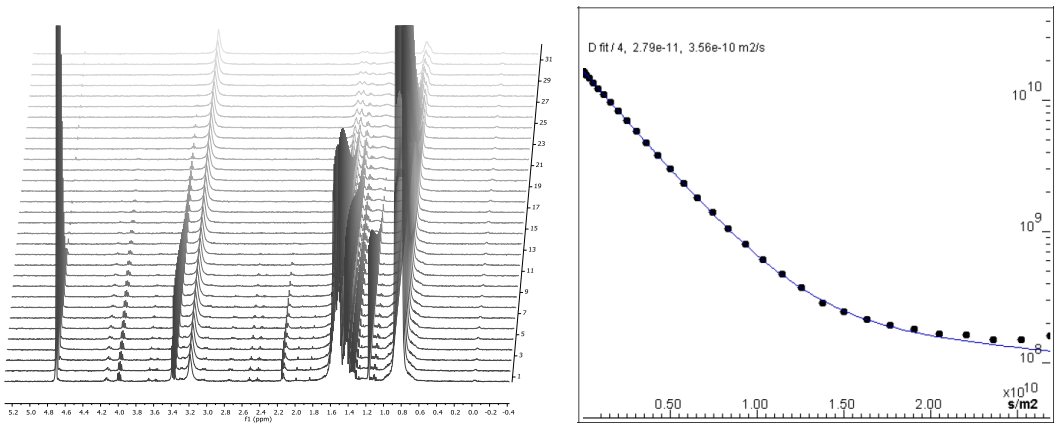


Figure 3.11 Diffusion ^1H NMR (D_2O, 500 MHz) of Leu_3 in the presence of 10 mM POPC LUVs. Left: stacked view of the 32 increments. The slowly diffusing peaks belong to the lipids. The intense, quickly suppressed peak at 4.7 ppm is residual water that was not suppressed. Right: Double exponential plot for the CH_2 + CH massif (1.7-1.3 ppm) of Leu_3 fitted in Bruker Dynamics Center 2.5.6.

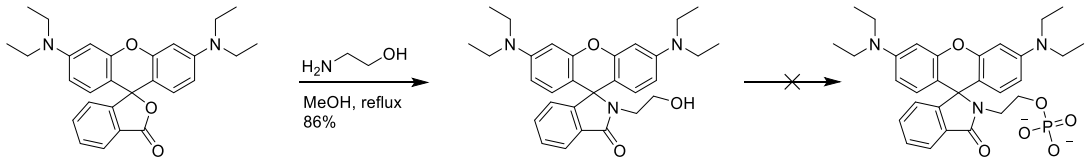
From this data, f_{bound} for Leu_3 was estimated to be in the order of 2%, which represents a membrane loading around 1 peptide molecule per 100 lipid molecules (or 2 peptide per 100 lipid molecules on the external leaflet only, assuming that Leu_3 did not readily cross the bilayer). This was consistent with the work of Wimley and White, who reported that hexapeptides with the general formula Ac-WLLXLL (X being any amino acid) partitioned to POPC membranes, with a f_{bound} in the range of 0.05-0.6 when X was a lipophilic amino acid ($f_{bound} = 0.14$ for X=L). In their study, the peptide was added with a peptide/lipid concentration ratio between 3:100 and 7:100, for which $f_{bound} = 0.14$ translates into membrane loadings in the order of one peptide molecule per lipid molecule.[371] As Leu_3 was

[a] The value of I_0^{bound} does not have to be the same between peaks, as it depends on its intrinsic intensity. Here, we want to emphasize the fact that uncertainties above 10% were obtained, yielding, at best, gross estimations of the fitting parameters.

soluble in water at concentrations around 10 mM, we envisioned to use it as a membranophilic peptide in model reactions.

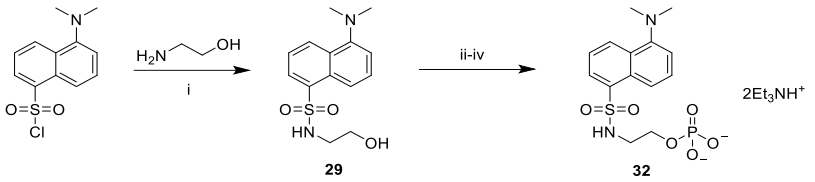
3.2.5. The design of membranophilic phosphates

AMP is inherently membranophobic because of its high polarity, as are all nucleotides. To perform a model coupling between membranophilic substrates, we therefore needed to use a different phosphate. Formally, any phosphoester with a reactivity comparable to that of AMP could have been used. However, we wanted to avoid strongly amphiphilic molecules that could behave as lipids. This narrowed the choice down to relatively compact, yet lipophilic, organic scaffolds. To open new perspectives, we chose to use a fluorescent compound. Fluorescence is indeed a very high sensitivity detection technique, and a range of experiments that were not possible, or more limited, with AMP, could be performed using a fluorophore. Among membranophilic fluorophores that could be easily functionalized, rhodamines attracted our attention first. Commercially available rhodamine B was easily substituted with an ethanolamine linker in one step.[442] However, the following phosphorylation step was problematic and we could not isolate practical yields of phosphorylated rhodamine (appendices, Scheme 3.12). Rhodamine, as most xanthene fluorophores, exists in equilibrium between an open and a closed form. The equilibrium is governed by pH changes. While this is a very useful tool for the design of fluorogenic probes,[442,443] it sometimes represents a synthetic burden.



Scheme 3.12 Attempts to synthesize a phosphorylated rhodamine. The failed phosphorylation conditions are fully reported in the appendices.

For this reason, we used a phosphorylated dansyl probe, that we prepared by modifications of literature conditions.[444,445] Dansyl chloride was first acylated with ethanolamine to give **(29)**, then the alcohol was activated as a mesylate (intermediate **(30)**) and substituted by di(tert-butyl)phosphate to give **(31)**. **(31)** was finally deprotected under acidic conditions to yield dansyl phosphoethanolamine (dansyl PE **(32)**, Scheme 3.13) in a very good overall yield.



Scheme 3.13 Synthesis of dansyl phosphoethanolamine from dansyl chloride. i: Et_3N, dichloromethane, rt, 2h, 99% after column chromatography; ii: MsCl, pyridine, rt, 1h, extracted; iii: $(tBuO)_2PO_2K$, TBAI, DMF, rt, 24h, extracted; iv: 5% TFA in dichloromethane, rt, 1h30, 95% over 3 steps after reverse-phase chromatography.

In particular, the original synthesis[445] involved the displacement of a tosylate with di(tert-butyl)phosphate as its tetrabutylammonium salt, that had to be prepared from the commercial potassium salt by careful protonation then deprotonation at low, controlled temperature. We avoided this tedious procedure by using the commercial potassium salt directly in the presence of a catalytic amount of tetrabutylammonium iodide in DMF. With careful exclusion of water, the conversion was sufficient to avoid purifying the intermediate. The final phosphate was successfully separated from the small amounts of hydrolyzed intermediate and other materials by ion-pairing chromatography.

Thanks to its fluorescence, the membranophilicity of dansyl PE could be studied by different independent methods. We used diffusion NMR in the presence of 10 mM POPC LUVs under the conditions already described for the peptides. The most intense peak, belonging to the NMe_2 group, was close to lipid signals (NMe_3^+ of the choline headgroup) and therefore difficult to use. We found that the aromatic signal belonging to H3 and H7 (appearing as a unified multiplet) represented the best compromise and was therefore used as the reference peak. Other peaks returned qualitatively similar results, but their SNR was too low in the last scans to properly fit the data. Diffusion NMR revealed the existence of two diffusion modes (Figure 3.12), the second one being associated with a diffusion rate comparable to that of the lipids, confirming that a fraction of the fluorophore was indeed anchored in the membranes. With the technical limitations already enumerated in the case of peptides, we found $D^{bound} = 4 \pm 2 \times 10^{-11}\, m^2.s^{-1}$ and $f_{bound} = \ 2 \pm 2\ \%$ which corresponded to a fluorophore/lipid ratio of 1:100.

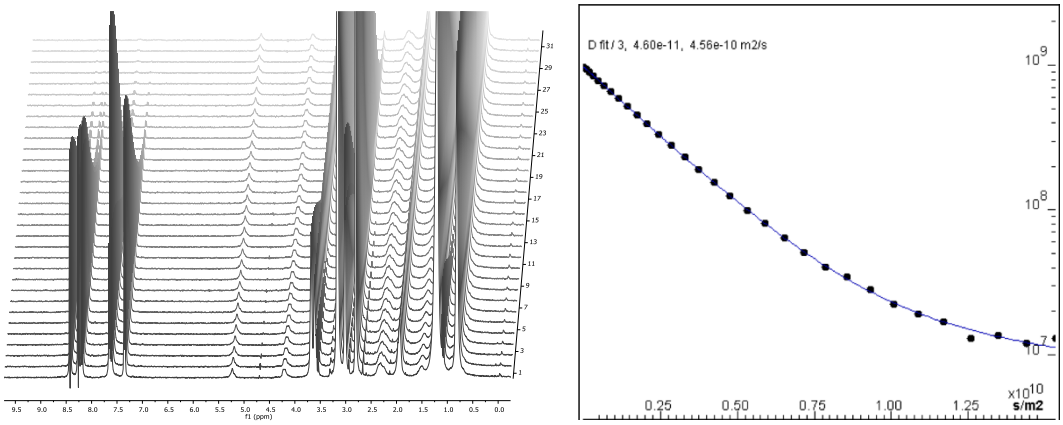
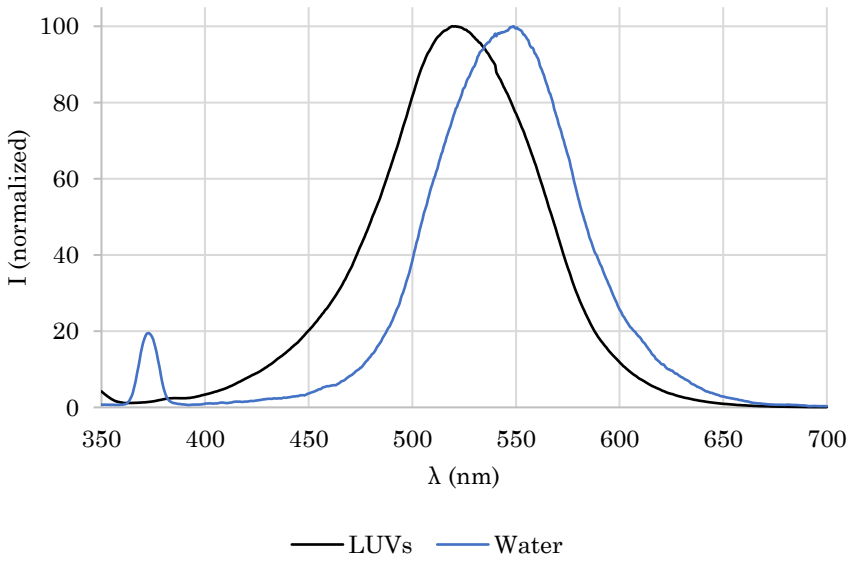


Figure 3.12 Diffusion ^1H NMR (D_2O, 500 MHz) of dansyl PE in the presence of 10 mM POPC LUVs. Left: stacked view of the 32 increments. The slowly diffusing peaks belong to the lipids (e.g. the broad signal at 5.2 ppm that corresponds to the unsaturation in the oleoyl chain). Right: Double exponential plot for the 7.58-7.47 ppm peak of dansyl PE fitted in Bruker Dynamics Center 2.5.6.

The spectral properties of fluorophores are often dependent on their environment. We expected that the fluorescence of dansyl PE would be different in membranes from what it was in water. A 5 µM sample of dansyl PE in sodium bicarbonate buffer (pH 8) had an excitation maximum of 331 nm. When excited at this wavelength, it emitted with a maximum at 550 nm (Graph 3.5). When a sample containing 5 mM dansyl PE and 10 mM POPC LUVs was diluted 1000 times to

reach 5 µM dansyl PE, a comparable spectrum was obtained (data not shown), illustrating that the fluorescence was mostly occurring from the aqueous phase.

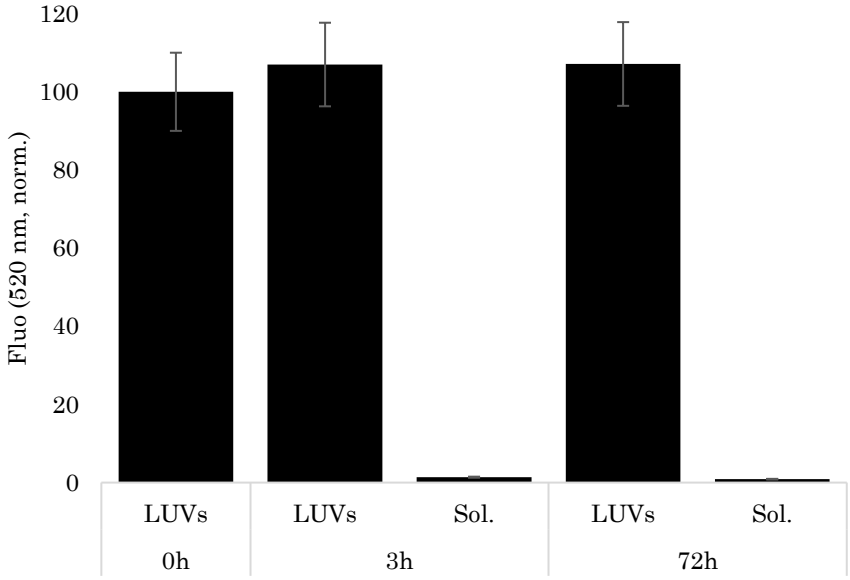
LUVs can be separated from the small molecules surrounding them by size-exclusion chromatography on gel columns. Indeed, the large liposomes elute through the gel without significant interaction and are collected in the dead volume, while small molecules diffuse through the material and elute later. Using this method, 10 mM LUVs mixed with 5 mM dansyl PE were separated from the aqueous bulk fluorophore by filtration on a column packed with Sephadex™ G-25 gel equilibrated with a sodium bicarbonate buffer. The front fraction contained all the lipids as confirmed by a Stewart assay, while the late fraction contained the large excess of non-encapsulated dansyl PE. A fluorescence spectrum of the undiluted front fraction was recorded and was representative of the fluorescence of dansyl PE in phospholipid membranes (Graph 3.5). The excitation maximum was found at 339 nm while the emission maximum was shifted to 520 nm. This confirmed the environment sensitivity of the probe. We note that the ratio of intensities $\frac{I_{LUV}}{I_{water}}$ reached values around 15 in the 400-470 nm region, implying that the fluorescence arising from the membranes may be measured with only a 5 to 10 % contribution of the background fluorescence by using this wavelength. While this could have been useful in quantitative fluorescence assays, it was probably insufficient for microscopy, where filters are usually poorly selective.



Graph 3.5 Normalized emission spectra of dansyl PE in water (excitation 331 nm) and in POPC LUVs (excitation 339 nm).

To verify the kinetic stability of the encapsulation, two 250 µL aliquots of the sample obtained by gel filtration were filtered a second time after 3 hours and 72 hours, respectively. The fluorescence of the front and tail fractions were measured. In both cases, the front fraction quantitatively retained its original fluorescence while the tail fractions gave signals below the detection limit (Graph

3.6), confirming that only negligible amounts of dansyl PE had leaked out of the membranes within three days.



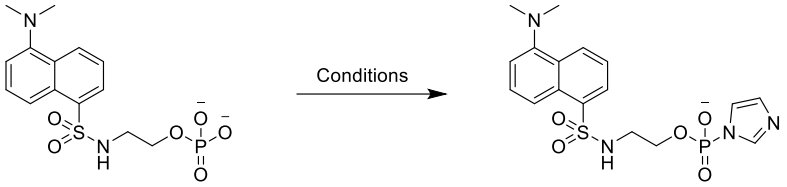
Graph 3.6 Retention of dansyl PE in LUVs after 3 hours and 72 hours of incubation as assessed by gel filtration. Fluorescence at 520 nm was measured (average of 3 measurements), corrected by the dilution factor caused by the gel filtration and normalized to the sample before incubation. The uncertainties on the filled samples was essentially due to the rough estimation of the eluted volumes.

3.2.6. The reactivity of membranophilic partners in the presence of lipid membranes

We first sought to reproduce the kinetic study carried out with the imidazolide displacement method, but replacing Val_2 and AMP by their respective membranophilic equivalents Leu_3 and dansyl PE. Thanks to the availability of the highly sensitive fluorescence HPLC, we could use much lower concentrations of substrates without detection issues. This permitted to lower the lipid concentration to more practical values, but also to study very dilute reactions where the substrates may be entirely membrane-bound without saturation of the bilayer.

A prerequisite to reproduce our standard kinetic conditions was to produce the imidazole-activated equivalent of dansyl PE, ImpD. When dansyl PE was subjected to the imidazolide activation under the Mukaiyama conditions, similarly to the synthesis of ImpA, the desired activated substrate ImpD was formed but could not be isolated by precipitation because of its high solubility in most organic solvents and to some extent in water. Reverse-phase chromatography was attempted but only returned mixtures of compounds including the desired ImpD and the pyrophosphate DppD along with other side products. When an aliquot of the phosphate in $CDCl_3$ was treated with an excess of CDI, the desired ImpD was quickly and quantitatively formed with imidazole as the only non-volatile by-product. However, we showed previously that the

presence of imidazole slowed down the hydrolysis of imidazolides, a competitive reaction that should dominate in dilute experiments. It was therefore desirable to keep imidazole pollution to a minimum in our studies. When activated in DMSO with a minimum of CDI (1.2 eq), dansyl PE was largely converted into DppD. It is likely that, without a significant excess of CDI, the activation was slow enough to allow remaining dansyl PE to displace the newly formed ImpD (Table 3.9).



Entry	Synthetic conditions	Purification	Outcome
1	10 eq imidazole, PySSPy, PPh_3, Et_3N, DMF, 18h	Precipitation	Did not precipitate from acetone and diethyl ether in the presence of $NaClO_4$
2		C18 chromatography	Degraded
3	10 eq CDI, $CDCl_3$, 1h	C18 chromatography	Degraded
4	10 eq CDI, $CDCl_3$, 1h	None	Excess of imidazole
5	1.2 eq CDI, DMSO-d_6, 2h	None	Mainly yields DppD

Table 3.9 Attempts to prepare and isolate ImpD in pure form.

Because of the difficulty to get pure ImpD, we decided to use a coupling reaction promoted by EDC and 1-ethylimidazole instead. We expected that, at such low concentrations and temperatures, the formation of peptide bonds would be very slow compared to the formation of the phosphoramidate, so that a restrained variety of species would be formed, among which Leu_3D would dominate by far. A test reaction was first performed at 10 mM concentration (dansyl PE, Leu_3) in 0.1 M MOPS buffer at pH 7.5. 40 mM EDC and 7.5 mM EtIm were added at 0°C, similarly to the simplified condensation conditions. The evolution was followed by HPLC with a fluorescence detector. As the polar C18 column that was used for peptido-nucleotides gave poorly reproducible chromatograms, we switched to a less retentive phenylhexyl column. In contrast to peptido-nucleotides, ammonium acetate as the eluent gave a resolution at least as good as formic acid. Furthermore, the acidic eluent resulted in altered fluorescence properties when compared to the data acquired in bicarbonate buffer. In practice, the signal was about 50 times stronger in ammonium acetate.

After less than 24 hours, the starting material was mostly consumed, and a massive precipitate had formed. HPLC showed the formation of two main products. Since the concentration of the sample was sufficient, NMR analysis was carried out to identify their nature. A sample of the crude mixture was lyophilized and solubilized in DMSO-*d*$_6$. Expectedly, the pyrophosphate DppD (δ_P = -8.8 ppm, 12 mol%; Figure 3.13) was found, accompanied with an unknown product (δ_P = +3.0 ppm) that possessed an ethylene phosphate group according to ^{1}H-^{31}P HMBC (Figure 3.13). At first, we envisioned that the peptide could induce a trans-amidation to yield *N*-dansyl-Leu_3, releasing *O*-phosphorylethanolamine that

would be responsible for the aforementioned ^{31}P peak. This hypothesis was quickly excluded by comparison with a synthetic standard of *N*-dansyl-Leu$_3$[446] by fluorescence HPLC. Furthermore, the complex multiplicity of the ethylene system suggested a constrained system, such as a cycle, whereas linear ethylene phosphates such as dansyl PE yield more simple signals.

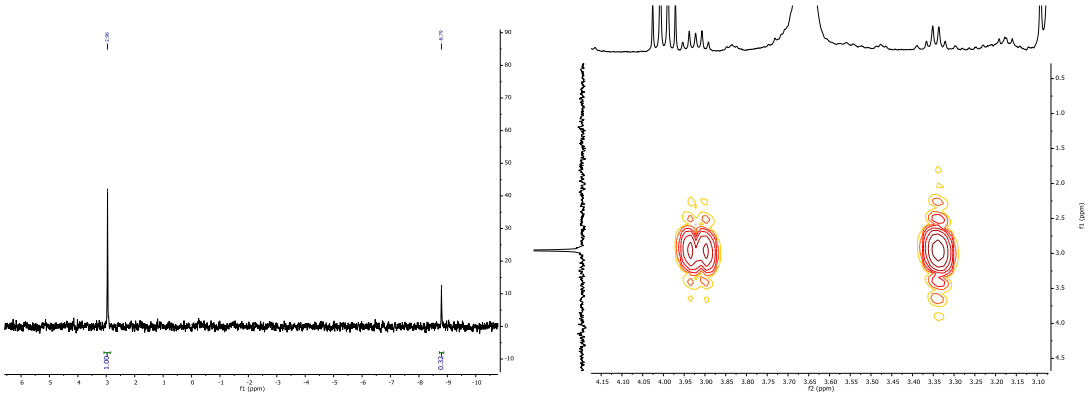
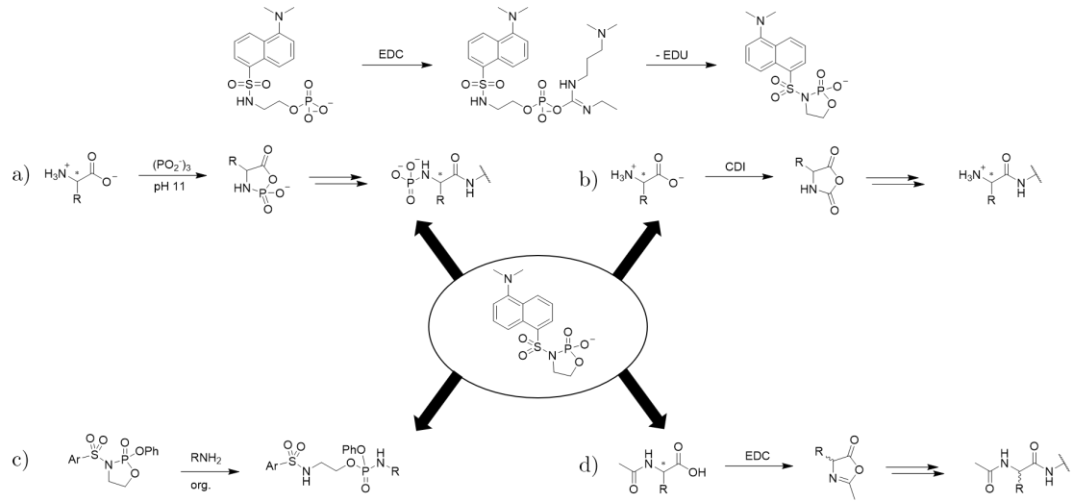


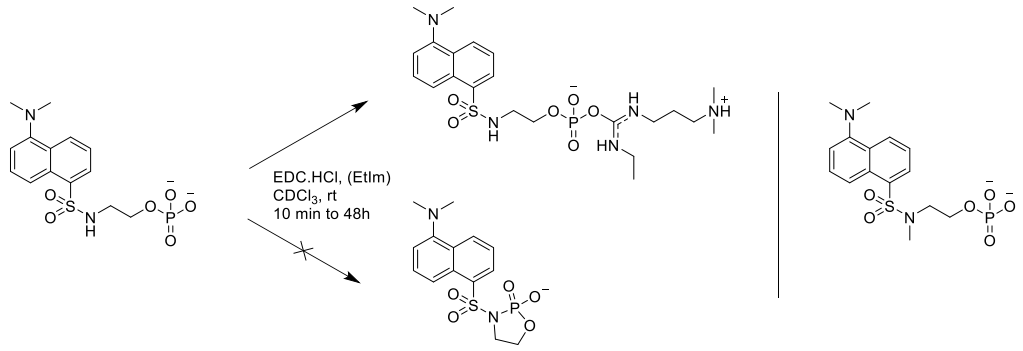
Figure 3.13 ^{31}P NMR (H_2O, 200 MHz, left) and ^{1}H-^{31}P HMBC NMR (H_2O, 500 MHz, right, area of interest) of a crude mixture obtained after the reaction of dansyl PE and Leu$_3$ in the EDC/1-ethylimidazole system.

Analysis of the crude mixture by mass spectrometry suggested that a species corresponding to dansyl PE – H_2O was present. This structure, dubbed dansyl cPS (cyclic phosphorylsulfonamide) would also explain the complex multiplicity observed in ^{1}H NMR. Indeed, cyclic phosphate species resulting from the attack of the nitrogen atom of a sulfonamide on an activated phosphate are known.[447] Furthermore, it resembles numerous cyclic anhydrides involved in primordial polymerization reactions (Scheme 3.14). Many of these species are reactive toward nucleophiles, notably amines, under basic conditions.[115,447] In one trial, we performed the same coupling reaction at pH 9 with the hope that the peptide could react with dansyl cPS. After consumption of the EDC, only the starting material was recovered. It is possible that dansyl cPS was formed then hydrolyzed, catalyzing the degradation of EDC. The presence of 10 mM POPC did not prevent this phenomenon.

Further experimental investigations are necessary to conclude on this chapter. One alternative would be to use pre-formed dansyl cPS as a weakly activated phosphate source then open it with a peptide under basic conditions. A preliminary attempt to synthesize pure dansyl cPS in $CDCl_3$ failed because the EDC-activated intermediate accumulated and did not react further within 48 hours. Performing the activation in the presence of EtIm gave the same results (Scheme 3.15, left). Alternatively, the whole cyclization process may be avoided by the use of a secondary sulfonamide, that should be accessible through the same synthetic pathway (Scheme 3.15, right).



Scheme 3.14 Top: proposed mechanism for the formation of the cyclic species dansyl cPS. Bottom: other cyclic mixed anhydrides known to occur in primordial oligomerization reactions. a) mixed phosphoric anhydrides formed during the N-phosphorylation of amino acids by metaphosphates [115]; b) N-carboxyanhydrides are intermediates in the CDI-promoted peptide growth [158]; c) organic mixed sulfonimide-phosphorimidate species can be prepared from N-sulfonylated ethanolamines and dichlorophosphates and are opened by nucleophiles under basic organic conditions [447]; d) oxazolones are intermediates in the EDC-promoted coupling of N-acetyl amino acids [162].



Scheme 3.15 Left: attempt to willingly produce dansyl cPS through EDC activation only returned the EDC-activated intermediate, irrespective to the use of 1-ethylimidazole. Right: potential phosphorylated probe that would not suffer cyclization upon activation.

3.2.7. Reactions between soluble and membranogenic matter

An alternative way of bringing polar nucleotides and amino acids to membranes is to couple them to lipophilic materials such as fatty chains, thus forming amphiphilic materials. Although this scenario has attracted relatively little attention from the prebiotic community, it is a major research topic in therapeutics, notably in the field of vectorization. Nucleotide-lipids and oligonucleotide-lipids were found to form a large variety of aggregates, to display enhanced membrane permeation properties and to be valuable tools to increase the potency of antiviral and anticancer drugs.[448,449]

Typical prebiotic sources of lipophilic materials include fatty alcohols and acids,[84–87] fatty alkyl phosphates[102] and acylglycerols.[96] Some examples of compounds that may be expected in a prebiotic setting are depicted in Table 3.10. We only investigated the possibility to couple amino acids, nucleotides, and peptido-

nucleotides to fatty alcohols. A reversible conjugation was desirable to enable the delivery of small biomolecules to the lumen of vesicles. While in the context of drug delivery, functions such as esters may be cleaved enzymatically, a prebiotic scenario calls for chemically labile groups. To stay in line with the concepts of the general condensation reaction, an esterified peptido-nucleotide (Table 3.10, entry 3) seemed ideal. This work was carried out at Universität Stuttgart, in the group of Prof. Clemens Richert.[a]

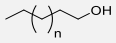
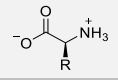
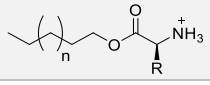
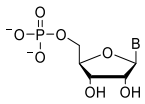
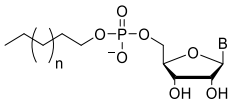
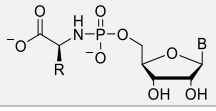
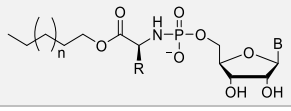
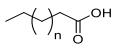
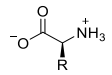
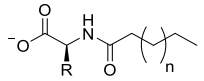
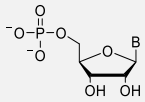
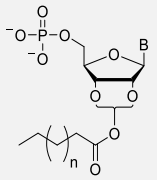
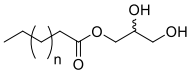
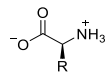
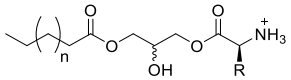
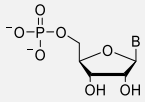
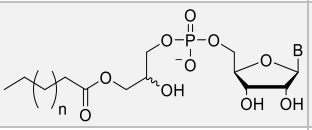
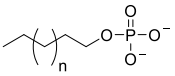
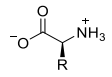
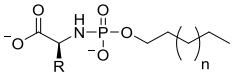
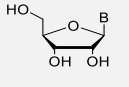
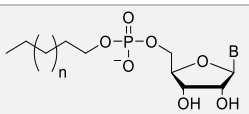
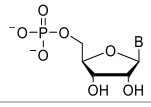
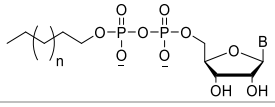
Entry	Lipophilic moiety	Polar moiety	Possible product	Refs
1				129
2				450,451
3				
4				
5				452
6				453
7				454
8				455
9				450,451
10				454

Table 3.10 Overview of typical amphiphilic molecules that may be produced from long chain material and amino acids, nucleosides and nucleotides. Other isomers and chemically different structures may also be envisioned. Some structures know synthetic precedents or analogs as indicated in the references.

[a] This work was made possible by a one-month research grant awarded by the Deutscher Akademischer Austauchdienst (DAAD) after a selection process. The work in Stuttgart was carried out under the direct supervision of Dr Peter Tremmel. The analytical equipment available on site differed from the equipment used in the rest of this work, as described in the appendices.

When studying the general condensation reaction, we observed the formation of phosphodiesters between AMP and the alcohol function of the HEPES buffer (Scheme 3.9 in section 3.2.1.3). As HEPES was more concentrated than AMP (500 mM and 200 mM respectively) and displayed a primary alcohol instead of the secondary 2'- and 3'-alcohols of AMP, a significant amount of it was formed when compared to diadenosyl phosphodiesters (quantitative ^{31}P NMR shown in Figure 3.7 in section 3.2.1.3). We envisioned that a similar reactivity could be observed with a long-chain alcohol instead of HEPES.

We began our investigations with the partially miscible 1-butanol. 200 mM AMP and 500 mM butanol were coupled in a buffer consisting of 500 mM MOPS pH 7.5, 80 mM $MgCl_2$, 800 mM EDC and 150 mM EtIm. These conditions corresponded to the general condensation buffer where HEPES would have been substituted for butanol and MOPS. The outcome was determined after 24 hours by ^{31}P NMR. The quick formation of AppA dominated the reaction (more than 80%) and the phosphodiester peaks were barely detected. The reaction was repeated with only 10 mM $MgCl_2$, which limited the formation of AppA to only 50% of the consumption of AMP. Under these conditions, 5 mM of the desired phosphodiester BuAMP were formed, whereas the sum of all other phosphodiesters (2'- and 3' with any length of RNA chain) was 17.5 mM (appendices, Figure S15). When excluding AppA and unreacted EtImpA, BuAMP represented about 10% of the mixture.

Optimal conditions were found when the aqueous phase was saturated with butanol.[a] Under these conditions, 14.8 mM BuAMP were formed, along with 13.6 mM other phosphodiesters. BuAMP represented 33% of species other than AppA (Figure 3.14). NMR control established that the butanol supernatant contained very little amounts of phosphorus-containing material. The same experiment was repeated with 1-hexanol and 1-octanol at their saturation limits.[b] As expected due to the low water solubility of these alcohols, only traces of hexyl AMP were obtained (1.1 mM, 1.7 % of non-AppA content, Figure S17), and octyl AMP was not detected (Figure S18). Water-miscible glycerol was also tested, as it is the backbone of most lipids. Its concentration was arbitrarily fixed to 1 M. 1- and 2-glyceryl phosphodiesters were formed efficiently (13.4 mM and 1.2 mM, respectively, Figure S19, Figure S20) and accounted for 22% of the non-AppA content. It was then evident that the efficiency of the reaction was conditioned by the concentration of the alcohol (Graph 3.7).

---

[a] A concentration of 0.8 M of butanol was measured by ^{1}H NMR in the aqueous phase using the lowest shifted CH_3 peak of endogenous EDU as a standard (appendices, Figure S21). This value was consistent with the solubility in pure water reported in PubChem (0.85 M).

[b] Following the same method as for butanol, these were found to be 65 mM for hexanol and 2.5 mM for octanol, while PubChem reported 0.06 M and between 0 and 4 mM respectively.

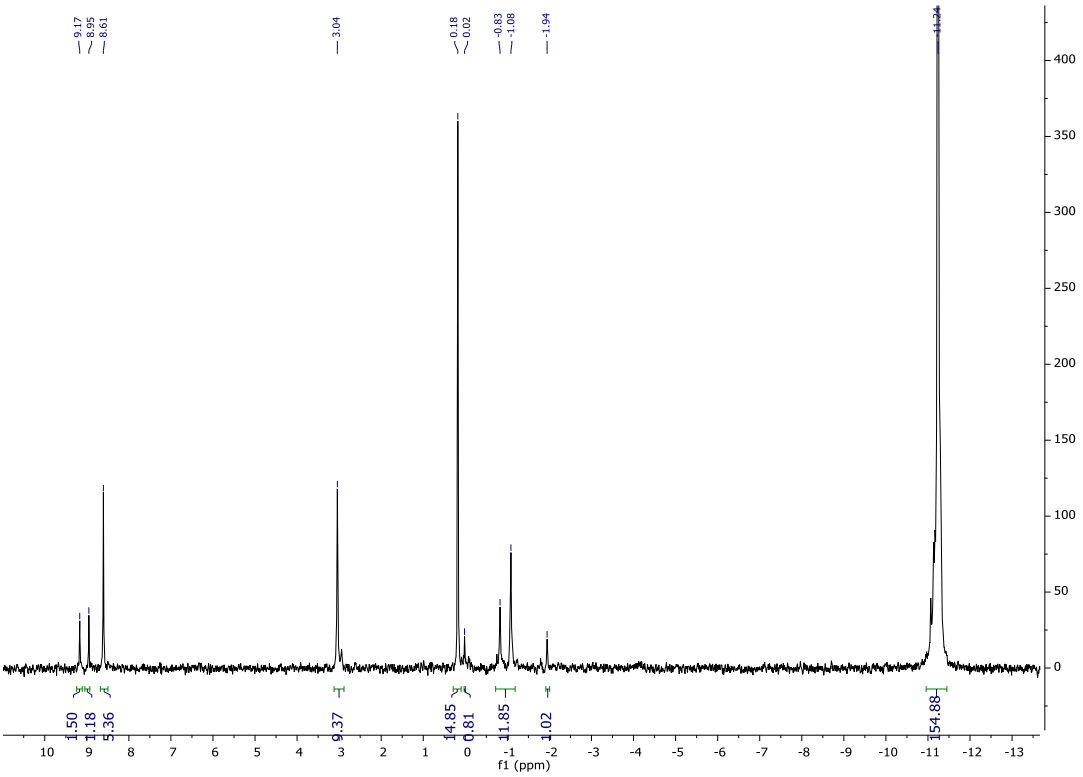
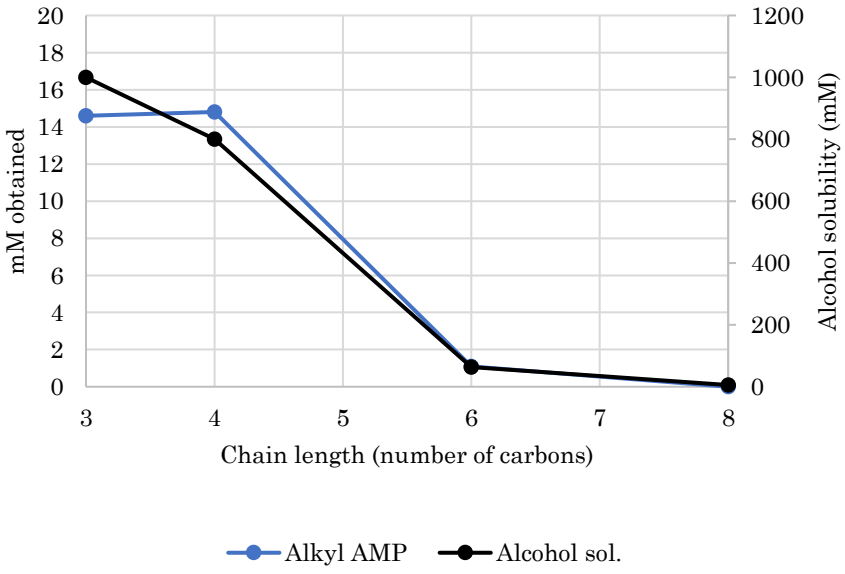


Figure 3.14 Quantitative ^{31}P NMR spectrum (H_2O, 160 MHz) of a general condensation reaction using 10 mM $MgCl_2$ and saturated butanol after completion. The integrals were summed up to 200 mM total phosphate content. Identity of the species: -11.2 ppm AppA; -2 to 0 ppm: 2'-5' and 3'-5' phosphodiesters; 0.18 ppm: BuAMP; 3.0 ppm: AMP; over 8 ppm: EDC by-products.

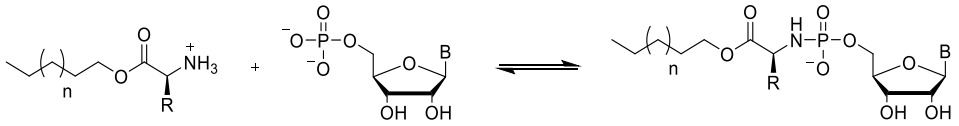


Graph 3.7 Pattern similarities between the solubility of alcohols (1M glycerol as C3, straight chain alcohols beyond, right vertical scale) and the formation of alkyl phosphodiesters with AMP (left vertical scale) in the general condensation buffer.

We envisioned that vesicles or micelles may enhance the availability of the most lipophilic alcohols and yield more significant amounts of the amphiphilic conjugates. When 50 mM monolaurin (*rac*-1-dodecanoylglycerol) was hydrated in a reaction mixture at pH 7.5, a floating precipitate quickly separated, as

monolaurin is a rigid lipid with a high T_m. Using a mixture of monolaurin and lauric acid (50 mM and 100 mM respectively) afforded a clear, slightly soapy solution after hydration at 40°C. Even under these conditions, the desired phosphodiester was not formed (Figure S22). The long-chain alcohols formed a layer when added above their solubility limit in water. The separation line was marked, and no emulsion or turbidity was observed, even with vigorous shaking, suggesting that only the soluble fraction was available to the other reagents.

We also investigated the formation of fatty esters of amino acids under the same general condensation conditions. These were of particular interest, as they may be coupled to nucleotides by phosphoramidate chemistry in a reversible manner (Scheme 3.16). In such hypothesis, a fatty amino acid ester would be used as a carrier to bring nucleotides across membranes. Furthermore, long-chain (C16) esters of various amino acids, and derivatives thereof were used as gelators,[129] which suggested that they may have interesting physical properties in the context of self-organizing matter.



Scheme 3.16 Reversible phosphoramidate formation between amino acid fatty esters and nucleotides as envisioned in this work.

The difficulty of forming amino acid esters by carbodiimide chemistry under neutral conditions has been discussed in section 3.2.1.2 and is mainly related to the massive formation of guanidines and *N*-acylureas. We used 1-^{13}C glycine to assess the formation of glycine butyl ester and by-products in the general condensation buffer saturated with butanol. The species could be identified based on ^{13}C NMR. When the amine of glycine was not protected by any group, the ester was not formed and a mixture of guanidine, *N*-acylureas and small peptides was recovered (Figure S23). Lowering the $MgCl_2$ concentration from 80 mM to 10 mM as for AMP did not influence the outcome of the reaction. We attempted the same reaction at pH 6 using a MES buffer, hoping that the protonation of the amine function may block the formation of guanidines. Unfortunately, the outcome was comparable to pH 7.5 (Figure S24).

A temporary protection of amines is possible through the reversible formation of imines. 500 mM acetaldehyde were added to the general condensation buffer saturated with butanol. A yellow color rapidly appeared and brownish tar accumulated over a few days. The desired ester was not detected (Figure S25). Danger et al. reported the formation of peptide bonds from *N*-acetylated and *N*-benzoylated amino acids through oxazolone intermediates.[162] When *N*-acetylglycine was treated with saturated butanol in the general condensation buffer, no guanidine could form, but the only products detectable by ^{13}C NMR were the *N*-acylureas of glycine (Figure S26). It was difficult to monitor the formation of the oxazolone intermediate by the method of Danger and coll., as it

relied on proton NMR, in areas that were overcrowded with EDC, EDU and buffer signals.

Our last attempt was to use a peptido-nucleotide as a protected source of amino acid. We chose ValA over GlyA for the sake of easier purification. Val*A was prepared from 1-^{13}C L-valine and ImpA and purified by manual ion-pairing chromatography. It was then reacted with saturated butanol in the general condensation buffer. Examination of the ^{1}H-^{31}P HMBC NMR spectrum, ^{13}C NMR spectrum and of MALDI-MS data did not evidence the formation of the ester (Figure S27, Figure S28). Only the products known to arise from the general condensation reaction (see section 3.2.1.3) were formed.

The general conclusion that we draw from these observations is that membranogenic matter is unlikely to efficiently react with soluble molecules in the context of carbodiimide-promoted chemistry. When the concentration of the fatty reagent is limited by its solubility, which is always the case of lipophilic material, side reactions take over and rapidly consume EDC forming stable by-products.

3.2.8. Conclusions

Lipids were probably present early in the geological history of Earth. They were either synthesized under thermal conditions or delivered during the crashes of carbonaceous comets and meteorites, and were likely present on the young Earth, when aqueous oligomerization of essential biomonomers started. There is a long-lived hypothesis that membranes may promote reactions in aqueous medium by bringing reagents in close spatial proximity and locking their conformations. The experimental examples related to oligomer formation have in common extremely lipophilic substrates and limited background reactivity when lipids are absent. To our knowledge, no study investigated the pertinence of membranes and amphiphiles in the broad context of a peptide-RNA world, notably because most oligomerization studies focus on the formation of homo-oligomers.

In this chapter, we demonstrated that the promotion of chemical reactions by lipid membranes was indeed difficult to generalize to a broad range of substrates, and in particular to nucleotides, which are membranophobic. Reactions that were already efficient in water tended to perform equally well in the presence of lipids. Furthermore, the amount of substrate that could accumulate in a bilayer or at its surface was limited, in the range of one substrate molecule per 100 lipid molecules in the case of lipophilic peptides. Since the effective concentration of lipids in a suspension is limited, this result suggested that the main opportunity offered by membranes would be to promote reactions between very dilute substrates, that would otherwise be extremely slow. Experiments employing membranophilic peptides and phosphates are ongoing to demonstrate the possibility of phosphoramidate formation inside membranes.

3.3. Perspective: first steps toward membrane anchoring of peptido-RNA and encapsulation of RNA

Reliable mechanisms for the encapsulation of small and medium molecules in the lumen of giant vesicles step remain elusive in the absence of enzymes and evolved biological channels. A classical argument, described in detail in section 2.3.1.3 is that hydrophilic substrates may be trapped spontaneously during the formation of vesicles upon hydration. The excess of non-entrapped reagent would then be washed away by natural dilution. Alternatively, sufficiently membranophilic species should be able to cross membranes by exploiting their defects. This point has also been detailed in section 2.3.1.3.

Nucleotides are inherently membranophobic and would require an ancillary fragment to permeate membranes. In this work, we suggest that simple lipophilic peptides may be suitable. In particular, phosphoramidate chemistry should provide a reversible conjugation of peptides to the nucleotides. Nucleic acids would be coupled to a membrane-permeating peptide outside of the vesicles, then cross the membrane and slowly hydrolyze inside. Only the peptide would be able to cross back and pick up the next nucleic acid, leading to progressive enrichment (Figure 3.15). The whole process would be dissipative, fueled by the coupling agent, and therefore irreversible. In this chapter, we discuss preliminary results and ongoing experiments that aim at demonstrating the feasibility of such encapsulation mechanism. In a first approach, we used small conjugates built from short peptides and mononucleotides. Such compounds could be produced in sufficient quantities to perform NMR studies. We also briefly envisioned the encapsulation of longer RNA strands through conjugation with longer cell-penetrating peptides.

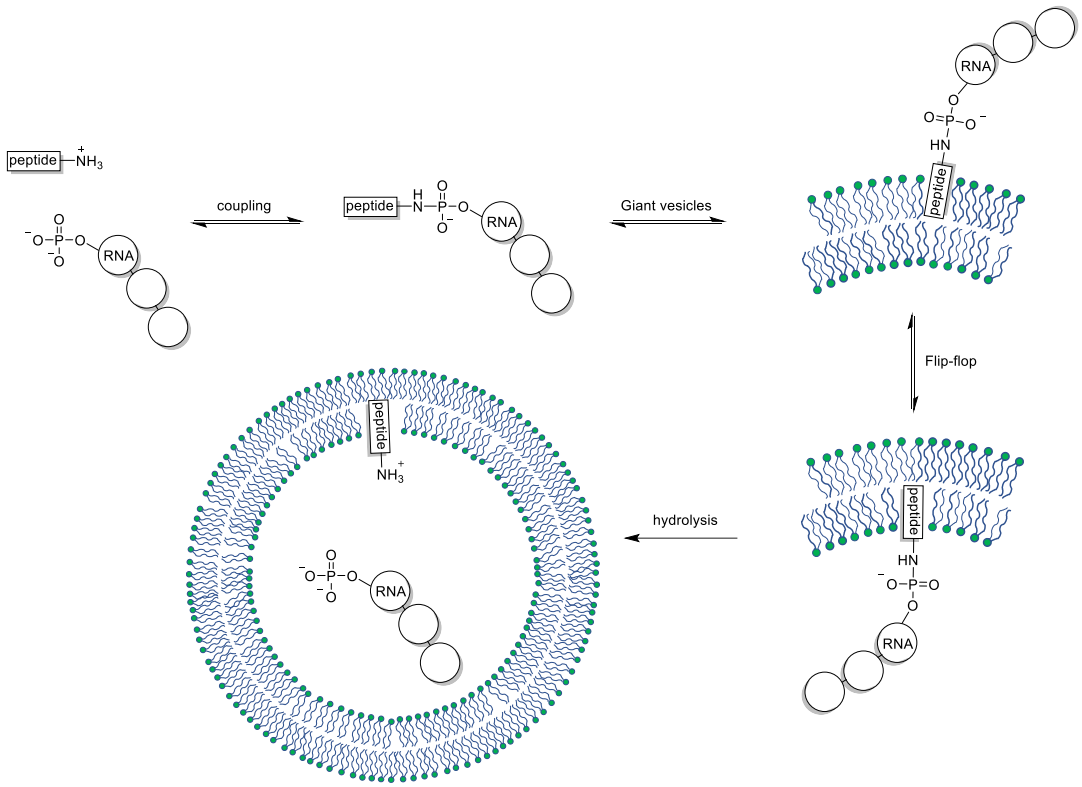


Figure 3.15 Encapsulation of RNA mediated by a reversible conjugation to a membranophilic peptide as envisioned in this work. The irreversible hydrolysis step is the driving force of the encapsulation.

3.3.1. The membranophilicity of peptido-nucleotides

The first step of the encapsulation scenario would be the binding of a peptido-nucleotide to the membrane of a vesicle. We only expected the peptide moiety to enter the bilayer, while the nucleotide would point out. Mechanisms for peptide-membrane association typically involve lipophilicity,[369] specific aggregation behavior[371] and ion-pairing interactions.[362] In our study, we only used peptides with aliphatic side chains, mostly because they reduced the opportunities for side-reactions.

3.3.1.1. Evaluation of the membranophilicity of small conjugates by diffusion NMR

Peptido-nucleotides with peptide moieties of increasing lipophilicity were prepared as described in section 3.1. We chose to prepare poly(valine)-based conjugates to corroborate the investigations of section 3.2.1 where we demonstrated that such species were formed under the general condensation conditions. Furthermore, we included the species Leu_3A because the Leu_3 moiety proved to be membranophilic (section 3.2.4).

The affinity for POPC membranes of these conjugates and control samples was then measured by diffusion NMR following the methods described in section 3.2.4. As expected, the results clearly showed that AMP did not bind to POPC

membranes. The species Val_2A and Val_3A did not bind as well, which was expected since their peptide moieties were themselves poor binders. Satisfactorily, Leu_3A exhibited binding with a substrate/lipid ratio comparable to the free peptide Leu_3, although this quantification must be mitigated by the technical limitations already explicated before. (Table 3.11).

The durability of the membrane permeation of Leu_3A was assessed by a gel filtration method similar to that employed with dansyl PE. Because Leu_3A was not fluorescent but could only be detected by UV, it was mandatory to use higher concentrations than with dansyl PE. 100 mM POPC vesicles were prepared, freeze-thawed and extruded as described before. The resulting suspension was thicker and more opaque than a typical 10 mM one. Leu_3A (10 mM) was then added from a concentrated stock solution the system was equilibrated overnight at 25°C. The vesicles were then collected within the void volume of a gel column and obtained as a turbid fraction. A UV spectrum was acquired after a 3/1000 dilution to avoid interferences caused by turbidity. No peak was found at 260 nm. ^{1}H NMR analysis of the undiluted fraction confirmed that only broad signals due to the lipids were present. We must conclude that Leu_3A did not anchor to POPC vesicles durably enough to be separated from the bulk medium.

More satisfying results were obtained with non-phospholipid vesicles. We used lauric acid and monolauroyl glycerol (LA/MLG 2:1) at the same 100 mM concentration. We prepared the vesicles using the procedure of Monnard and Deamer.[294] Lauric acid was first molten in water at 50°C then NaOH was added to reach pH 10. At this point a micellar suspension was obtained and shaken at 50°C until translucent. MLG was added to it, then the pH was slowly brought back to 7.5 with strong shaking to form vesicles. Freeze-thaw cycles and extrusion were performed as usual. The resulting 100 mM vesicles were also incubated with 10 mM Leu_3A overnight at 25°C. The vesicles were then separated from the non-encapsulated content by gel filtration, eluting with 5 mM lauric acid adjusted to pH 7.5. It was necessary to include the fatty acid above the CVC in the elution buffer to avoid vesicle rupture during the filtration.[294] The first fractions containing the lipid vesicles were assayed by UV at 3/1000 dilution and returned a concentration of 0.95 mM Leu_3A. The concentration of the lipids was determined by extraction of an aliquot and quantitative NMR using acetonitrile as an internal calibrant (appendices, Figure S 33). A value of 37 mM was found. This gave a value of $f_{bound} = 0.026$. Since all of the Leu_3A was expectedly in the bound form thanks to the purification, this value directly implies that around 2.6 Leu_3A molecules could bind 100 lipid molecule. ^{1}H NMR of the vesicle-containing fraction evidenced that Leu_3A was intact in the sample (appendices, Figure S 31 and Figure S 32). More elaborate analyses are under way to characterize more precisely the interactions that allow Leu_3A to bind LA/MLG vesicles. This experiment supports that peptido-nucleotides may adhere to "early" prebiotic membranes better than to late POPC membranes, which is consistent with the fact that permeability through these membranes is also better, as already discussed. The capacity of the LA/MLG membrane, in terms of the number of substrates per lipid molecule, seems slightly higher than the typical values that

we observed with POPC membranes but must be confirmed by more precise determinations of the latter. Finally, we note that our experiment was performed in the absence of salts, and in particular of the divalent cations (Mg^{2+}, Ca^{2+}) that are usually believed to create anion-anion salt bridges between, for example, the phosphates of RNA and phospholipids.[30] The 80 mM $MgCl_2$ included in the general condensation buffer may represent an excessive stress to such simple vesicles, but divalent cations in the low-mM range may provide even better binding.

Entry	Species	Lipid vesicles	Retention (gel permeation)	Binding (diffusion NMR)
1	AMP	POPC LUVs 10 mM	-	No
2	Val_2A	POPC LUVs 10 mM	-	No
3	Val_3A	POPC LUVs 10 mM	-	No
4	Leu_3A	POPC LUVs 100 mM	No	1:100 [1]
5	Leu_3A	LA/MLG LUVs 100 mM	2.6:100	-

Table 3.11 Association of peptido-nucleotides with POPC membranes as determined by diffusion NMR (complete description in section 3.2.4 and in the appendices) or by gel permeation chromatography. When binding was observed, a binding ratio (in number of peptido-nucleotide molecules for 100 lipid molecules) is given. For diffusion NMR, 5 mM species were dissolved in a 10 mM suspension of POPC LUVs in bicarbonate buffer at pH 8. For gel filtration, 10 mM species were added to 100 mM LUVs then separated. [1] precision unsatisfactory ($f_{bound} = 2 \pm 6\%$) due to technical limitations as stated in the appendices.

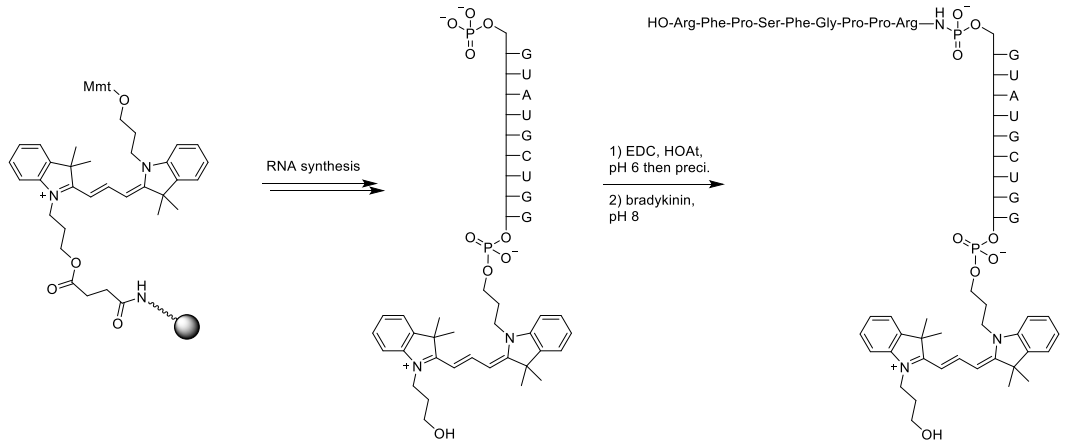
3.3.1.2. Membrane anchoring of fluorescent peptido-RNA

Following previous studies in the team,[23,25,26] we investigated the possibility to localize short RNA strands to vesicle membranes through conjugation with a membranophilic peptide, this time using phosphoramidate as a reversible covalent link. For this purpose, a peptido-RNA with longer sequences than what we described in the previous chapters was required.

This work was initiated as a part of our collaboration with the group of Prof. Clemens Richert (Universität Stuttgart, Germany), where the sample of peptido-RNA was produced. It was then studied in our group. The nonapeptide bradykinin was chosen, as its affinity for phospholipid membranes has been studied,[375] and because it is widely available. It combines hydrophobic residues (2×Phe, 3×Pro) with cationic ones (2×Arg). Bradykinin's membranophilicity is not solely based on lipophilicity or aggregation, and it is therefore highly soluble in water and in DMF, a practical advantage for synthesis. The RNA moiety was a 5'-phosphorylated non-self-complementary nonamer. The fluorescent tag Cy3 was added to the 3'-end of the RNA strand through an additional phosphodiester to allow microscopic imaging. The RNA strand was synthesized by conventional CPG-supported automated chemistry, using Mmt-Cy3 CPG to introduce the fluorophore and phosphate-on for the 5'-phosphorylation. After purification, the

5'-phosphate was activated as an OAt ester, isolated by precipitation and coupled in water with an excess of bradykinin. The product bradykinin-RNA-Cy3 was obtained in low yields (13%) and small amounts (Scheme 3.17).[456] It was then handled as a stock solution in sterile ammonium acetate buffer.

We prepared POPC giant vesicles in a bicine buffer (pH 8.2) and transferred them to a glass slide for confocal microscope observation. An aliquot of the stock of bradykinin-RNA-Cy3 was pipetted in and its progression was monitored by fluorescence imaging. At first, an intense local brightness was observed due to the high local concentration. When the compound started to diffuse, we observed that it located to the membranes (Figure 3.16).



Scheme 3.17 Synthesis of bradykinin-RNA-Cy3 by the coupling approach between OAt-RNA-Cy3 and bradykinin.

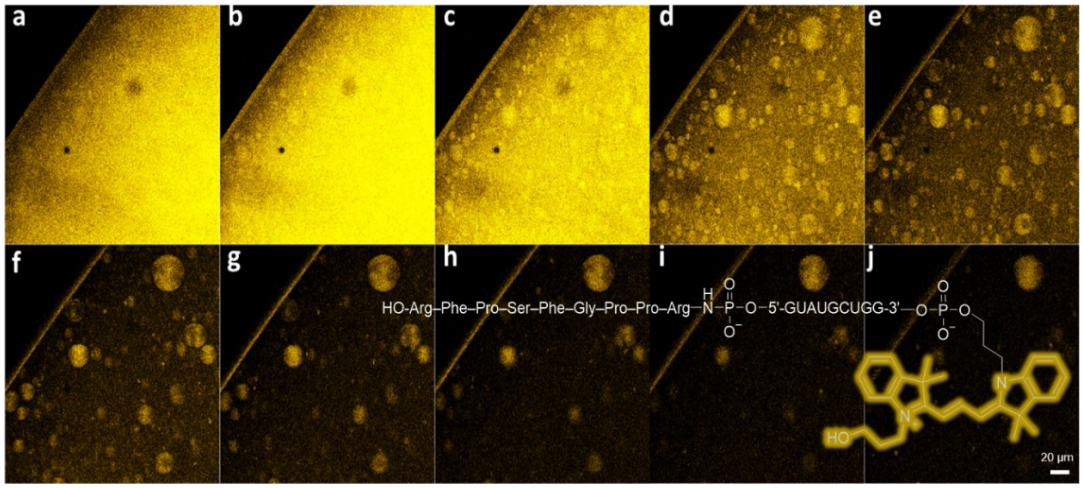


Figure 3.16 Fluorescence imaging (confocal microscope, laser scanning excitation, λ_{ex} = 547 nm, λ_{em}= 565 nm) of giant vesicles made of 1 mM POPC in a 125 mM Na-bicine buffer (pH 8.2) upon addition of the conjugate bradykinin-RNA-Cy3 (slide a). With time (slides b-j, 0-10 minutes), the conjugate migrated to the membranes of the vesicles. Scale bar: 20 µm

Unfortunately, the amounts of bradykinin-RNA-Cy3 available only covered preliminary experiments and this timelapse observation. The synthesis of a new batch is in progress to achieve quantitative measurements such as the partition coefficient of the conjugate between water and POPC.[26]

3.3.1.3. Perspective: peptido-RNA cleavage and RNA encapsulation

Exchanges between the inner and the outer layer of a lipid bilayer are known under the name "flip-flop".[337] Modern cells use flippases to control flip-flop exchanges and allow it to proceed through rigid membranes. It seems, however, that spontaneous flip-flop occurs in fatty acid membranes and in phospholipid membranes when the acyl chains are sufficiently short.[338] Cationic single-chain surfactants are known to be able to cross POPC membranes, likely through a flip-flop-like mechanism, and indirect evidence shows that small cationic peptides also have this ability.[362] It must be noted that the interior of a bilayer, where the lipid chains are stacked, is an extremely lipophilic environment that may be assimilated to an aliphatic solvent. It is therefore highly unlikely that any water-soluble material goes through it directly. For the same reason, membranophilic peptides, no matter how lipophilic they are, stay at the surface of vesicles or at best penetrate the headgroup zone up to the glycerol backbone.[335,374,457,458]

We then reason that small peptido-nucleotides should also be capable of flipping through POPC membranes. Therefore, the RNA moiety would be exposed to the interior of the vesicle. Under hydrolytic conditions (acidic medium[390] or presence of a nucleophilic catalyst such as 1-ethylimidazole but without any coupling agent[a]), the phosphoramidate would be cleaved and the nucleotides that pointed away from the inner layer trapped inside the vesicle. An alternative, more selective path would be the prior encapsulation of a hydrophilic hydrolysis catalyst. The hydrolytic abilities of the peptides Ser-His and Ser-His-Asp are known[459] and used in the active site of a wide range of hydrolases.[182] This catalytic activity has been exploited in the reverse direction to accelerate the formation of peptide bonds from amino acid methyl esters[143] and the oligomerization of nucleotide imidazolides.[460] When encapsulated in vesicles, its ability to catalyze the formation of insoluble dipeptides could be exploited as a selection mechanism in growth and division experiments.[355] Although Ser-His has never been evaluated in the context of phosphoramidates, it may be a promoter to their hydrolysis by analogy with amide hydrolysis.

Our results suggest that peptides as small as Leu_3 may be sufficient to allow mononucleotides to permeate soft membranes. Phospholipid membranes may require more lipophilic peptides. The hexapeptides used by Wimley and White are attractive candidates.[371] These peptides are poorly soluble in water but the conjugation to nucleotides may solve this problem.[20] After membrane anchoring, it is hoped that a significant population of peptido-nucleotides will be able to flip to the inner layer of a membrane and, upon acidification, to release the nucleotide. Further experiments are required to evaluate the experimental feasibility of such mechanism.

---

[a] C. Richert, unpublished results

4. Conclusions

The initial goal of this research was to investigate whether lipid membranes were able to promote, or at least influence, the formation of amphiphilic peptido-RNA in a general condensation buffer. Such effect was expected to arise from crowding mechanisms at the periphery of the membrane. We first used the general condensation reaction applied to valine and AMP as a model reaction. Valine was chosen as a compromise between lipophilicity, solubility (the experiments in the general condensation buffer were performed at 200 mM) and reactivity according to screening experiments published by Richert and coll. prior to this work.[31] Despite extensive testing, we were unable to evidence any significant effect of the presence of lipids on the outcome of the oligomerizations. In contrast, slight changes in the composition of the condensation buffer had much more noticeable consequences. Similar conclusions were drawn from investigating the formation of Val_2A from ImpA and Val_2. Three hypotheses may be formulated to explain this *non-lieu*: (i) oligovaline peptides and peptido-nucleotides did not significantly interact with lipid membranes; (ii) the availability of the membrane limited the amounts of substrate susceptible of membrane-promoted reaction, so that the processes occurring in membranes did not account for a detectable fraction of the total production of oligomers; (iii) membrane crowding does not bring about privileged reactivity pathways in the case of peptido-RNA.

The point (i) was investigated by probing polyvaline-membrane interactions using diffusion NMR. We found that, in the presence of model POPC vesicles, typical peptides such as Val_3 did not exhibit any diffusion regime consistent with membrane anchoring. Under the same conditions, the even more lipophilic peptide Leu_3 displayed two distinct diffusion regimes, one originating from membrane-bound peptide molecules. Quantitative evaluations of this binding were not possible with the NMR equipment at our disposal. We are in expectation of an access to a very high field spectrometer to perform precise measurements and probably more elaborate characterization of the interactions between peptido-nucleotides and membranes. Yet, our first results suggest that peptido-RNA chemistry is possibly sensitive to lipid membranes when very lipophilic amino acids are used but not in most other cases. Positively charged amino acids were not investigated.

The point (ii) predicted that only reactions under dilute conditions should benefit from membrane promotion to a noticeable extent. In other terms, the expected benefit of lipid membranes would be for bimolecular condensations to proceed efficiently despite high dilutions and competition with concentration-independent unimolecular hydrolysis reactions. Our data suggests that phospholipid membranes can accommodate around one peptido-nucleotide molecule for 100 lipid molecules. The value must still be confirmed by precise measurements, but the order of magnitude is consistent with literature data on peptides.[371] We performed a kinetic study of the formation of Val_2A at various concentrations but were unable to find a significant effect of lipid membranes. The point (i) suggests that this experimental design should be extended to more membranophilic substrates.

The point (iii) would stand against literature reports of chemical reactions that performed well in the presence of lipids but not, or not efficiently, in water. We attempted to address point (iii) in the specific case of phosphoramidate formation using model substrates. This investigation is still in course and requires the synthesis of new model reagents.

From our observations, we conclude that a large body of oligomerization reactions are undisturbed by the presence of lipid membranes. Jointly, we observed the formation of giant vesicles under the far-from-ideal conditions of the general condensation reaction, not only from evolved phospholipids but also from prebiotic mixtures of lipids.

Peptido-RNA was also envisioned as a means to localize RNA to the proximity of lipid membranes. We provided a preliminary experimental insight into this aspect using a synthetic, fluorescent-labelled peptido-RNA that adhered to phospholipid vesicles as observed by confocal microscopy. The range of benefits from such anchoring was theoretically exposed before the beginning of the project[13,20] but is still to be explored experimentally. In addition to these hypotheses, experiments in other groups pointed out how physically immobilizing a growing oligomer and regularly providing it with activated monomers improved the formation of long strands when compared to solution-phase reactions,[216] and how immobilized RNA templates were copied more efficiently that dissolved ones.[239] Membrane-bound peptido-RNA may then be envisioned as a way to improve RNA growth, or its replication. In such experiments, a membrane-bound RNA strand would either be randomly elongated or serve as a template for complementary copying. The vesicles would be physically kept in a small volume, for example above a filtration device,[315] and fresh reagents would be periodically added, while washing out waste and inactivated species.

Finally, peptido-nucleotides may enable an irreversible, energy-driven accumulation of nucleotides inside the lumen of vesicles, exploiting the easy hydrolysis of the phosphoramidate bond to release nucleotides from the end of membrane-penetrating peptide once they have crossed the membrane. The experimental feasibility of such complex machinery is still to be demonstrated but is a very exciting perspective for molecules that look like so many others at first glance.

5. Experimental methods and appendices

Table of contents of the appendices

1. Abbreviations

(Ac-)aaAMP	(*N*-acetyl-)aminoacyl-adenosine 5'-phosphoric anhydride
aa-tRNA	aminoacyl-transfer RNA
Ac-CoA	acetyl coenzyme A
AMP	adenosine 5'-monophosphphosphate
AmTP	amidotriphosphate
ATP	adenosine 5'-triphosphate
Bn/Bzl	benzyl
BVdU	6-((*E*)-bromovinyl)-2'-deoxyuridine
CAPA	cyclic acylphosphoramidate
CDI	1,1'-carbonyldiimidazole
CE	2-cyanoethyl
CMP	cytidine 5'-monophosphate
CPG	controlled-porosity glass
CTAB	hexadecyltrimethylammonium bromide (cetyl trimethylammonium bromide)
cTMP	cyclotrimetaphosphate
CVC	critical vesiculation concentration
Cy3	cyanine 3 fluorescent tag
dansyl	5-dimethylaminonaphthalenesulfonyl
DAP	sodium phosphorodiamidate (diamidophosphate)
DIPEA	ethyl diisopropylamine
DKP	diketopiperazine
DLS	dynamic light scattering
DMAP	4-dimethylaminopyridine
DMAPO	4-dimethylaminopyridine-1-oxide
Dmt	4,4'-dimethoxytriphenylmethyl (dimethoxytrityl)
DOSY	diffusion-ordered spectroscopy
EDC (·HCl)[a]	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (hydrochloride)
EDU	1-ethyl-3-(3-dimethylaminopropyl)urea
ee	enantiomeric excess
EEDQ	*N*-Ethoxycarbonyl-2-ethoxy-1,2-dihydroquinoline
ELSD	evaporative light-scattering detector
ESI	Electrospray ionization
EtIm	1-ethylimidazole
FAD	flavine adenine dinucleotide
FaPy	formamidopyrimidine
FFE	free-flow electrophoresis

[a] Also frequently known in the literature as EDCI or EDAC. Older papers may refer to it as CDI, which should be avoided as it is the current abbreviation for 1,1'-carbonyldiimidazole.

Fm	fluorenemethyl
FMN	flavine mononucleotide
Fmoc	fluorenemethoxycarbonyl
FRET	Förstner resonance energy transfer
(e)GFP	(enhanced) green fluorescent protein
GMP	guanosine 5'-monophosphate
G(U)V	giant unilamellar vesicle (diameter > 1 μm)
HATU	1-[bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate
HBTU	2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonate
HFIP	1,1,1,3,3,3-hexafluopropan-2-ol
HMBC	heteronuclear multiple-bond correlation
HOAt	1-hydroxy-7-azabenzotriazole
HOBt	1-hydroxybenzotriazole
Im	imidazole
IM	ion mobility
IMP	inosine 5'-monophosphate
LED	longitudinal eddy current
LUT	lookup table
LUV	large unilamellar vesicle (diameter > 100 nm)
MALDI	matrix-assisted laser desorption-ionization
MES	2-morpholinoethane-1-sulfonate
MeIm	1-methylimidazole
MLV	multilamellar vesicle
Mmt	4-methoxytriphenylmethyl (monomethoxytrityl)
MOPS	3-morpholinopropane-1-sulfonate
NAD	nicotinamide dinucleotide
NAM	nicotinamide mononucleotide
NBD	[2-(4-nitro-2,1,3-benzoxadiazol-7-yl)amine
NCA	*N*-carboxyanhydride
NCL	native chemical ligation
NHS	*N*-hydroxysuccinimide
PCR	polymerase chain reaction
PE	*O*-phosphoethanolamine
PEG	poly(oxyethylene) (poly(ethylene glycol))
PFG	pulsed field gradient
Pfp	pentafluorophenoxy
Pi	inorganic orthophosphate (irrespective of the counterions and protonation state)
PNA	peptide-nucleic acid
PPi/pp	inorganic pyrophosphate/pyrophosphate substituent (on a molecule)
pRNA	pyranoribonucleic acid
rTCA	reverse tricarboxylic acid cycle

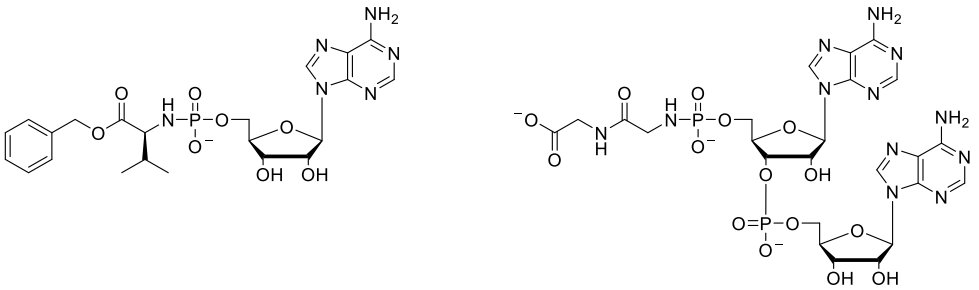
REDOR	rotational-echo double-resonance
SAX	strong anion exchange
siRNA	small interfering RNA
SNR	signal-to-noise ratio
SPE	solid-phase extraction
SUV	small unilamellar vesicle (diameter < 100 nm)
(D)STE	(dynamic) stimulated echo
TBA(F/I)	tetra-*n*-butylammonium (fluoride/iodide)
TEA	triethylammonium
(cryo)-TEM	transmission electron (cryo)microscopy
TFA	trifluoroacetic acid
TMP	thymidine 5'-monophosphate
TNA	threonucleic acid
TOF	time-of-flight
TsOH	4-toluenesulfonic acid hydrate
UMP	uridine 5'-monophosphosphate

2. Nomenclature of peptido-RNA

To this day, no specific nomenclature officially exists for peptide-RNA conjugates. Richert and coll. proposed a shorthand notation[31] that we modified slightly to produce a long and a short nomenclature. One main difficulty is that the peptido-RNA conjugation operates between the N-terminus of the peptide and the 5'-terminus of the RNA. Therefore, the peptide moiety, which is usually spelled from the N-terminus to the C-terminus, had to be reversed to take the "reading order" of the peptido-RNA sequence. Simple peptides and RNA sequences were named according to regular rules.[461,462] Additional products and reagents were designated according to the most common abbreviations in literature.

Long nomenclature

The ribonucleotide residues were designated by their 1-letter notation (A, C, G, U). The peptides were designated by their 3-letter notation (Gly, Ala, Arg etc) to avoid confusion between each part, especially when the 1-letter notations are overlapping (for example, A may designate alanine or adenosine residues). The terminations were as usual for peptide nomenclature, and each phosphate connection other than 3'-5' was designated as "p". A typical peptido-RNA therefore bears a name such as HO-Pep-Pep-Pep-pNNN (where "Pep" and "N" may be substituted by the appropriate residue abbreviations). Two examples are displayed below.

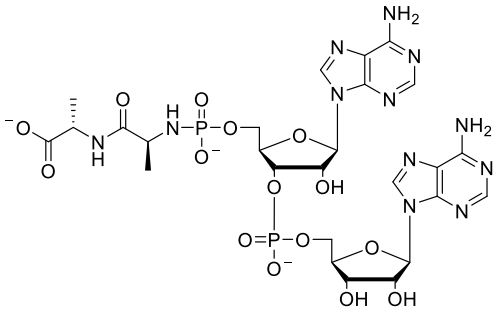


Scheme S1 Long nomenclature examples. The molecule on the left may be designated as BnO-ValpA. The molecule on the right bears the name HO-Gly-Gly-pAA

This notation gives a comfortable amount of structural details and is ideal when protective groups are employed, such as the benzylated example above. However, a shorter notation was desirable for simple cases to make the reading more fluid.

Short nomenclature

The terminations are omitted, as well as the intermediate phosphoramidate. Identical residues are grouped by a lowercase number. A generic example would be Pep_nN_p. Peptido-RNA with modifications should be named in the long form. An example is shown below.



Scheme S2 This peptido-RNA bears the short notation Ala_2A_2.

The other related products used in this work are noted as below:

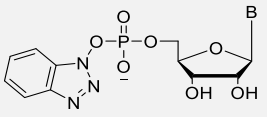
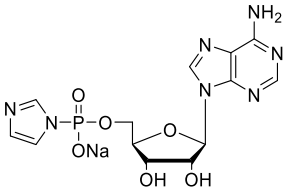
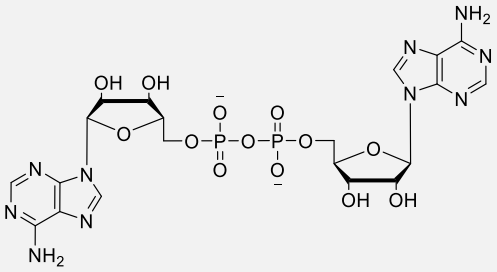
Structure	Short name	Litterary name
	BtOpN	Benzotriazolide of NMP
	ImpA	Imidazolide of AMP
	AppA	Diadenosine 5'-pyrophosphate

Table S1 Short and literary nomenclature of additional compounds.

The phosphorylated fluorophore dansyl phosphoethanolamine was in replacement of nucleotide in some experiments, and is noted as D, following the same conventions as for nucleotides. Therefore, notations such as ImpD or Leu_3D were employed.

3. Structure of the lipids used or mentioned in this work

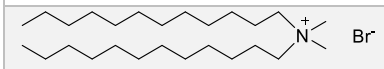
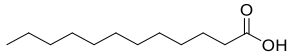
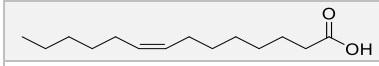
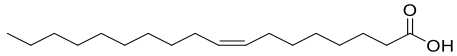
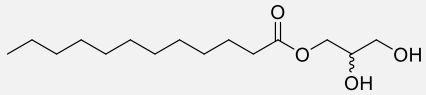
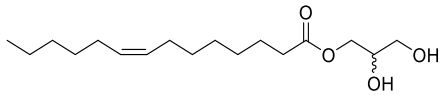
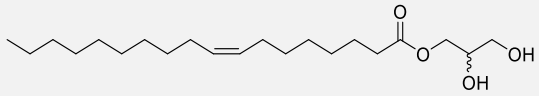
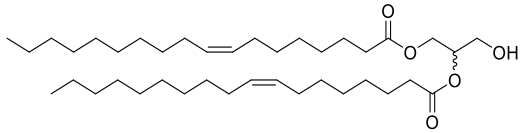
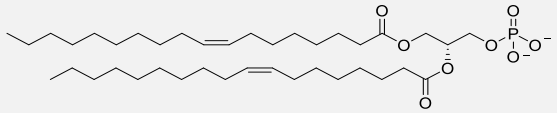
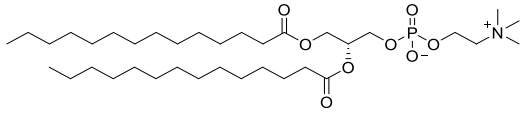
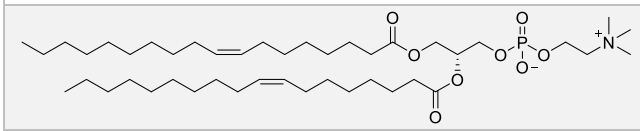
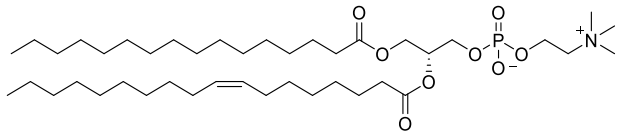
Structure	Name	Abbreviation
	Didodecyl dimethylammonium bromide	DDAB
	Lauric acid	-
	Myristoleic acid	-
	Oleic acid	OA
	*rac*-1-lauroylglycerol	MLG
	*rac*-1-myristoleoylglycerol	MMG
	*rac*-1-oleoylglycerol	MOG
	*rac*-1,2-dioleoylglycerol	DOG
	1,2-dioleoyl-*sn*-glycero-3-phosphate	DOPA
	1,2-dimyristoyl-*sn*-glycero-3-phosphocholine	DMPC
	1,2-dioleoyl-*sn*-glycero-3-phosphatidylcholine	DOPC
	1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphatidylcholine	POPC

Table S2 Structure of the lipids used and mentioned in this work.

4. Materials and general methods

Chemicals. 1-(3-dimethylaminopropyl)-3-ethyl carbodiimide hydrochloride (EDC·HCl), nucleosides and nucleotides were from Carbosynth. The fatty acids and their derivatives were from TCI Europe. The peptides Val_2, Val_3, Leu_3, Ala_3 (all-L residues throughout the entire work) and bradykinin (acetate salt) were from Bachem Germany. Lipid Removal Agent™ (LRA, Supelco) was purchased in the bulk form from Sigma-Aldrich. All the other reagents and ordinary grade solvents were from Fisher Scientific or Sigma-Aldrich. Triphenylphosphine was recrystallized from absolute ethanol. Imidazole was recrystallized from dichloromethane. All the other reagents were used as received. The tetra(n-butylammonium) salts of phosphates (nucleotides, Pi) were prepared by mixing the free acid form of the phosphate in aqueous solution with 1 or 2 equivalents of tetra(n-butylammonium) hydroxide as appropriate and lyophilizing the solution to yield a white powder that contained little water and was not hygroscopic.

Atmosphere-sensitive reactions were performed under argon in oven-dried glassware. Commercially anhydrous solvents (10 to 50 ppm water) were purchased under septum from Acros Organics and transferred by standard syringe and cannula techniques.

Thin layer chromatography was carried out on silica-coated plates (Macherey-Nagel Fluorosil 60). The plates were inspected under UV (260 or 315 nm as appropriate) and developed with either 20% sulfuric acid in ethanol, basic aqueous potassium permanganate, 2% ninhydrine in *n*-butanol or acidic aqueous ammonium molybdate/cerium sulfate.

Unless otherwise specified, **conversions** were determined by an appropriate quantitative analysis of a reaction medium and given in mole percent (mol%), and **yields** were calculated from the weight of isolated samples that were deemed pure. When traces of solvents or salts could not be removed by conventional techniques, they were deduced from the yield according to ^{1}H NMR integrations.

Distilled water and ultrapure type-1 water were obtained from a Fisher Scientific Accu 20 unit.

The pH of solutions was measured using a Mettler-Toledo Five-go pH-meter connected to a 3 mm micro-electrode and was calibrated before use with NIST standard solutions.

Unless specified, the composition of eluents and other solvent mixtures is given in volume percent (vol%).

UV-visible spectra were acquired using a variety of cuvette spectrometers with a path length of 10.00 mm or with a Thermo-Fisher Nanodrop 2000 with a path length of 1.00 mm.

1.1. NMR spectroscopy

NMR spectra were acquired on Bruker Avance 300, 400 and 500 spectrometers with BBFO, PABBO or BBI 5mm room-temperature gradient probes. The samples were thermostated to 298.2 K unless otherwise mentioned. High-resolution NMR was accomplished on a Bruker Avance III HD 700 MHz spectrometer equipped with a $^{1}H/^{13}C/^{31}P$ 5 mm QCI gradient cryoprobe cooled in liquid helium in 5 mm or 3 mm tubes.

Deuterated solvents were from Eurisotop. $CDCl_3$ was stored over basic alumina and sodium sulfate. When D_2O was used alone, pD values are reported as read from a pH-meter without correction. When H_2O/D_2O 9:1 or higher was used, pH values are reported as read and the small influence of deuterium was neglected.

Several pulse programs were used for residual water presaturation but excitation sculpting[440] was found to be preferable in most situations, especially when intense peaks needed to be saturated.

In most cases, H2' signals of nucleotides in water were masked by the water residual peak and are therefore not reported.

Standard sequences were used for the one-dimension acquisitions. For ^{31}P NMR, a relaxation delay `d1` of 5 seconds for qualitative spectra and 10 seconds for quantitation were applied. Broad-band proton decoupling was applied unless coupling constants are reported. ^{1}H-^{31}P HMBC acquisitions with or without water presaturation were acquired to help the attribution of ^{31}P peaks when needed. It used the `hmbcgpndqf` pulse program from the regular Bruker catalog, a sequence initially developed for ^{1}H-^{13}C HMBC. The z-gradients were adjusted for ^{31}P (`gpz1 = 70%, gpz2 = 30%, gpz3 = 80.5%`). The spectral width was 13 ppm for ^{1}H (centered on 3.5) and 60 ppm for ^{31}P (centered on 0) and the respective FID resolutions were 5 and 76 Hz. In practice, after data processing, cross-peaks as close at 30 to 50 Hz in ^{31}P could be distinguished, depending on the context.

Diffusion NMR was carried out on a Bruker Avance 500 MHz spectrometer with a 5 mm BBI gradient probe. The sample was not spun in the probe. The pulse program `stebpesgp1s` (Bruker, graphical representation in Figure S1), featuring bipolar shaped gradients, excitation sculpting saturation and one spoil gradient, was used. The hard pulse was calibrated to reach a precise 90° tilt using the `pulsecal` utility before each acquisition. 32 increments with a linear increase of the gradient strength from 2 to 98% were performed, with 8 dummy scans and 64 scans each, resulting in a total acquisition time of approximately 2 hours. Applying more scans failed to improve the SNR significantly and generated artefacts that are likely due to convection, especially since the intense saturation irradiation generated substantial heat in the tube. The spectral width was 12 ppm. The following delays and pulses were set: `d1` 2 s; `d20` 100 ms; `p30` adjusted between 1200 and 2000 µs for each sample by manual trials (the actual pulse length is twice this value because of the use of bipolar gradients). Shaped gradients were used. The spectra stack was Fourier-transformed and phased in

Bruker Topspin 3.5, then the data was imported into Bruker Dynamics Center 2.5.6. The peaks were manually integrated then fitted to single or double exponentials with a confidence level of 98%. The values of I_0^{bound} and I_0^{aq} , and their uncertainties ΔI_0^{bound} and ΔI_0^{aq} were used as delivered by the software to calculate the bound fraction:

$$f_{bound} = \frac{I_0^{bound}}{I_0^{bound} + I_0^{aq}}$$

Several sources of error are potentially cumulated. Below 64 scans, repeated acquisitions on a sample showed superimposable spectra and we therefore neglected errors from the acquisition itself. The processing is a source of uncertainties due to the manual integration of the peaks, the lack of SNR and the fitting itself. In principle, errors in the data processing will have consequences in the quality of the fit. We therefore considered that the fitting errors returned by Dynamics Center were representative of the error on each parameter. The error on f_{bound} was estimated by a linear combination of first order partial derivatives:

$$\Delta f_{bound} = \Delta I_0^{bound} \left| \frac{\partial f_{bound}}{\partial I_0^{bound}} \right| + \Delta I_0^{aq} \left| \frac{\partial f_{bound}}{\partial I_0^{aq}} \right|$$

$$\Leftrightarrow \Delta f_{bound} = \Delta I_0^{bound} \left(\frac{I_0^{aq}}{\left(I_0^{aq} + I_0^{bound}\right)^2} \right) + \Delta I_0^{aq} \left(\frac{I_0^{bound}}{I_0^{aq} + I_0^{bound}} \right)$$

This estimation is pessimistic because it suggests that all errors are cumulated. In practice, errors statistically compensate for each other partially, because they push in different directions. A quadratic average (square root of the sum of the squares of each contribution) is a more realistic evaluation of this phenomenon. However, with the values in this work, quadratic averages and linear cumulations gave almost identical results.

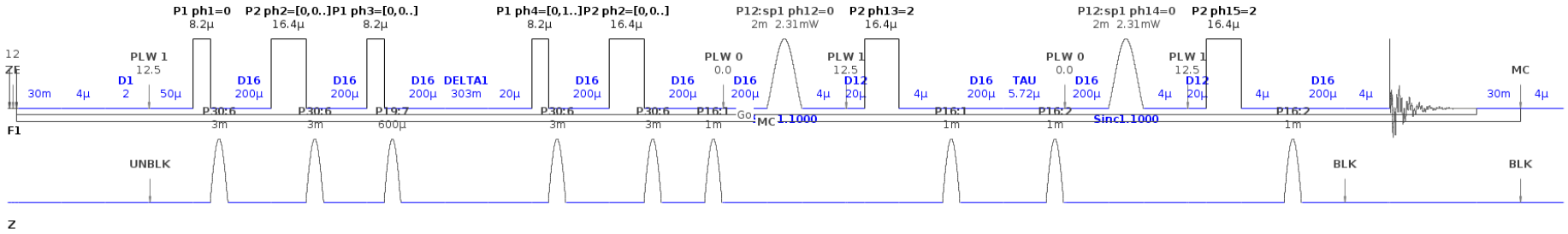


Figure S1 Graphical representation of the pulse program `stebpesgp1s` *used for diffusion NMR in this work. The top line is the* ^{1}H *channel, the bottom line is the gradient channel. The values of delays are specific of one particular experiment and are not necessarily representative of the values actually used for each experiment.*

1.2. Liquid chromatography

HPLC-grade solvents (reverse-phase flash chromatography and preparative HPLC), LC/MS grade solvents (analytical HPLC, UHPLC, LC/MS), LC/MS grade formic and acetic acids and HPLC-grade buffers and additives were from Fisher Scientific. The mobile phases were prepared daily using either pure buffers and additives or concentrated stock solutions that were filtered to 0.2 µm. Type-1 ultrapure water was used. The pH of the aqueous mobile phases was adjusted for each preparation using the appropriate acid or base.

The samples were filtered through 0.45 µm membranes (Millex FH syringe filters for organic samples and Millex PVDF for aqueous samples) before injection.

Several chromatographic systems, equipped with various columns and detectors, were used in this work. For clarity, they are designated with letters as follows.

Flash chromatography was performed on an Interchim Puriflash 210 MPLC apparatus equipped with a binary gradient pump and a UV detector. Various columns and eluents were used as described in the experimental section.

Preparative high-performance liquid chromatography (**preparative HPLC, system A**) was carried out on a Shimadzu LC20 semi-preparative system with a two-pump gradient unit (high-pressure mixer), dual-wavelength UV detection and automated fraction collection. The column used was a Phenomenex Luna Omega Polar C18, 10*250 mm, 5 µm, 100 Å.

Analytical HPLC (**system B**) was performed on the Shimadzu LC20 semi-preparative system described above, using one or a combination of the following detectors: dual-band UV detector, evaporative light-scattering detector (ELSD, the nebulizer was set at 40°C, the drying gas was nitrogen with 3.5 bar backpressure, the gain was adjusted as needed), fluorescence detector. The mobile phases were degassed by pulsed sonication before use. The columns used were a Phenomenex Luna Omega Polar C18, 3*150 mm, 3 µm, 100 Å (**column B1**), operated at 0.65 mL/min, an Agilent RPmAB C4, 2.1*100 mm, 3 µm, 300 Å (**column B2**) operated at 0.5 mL/min or a Phenomenex Kinetex Phenylhexyl, 3*100 mm, 2.6 µm, 100 Å (**column B3**) operated at 0.8 mL/min. The data was treated in the LabSolutions software.

UHPLC (**system C**) was carried out on a Thermo-Fisher U3000 system featuring a quaternary pump unit preceded by a vacuum degassing system, an automated sample injector and a quad-band UV detector. The column was thermostated at 30°C in an oven. The same columns as for system B were used. The data was treated in Chromeleon 7.

UHPLC/MS/MS (**system D**) was carried out on the same UHPLC system described above, hyphenated to a Bruker Impact II quadrupole/time of flight (Q-TOF) high-resolution mass spectrometer. Internal calibration was automatically performed before each run. The acquisition was operated at 10 Hz in dual mode: a high-resolution full MS spectrum was first acquired, then the main peak was

automatically identified and fragmented by the quadrupole unit to yield a high-resolution MS^2 spectrum. This procedure offers high-resolution MS/MS with an overall frequency of 5 Hz using a simple Q-TOF detector. The column used was a Phenomenex Luna Omega Polar C18, 2.1*100 mm, 1.7 µm, 100 Å (**column D**). The data was treated in Bruker Compass.

1.3. Mass spectrometry

Low-resolution mass spectrometry was carried out on a Bruker AmaZon SL ion trap spectrometer equipped with an electrospray source, operated in positive or negative ion mode within a mass range appropriate to the molecules to be analyzed. High-resolution mass spectrometry was carried out on a Bruker Impact II Q-TOF spectrometer equipped with an electrospray source, operated in positive or negative ion mode within a mass range appropriate to the molecules to be analyzed. MALDI-MS was carried out on a Bruker Microflex MALDI-TOF spectrometer operated in negative ion mode in the 500-2000 Da mass range. The sample (0.5 µL) was mixed with a matrix consisting of 2'-4'-6'-trixydroxyacetophenone and triammonium citrate (TC matrix, 0.5 µL) and spotted on a stainless-steel plate then dried under vacuum for 5 minutes.

1.4. Vesicle preparation and imaging

Natural and fluorescent phospholipids were purchased as stock solutions in chloroform from Avanti Polar Lipids and stored at -20°C. *rac*-1-oleoylglycerol (MOG) and *rac*-1,2-dioleoylglycerol (DOG) were prepared as described previously by us[99] and handled as stock solutions in chloroform/methanol.

Giant vesicles (GVs) of phospholipids were prepared by the natural swelling method. Appropriate amounts of lipid stock solutions were diluted with a few drops of chloroform in a glass tube. The solution was evaporated at high rotation speed in a rotary evaporator to leave a thin film that was further dried under vacuum. An appropriate aqueous solution was added to the film, the tube was gently shaken for a few seconds then allowed to hydrate without shaking for several hours. If lipids with high T_m were used, this step was carried out at an appropriately elevated temperature.

Monodisperse large unilamellar vesicles (LUVs). GVs were prepared as described above then repeatedly frozen in liquid nitrogen and thawed in warm water, 6 to 10 times. This step leads to unilamellar vesicles.[309] The resulting suspension was extruded through 0.1 µm polycarbonate membranes using an Avanti Mini Extruder. Concentrated samples (typically above 2 mM) were first extruded at 1 µm to facilitate the process. The resulting LUVs were suitable for spectroscopy, including NMR, and did not sediment over several weeks. Trace amounts of ethanol (a common stabilizer, tracked down to the commercial stock solutions) were detected by NMR in LUV samples, only in ^1H spectra when many

scans were performed. This was not removed by prolonged evaporation of the films and the quantity was too low to disturb any experiment.

Encapsulation of small molecules in the lumen or membrane of LUVs. This protocol was developed for 250 µL of 10 mM POPC LUVs extruded to 100 nm and may require adjustments for other samples. A GE Healthcare NAP-10 size-exclusion column (originally developed for DNA desalting) was equilibrated with 10 mL of the buffer used to prepare the vesicles. Gravity flow was used exclusively throughout the procedure. When the flow of equilibration buffer had stopped spontaneously, the LUV sample (250 µL) was applied gently and allowed to penetrate the column completely. 500 µL of buffer were further added and allowed to penetrate the column completely. A collection tube was placed under the column and 1000 µL of buffer were added to elute the vesicles. The column was then eluted in separated 750 µL fractions, small molecules typically eluted in fractions 4 through 6.

Optical fluorescence microscopy was carried out on a Zeiss LSM 800 microscope equipped with a Zeiss AxioCam sensor and 20x, 50x and 100x lenses. The latter was used with oil immersion. Bright field and fluorescence images were acquired through the AxioVision software then processed using ImageJ for quantitative measurements and with GIMP 2.10.8 when image enhancement was necessary for qualitative purposes only. The native aspect ratio of 660*492 px was never altered.

1.5. Peptide synthesis

Peptides supported on 2-chlorotrityl resins for supported synthesis trials were prepared by conventional solid-support techniques using an ABI 433a peptide synthesizer. The manufacturer's "FastMoc" procedures, on 0.1 or 0.25 mmol scales, were used without modification. 1 mmol of each Fmoc-amino acid (Fluka) was used, the activation solution was 0.5M HBTU/HOBt in DMF with DIPEA as a base and NMP as the solvent for all steps. The iterative Fmoc group was cleaved with 20% piperidine in DMF and the yield of each coupling was determined based on the absorbance of the piperidine-DBF adduct at 290.0 nm (the peak is wider and therefore more reliable than the conventional 301 nm peak[463]) using an in-line UV spectrometer. 2-chlorotrityl resin loaded with the appropriate Fmoc-amino acid (Bachem) was used. After the synthesis, the resin was recovered, washed thoroughly with dichloromethane and dried. The cleavage was induced using 20% HFIP in dichloromethane, 10 mL, for 30 minutes at room temperature. The eluate was then gently concentrated to 1-2 mL and poured into cold diethyl ether. The precipitate was collected, washed with diethyl ether, dried and analyzed.

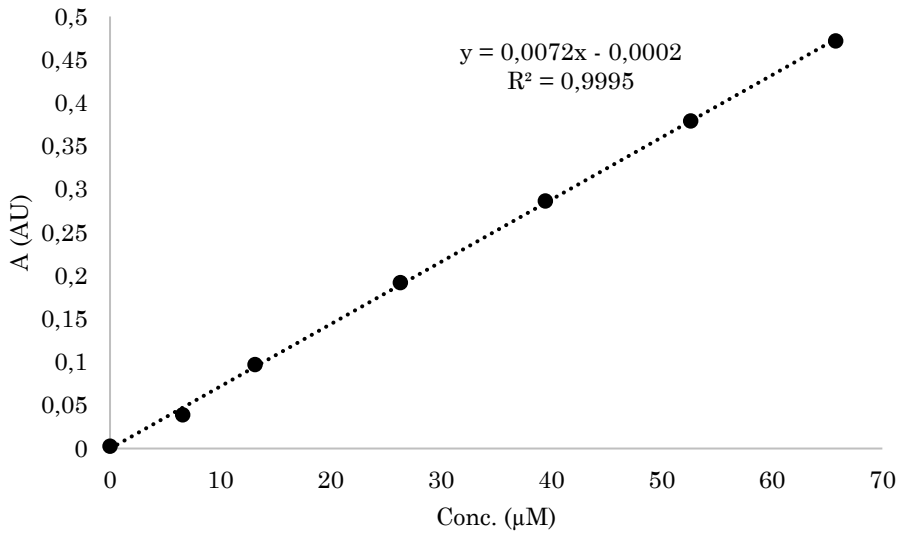
1.6. Oligonucleotide synthesis

RNA strands were prepared on an Expedite 8909 synthesizer using the manufacturer's cycles. Cy3-CPG (Glen Research) was used for the synthesis of 3'-Cy3 RNA. The monomers were TAC-protected, and the 5'-phosphorylation was accomplished with the Phosphate-On reagent (synthesized by undergraduate students in Universität Stuttgart). After the RNA synthesis was complete, the CPG (typically 40-50 mg for a 1 μmol synthesis) was recovered from the synthesis column and dried briefly under vacuum. It was treated with a mixture of aqueous ammonia (25%, ca. 1 mL) and aqueous methylamine (40%, ca. 1 mL) at 55°C for 12 minutes. The CPG was filtered off, washed with a minimum of water and the filtrate was sparged with nitrogen or submitted to vacuum evaporation (Savant SpeedVac) until the amine smell had disappeared and lyophilized. The residue was dissolved in 250 μL of $Et_3N.3HF$ and shaken at room temperature for 4h. After this time, methoxytrimethylsilane (1.5 mL) was added and large amounts of gas (Me_3SiF) evolved in the first minutes. When the gas evolution slowed down, the deprotected RNA was allowed to precipitate at 0°C for 1 hour, was collected by centrifugation, dried under vacuum and dissolved in sterile ammonium acetate (10 mM). The purification was carried out by HPLC, on system A, with a gradient of acetonitrile in 10 mM ammonium acetate.

1.7. Other methods

Stewart's quantification of phospholipids (Stewart assay)

Based on the literature procedure.[318] This protocol relies on the formation of a red, chloroform-soluble complex between phospholipids and ferrothiocyanate ions. 0.1 M ammonium ferrothiocyanate (the authors report that a mixture of $Fe(SCN)_3$ and $NH_4Fe(SCN)_4$ was prepared by mixing 0.1 mol iron (III) chloride hexahydrate and 0.3 mol ammonium thiocyanate in water and making up to 1 liter. The lipid suspension in aqueous buffer (100 μL, dilutions were made as appropriate) was mixed with ammonium ferrothiocyanate (500 μL) and chloroform (500 μL) and the biphasic sample was vigorously vortexed. The phases were separated by centrifugation and the absorbance of the chloroform phase was read at 480 nm. A calibration curve was established using POPC as a standard and fitted to a linear model.



Graph S1 Calibration curve for the Stewart phospholipid assay established with standard POPC.

5. Additional figures to section 3.1 The synthesis and the purification of nucleotide phosphoramidates

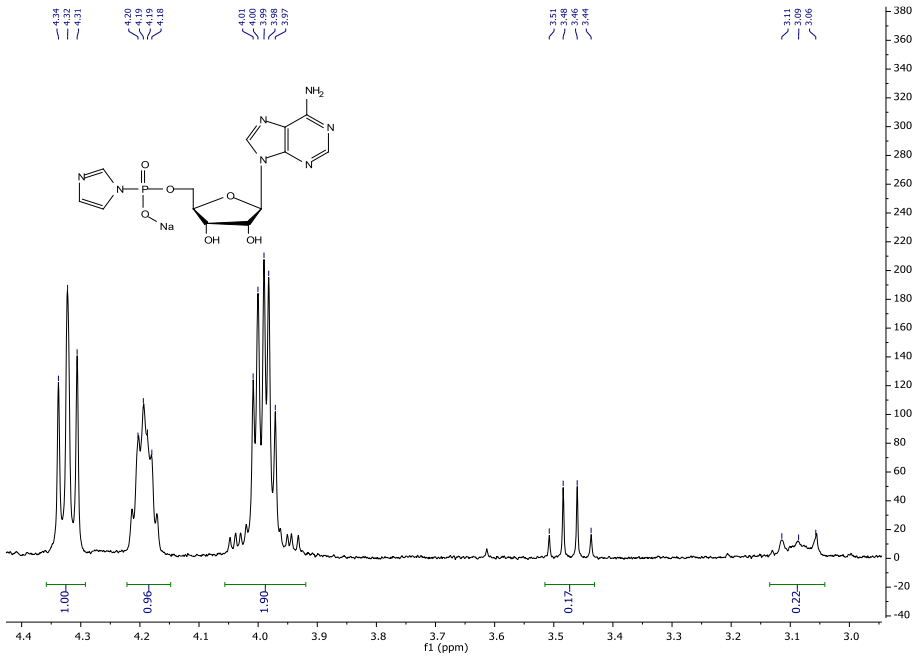


Figure S2 Enlarged view of the ^{1}H NMR of ImpA (300 MHz, D_2O) showing the residual signals for triethylamine (3.47 ppm, integral = 0.17 for 6H, 2.8 mol%) and tetrabutylammonium (3.09 ppm, integral = 0.22 for 8H, 2.7%), confirming that the exchange to the sodium salt by precipitation is efficient.

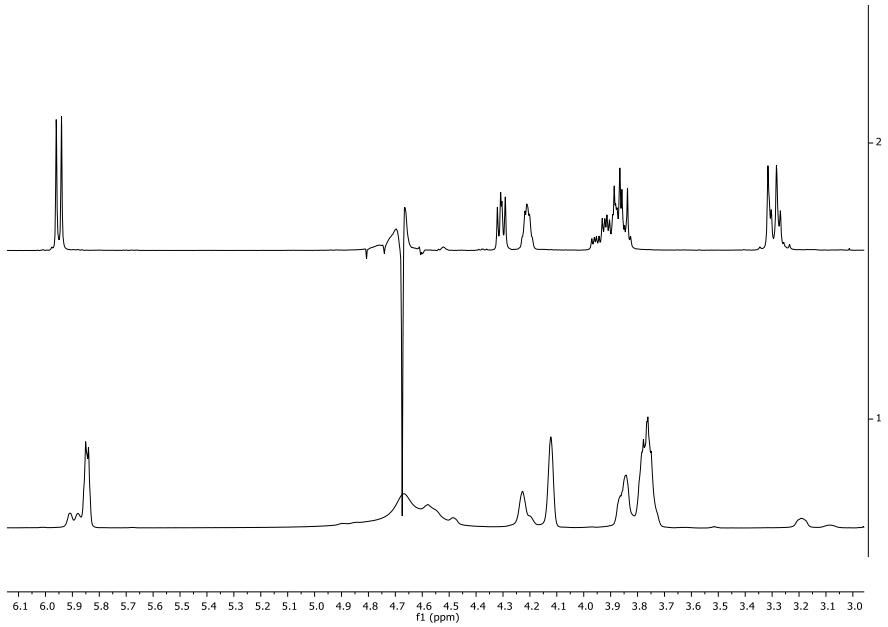


Figure S3 Stacked view of ^{1}H NMR spectra (D_2O, 500 MHz) of pure Val_2A in bicarbonate buffer (top) and in borate buffer (bottom). The broadening of the signals in borate is caused by the formation of a complex between borate ions and the 2'-3' diol of the nucleotide. This effect was also evidenced by NMR by Benner and coll.[464]

6. Additional information and procedures to section 3.2 On the role of membranes in oligomerization and conjugation of oligomers

Additional information to sections 3.2.1.1 through 3.2.1.3: general condensation reaction

AMP disodium salt (39.4 mg, 0.1 mmol, 200 mM) and L-valine (11.7 mg, 0.1 mmol, 200 mM) were dissolved in HEPES buffer (1M, 250 µL, pH 7.5, containing 10 mol% D_2O). $MgCl_2$ (40 µL of a 1M solution, 0.04 mmol, 80 mM) and 1-ethylimidazole (7.8 µL, 0.075 mmol, 150 mM) were added, followed by EDC·HCl (76 mg, 0.4 mmol, 800 mM) and the volume was adjusted to 500 µL with water. The sample was then allowed to stand for 7 days. For LC/MS analyses, a 10 µL aliquot was drawn and diluted 100 times in water. The rest of the sample was analyzed directly by NMR.

The LC/MS separation was performed on system D, column D, using a linear gradient of water and 1:1 methanol/acetonitrile (0 to 100%) with 0.1% formic acid. The products were detected in negative mode, on the mass range 150-3000 Da. When the peaks were poorly resolved, mass-directed deconvolution algorithms were used. A total of 54 species were separated by this procedure, as shown on the deconvoluted chromatogram (Figure S4, Table S4). The identity of each species was determined from its MS/MS information. Some neutral losses were characteristic from the presence of specific fragments and helped the identification of the species that arose from the side-reactivity of EDC. For more clarity, a short notation for these fragments was established as depicted in Table S3.

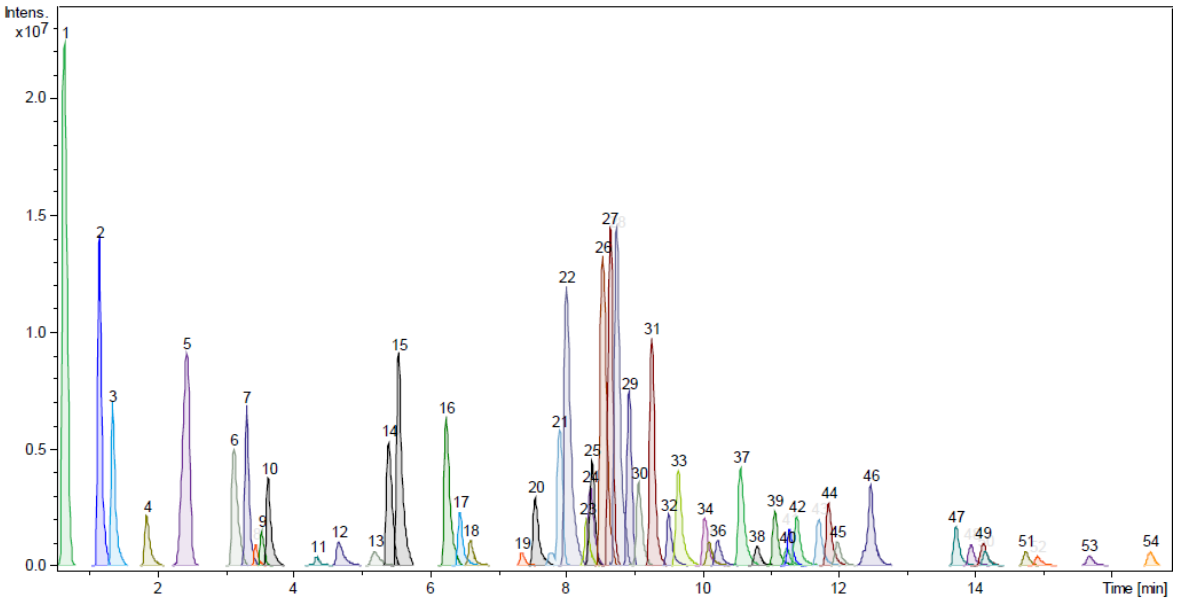


Figure S4 Total ion count chromatogram after mass-directed deconvolution of the outcome of a general condensation reaction between valine and AMP.

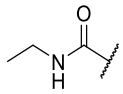
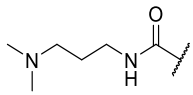
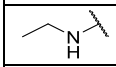
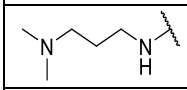
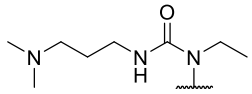
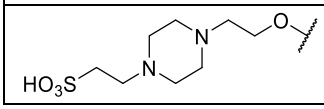
Fragment	Short name	Formula[1]	Neutral loss[1]
	EIC (ethyl isocyanate)	C_3H_5NO	71.03
	DPIC (dimethylaminopropyl isocyanate)	$C_6H_{12}N_2O$	128.10
	E (ethylamino)	$C_2H_5NO_{-1}$	nd
	DP (dimethylaminopropylamino)	$C_5H_{12}N_2O_{-1}$	nd
	EDC	$C_8H_{19}N_3O$	71.03 and 128.10 separately
	HEPES	$C_8H_{16}N_2O_3S$	220.09

Table S3 Typical fragments found in the products of the general condensation reaction. [1] the addition of these molecular fragments is accompanied by the loss of either a hydrogen atom, or a molecule of water. The neutral loss and the change in chemical formula are therefore different from the exact mass and the chemical formula of the fragment itself. nd = non detected

Peak number	Identification	Peak number	Identification	Peak number	Identification
1	HEPES	19	Val_2A_3	37	Val_2A+EDC
2	AMP	20	Val_2A_2	38	Val_3A+DP
3	AMP+HEPES	21	Val_2A_3	39	Val_3A+HEPES
4	A_2	22	Val_2A	40	Val_3A+EDC
5	A_2	23	Val_2A_2	41	Val_2A+EDC
6	AMP+DP	24	Val_3A	42	Val_3A
7	A_3	25	Val_2A_3	43	Val_3A+EDC
8	A_2+DPIC	26	Val_2A-H_2O	44	Val_4A
9	A_3	27	Val_2A	45	Val_3A
10	A_2+DPIC	28	ValA+EDC	46	Val_3A-H_2O
11	A_2+EIC	29	Val_2A+HEPES	47	Val_3A_2+EDC
12	A_2+EIC	30	Val_2A_2-O	48	Val_4A_2
13	Unknown	31	Val_2A+HEPES	49	Val_4A
14	Val_2A	32	Val_3A+HEPES	50	Val_4A_2+HEPES
15	ValA	33	Val_3A	51	Unknown
16	ValA+E	34	Val_2A+EDC	52	Val_5A+EDC
17	ValA+HEPES	35	Val_3A	53	Val_5A
18	ValA+CO_2	36	ValA+EIC+DPIC-H_2O	54	Unknown

Table S4 Identification of the species produced by a general coupling reaction between valine and AMP, as determined by LC/MS.

The NMR analysis was performed at 400 MHz (^{1}H frequency). A representative ^{1}H-^{31}P HMBC spectrum is shown below.

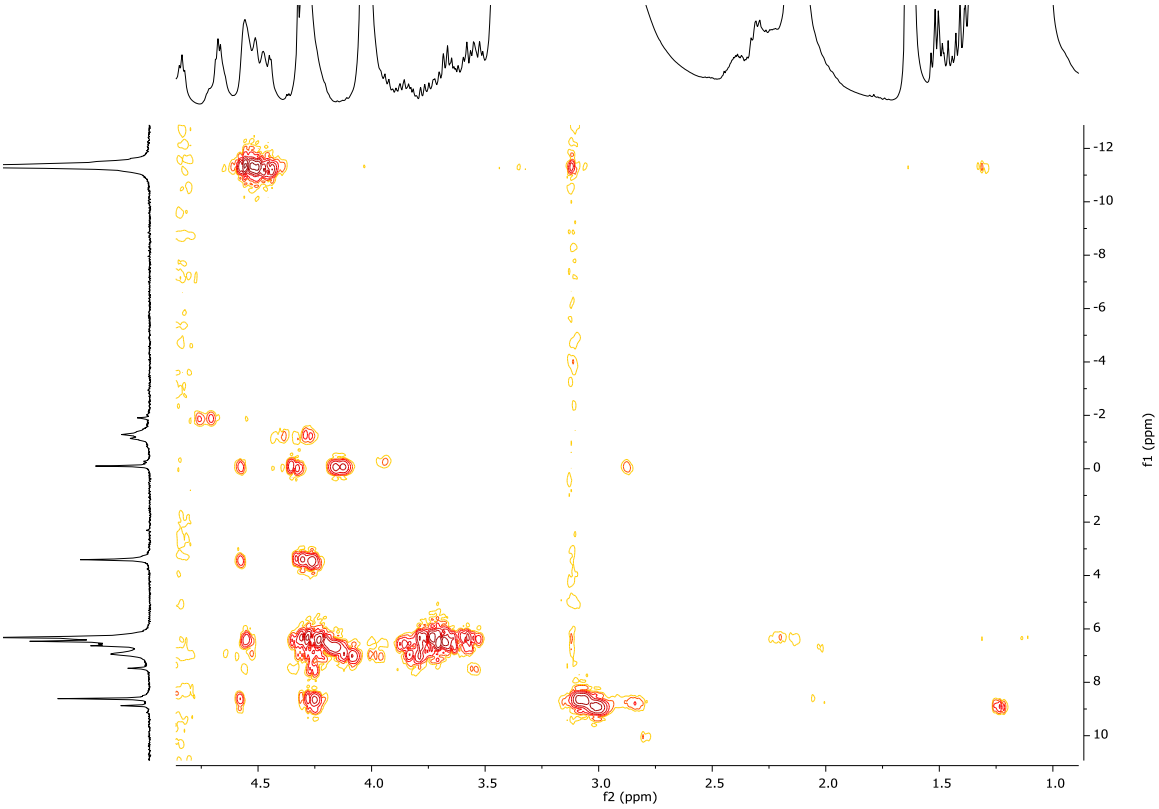


Figure S5 ^{1}H-^{31}P HMBC NMR spectrum (H_2O, 400 MHz) of a crude general condensation mixture obtained from valine and AMP in the absence of lipids. The massive peaks overwhelming the proton trace are due to the excess of EDU and HEPES.

Additional information to section 3.2.2: Simplified condensation reaction in the presence of lipids

An appropriate mixture of lipids was evaporated from stock solutions in chloroform (or chloroform/methanol for MOG and DOG) to form a thin, uniform lipid film. It was further dried under vacuum to remove traces of organic solvents. AMP disodium salt (39.4 mg, 0.1 mmol, 200 mM) and L-valine (11.7 mg, 0.1 mmol, 200 mM) were dissolved in water (250 µL). 1-ethylimidazole (7.8 µL, 0.075 mmol, 150 mM) was added, followed by EDC·HCl (76 mg, 0.4 mmol, 800 mM) and the volume was adjusted to 500 µL with water. If necessary, the pH was corrected to 7.5 using concentrated NaOH or HCl. The solution was transferred to the lipid film with gentle tumbling then the sample was allowed to stand for 7 days. For LC/MS analyses, a 10 µL aliquot was drawn and diluted 100 times in water/methanol 9:1. For NMR analyses, the rest of the sample was diluted with MeOD until homogenous.

Additional information to section 3.2.2.2: simplified condensation reaction in the presence of various concentrations of DOPA

The protocol above was followed using DOPA (1,2-dioleoyl-*sn*-glycero-3-phosphate) at concentrations of 0 (control), 0.5, 1, 2 and 5 mM. Each sample was

denatured with MeOD and analyzed by ^{31}P NMR. The peaks corresponding to each class of compounds were integrated and the integrals summed up to 100. The spectrum of the sample containing 1 mM DOPA is shown as a representative example.

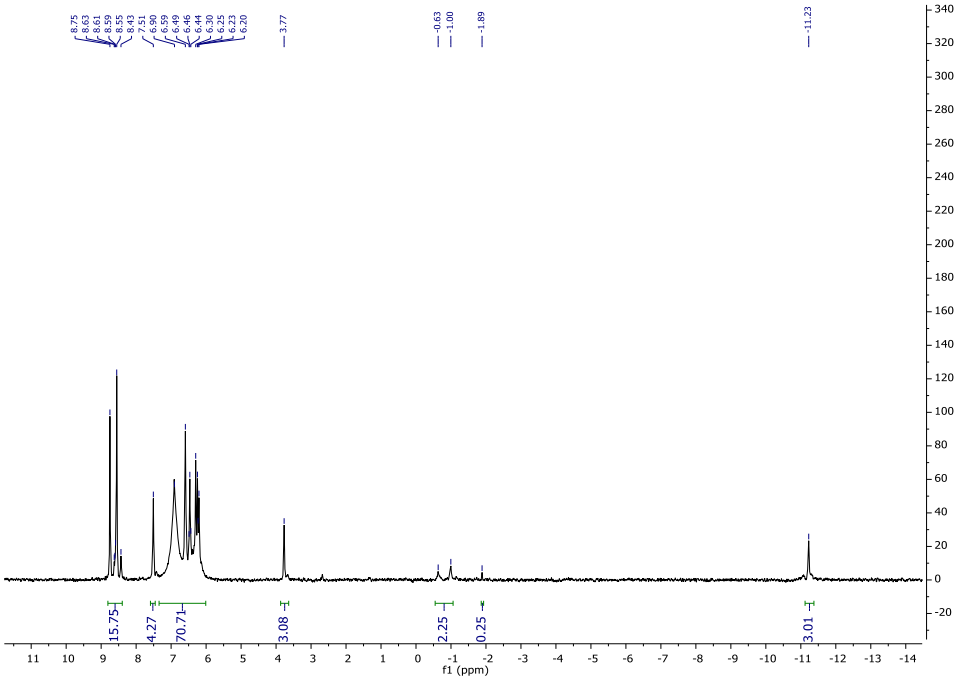


Figure S6 ^{31}P NMR spectrum (MeOD/H_2O, 160 MHz) of a crude general condensation mixture obtained from valine and AMP in the presence of 1 mM DOPA for 7 days at 25°C.

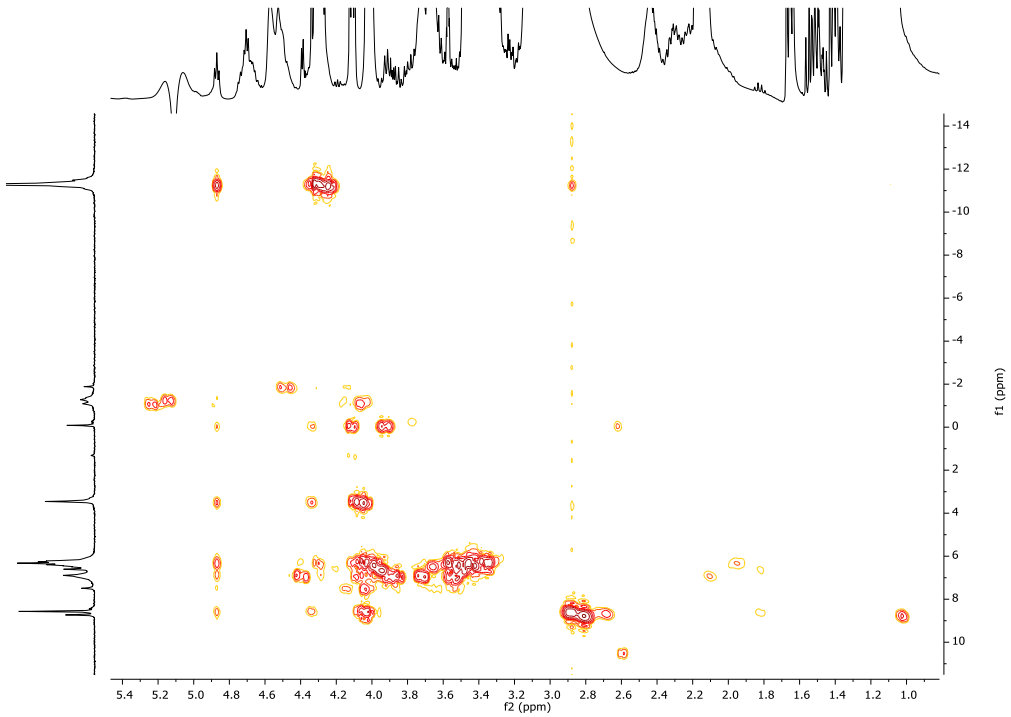


Figure S7 ^{1}H-^{31}P HMBC NMR spectrum (MeOD/H_2O, 400 MHz) of a crude general condensation mixture obtained from valine and AMP in the presence of 1 mM DOPA for 7 days at 25°C. The massive peaks overwhelming the proton trace are due to the excess of EDU and HEPES.

The quantitative distribution of species was calculated and plotted for each experiment.

Additional information to section 3.2.2.3: simplified condensation reaction in the presence of various lipid mixtures

The protocol described above was followed using lipid mixtures A, B, C and D in the table below. A control experiment without lipids was performed the same way. The total concentration of lipids was 2 mM.

	Lipid proportion				
Mixture	Oleic acid	MOG	DOG	DOPA	DOPC
A	8	2	-	-	-
B	1	1	2	6	-
C	-	-	-	10	-
D	-	-	-	3	7

Table S5 Lipid mixtures (molar fractions) used in the simplified condensation reaction. MOG: 1-oleoylglycerol; DOG: 1,2-dioleoylglycerol; DOPA: 1,2-dioleoyl-glycero-1-phosphate; DOPC: 1,2-dioleoyl-glycero-1-phosphatidylcholine. The total amount of lipids was 2 mM in each experiment.

Each sample was denatured with MeOD and analyzed by ^{31}P NMR. The peaks corresponding to each class of compounds were integrated and the integrals summed up to 100. The spectrum of the sample A is shown as a representative example (Figure S8).

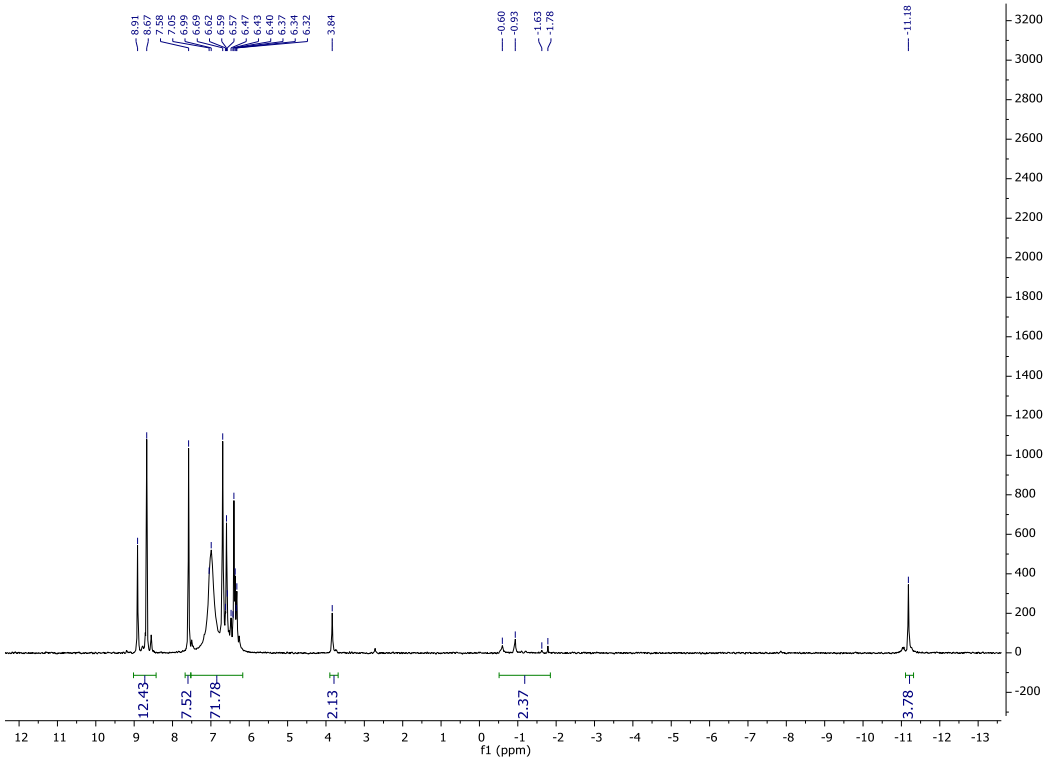


Figure S8 ^{31}P NMR spectrum (MeOD/H_2O, 160 MHz) of a simplified condensation reaction performed in the presence of the lipid mixture A for 7 days at 25°C.

The quantitative distribution of species was calculated and plotted for each experiment.

Additional information to section 3.2.3: imidazolide displacement reaction in the presence of lipids

In a glass tube, an appropriate mixture of lipids was evaporated from its chloroform solution to a thin, homogenous film and the film was further dried under vacuum. ImpA and Val_2 were diluted to the appropriate concentrations in ice-cold 100 mM MOPS, pH 7.5, and the solution was transferred to the lipid film with gentle tumbling. The samples were then allowed to stand without stirring at 0°C and the course of the experiment was followed by analytical HPLC (system C, column B1). Control experiments without lipids were performed in a similar manner with omission of the film formation step. For experiments involving lipids up to 1 mM, a 10 µL sample was drawn, diluted 100 times in water, and injected. For experiments involving more than 1 mM lipids, the lipids were removed using Lipid Removal Agent™ (Supelco) as follows. Lipid removal agent™ was wetted with water and centrifuged. The supernatant water was discarded, and a slurry was obtained, that could be pipetted with precision. 10 µL of the sample containing lipids was diluted to 50 µL with water, then 10 µL of the slurry were added and the suspension was vigorously shaken. The powder was separated by centrifugation and 50 µL of supernatant collected. 50 µL of water were further added to the power residue, the extraction was repeated, and the supernatants merged to a 100 µL sample that was diluted appropriately and micro-filtered before the injection.

Validation of the HPLC quantification by ^{31}P NMR

A control experiment involving 50 mM Val_2 and 50 mM ImpA without lipids as described above was analyzed after 15 days by analytical HPLC with UV detection at 260 nm (representative example in Figure S9) and by quantitative ^{31}P NMR. For the main species of interest, HPLC and NMR quantifications matched (Table S6). The NMR analysis also confirmed that phosphodiesters and other minor products formed to a very limited extent. The error on the estimation of AppA is justified by its very different absorbance coefficient at 260 nm. pApA could not be clearly located on the chromatogram. Overall, the two products of highest quantitative interest, AMP and Val_2A, were highly consistent.

Product	NMR	UV HPLC	Average	Stdev	Error (%)
ImpA	13.4	15.8	14.6	1.70	12
AMP	59.7	57.7	58.7	1.41	2.4
Val_2A	21.7	22.5	22.1	0.57	2.6
AppA	2.4	4.04	3.22	1.16	36
pApA	0.7	nd	-	-	

Table S6 Quantification of the species obtained after a displacement reaction by HPLC (absorbance at 260 nm) and ^{31}P NMR. Standard deviations and error fractions are calculated to estimate the level of consistency between the two methods.

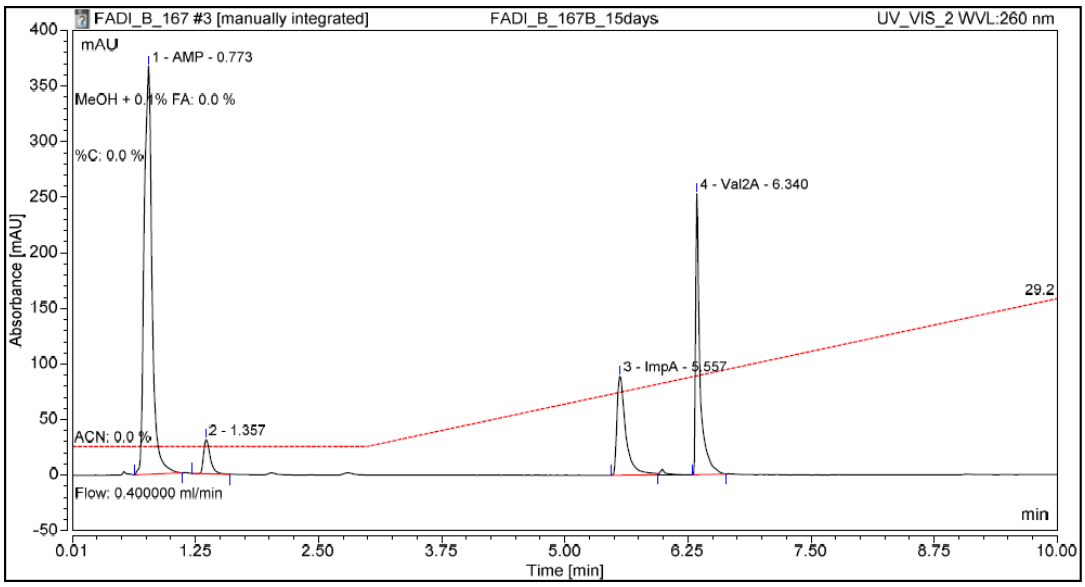


Figure S9 Analytical HPLC chromatogram of an aliquot of a displacement reaction without lipids after 15 days. Detection: UV at 260 nm. Eluents: A. ammonium acetate 10 mM pH 7.5; B. acetonitrile. The limited resolution and peak tailing motivated us to use acidic eluents (0.1 % formic acid) in further studies.

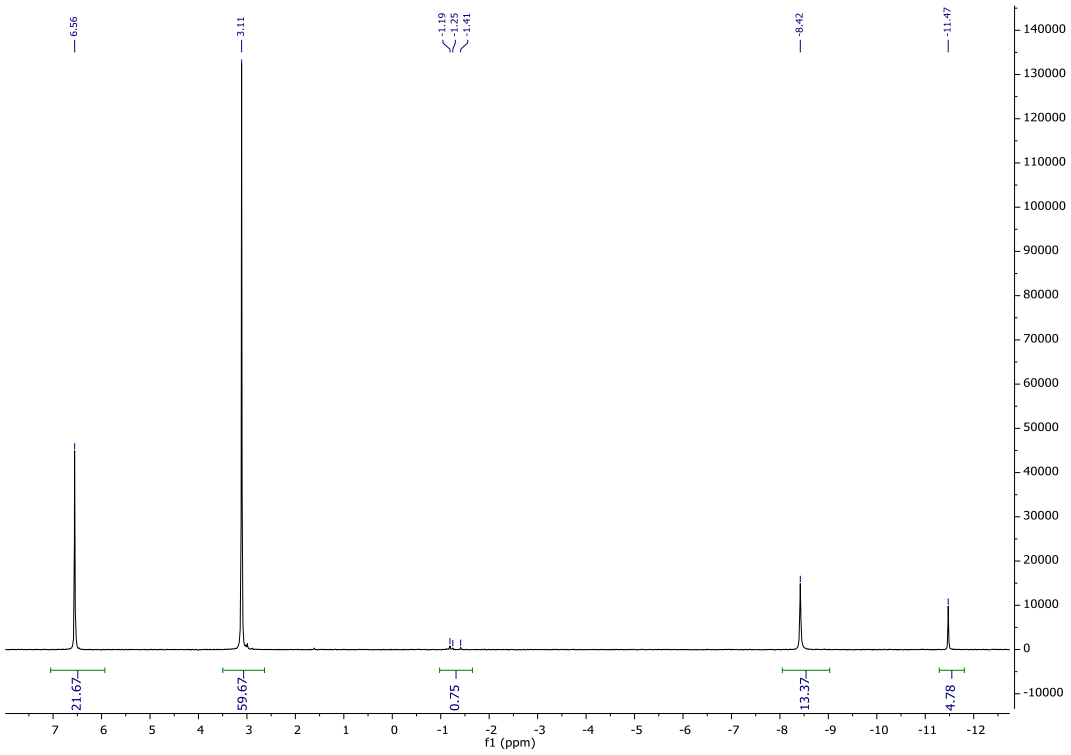


Figure S10 ^{31}P NMR (H_2O , 120 MHz) of an aliquot of a displacement reaction without lipids after 15 days. Integrals normalized to 100.

Imidazolide displacement reaction in imidazole buffer

The reaction was performed under the same conditions except that 100 mM imidazole were added. The outcome was determined after 15 days by HPLC and NMR and compared as described above. The HPLC analysis was repeated after 28 days.

Product	NMR	UV HPLC	Average	Stdev	Error (%)
ImpA	55.3	52.6	53.95	1.9	3.5
AMP	18.0	17.3	17.65	0.5	2.8
Val_2A	22.8	23.2	23	0.3	1.2
AppA	1.02	nd	-	-	-
pApA	1.53	nd	-	-	-

Table S7 Quantification of the species obtained after a displacement reaction in the presence of imidazole by HPLC (absorbance at 260 nm) and ^{31}P NMR. Standard deviations and error fractions are calculated to estimate the level of consistency between the two methods.

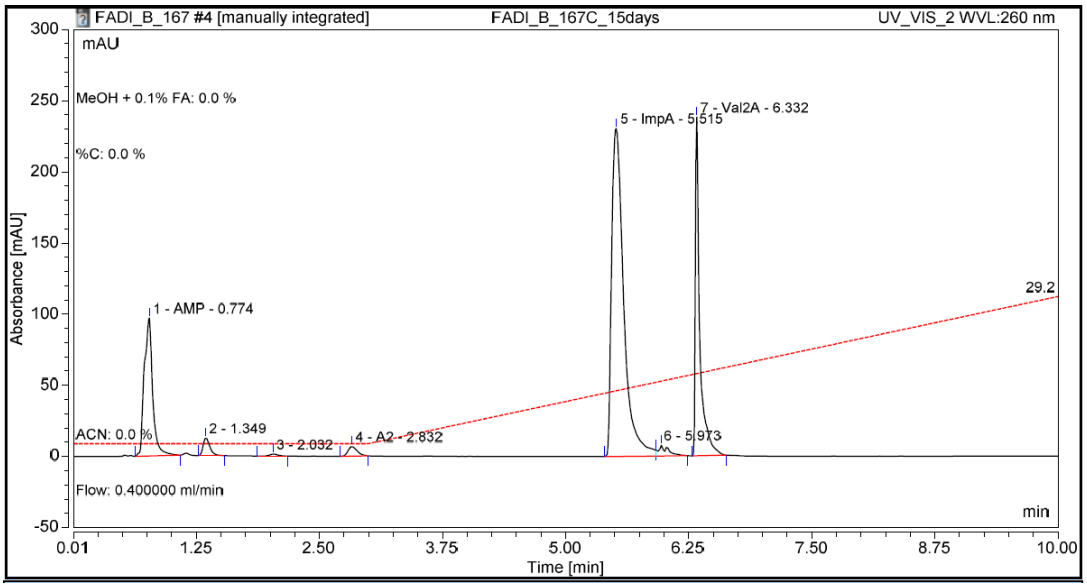


Figure S11 Analytical HPLC chromatogram of an aliquot of a displacement reaction without lipids, including 100 mM imidazole after 15 days. Detection: UV at 260 nm. Eluents: A. ammonium acetate 10 mM pH 7.5; B. acetonitrile. The limited resolution and peak tailing motivated us to use acidic eluents (0.1 % formic acid) in further studies.

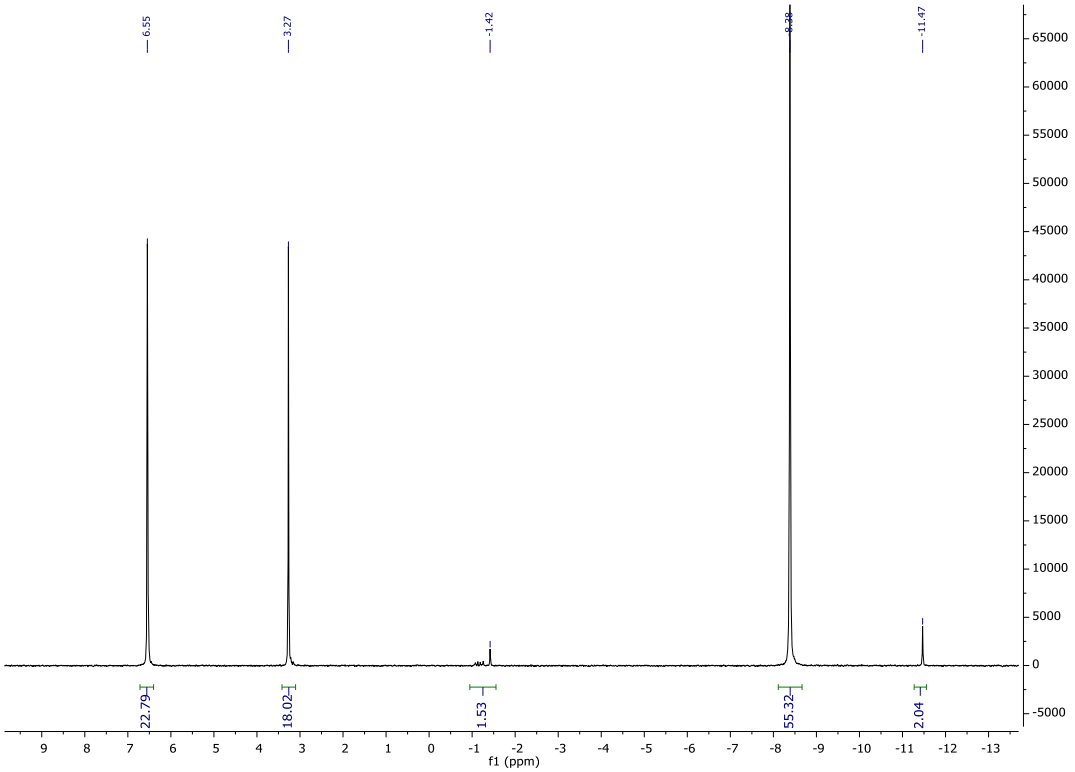


Figure S12 ^{31}P NMR (H_2O , 120 MHz) of an aliquot of a displacement reaction without lipids, including 100 mM imidazole, after 15 days. Integrals normalized to 100.

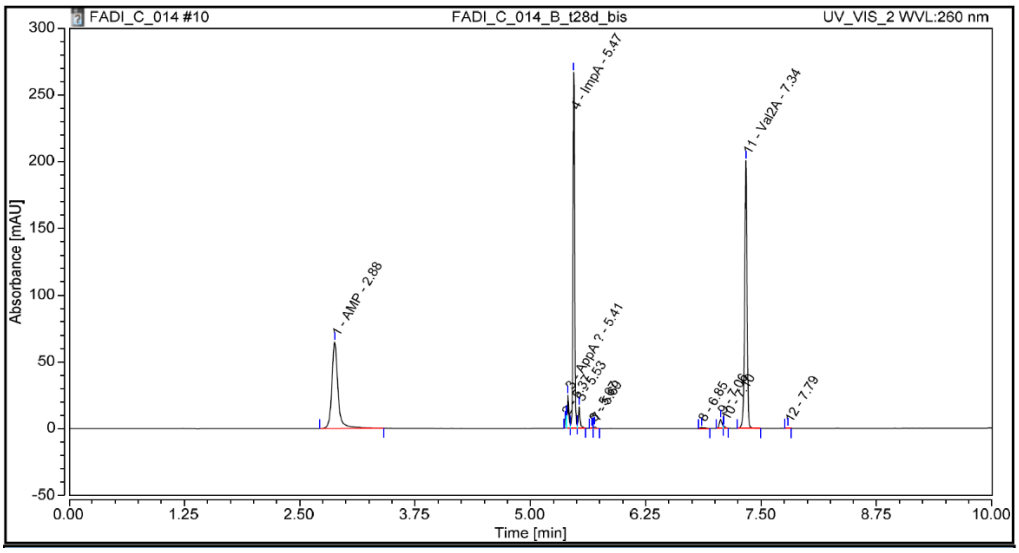
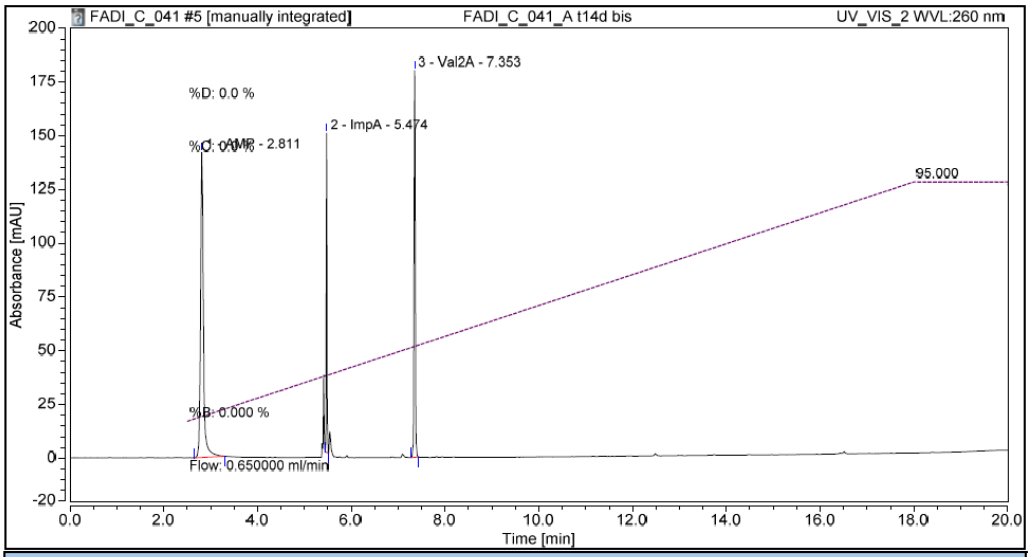


Figure S13 Analytical HPLC chromatogram of an aliquot of a displacement reaction without lipids, including 100 mM imidazole after 28 days. Detection: UV at 260 nm. Eluents: A. 0.1 % formic acid in water; B. 0.1 % formic acid in acetonitrile. Relative integrals: AMP 28.6 %, ImpA 29.3 %, Val₂A 34.2 %. The gradient is the same as depicted above.

Imidazolide displacement reaction in the presence of various concentrations of POPC

ImpA (10.5 mg, 50 mM) and Val_2 (5.40 mg, 50 mM) in 500 µL of MOPS buffer were used. A representative chromatogram with the eluent gradient and the integration marks is shown below.



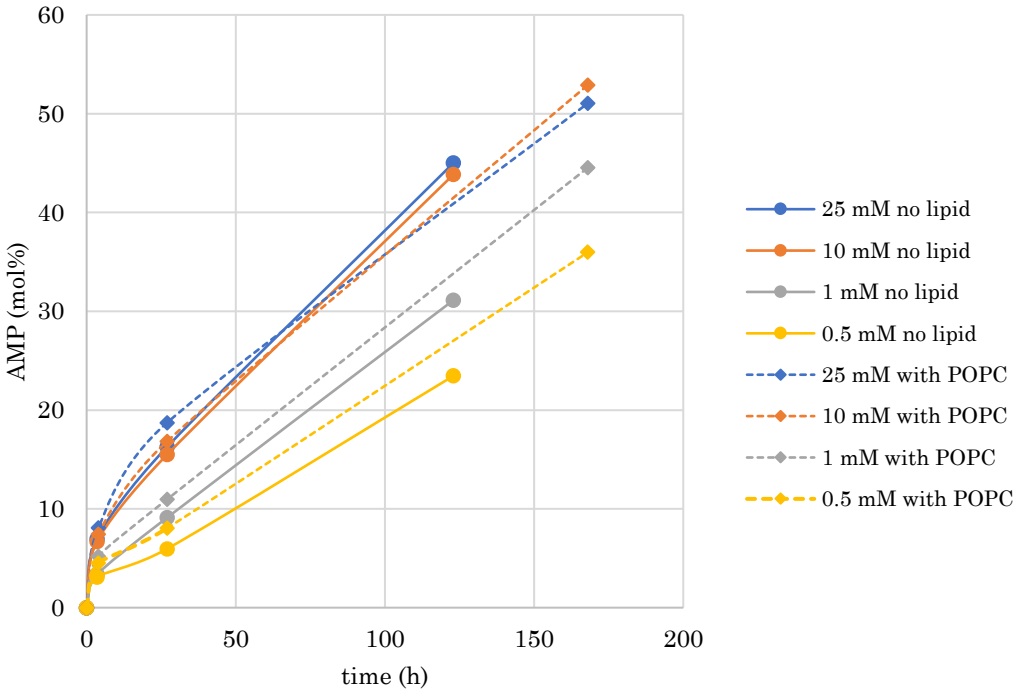
Integration Results

No.	Peak Name	Retention Time min	Area mAU*min	Height mAU	Relative Area %	Relative Height %	Amount
1	AMP	2.811	10.176	141.902	56.13	30.18	n.a.
2	ImpA	5.474	2.747	148.384	15.15	31.56	n.a.
3	Val2A	7.353	5.206	179.914	28.72	38.26	n.a.
Total:			**18.129**	**470.200**	**100.00**	**100.00**	

Figure S14 Representative chromatogram of an aliquot of a displacement reaction in the presence of 50 mM POPC after 14 days. The same method was used for the other samples. The eluents were A. 0.1% formic acid in water and B. 0.1% formic acid in acetonitrile.

Imidazolide displacement reaction under dilute conditions

A 50 mM solution of ImpA and Val_2 was prepared in cold MOPS buffer then diluted to the appropriate values in cold MOPS buffer to prepare 500 µL solutions. The course of the reactions was followed by HPLC using the same method depicted above. The formation of AMP is depicted below



Graph S2 Formation of AMP upon hydrolysis of ImpA in the displacement reaction at various concentrations of reagents, with (dashed lines) or without (solid lines) 50 mM POPC, as followed by analytical HPLC, using the UV trace at 260 nm.

Complement to section 3.2.7: general condensation reaction of AMP with alkanols

Representative procedure: AMP (34.7 mg, 200 mM) was dissolved in 250 µL of 1 M MOPS buffer pH 7.5 containing 10 % D_2O then magnesium chloride (10 mM), 1-ethylimidazole (7.2 µL, 150 mM) were added and the volume was made up to 500 µL with water containing 10% D_2O. 500 µL of the desired alcohol was added (in the case of glycerol, 46 mg were dissolved in the aqueous layer to make a 1 M solution) and the biphasic system was strongly vortexed to saturate the aqueous phase with the alcohol. EDC (77 mg, 800 mM) was added in one portion and the mixture was shaken for 48 hours. The phases were then separated by centrifugation and the aqueous layer was analyzed by ^{31}P NMR. The total of the integrals was set to 200 to directly read the overall concentration of phosphate species. A control analysis of the alcoholic phases (50 µL DMSO-d_6 added for the lock procedure) revealed negligible concentrations of species other than the alcohol in these.

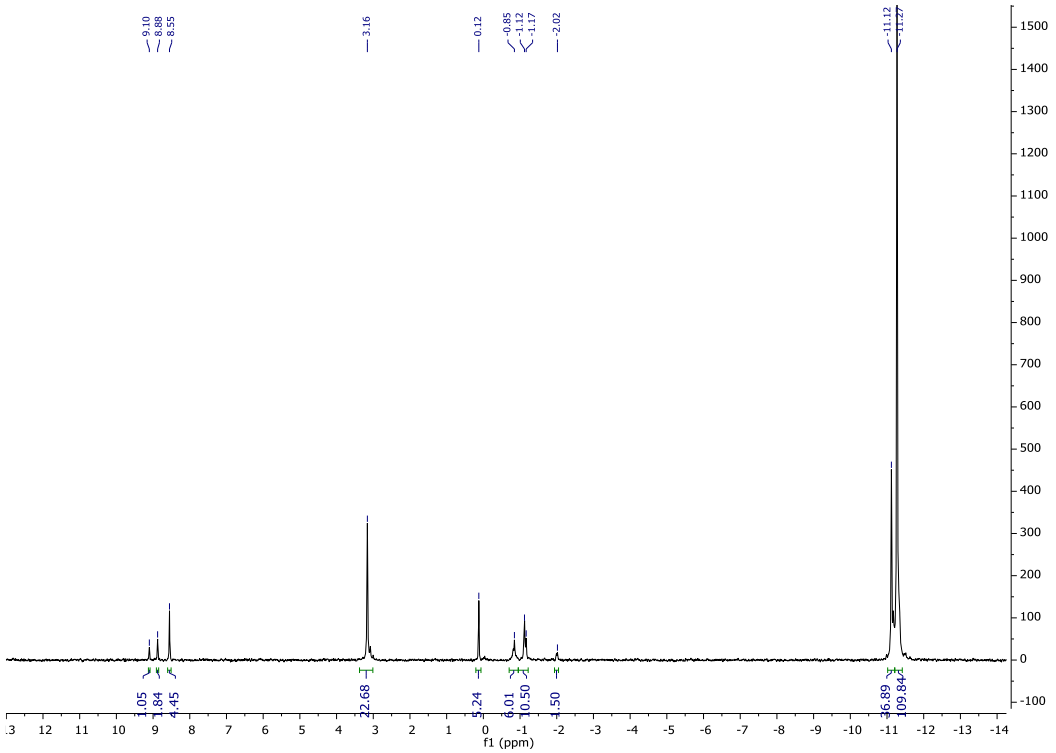


Figure S15 ^{31}P NMR spectrum (H_2O, 160 MHz) for an assay involving 0.5 M 1-butanol and 200 mM AMP. The desired phosphodiester has a shift of 0.12 ppm.

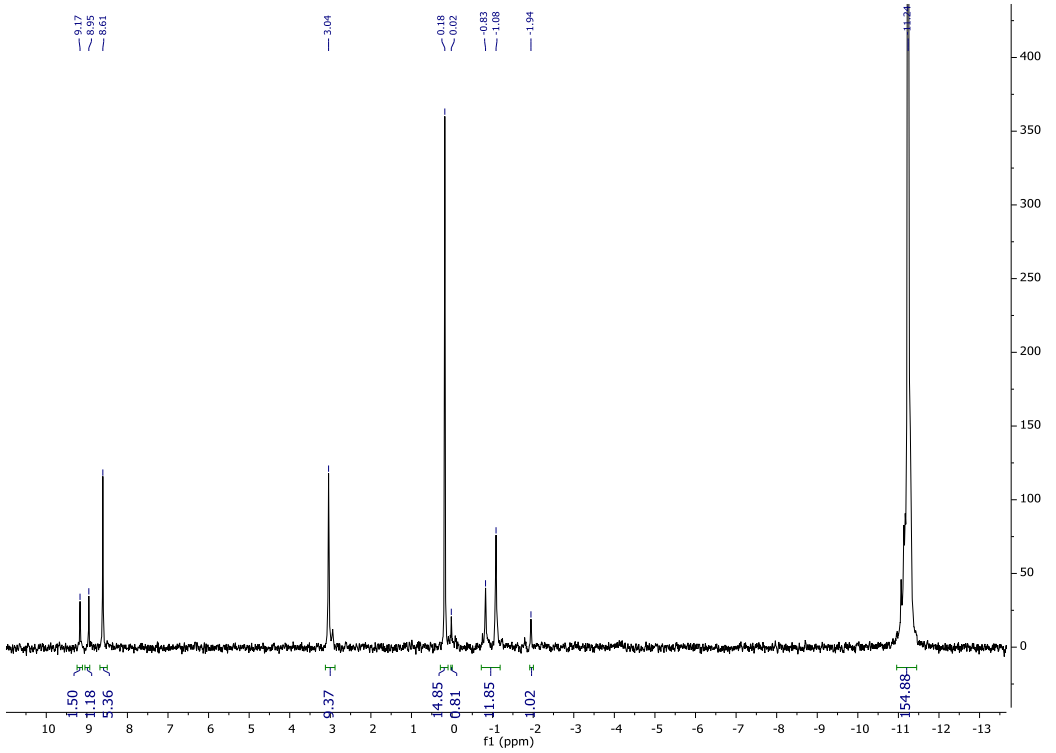


Figure S16 ^{31}P NMR spectrum (H_2O, 160 MHz) for an assay involving saturated 1-butanol and 200 mM AMP. The desired phosphodiester has a shift of 0.18 ppm.

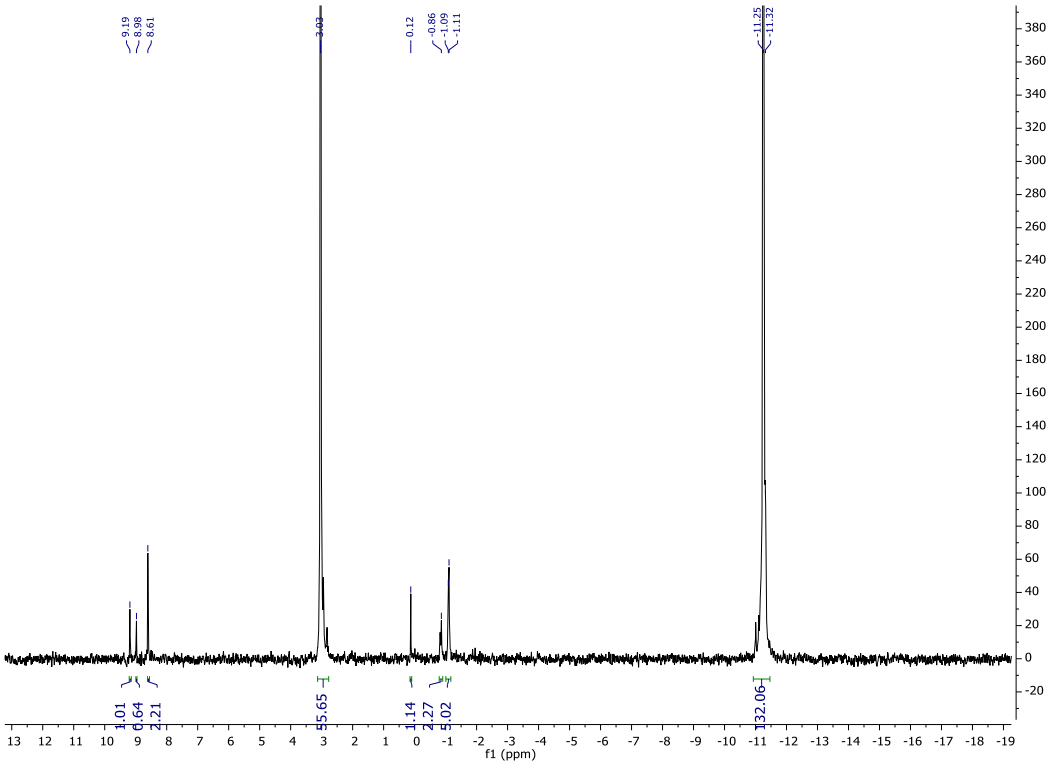


Figure S17 ^{31}P NMR spectrum (H_2O, 160 MHz) for an assay involving saturated 1-hexanol and 200 mM AMP. The desired phosphodiester has a shift of 0.12 ppm.

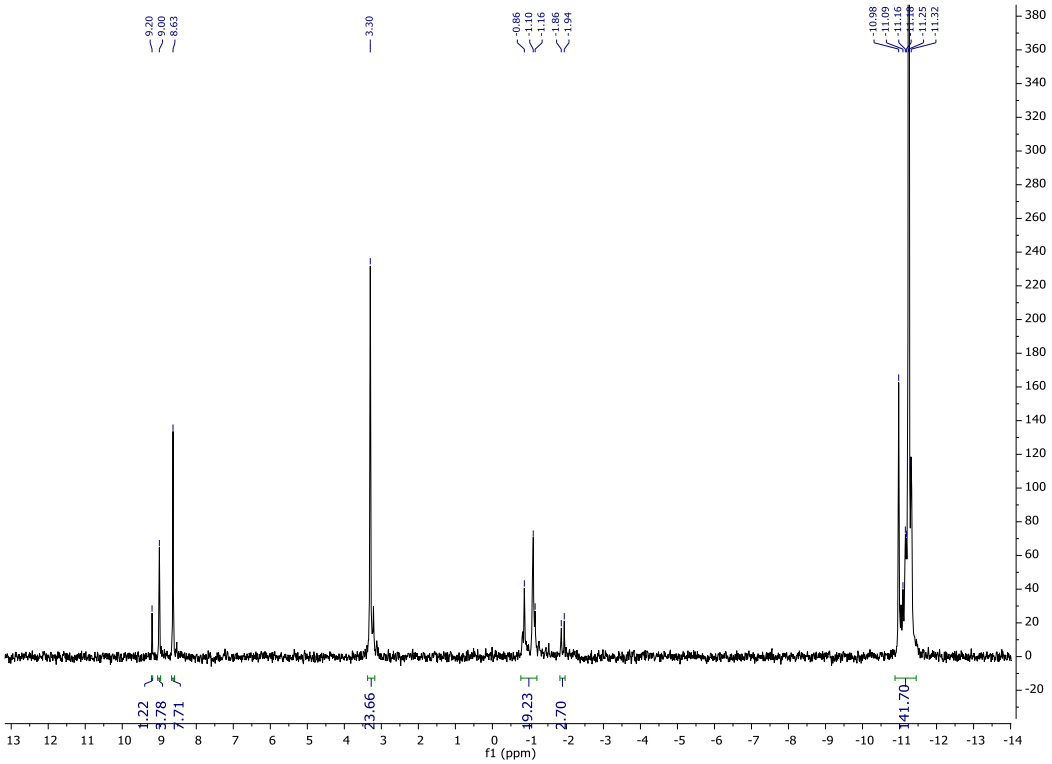


Figure S18 ^{31}P NMR spectrum (H_2O, 160 MHz) for an assay involving saturated 1-octanol and 200 mM AMP. The desired phosphodiester was below the detection limits.

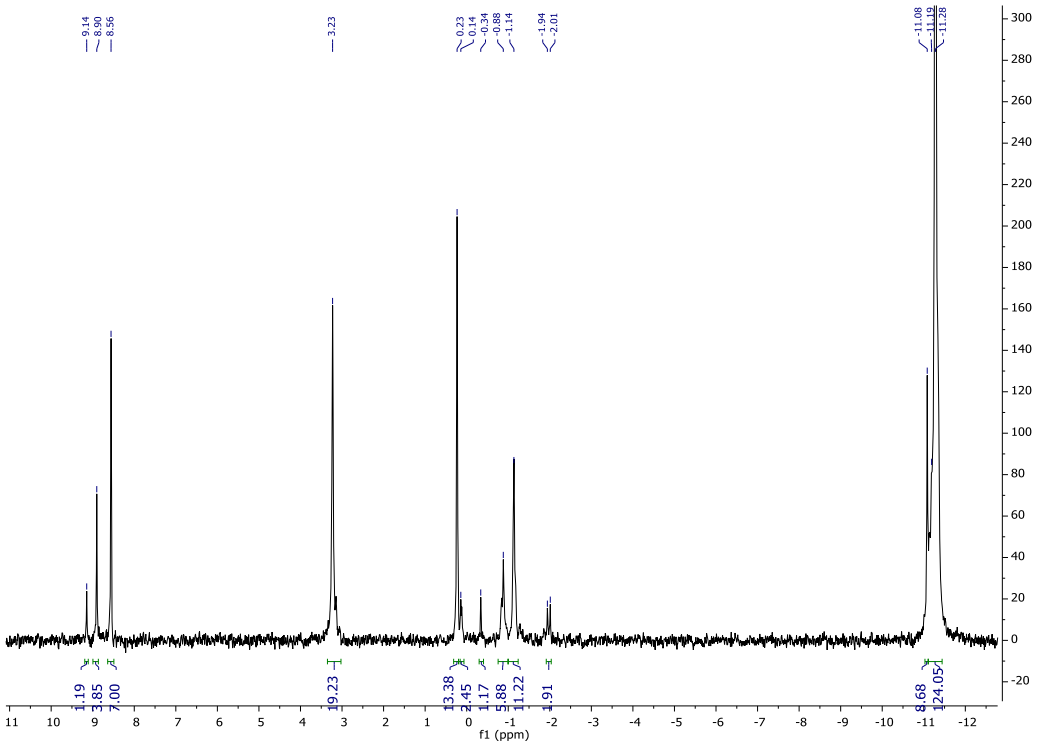


Figure S19 ^{31}P NMR spectrum (H_2O, 160 MHz) for an assay involving 1 M glycerol and 200 mM AMP. The 1-glyceryl phosphodiester is at 0.23 ppm while the 2-glyceryl phosphodiester is at -0.34 as determined by ^{1}H/^{31}P HMBC.

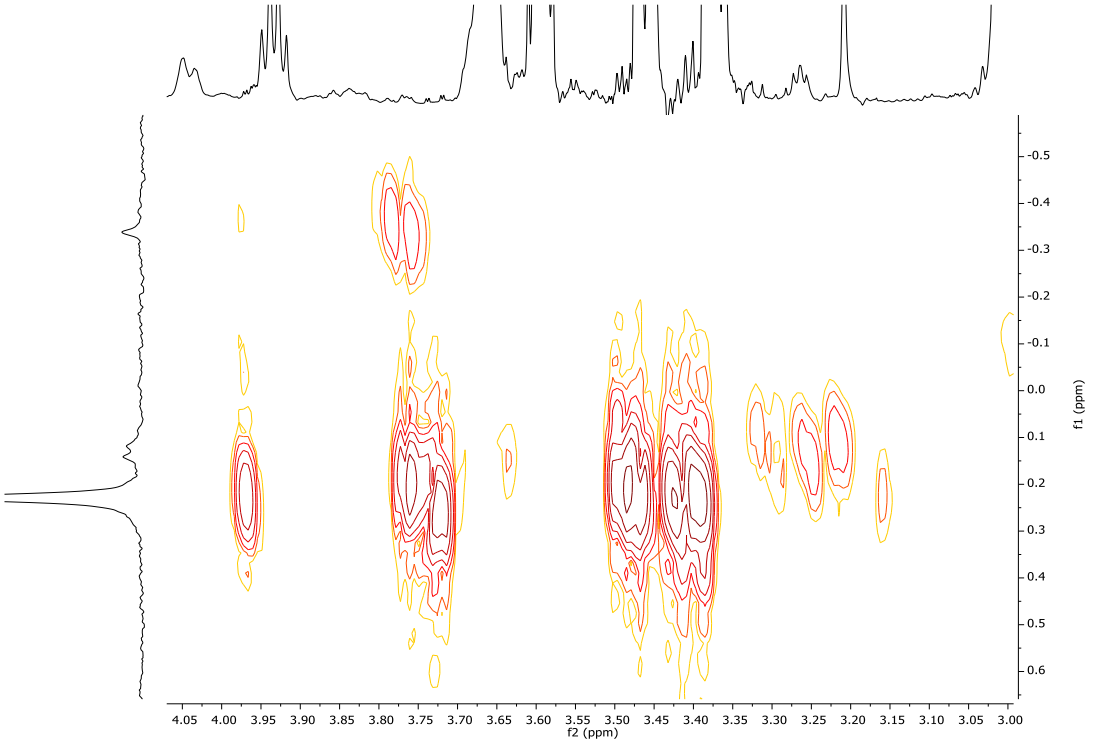


Figure S20 ^{1}H/^{31}P HMBC NMR spectrum (H_2O, 700 MHz) for an assay involving 1 M glycerol and 200 mM AMP. The 1-glyceryl phosphodiester is at 0.23 ppm while the 2-glyceryl phosphodiester is at -0.34. The cross-peaks are small in ^{1}H and largely covered by the excess of glycerol.

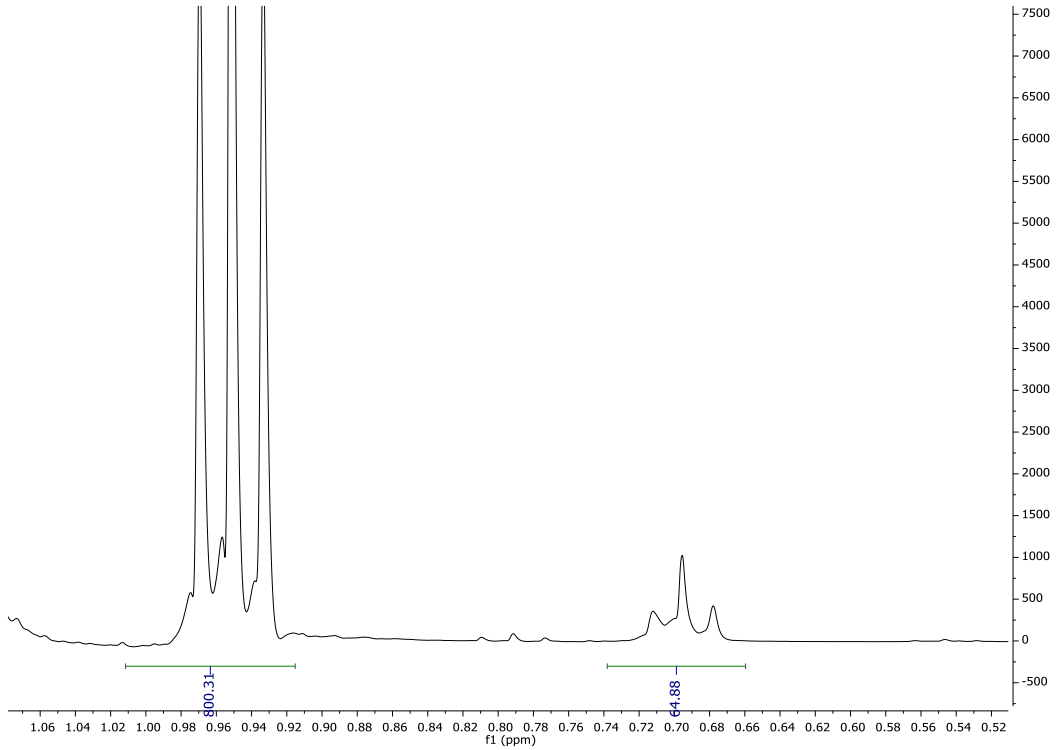


Figure S21 Enlargement of the ^{1}H NMR spectrum (H_2O, 400 MHz) for an assay involving saturated 1-hexanol and 200 mM AMP, as used to determine the actual concentration of the alcohol (0.70 ppm) using 800 mM EDU (0.95 ppm) as an internal standard. The other samples were measured in a similar manner.

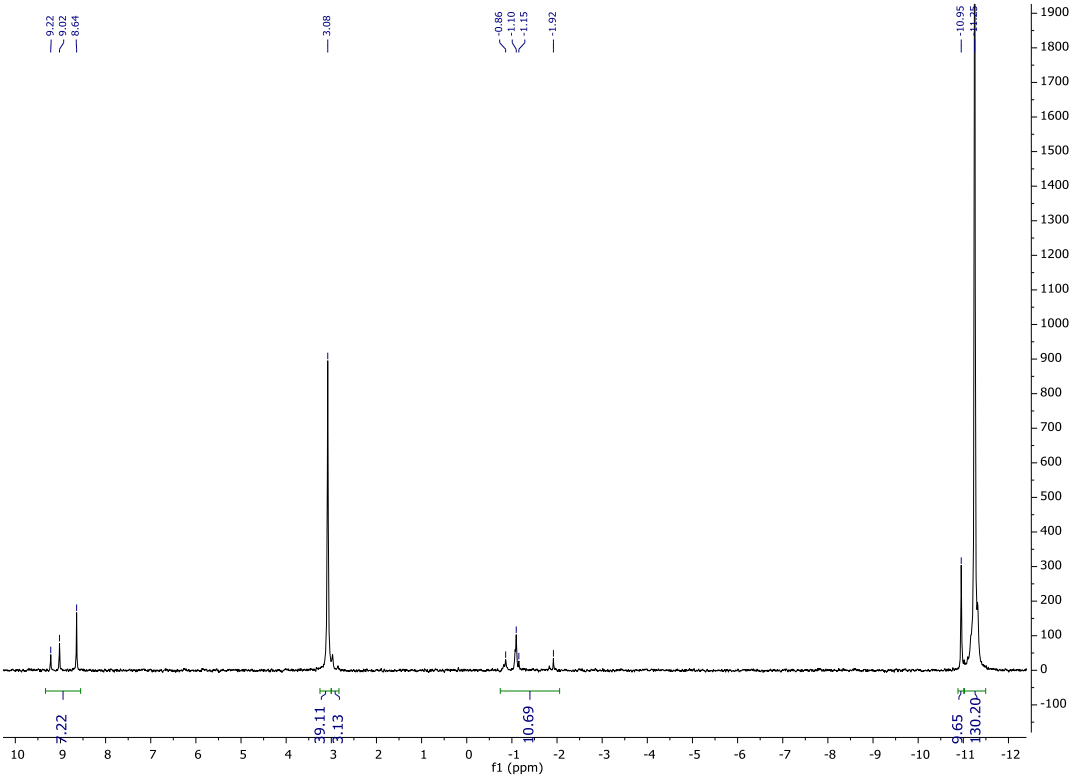


Figure S22 ^{31}P NMR spectrum (H_2O, 160 MHz) for an assay involving 50 mM monolaurin, 100 mM lauric acid and 200 mM AMP.

Complement to section 3.2.7: general condensation reaction of glycine with alkanols

Typical procedure (modifications were applied as described in the main manuscript): 1-^{13}C glycine (7.6 mg, 200 mM) was dissolved in 250 µL of 1 M MOPS buffer pH 7.5 containing 10 % D_2O then magnesium chloride (10 mM), 1-ethylimidazole (7.2 µL, 150 mM) were added and the volume was made up to 500 µL with water containing 10% D_2O. 500 µL of 1-butanol was added and the biphasic system was strongly vortexed to saturate the aqueous phase with the alcohol. EDC (77 mg, 800 mM) was added in one portion and the mixture was shaken for 48 hours. The phases were then separated by centrifugation and the aqueous layer was analyzed by ^{13}C NMR. A control analysis of the alcoholic phases (50 µL DMSO-d_6 added for the lock procedure) revealed negligible concentrations of species other than the alcohol in these. Brotsman et al[465] report the desired ^{13}C ester peak at 167.4 ppm (*p*-toluenesulfonate salt in $CDCl_3$).

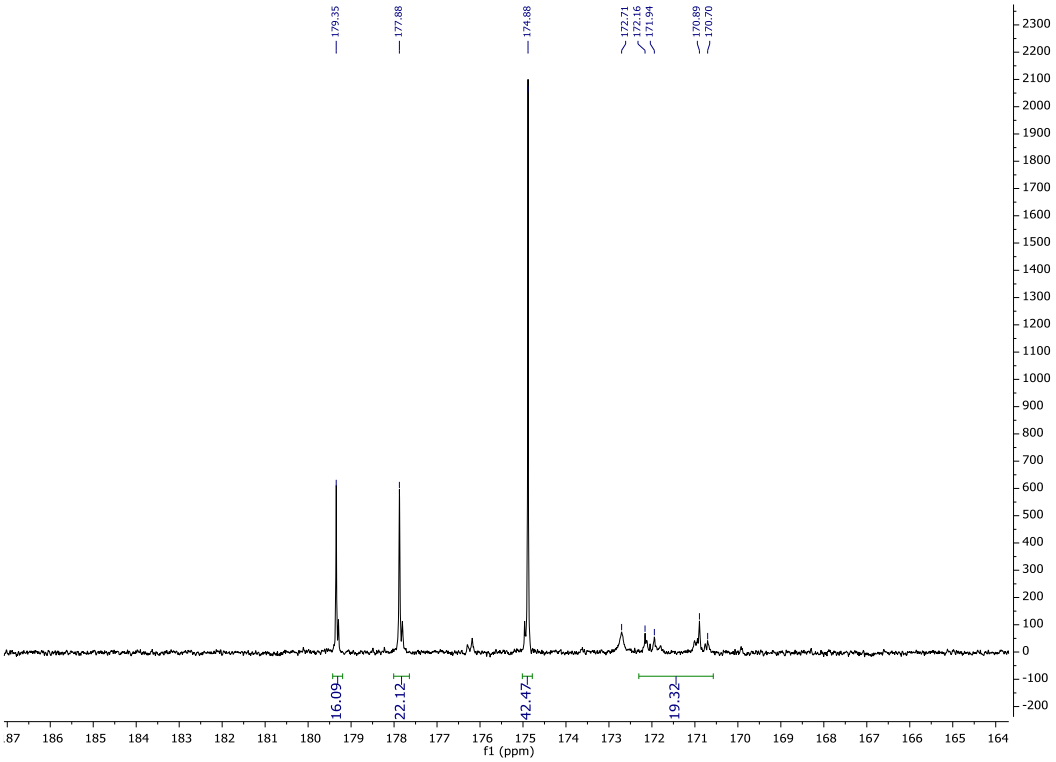


Figure S23 ^{13}C NMR (H_2O, 100 MHz) for an assay involving saturated butanol and 200 mM 1-^{13}C glycine. The butyl ester peak was not detected.

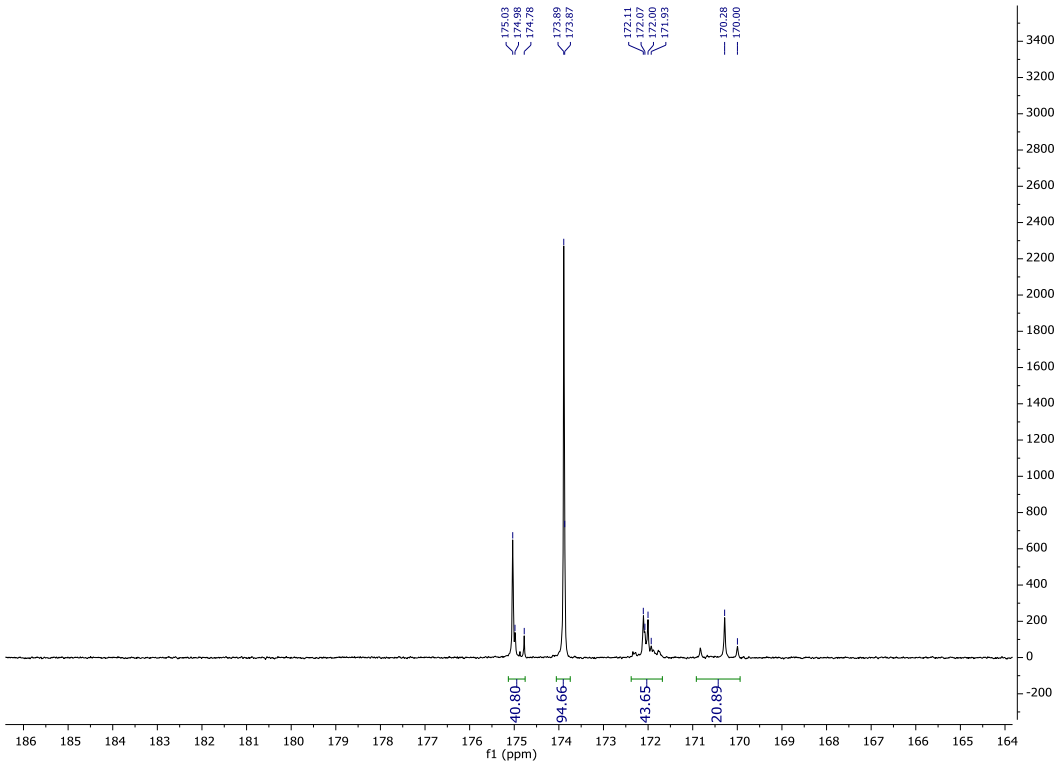


Figure S24 ^{13}C NMR (H_2O, 100 MHz) for an assay involving saturated butanol and 200 mM 1-^{13}C glycine in MES buffer at pH 6. The butyl ester peak was not detected.

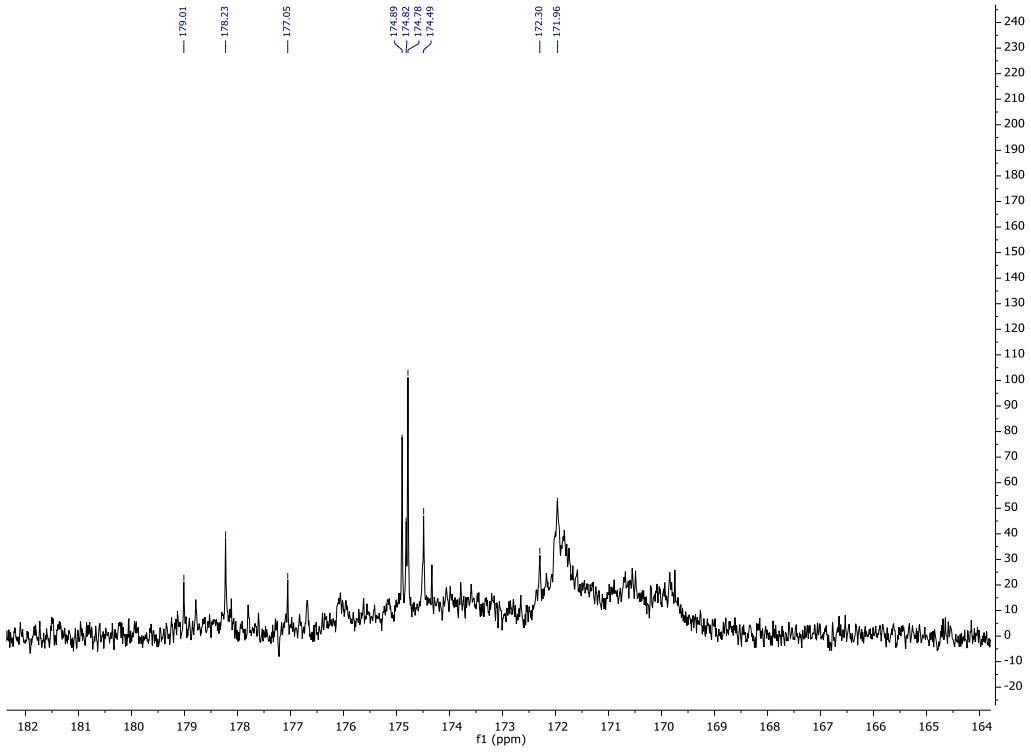


Figure S25 ^{13}C NMR (H_2O, 100 MHz) for an assay involving saturated butanol and 200 mM 1-^{13}C glycine in the presence of 500 mM acetaldehyde. The butyl ester peak was not detected. The broad, unresolved aspect of the spectrum is consistent with the formation of a dark brown tar in the sample.

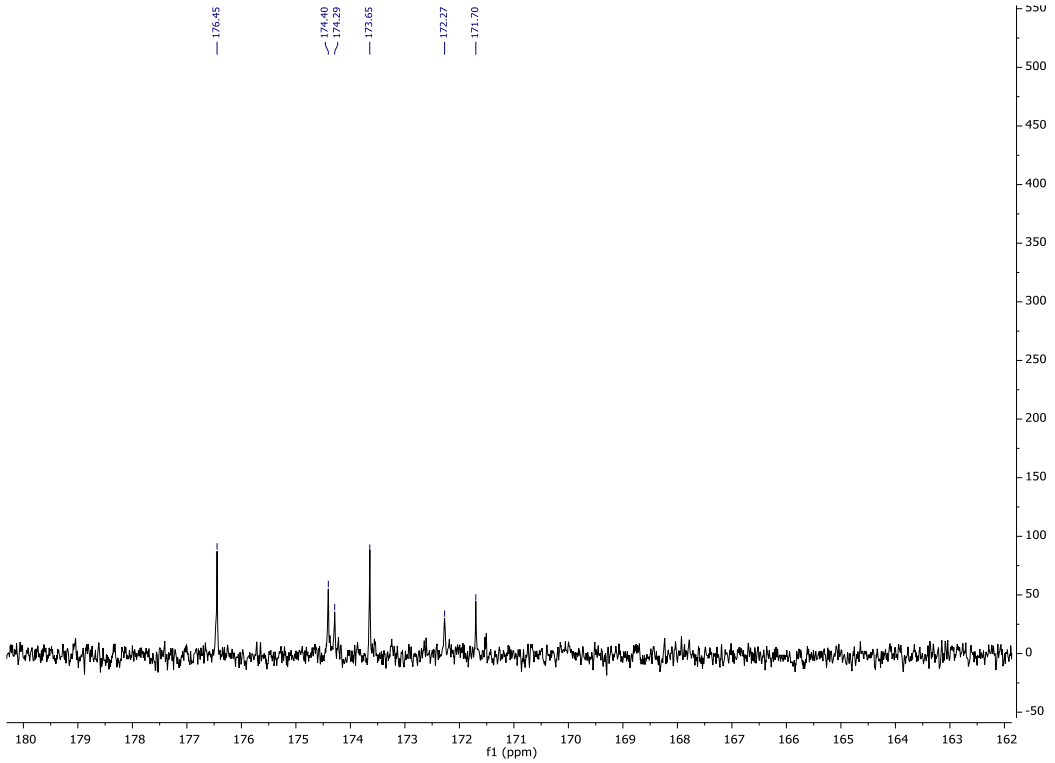


Figure S26 ^{13}C NMR (H_2O, 100 MHz) for an assay involving saturated butanol and 200 mM N-acetylglycine. The butyl ester peak was not detected.

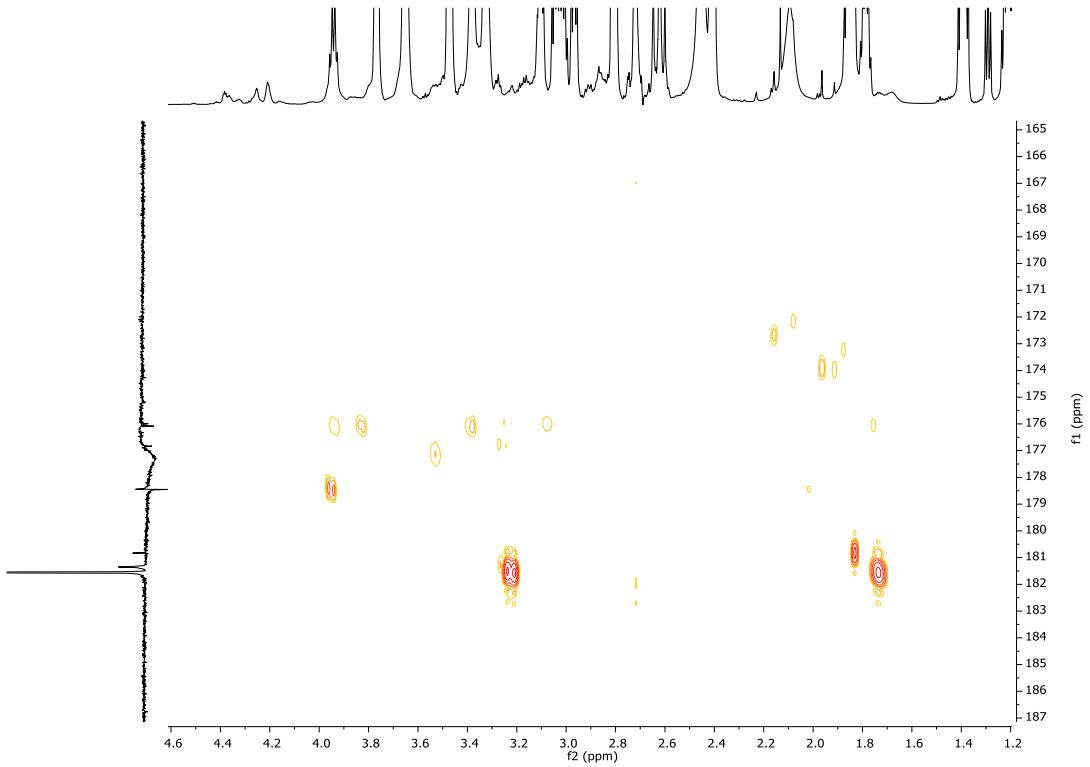


*Figure S27 $^{1}H/^{13}C$ HMBC NMR spectrum (H_2O, 700 MHz) for an assay involving saturated butanol and 200 mM Val*A. The butyl ester was not detected.*

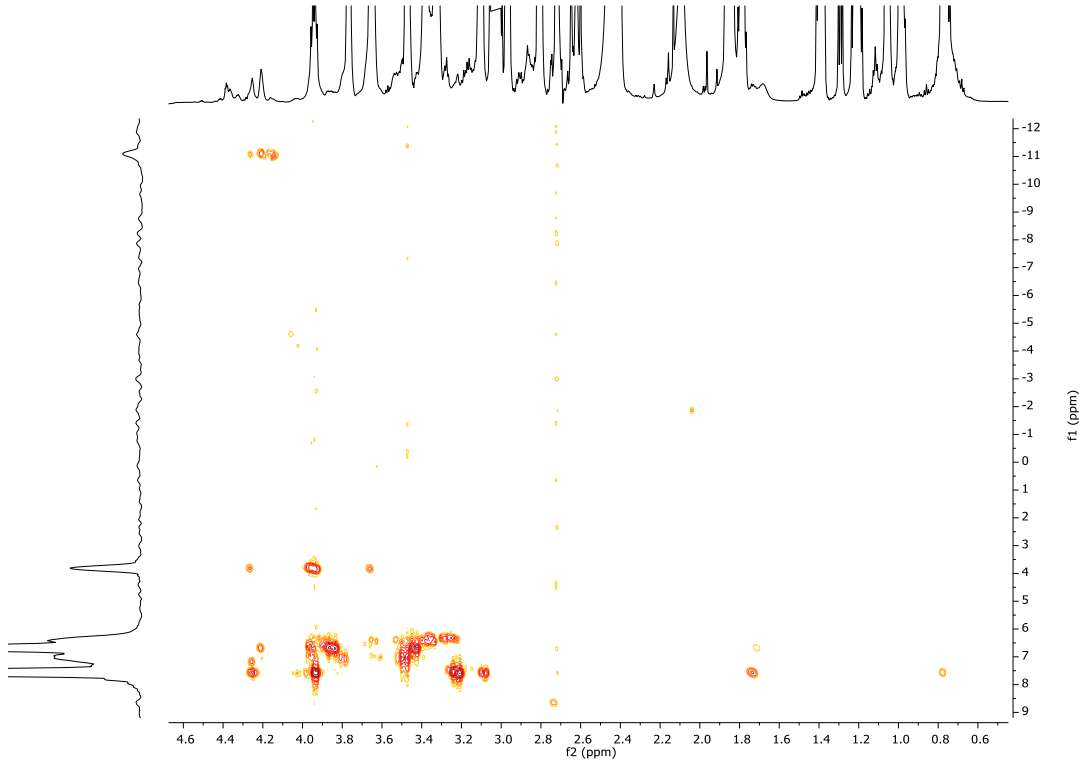


*Figure S28 $^{1}H/^{31}P$ HMBC NMR spectrum (H_2O, 700 MHz) for an assay involving saturated butanol and 200 mM Val*A. Only the formation of EDC-derived products was detected (small amounts of AMP (4 ppm in ^{31}P dimension) and AppA (-11 ppm in ^{31}P dimension) were present in the starting material).*

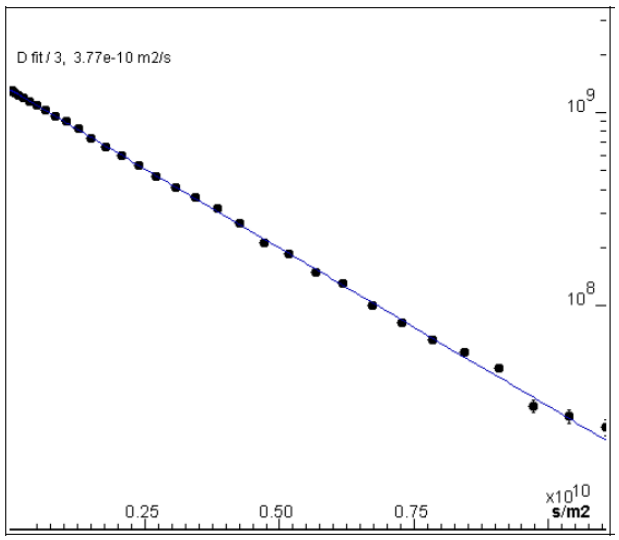
7. Measurements of membrane permeation

Here are reported the measurements obtained by diffusion NMR on 5 mM samples prepared in 10 mM POPC LUVs (or in buffer without lipids for the controls). The value of I_0 coefficients is context-dependent and is only useful to calculate f_{bound} in samples that show membrane adherence.

AMP

Reference peak: H1' (all peaks are consistent)

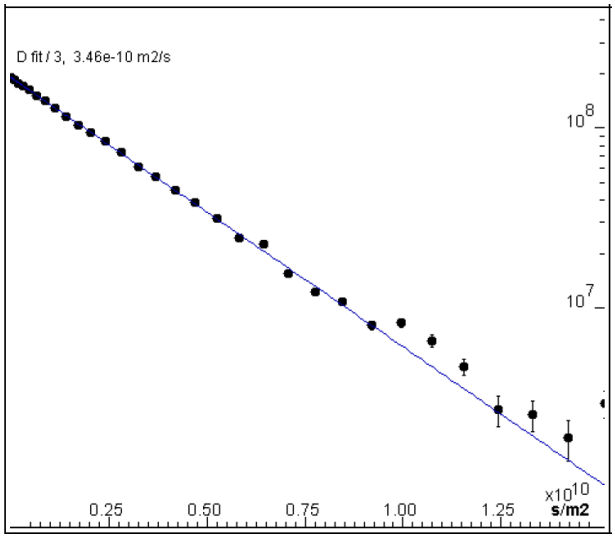
$D^{aq} = 3.77 \pm 0.03 \times 10^{-10}\ m^2 s^{-1}$



Val$_3$

Reference peak: CH at 3.80 ppm (all alpha protons are consistent, isopropyl peaks are overlapped with lipid signals)

$D^{aq} = 3.46 \pm 0.04 \times 10^{-10}\ m^2 s^{-1}$



Leu$_3$

Reference peak: sum of CH_2 and CH

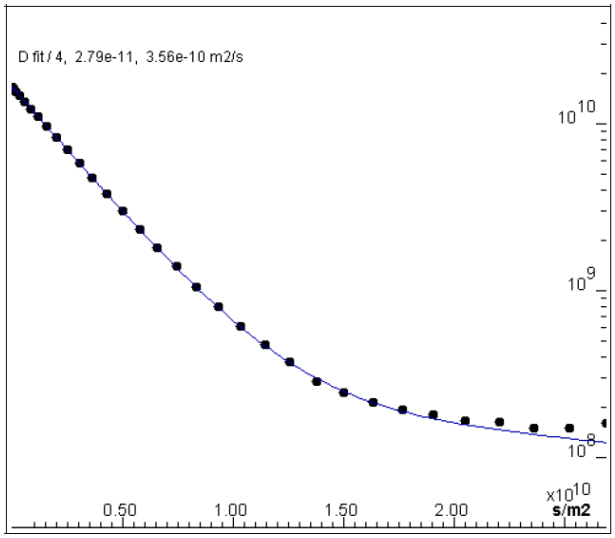
$D^{aq} = 3.56 \pm 0.03 \times 10^{-10}\ m^2 s^{-1}$

$D^{bound} = 3 \pm 2 \times 10^{-11}\ m^2 s^{-1}$

$I_0^{aq} = 1.64 \pm 0.01 \times 10^{10}$

$I_0^{bound} = 2.6 \pm 0.7 \times 10^{8}$

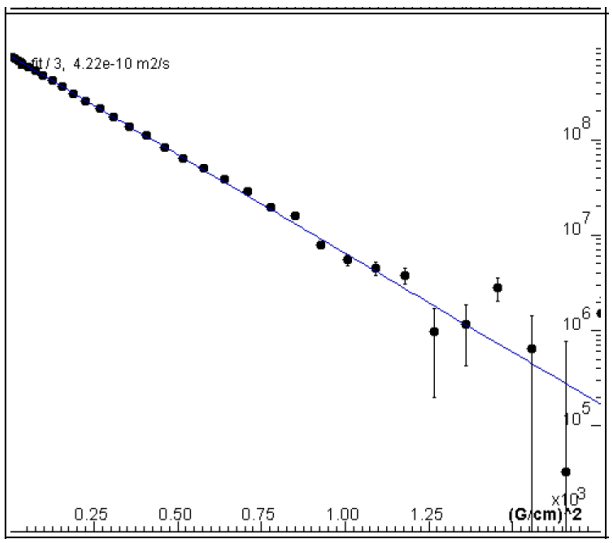
$f_{bound} = 2 \pm 2\ \%$



Val$_2$A

Reference peak: H1' (all nucleoside protons are consistent, isopropyl peaks are overlapped with lipid signals)

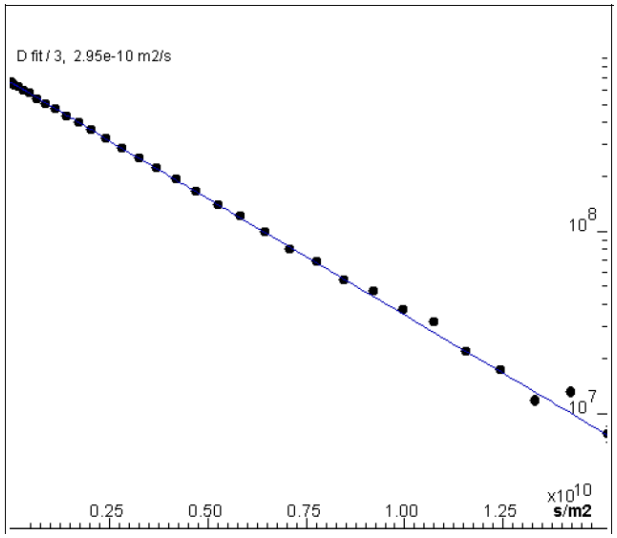
$D^{aq} = 4.22 \pm 0.02 \times 10^{-10}\ m^2 s^{-1}$



Val$_3$A

Reference peak: H1' (all nucleoside protons are consistent, isopropyl peaks are overlapped with lipid signals)

$D^{aq} = 2.95 \pm 0.02 \times 10^{-10}\ m^2 s^{-1}$



Leu$_3$A

Reference peak: sum of CH_2 and CH from the leucine sidechains

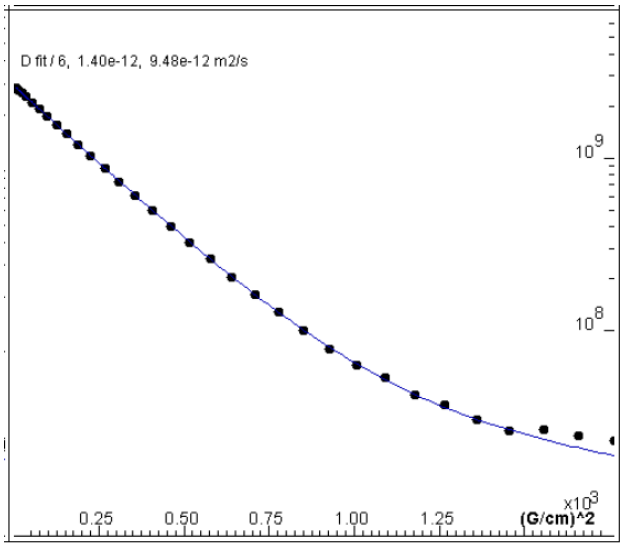
$D^{aq} = 9.5 \pm 0.1 \times 10^{-12}\ m^2 s^{-1}$

$D^{bound} = 1.4 \pm 0.8 \times 10^{-12}\ m^2 s^{-1}$

$I_0^{aq} = 2.59 \pm 0.02 \times 10^9$

$I_0^{bound} = 5 \pm 3 \times 10^7$

$f_{bound} = 2 \pm 6\ \%$



Dansyl PE

Reference peak: sum of H3 and H7

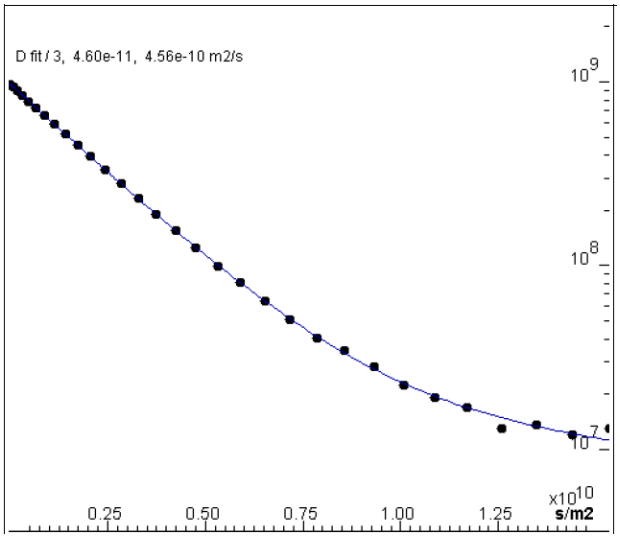
$D^{aq} = 4.56 \pm 0.03 \times 10^{-10}\ m^2 s^{-1}$

$D^{bound} = 4 \pm 2 \times 10^{-11}\ m^2 s^{-1}$

$I_0^{aq} = 9.67 \pm 0.05 \times 10^8$

$I_0^{bound} = 2.1 \pm 0.5 \times 10^7$

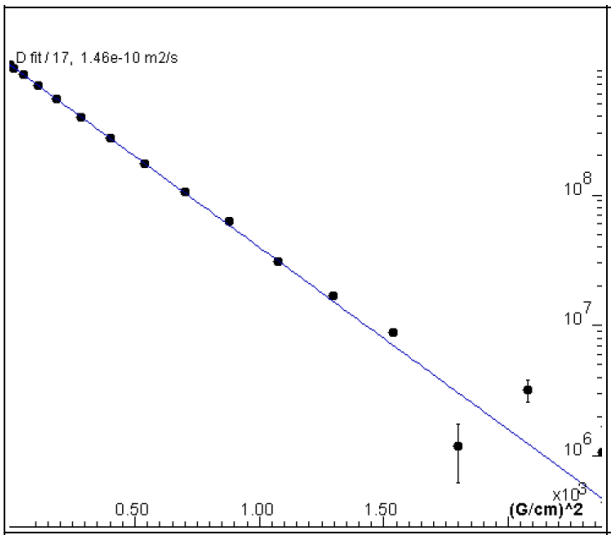
$f_{bound} = 2 \pm 1.5\ \%$



Negative control: Val_3A in the absence of lipids

Reference peak: H1'. The experiment was performed using 16 increments instead of 32.

$D^{aq} = 1.46 \pm 0.01 \times 10^{-10}\ m^2 s^{-1}$

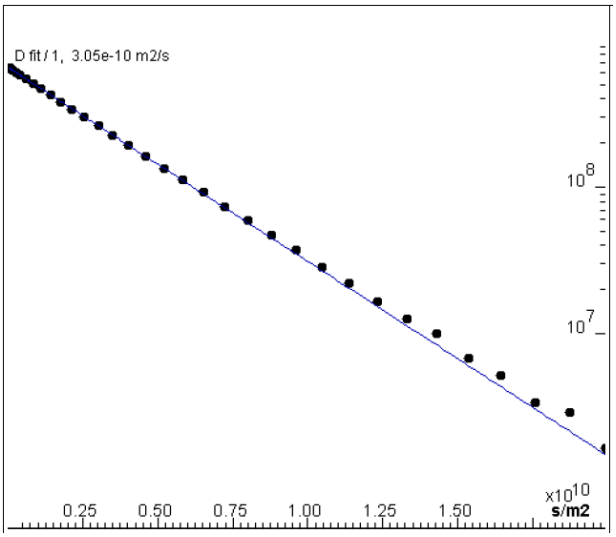


Negative control: Leu_3A in the absence of lipids

Experiment performed using the crude mixture after one month of reaction. As expected, no slow diffusion regime was visible in the absence of lipids.

Reference peak: H1'

$D^{aq} = 3.05 \pm 0.01 \times 10^{-10}\ m^2 s^{-1}$



Limitations to the diffusion measurements induced by insufficient SNR

With the equipment at our disposal, we could not get satisfying SNR for the last increments of diffusion experiments at the concentrations studied. As an illustration, we show below the increments 1, 28 and 32 from the DSTE acquisition of dansyl PE as they were used in this work. In the late increments, the signal can still be distinguished by eyes, but its integration is highly uncertain.

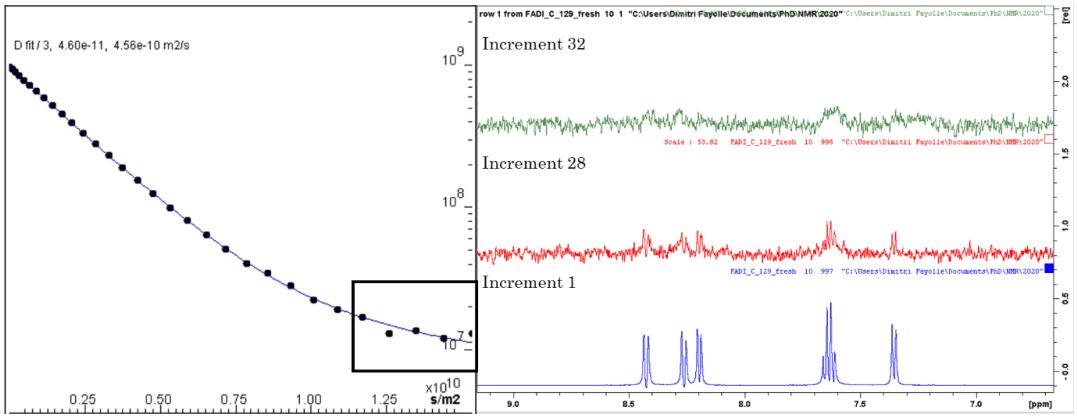


Figure S29 Stacked view of DSTE increments 1, 28, 32 (from bottom to top) of 5 mM dansyl PE in the presence of 10 mM POPC LUVs.

Anchoring of Leu_3A to POPC membranes by gel filtration.

100 mM POPC LUVs (500 µL) were prepared in 0.1 M $NaHCO_3$ in D_2O by hydration, freeze-thawing and extrusion following the general procedures. Leu_3A was added from a concentrated stock solution in the same buffer to reach 10 mM. The vesicles were then gently agitated overnight at 25°C. A gel filtration column was then equilibrated with the same buffer in D_2O, the vesicles were loaded in one portion and 500 µL fractions were eluted. 3 µL samples were drawn from the turbid fractions, diluted to 1000 µL with water and a UV spectrum recorded. No peak was observed at 260 nm. The whole undiluted fraction was also examined by ^1H NMR and no signal other than the broad phospholipid resonances was observed (Figure S 30).

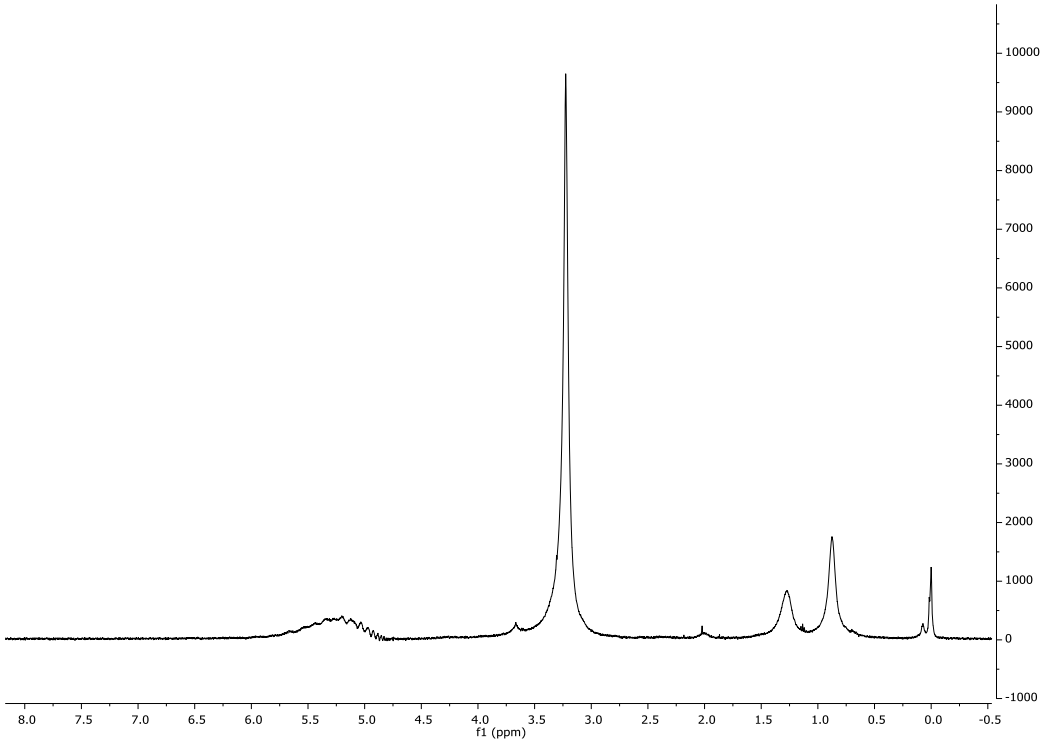


Figure S 30 ^1H NMR spectrum (D_2O, 500 MHz) of concentrated POPC LUVs after gel filtration, that did not retain Leu_3A. The choline peak at 3.2 ppm dominates. The very poor response of the long chain signals may be explained by a very fast translational relaxation (short T_2) in such homogenous sample. Artefacts were generated by the excitation sculpting sequence.

Anchoring of Leu_3A to LA/MLG membranes by gel filtration.

Lauric acid (6.6 mg, 0.033 mmol, 66 mM final concentration) was dispersed in 500 µL of D_2O at 50°C and 5 µL of 30% NaOH was added to reach a pD of 10. *rac*-1-lauroylglycerol (5.5 mg, 0.017 mmol, 33 mM final concentration) was added and the suspension was shaken vigorously at 50°C for 30 minutes. The sample was then cooled to room temperature and titrated back to pD 7.5 with aliquots of 1 M HCl, with vigorous vortexing between each addition to homogenize the vesicles that form. The suspension was then immediately submitted to freeze-thawing and extrusions as described in the general procedures. Leu_3A was added from a

concentrated stock solution to reach a concentration of 10 mM and the sample was gently shaken overnight at 25°C. The vesicles were then separated from the bulk by gel filtration using 5 mM lauric acid in D_2O, adjusted to pD 7.5, as the eluent. 500 µL fractions were collected. A 3 µL aliquot was drawn from the main turbid fraction, diluted to 1000 µL with water and a UV spectrum was recorded, with no detectable perturbation arising from turbidity. The absorbance at 260 nm indicated a concentration of 0.95 mM in Leu_3A. The rest of the sample was analyzed directly by ^1H NMR and the integrity of Leu_3A was confirmed (Figure S 31, Figure S 32). 100 µL of the solution was then recovered un a tube for lipid extraction. Chloroform/methanol (2:1, 500 µL) was added, and after vigorous mixing and centrifugation the organic layer was recovered. The aqueous layer was diluted to 500 µL with water and an identical extraction was repeated. The total extract was evaporated to dryness. $CDCl_3$ containing 2.93 mg/g CH_3CN as an internal standard (115 mg, corresponding to 0.34 mg or 8.28 µmol of CH_3CN) was then added, the sample was transferred to an NMR tube and two small volumes of $CDCl_3$ were used to quantitatively transfer the sample to the NMR tube. A ^1H spectrum (64 scans, `d1` = 60s, 90° pulse) was then acquired (Figure S 33) and the acetonitrile peak (2.01 ppm) was used as a reference. The CH_2 peaks of both LA and MLG (closely overlapped) at 2.35 ppm was chosen to quantify both lipids. The rest of the spectrum confirmed the integrity of the lipid extract.

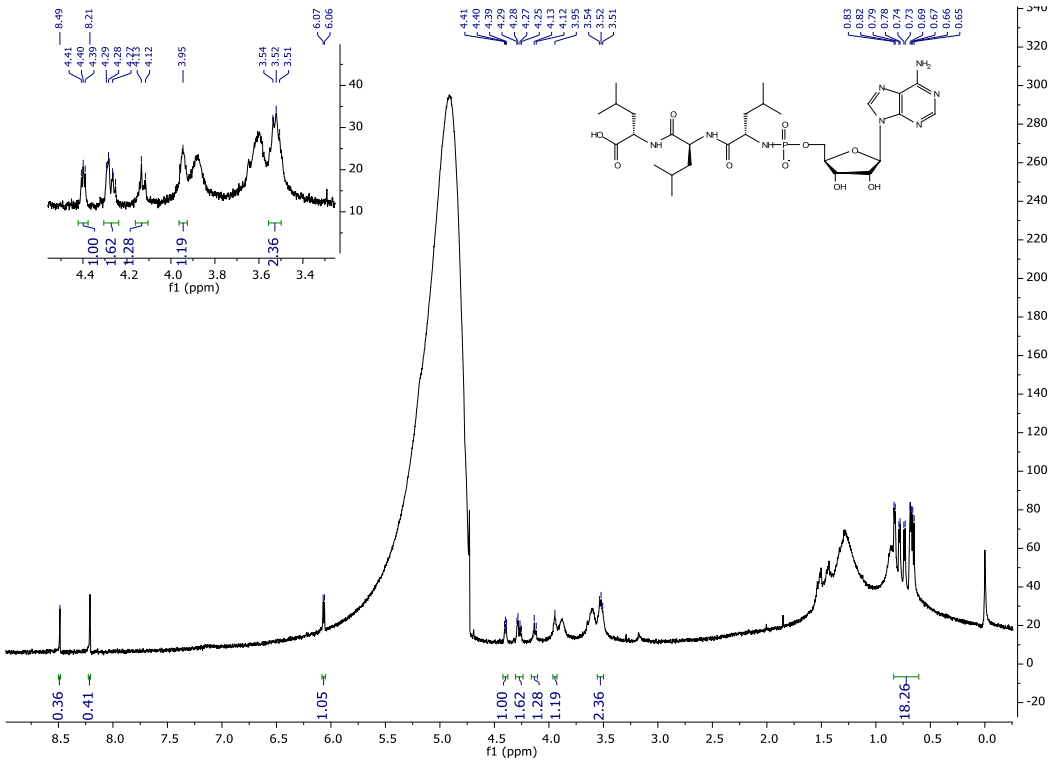


Figure S 31 ^1H NMR (D_2O, 500 MHz) of Leu_3A embedded in LA/MLG membranes after gel filtration separation. Only the peaks belonging to Leu_3A are picked and integrated. A large artefact was generated despite the application of presaturation. The lipid chains are very weak in integral and may be disfavored by very quick translational relaxations (short T_2)

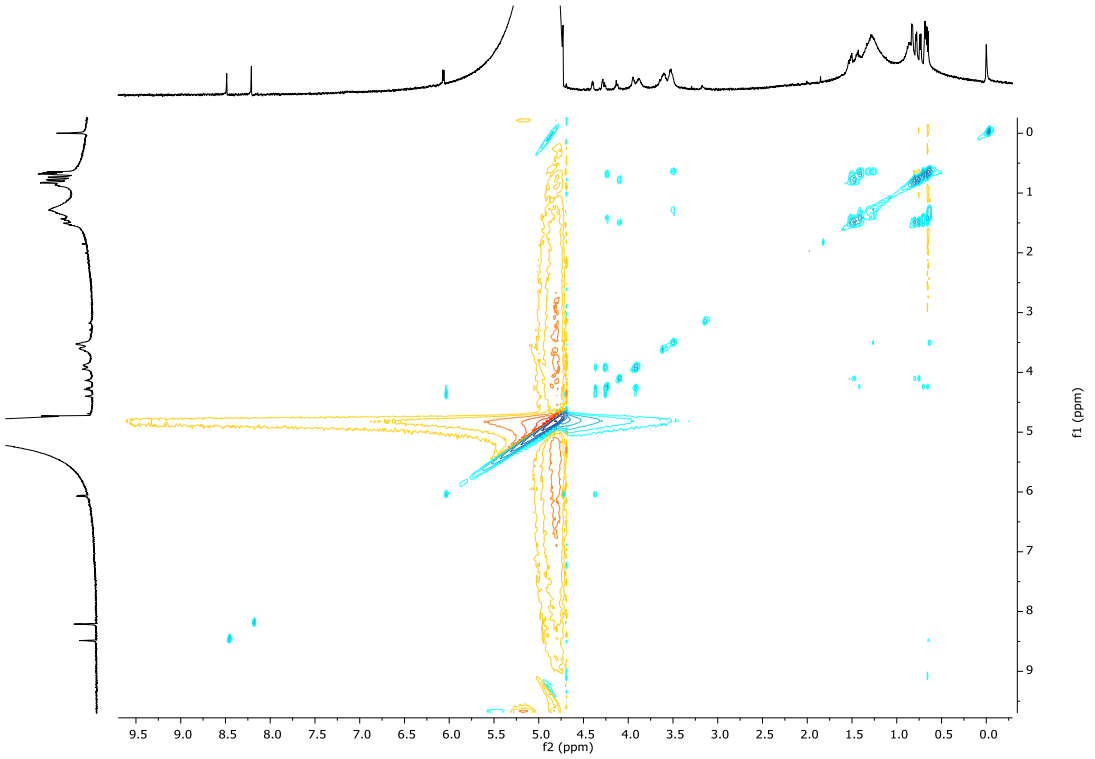


Figure S 32 ^{1}H-^{1}H DIPSI NMR (D_2O, 500 MHz) of Leu_3A embedded in LA/MLG membranes after gel filtration separation. Lipids did not give cross-signals likely because of too fast translational relaxation (short T_2).

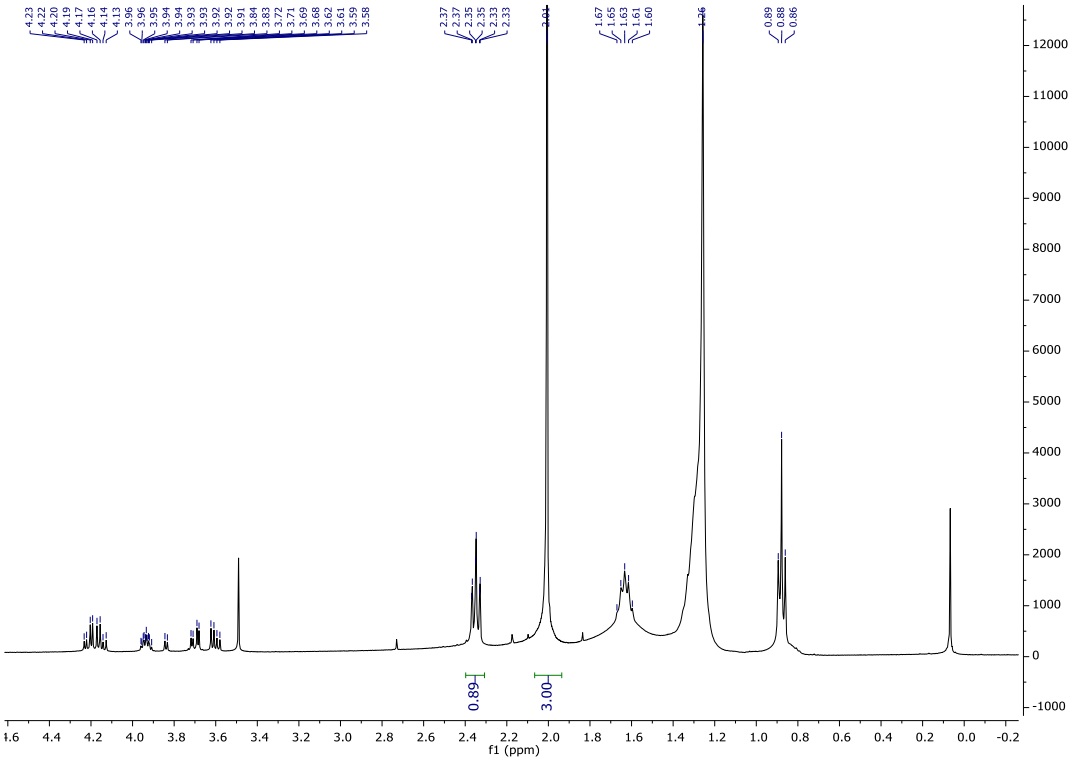
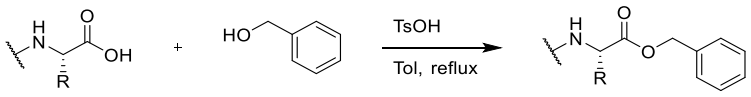


Figure S 33 Quantitative ^{1}H NMR ($CDCl_3$, 400 MHz) of the lipids extracted from LA/MLG LUVs embedding Leu_3A using acetonitrile (0.34 mg, 8.28 µmol) as an internal standard. The CH_2 signal at 2.35 ppm (both LA and MLG, overlapped) was used as a lipid reference.

8. Synthesis of chemical compounds

Unless otherwise specified, all amino acid residues in this work are of the L configuration. Commercially enantiopure amino acids were used and we have no analytical reason to suspect significant racemization reactions. For clarity reasons, the stereochemistry of amino acids is therefore not explicitly mentioned in the following schemes.

Benzylation of amino acids and dipeptides by the Dean-Stark method



The desired amino acid or dipeptide (1 eq), benzyl alcohol (5 eq) and *p*-toluenesulfonic acid (1.2 eq) were refluxed for 24 in anhydrous toluene (30 mL/g of substrate) in a Dean-Stark apparatus. The mixture was then cooled down to room temperature, concentrated to 2-3 mL and added dropwise to cold diethyl ether to induce the precipitation of the TsOH salt of the desired ester. The precipitate was washed with diethyl ether and dried under vacuum.

L-valine benzyl ester was obtained from L-valine (1.00 g, 8.5 mmol) as a white powder (2.2 g, 66%) and was identical to a commercial sample.

^{1}H NMR (DMSO-d_6, 300 MHz): δ = 8.31 (br, 3H, NH_3^+), 7.51-7.44 (m, 2H, tosyl), 7.44-7.28 (m, 5H, benzyl), 7.14-7.07 (m, 2H, tosyl), 5.26 (dd, J = 15.8 and 12.2 Hz, 2H, benzyl CH_2), 4.00 (br, 1H, α), 2.29 (s, 3H, Me from tosyl), 2.20-2.05 (m, 1H, β), 0.93 (app. t, J = 7.5 Hz, 6H, γ) ppm.

ESI-MS (M+H): m/z = 208.1

L-glycylglycine benzyl ester was obtained from L-glycylglycine (61 mg, 0.46 mmol) as a white powder (162 mg, 65%) and was identical to a commercial sample.

^{1}H NMR (MeOD, 300 MHz): δ = 7.75-7.70 (m, 2H, tosyl), 7.44-7.32 (m, 5H, benzyl), 7.28-7.22 (m, 2H, tosyl), 5.21 (s, 2H, benzyl CH_2), 4.10 (s, 2H, CH_2 from Gly2), 3.76 (s, 2H, CH_2 from Gly1), 2.39 (s, 2H, Me from tosyl) ppm.

ESI-MS (M+H): m/z = 223.1

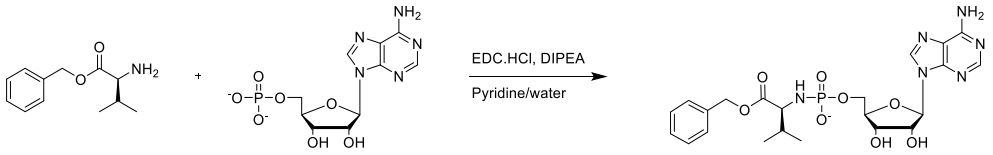
L-valylvaline benzyl ester was formed from L-valylvaline (101 mg, 0.46 mmol). A sticky residue was recovered by centrifugation after the precipitation, and ^{1}H NMR evidenced the presence of the desired ester (15% yield) alongside with L-

valylvaline tosylate and an excess of reagents. The mixture was not further separated.

^{1}H NMR (DMSO-d_6, 300 MHz, pertinent signals): δ = 8.58 (d, J = 7.7 Hz, 1H, NH amide), 8.05 (br, 3H, NH_3^+), 7.50-7.43 (m, 2H, tosyl), 7.40-7.31 (m, 5H, benzyl), 7.13-7.08 (m, 2H, tosyl), 5.14 (s, 2H, benzyl CH_2), 4.27 (dd, J = 7.7 and 6.0 Hz, 1H, α of Val2), 3.76-3.64 (m, 1H, α of Val1), 2.29 (s, 3H, Me from tosyl), 2.20-1.933 (m, 2H, β), 1.00-0.81 (m, 12H, γ) ppm.

ESI-MS (M+H): m/z = 307.2

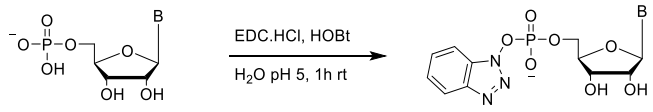
Synthesis of BnO-Val-p-A by the direct coupling method



L-valine benzyl ester hydrochloride (131 mg, 0.54 mmol), adenosine monophosphate (187 mg, 0.54 mmol), EDC hydrochloride (103 mg, 0.54mmol) and DIPEA (370 µL, 2.16 mmol) were dissolved in 4.5 mL of pyridine and 1.5 mL of water. The reaction was followed by ^{31}P NMR of aliquots and EDC hydrochloride was added periodically as required (2 additions of 1 eq. each). When the conversion reached a stable value (83% by integration of ^{31}P NMR signals), the crude mixture was poured slowly into 60 mL of cold diethyl ether containing sodium perchlorate (500 mg). The precipitate was collected, washed with ether and dried to afford a pale-yellow powder (180 mg) that contained a mixture of the desired phosphoramidate (60% yield) and unreacted AMP. No further purification was performed.

^{1}H NMR (MeOD, 300 MHz, pertinent signals): δ = 8.52 (s, 1H, H2), 8.18 (s, 1H, H8), 7.33-7.26 (m, 5H, benzyl), 6.06 (d, J = 6.3 Hz, 1H, H1'), 4.69-4.64 (overlapped, 1H, H3'), 4.32 (dd, J = 5.1 and 2.8 Hz, 1H, H4'), 3.98-3.93 (m, 2H, H5' and H5"), 3.67 (dd, J = 9.4 and 5.3 Hz, 1H, α), 1.96-1.87 (m, 1H, β), 0.85 (dd, J = 13.0 and 6.8 Hz, 6H, γ) ppm.

Activation of nucleotides as benzotriazolides



AMP (free acid), UMP or GMP (disodium salts) (1 mmol), HOBt hydrate (3 mmol) and EDC hydrochloride (5 mmol) were dissolved in water (2.5 mL) and the pH was adjusted to 5 using dilute hydrochloric acid. After the reaction was complete as determined by ^{31}P NMR, the crude mixture was poured into an ice-cold solution

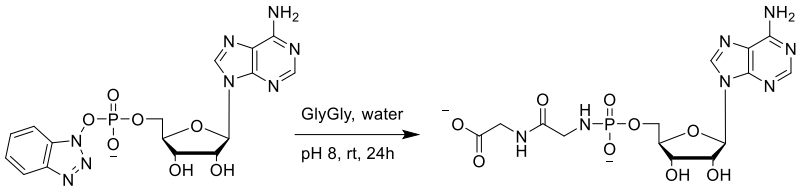
of sodium perchlorate (1g) in acetone (20 mL). A sticky residue was recovered, washed with acetone and dessicated overnight under vacuum over phosphorus pentoxide. A white solid was obtained, and an accurately weighted aliquot (ca. 50 mg) was titrated by ^{31}P NMR against triethyl phosphate (10.0 µL). The crude mixture was used without further purification on the basis of this titration. The conversion was always complete, and the recovery usually reached 60-80%. The ^{1}H spectrum was excessively polluted with HOBt and EDU signals to be properly interpreted.

AMP-OBt ^{31}P NMR (D_2O, 120 MHz) δ = -1.1 ppm.

UMP-OBt ^{31}P NMR (D_2O, 120 MHz) δ = -0.6 ppm.

GMP-OBt ^{31}P NMR (D_2O, 120 MHz) δ = -0.6 ppm.

HO-Gly-Gly-pA



Glycylglycine (132 mg, 1 mmol) was dissolved in water (1 mL) and adjusted to pH 8 by addition of a NaOH solution. Crude AMP benzotriazolide (100 mg, 0.2 mmol) was added and the evolution of the reaction was followed by ^{31}P NMR, showing disappearance of the starting material (-1.06 ppm), formation of the desired product (7.55 ppm) and after some time, hydrolysis to AMP (1.26 ppm). After 24h of reaction, the starting material was consumed, and the title compound was isolated by preparative strong anion-exchange HPLC (system A, column A3) using a growing concentration of ammonium formate pH 6.0. The compound could not be entirely separated from AMP.

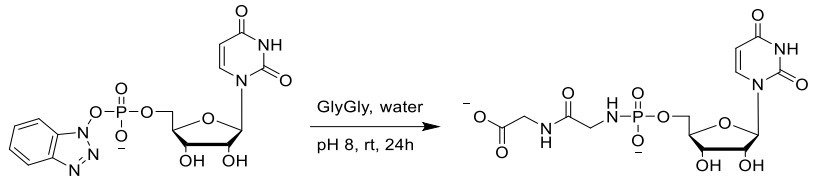
^{1}H NMR (H_2O, 500 MHz): δ = 8.40 (s, 1H, H8), 8.18 (s, 1H, H2), 6.04 (d, *J* = 5.6 Hz, 1H, H1'), 4.44-4.38 (m, 1H, H3'), 4.32-4.23 (m, 1H, H4'), 4.02-3.91 (m, 2H, H5' and H5"), 3.62 (s, 2H, Gly1), 3.39 (d, *J* = 10.6 Hz, 2H, Gly2) ppm.

^{13}C NMR (H_2O, 125 MHz): δ = 176.1 (COO^-), 174.2 (d, *J* = 6.9 Hz, CONH), 154.6 (C6), 151.5 (C2), 148.8 (C4), 140.1 (C8), 118.6 (C5), 87.2 (C1'), 83.9 (d, *J* = 8.8 Hz, C4'), 74.1 (C2'), 70.3 (C3'), 63.8 (d, *J* = 4.8 Hz, C5'), 44.4 (CH_2 from Gly1), 42.7 (d, *J* = 2.7 Hz, CH_2 from Gly2) ppm.

^{31}P NMR (H_2O, 200 MHz): δ = 7.69 ppm.

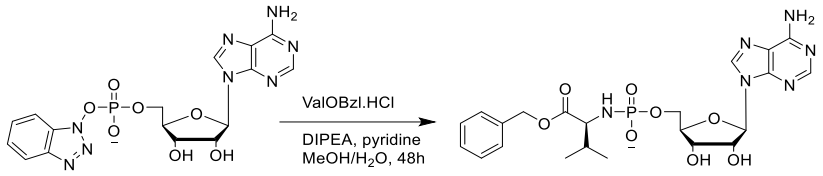
Exact mass ($[M+H]^+$) calcd. for $C_{14}H_{21}N_7O_9P$ 462.1133, found 462.1132 (0.2 ppm)

HO-Gly-Gly-pU



The procedure described for Gly_2A was followed using UMP-OBt (300 mg, 0.2 mmol). A purification was attempted by reverse-phase HPLC (system A, column A2) using 0.1 M ammonium acetate and acetonitrile but only impure fractions were recovered.

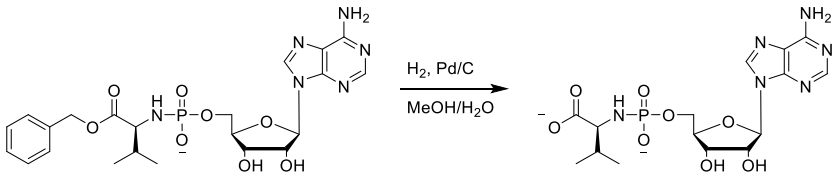
Synthesis of BnO-Val-pA by the displacement method.



L-valine benzyl ester hydrochloride (267 mg, 1.1 mmol) and AMP-OBt (1.2g, 1.0 mmol) were dissolved in methanol (10 mL) and water (1 mL). Pyridine (80 µL, 1 mmol) and DIPEA (680 µL, 4 mmol) were added and the solution was stirred at room temperature for 2 days. The desired compound was purified by preparative flash chromatography on a C18 column using a gradient of water and acetonitrile containing 0.1% of formic acid. The desired fractions were pooled, immediately neutralized with NaOH, and lyophilized. Despite the quenching, degradation occurred.

^{1}H NMR (D_2O, 300 MHz, pertinent signals): δ = 8.49 (s, 1H, H2), 8.29 (s, 1H, H8), 7.39-7.27 (m, 5H, benzyl), 6.14 (d, J = 5.2 Hz, 1H, H1'), 4.76 (t, J = 5.3 Hz, 1H, H3'), 4.49 (app. t, J = 4.8 Hz, 1H, H4'), 4.41-4.35 (m, 2H, H5' and H5"), 3.57 (dd, J = 8.9 and 6.4 Hz, 1H, α), 2.03-1.92 (m, 1H, β), 0.92 (dd, J = 6.8 and 1.3 Hz, 6H, γ) ppm.

Deprotection of BnO-Val-pA

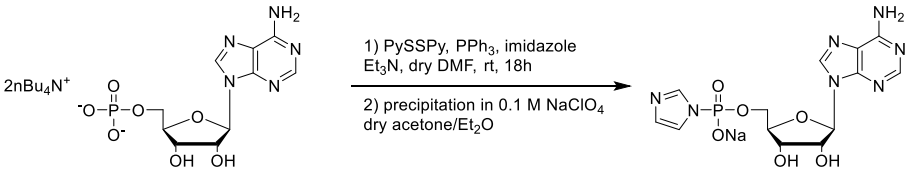


BnOValpA (250 mg, 0.38 mmol) was dissolved in methanol (10 mL) and water (1 mL). 10% palladium on activated charcoal (100 mg) was added and the solution was strongly stirred overnight under 2 bars of hydrogen. The catalyst was

removed by filtration on celite, and the crude mixture was concentrated then lyophilized. A purification by preparative flash chromatography on a C18 column using a gradient of 0.1 M ammonium acetate and acetonitrile provided insufficient resolution. The purification was repeated using a gradient of water and acetonitrile containing 0.1% of formic acid. The desired fractions were pooled, immediately neutralized with NaOH and lyophilized. Despite this precaution, a partial hydrolysis of the compound occurred during the purification.

^{1}H NMR (D_2O, 300 MHz): δ = 8.44 (s, 1H, H2), 8.17 (s, 1H, H8), 6.01 (d, *J* = 6.1 Hz, H1'), 4.37-4.32 (m, 1H, H3'), 4.30-4.32 (m, 1H, H4'), 3.91 (br dd, *J* = 4.0 and 3.0 Hz, 2H, H5' and H5"), 3.19 (dd, *J* = 9.5 and 5.6 Hz, 1H, α), 1.76-1.68 (m, 1H, β), 0.77 (dd, *J* = 9.2 and 6.9 Hz, γ) ppm.

Activation of AMP as imidazolide

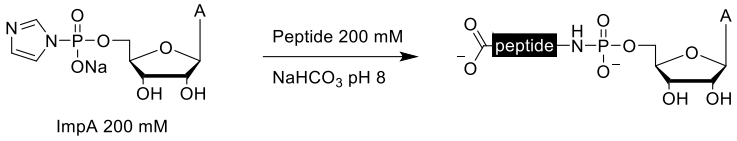


AMP (as a bis(tetrabutylammonium) salt, 1.66 g, 2 mmol), imidazole (1.36 g, 20 mmol) and triphenylphosphine (1.05 g, 4 mmol) were dissolved in anhydrous DMF (12 mL) under argon. Triethylamine (1.1 mL, 8 mmol) was added, followed in one portion by pyridyl disulfide (1.76 g, 8 mmol). The solution immediately turned bright yellow and was stirred overnight under argon. Sodium perchlorate (1.22 g, 10 mmol) was dissolved in anhydrous acetone (100 mL) and anhydrous diethyl ether (50 mL) and the solution was cooled in ice. The crude reaction mixture was cannulated to it, strongly triturated for 30 minutes then quickly filtered. The residue was washed thoroughly with anhydrous acetone until the yellow color completely disappeared, then with anhydrous diethyl ether. Air contact was avoided as much as possible during this step. The resulting white solid (937 mg, quantitative) was dried under high vacuum overnight. The analysis was consistent with literature reports.[466]

^{1}H NMR (D_2O, 300 MHz): δ = 8.15 (s, 1H, H2), 8.13 (s, 1H, H8), 7.71 (br t, *J* = 1.1 Hz, 1H, H2(imidazole)), 7.01 (app. q, *J* = 1.5 Hz, 1H, H4(imidazole)), 6.81 (br quint, *J* = 1.1 Hz, 1H, H5(imidazole)), 5.95 (d, *J* = 5.1 Hz, 1H, H1'), 4.32 (app t, *J* = 4.6 Hz, 1H, H3'), 4.22-4.16 (m, 1H, H4'), 4.05-3.93 (m, 2H, H5' and H5") ppm. Traces of solvents that could not be removed under vacuum are visible.

^{31}P NMR (D_2O, 120 MHz): δ = -8.08 ppm.

Synthesis of peptido-nucleotides by the imidazolide method



The desired peptide (0.1 mmol) was dissolved in sodium bicarbonate buffer (500 µL, 0.5 M) and the pH was adjusted to 8.5 using concentrated NaOH. The desired nucleotide imidazolide (0.1 mmol) was added and the pH was again adjusted to 8.5, then the reaction was allowed to proceed at room temperature and was followed by analytical HPLC (system C, column B1). A sufficient conversion was usually reached after 5 to 7 days. The sample was then diluted to 2 mL in water, micro-filtered and purified by preparative HPLC (system A, column A1) with a gradient of acetonitrile in 10 mM ammonium acetate (typically 0-30%). The fractions containing the product were identified by mass spectrometry, concentrated and lyophilized. The resulting products were then stored as stock solutions in sodium bicarbonate buffers (0.1 M, pH 8) and their concentration was determined by UV absorbance at 260 nm, assuming an extinction coefficient of 15200 $M^{-1}.cm^{-1}$.

Val*A (1-^{13}C-ValpA) was prepared on a 0.5 mmol scale from 1-^{13}C L-valine and gave a 44% conversion in 7 days. It was purified manually by ion-pairing chromatography on hand-packed C18 silica eluting with 0-10% acetonitrile in 0.1 M triethylammonium acetate. The fractions were pooled by UV, identified by MALDI-MS and lyophilized. The residue was taken in a minimum of water and precipitated with 10 volumes of 0.1 M sodium perchlorate in acetone, collected by centrifugation, washed with acetone twice and dried. Isolated yield 17%.

^{1}H NMR (H_2O, 400 MHz): δ = 8.39 (s, 1H, H2), 8.03 (s, 1H, H8), 5.99 (d, *J* = 5.9 Hz, 1H, H1'), 4.38 (dd, *J* = 4.9 and 3.5 Hz, 1H, H3'), 4.30-4.25 (m, 1H, H4'), 3.98-3.91 (m, 2H, H5' and H5"), 3.21 (br, 1H, α), 1.78-1.67 (m, 1H, β), 0.77 (dd, *J* = 11.3 and 6.8 Hz, 6H, γ) ppm. Traces of acetone (2.12 ppm) were not removed by successive lyophilizations, a small amount of sodium acetate (1.84 ppm) formed during the precipitation is visible

^{13}C NMR (H_2O, 100 MHz): δ = 181.8 (d, *J* = 3.0 Hz, $^{13}COO^-$), 155.4 (C6), 152.7 (C2), 148.8 (C4), 139.8 (C8), 118.4 (C5), 86.8 (C1'), 84.4 (d, *J* = 9.8 Hz, C4'), 74.5 (C2'), 70.6 (C3'), 63.7 (d, *J* = 4.9 Hz, C5'), 62.6 (α), 32.2 (d, *J* = 6.2 Hz, β), 18.8 (d, *J* = 2.0 Hz, γ), 17.6 (d, *J* = 1.7 Hz, γ) ppm. Traces of acetone (215.4, 30.3 ppm) were not removed by repeated lyophilizations, a small amount of sodium acetate (174.5, 23.4 ppm) formed during the precipitation is visible

^{31}P NMR (160 MHz, H_2O): δ = 7.66 (d, *J* = 3.0 Hz) ppm.

MALDI-MS ($[M-H]^-$) calculated for $C_{14}{}^{13}CH_{22}N_6O_8P$ 446.1 found 445.9

Val$_2$A. 71% conversion, 52% isolated yield.

^{1}H NMR (H_2O, 500 MHz): δ = 8.34 (s, 1H, H8), 8.06 (s, 1H, H2), 7.69 (d, *J* = 8.6 Hz, 1H, NH[Val1]), 5.94 (d, *J* = 5.9 Hz, 1H, H1'), 4.30 (dd, *J* = 3.6 and 5 Hz, 1H, H3'), 4.23-4.19 (m, 1H, H4'), 3.91 (ddd, *J* = 11.8, 5.2 and 3.0 Hz, 1H, α [Val1]), 3.89-3.83 (m, 2H, 5' and H5"), 3.34-3.24 (m, 2H, α [Val2] and NH[Val2]), 1.91-1.78 (m, 2H, β[Val1 and Val2] overlapped with acetate), 0.72 (dd, *J* = 6.9 and 1.3 Hz, γ[Val2]), 0.65 (dd, *J* = 20.8 and 6.8 Hz, γ[Val1]) ppm. H2' is hidden by the water signal.

^{13}C NMR (H_2O, 125 MHz): δ = 178.2 (COO-[Val1]), 176.0 (d, *J* = 3.4 Hz, CO[Val2]), 155.5 (C6), 152.8 (C2), 149.0 (C4), 139.7 (C8), 118.6 (C5), 86.9 (C1'), 84.2 (d, *J* = 8.9 Hz, C4'), 74.0 (C2'), 70.6 (C3'), 64.0 (d, *J* = 5.0 Hz, C5'), 61.4 (α[Val2]), 61.0 (α[Val1]), 31.6 (d, *J* = 5.1 Hz, β[Val2]), 30.6 (β[Val1]), 18.8 (d, *J* = 9.1 Hz, γ[Val2]), 17.7 (γ[Val1]), 16.9 (γ [Val1]) ppm.

^{31}P NMR (H_2O, 200 MHz): δ = 6.78 ppm.

Exact mass ([M-H]$^-$) calculated for $C_{20}H_{31}N_7O_9P$ 544.1926 found 544.1901.

Val$_3$A. 70% conversion, 40% isolated yield.

^{1}H NMR (H_2O, 500 MHz): δ = 8.38 (s, 1H, H8), 8.10 (s, 1H, H2), 8.00 (d, *J* = 7.9 Hz, 1H, NH[Val2]), 7.82 (d, *J* = 8.9 Hz, 1H, NH[Val1]), 5.97 (d, *J* = 5.9 Hz, 1H, H1'), 4.34 (dd, *J* = 5.0 and 3.7 Hz, 1H, H3'), 4.19-4.23 (m, 1H, H4'), 3.91 (dd, *J* = 8.9 and 6.4 Hz, 1H, α[Val1]), 3.83-3.88 (m, 2H, H5' and H5"), 3.77-3.83 (m, 1H, α[Val2]), 3.19-3.30 (m, 2H, α[Val3] and NH[Val3]), 1.98-1.88 (m, 1H, β[Val1]), 1.83-1.76 (m, 1H, β[Val2] overlapped with acetate), 1.76-1.68 (m, 1H, β[Val3]), 0.77 (dd, *J* = 6.8 and 4.3 Hz, γ[Val1]), 0.69 (dd, *J* = 16.2 and 6.8 Hz, γ[Val3]), 0.64 (dd, *J* = 27.8 and 6.7 Hz, γ[Val2]) ppm. H2' is hidden by the water signal.

^{13}C NMR (H_2O, 125 MHz): δ = 178.2 (COO-[Val1]), 176.3 (d, *J* = 2.3 Hz, CO[Val3]), 172.7 (CO[Val2]), 155.7 (C6), 153.0 (C2), 149.1 (C4), 140.0 (C8), 118.7 (C5), 86.8 (C1'), 84.2 (d, *J* = 8.8 Hz, C4'), 73.8 (C2'), 70.7 (C3'), 64.1 (d, *J* = 4.8 Hz, C5'), 61.0 (α[Val1]), 60.9 (α[Val3]), 60.0 (α[Val2]), 32.0 (d, *J* = 6.5 Hz, β[Val3]), 30.6 (β[Val1]), 29.7 (β[Val2]), 18.8 (γ[Val3]), 18.6 (γ[Val3]), 18.3 (γ[Val2], 18.2 (γ[Val2]), 17.6 (γ[Val1]), 17.3 (γ [Val1]) ppm.

^{31}P NMR (H_2O, 200 MHz): δ = 6.87 ppm.

Exact mass ([M-H]$^-$) calculated for $C_{25}H_{40}N_8O_{10}P$ 643.2610 found 643.2620.

Ala$_3$A. 63% conversion. 35% isolated yield by MPLC.

^{1}H NMR (D_2O, 500 MHz): δ = 8.30 (s, 1H, H8), 8.05 (s, 1H, H2), 5.96 (d, *J* = 5.6 Hz, 1H, H1'), 4.33 (dd, *J* = 5.0 and 3.8 Hz, 1H, H3'), 4.24-4.21 (m, 1H, H4'), 4.12

(app. q, *J* = 7.3 Hz, 1H, α[Ala2]), 3.95-3.85 (m, 3H, H5', H5" and α[Ala1]), 3.52 (app. dq, 9.6 and 7.2 Hz, 1H, α[Ala3]), 1.20 (d, *J* = 7.3 Hz, ß[Ala2]), 1.17 (d, *J* = 7.3 Hz, ß[Ala1]), 1.07 (d, *J* = 7.3 Hz, ß[Ala3]) ppm. H2' is hidden by the water signal.

^{13}C NMR (H_2O, 125 MHz): δ = 179.9 (COO-[Ala1]), 177.2 (d, *J* = 2.3 Hz, CO[Ala3]), 173.6 (CO[Ala2]), 155.4 (C6), 152.7 (C2), 148.9 (C4), 139.7 (C8), 118.5 (C5), 87.0 (C1'), 84.0 (d, *J* = 9.1 Hz, C4'), 74.2 (C2'), 70.5 (C3'), 63.9 (d, *J* = 5.0 Hz, C5'), 51.0 (α[Ala1]), 50.9 (α[Ala3]), 49.3 (α[Ala2]), 19.8 (D, *J* = 5.3 Hz, ß[Ala3]), 17.3 (ß[Ala1]), 16.8 (ß[Ala2]) ppm.

^{31}P NMR (H_2O, 200 MHz): δ = 6.50 ppm.

Exact mass ([M-H]$^-$) calculated for $C_{19}H_{29}N_8O_{10}P$ 559.1671 found 559.1678.

Leu$_3$A. Because of the lower solubility of Leu$_3$, the method was slightly modified. Leu$_3$ (55 mg, 0.15 mmol) was lyophilized twice from 0.1 M HCl (15 mL) to yield the hydrochloride salt that was more soluble than the zwitterion. It was dissolved in 2 mL of sodium bicarbonate buffer (pH 8, 0.5 M) then ImpA (0.1 mmol) was added and the mixture was stirred for 4 months. Because of the dilution, the reaction was slower, and the yield lowered by competition of the hydrolysis.

50% conversion, 45% isolated yield.

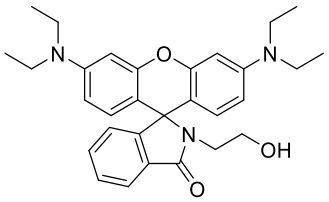
^{1}H NMR (D_2O, 400 MHz): δ = 8.37 (s, 1H, H8), 8.08 (s, 1H, H2), 5.97 (d, *J* = 5.6 Hz, 1H, H1'), 4.32 (dd, *J* = 5.1 and 3.9 Hz, 1H, H3'), 4.23-4.17 (overlapped m, 2H, H4' and α[Leu2]), 4.08-4.00 (m, 1H, α[Leu1]), 3.93-3.81 (m, 2H, H5' and H5"), 3.45 (ddd, *J* = 14.6, 8.4 and 6.2 Hz, 1H, α[Leu3]), 1.53-1.32 (m, 6H, ß[Leu1, 2 and 3] overlapped), 1.32-1.15 (m, 3H, γ [Leu1, 2 and 3] overlapped), 0.72 (dd, *J* = 17.8 and 6.0 Hz, δ [Leu1]), 0.69 (dd, *J* = 21.5 and 6.1 Hz, δ [Leu1]), 0.59 (dd, *J* = 9.3 and 6.6 Hz, δ [Leu3]) ppm. H2' is hidden by the water signal.

^{13}C NMR (D_2O, 100 MHz): δ = 178.2 (COO-[Leu1]), 177.2 (d, *J* = 2.7 Hz, CO[Leu3]), 173.3 (CO[Leu2]), 155.6 (C6), 152.9 (C2), 149.1 (C4), 140.0 (C8), 118.8 (C5), 87.1 (C1'), 84.1 (d, *J* = 8.8 Hz, C4'), 74.0 (C2'), 70.5 (C3'), 64.1 (d, *J* = 4.8 Hz, C5'), 63.8 (app t, *J* = 3 Hz, C5'), 54.1 (m, α [Leu3] and α [LeuX]), 52.1 (α [LeuX]), 43.3 (br d, *J* = 6 Hz, ß[Leu3]), 40.8 (β [LeuX]), 39.7 (β [LeuX]), 24.5, 24.2, 24.0 (γ [Leu], not discriminated), 22.5, 22.0, 21.9, 21.4, 21.0, 20.8 (δ [Leu], not discriminated) ppm. H2' is hidden by the water signal.

^{31}P NMR (D_2O, 160 MHz): δ = 6.21 ppm.

Exact mass ([M-H]$^-$) calculated for $C_{28}H_{46}N_8O_{10}P$ 685.3080, found 685.3061

N-(Rhodamine B)-lactam-2-aminoethanol



Procedure adapted from literature.[442] Rhodamine B (4.73 g, 10 mmol) and ethanolamine (20 mL) were refluxed in methanol (20 mL) overnight. A precipitate formed spontaneously, that was collected and washed extensively with cold methanol until the pink color had faded off. After drying, the title compound was obtained in pure form (4.2 g, 8.6 mmol, 86%) and the analytical data matched the report of the authors.

^{1}H NMR ($CDCl_3$, 300 MHz): δ = 7.84-7.81 (m, 1H, H3), 7.40-7.34 (m, 2H, H4 and H5), 7.01-6.98 (m, 1H, H6), 6.41 (d, *J* = 8.9 Hz, 2H, H1' and H8'), 6.30 (d, *J* = 2.4 Hz, 2H, H4' and H5'), 6.22 (dd, *J* = 8.9 and 2.4 Hz, 2H, H2' and H7'), 4.09 (br t, *J* = 5.8 Hz, 1H, OH), 3.39 (br q, *J* = 5.0 Hz, 2H, C$\underline{H_2}$OH), 3.26 (q, *J* = 7.1 Hz, 8H, $(CH_2)_4$), 3.21 (overlapped, 2H, C$\underline{H_2}CH_2$OH), 1.09 (t, *J* = 7.1 Hz, 12H, $(CH_3)_4$) ppm.

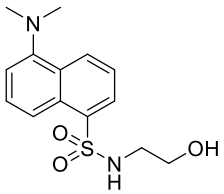
Failed attempts to phosphorylate the substituted rhodamine

The following phosphorylation conditions were attempted without success:

- the reaction with tetrabutylammonium dihydrogen phosphate and trichloroacetonitrile[467] in anhydrous dichloromethane afforded an inseparable mixture of more than 5 products.
- the starting material was recovered as the open tautomer when treated with di tert-butyl chlorophosphate (freshly prepared from the corresponding phosphite and NCS) in dichloromethane, using either triethylamine or triethylamine/pyridine as a base.
- the starting material was recovered as the open tautomer when treated with dibenzyl chlorophosphate (freshly prepared from the corresponding phosphite and NCS) in dichloromethane, using either triethylamine or triethylamine/pyridine as a base.
- the starting material was consumed when first deprotonated with NaH, followed by treatment with dibenzyl chlorophosphate in anhydrous THF or anhydrous DMF. The desired product did not form to significant extents.
- the starting material did not react when deprotonated with n-butyllithium then treated with dibenzyl chlorophosphate in anhydrous THF.
- the reaction with dibenzyl *N,N*-diisopropyl phosphoramidite and 1-ethylthiotetrazole in dichloromethane, followed by mCPBA oxidation gave no trace of the desired product.
- the starting material was consumed when treated with $POCl_3$ and triethylamine in anhydrous THF, but the desired product was not formed.
- the phosphorylation with diphenyl chlorophosphate in pyridine afforded the desired diphenyl phosphate in 87% yield. However, the phenyl groups

could not be removed by hydrogenation over PtO_2 or over Pt/C. Saponification with NaOH in water/dioxane cleaved the CH_2O-P bond before the PhO-P bonds.

5-(dimethylamino)-N-(2-hydroxyethyl)naphthalene-1-sulfonamide (29)

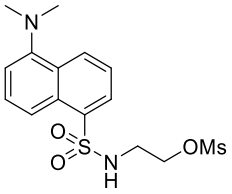


Procedure adapted from the literature.[444] A solution of dansyl chloride (810 mg, 3.0 mmol) in anhydrous dichloromethane (100 mL) was added dropwise over 30 minutes at room temperature to a stirred solution of ethanolamine (300 µL, 4 mmol) and triethylamine (1.35 mL, 9 mmol) in anhydrous dichloromethane (10 mL). After one hour of further stirring at room temperature, the solution was concentrated to 2-3 mL and acetone (10 mL) was added. Triethylamine hydrochloride was filtered off, washed with acetone, and the filtrate was concentrated to dryness. The residue was chromatographied on silica gel, eluting with ethyl acetate/cyclohexane 6:4 and the title compound was obtained as a yellow-green solid (1.07 g, 2.97 mmol, 99%). The analytical data matches the author's report.

^{1}H NMR (MeOD, 500 MHz): δ = 8.56 (dt, *J* = 8.6 and 0.9 Hz, 1H, H4), 8.37 (dt, *J* = 8.6 and 0.9 Hz, 1H, H8), 8.21 (dd, *J* = 7.3 and 1.2 Hz, 1H, H2), 7.62-7.54 (m, 2H, H3 and H7), 7.27 (dd, *J* = 7.6 and 0.7 Hz, 1H, H6), 3.49 (t, *J* = 6.1 Hz, 2H, C$\underline{H_2}$OH), 2.95 (t, *J* = 6.1 Hz, 2H, C$\underline{H_2}$NHSO$_2$), 2.87 (s, 6H, NMe$_2$) ppm.

^{13}C NMR (MeOD, 125 MHz): δ = 151.8 (C5), 135.5 (C1), 129.84 (C8a), 129.78 (C4), 129.6 (C4a), 128.7 (C2), 127.7 (C7), 122.9 (C3), 119.1 (C8), 115.1 (C6), 60.5 (CH_2O), 44.6 (CH_2N), 44.4 (NMe$_2$) ppm.

2-((5-(dimethylamino)naphthalene)-1-sulfonamido)ethyl methanesulfonate (30)

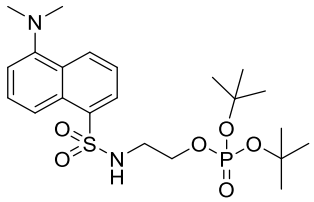


To a solution of dansyl ehtanolamine **(29)** (355 mg, 1.0 mmol) in anhydrous pyridine (5 mL) was added methanesulfonyl chloride (385 µL, 5 mmol) dropwise at room temperature. After one hour of reaction, 10 mL of water were added and the mixture was further stirred for 30 min, then extracted twice with ethyl

acetate. The organic phase was dried on sodium sulfate and concentrated. ^{1}H NMR of the crude extract confirmed that the desired compound was obtained in quantitative yield (372 mg, 1.0 mmol) with sufficient purity to be used without further purification.

^{1}H NMR (MeOD, 300 MHz): 8.56 (dt, *J* = 8.6 and 1.0 Hz, 1H), 8.33 (dt, *J* = 8.6 and 1.3 Hz, 2H), 8.21 (dd, *J* = 7.3 and 1.3 Hz, 1H), 7.58 (ddd, *J* = 8.6, 7.3 and 4.6 Hz, 2H), 7.27 (dd, *J* = 7.6 and 0.7 Hz, 1H), 4.07 (t, *J* = 7.3 Hz, 2H), 3.20 (t, *J* = 7.3 Hz, 2H), 2.87 (s, 6H, NMe_2), 2.84 (s, 3H, Ms) ppm. Measurable amounts of pyridine (8.52, 7.84, 7.27 ppm) and ethyl acetate (4.09, 2.00 1.22 ppm) were deduced for yield calculation.

Di-tert-butyl (2-((5-(dimethylamino)naphthalene)-1-sulfonamido)ethyl) phosphate (31)

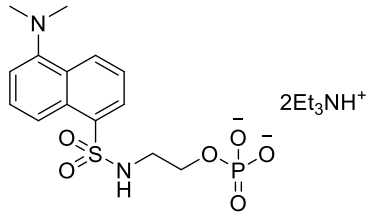


The mesylate **(30)** (342 mg, 1.0 mmol) was dissolved in anhydrous DMF (10 mL) under argon, and commercial potassium di-tert-butyl phosphate (360 mg, 1.5 mmol) and tetrabutylammonium iodide (50 mg, 0.13 mmol, cat.) were added. The thick mixture was stirred for 24h at room temperature then diluted with 100 mL of water and extracted with ethyl acetate (3×50 mL). The organic layer was dried over sodium sulfate and concentrated to afford the desired product (490 mg, quantitative) as a dark yellow oil. TLC in ethyl acetate/cyclohexane 7:3 showed a single product (R_f = 0.5). The analysis matched the report of Crane et al.[445]

^{1}H NMR (MeOD, 300 MHz): δ = 8.54 (dt, *J* = 8.6 and 1.1 Hz, 1H, H4), 8.33 (dt, *J* = 8.7 and 0.9 Hz, 1H, H8), 8.19 (dd, *J* = 7.3 and 1.2 Hz, 1H, H2), 7.58-7.47 (m, 2H, H3 and H7), 7.24 (dd, *J* = 7.6 and 0.8 Hz, 1H, H6), 3.84 (app. q, *J* = 6.0 Hz, 2H, C$\underline{H_2}$OP), 3.09 (app. t, *J* = 6.0 Hz, 2H, C$\underline{H_2}$NHSO$_2$), 2.85 (s, 6H, NMe_2), 1.38 (d, *J* = 0.6 Hz, 18H, tBu) ppm.

^{31}P NMR (MeOD, 120 MHz): δ = -10.7 (t, *J* = 7.0 Hz) ppm

Bis(triethylammonium)(2-((5-(dimethylamino)naphthalene)-1-sulfonamido) ethyl)phosphate (32)



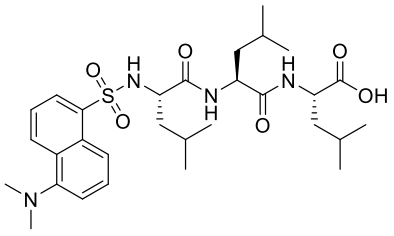
The phosphate **(31)** (490 mg, 1 mmol) was dissolved in dichloromethane (20 mL), and trifluoroacetic acid (1 mL) was added. After stirring for 2 hours at room temperature, the solution was concentrated, co-evaporated with dichloromethane three times and the resulting oil was diluted in methanol (1 mL) and water (4 mL). Triethylamine (300 µL, 2 mmol) was added slowly and the bis(triethylammonium) salt of the title compound was purified by ion pairing flash chromatography (C18 silica), eluting with 50 mM triethylammonium acetate/acetonitrile (20 to 80%) to afford the desired compound as a yellow oil (570 mg, > 95% over 3 steps). The analysis matched the report of Crane et al[445] with the exception of the presence of the triethylammonium signals (the authors isolated the free acid). Preforming the bis(triethylammonium) before chromatography is important to avoid the partial elution of other ionic forms (when loading 1 mmol onto the column, the concentration of triethylammonium ions is not sufficient to generate the ion pair before significant elution begins).

^{1}H NMR (DMSO-*d*$_6$, 300 MHz): δ = 8.45 (d, *J* = 8.5 Hz, 1H, H4), 8.30 (d, *J* = 8.7 Hz, 1H, H8), 8.11 (dd, *J* = 7.3 and 1.1 Hz, 1H, H2), 7.65-7.53 (m, 2H, H3 and H7), 7.25 (d, *J* = 7.3 Hz, 1H, H6), 3.68 (dt, *J* = 9.6 and 5.9 Hz, 2H, C$\underline{H_2}$OP), 2.94 (t, *J* = 5.9 Hz, 2H, C$\underline{H_2}$NHSO$_2$), 2.83 (s, 6H, NMe$_2$) ppm, 2.79 (q, *J* = 7.3 Hz, 12H, 2×N(C$\underline{H_2}$CH$_3$)$_3$), 1.07 (t, *J* = 7.3 Hz, 18H, 2×N(CH$_2$C$\underline{H_3}$)$_3$).

^{31}P NMR (DMSO-*d*$_6$, 120 MHz): δ = 0.0 ppm

ESI-MS $[M+H]^+$ = 375.0

***N*-dansyl-Leu$_3$**



Leu$_3$ (8 mg) and sodium carbonate (40 mg) were dissolved in 1 mL of water, then dansyl chloride (16 mg) in acetone (0.5 mL) was added and the mixture was strongly stirred for 1h. It was then extracted with ethyl acetate (2 mL) and concentrated. Purification by column chromatography on silica gel eluting with

10% methanol in ethyl acetate afforded the desired compound as a yellow oil (yield not determined).

^{1}H NMR (DMSO-d_6, 300 MHz): δ = 8.57 (dt, J = 8.5 and 1.0 Hz, 1H, H4), 8.36 (dt, J = 8.7 and 1.0 Hz, 1H, H8), 8.24 (dd, J = 7.4 and 1.3 Hz, 1H, H2), 7.58 (app. q, J = 8.1 Hz, 2H, H3 and H7), 7.27 (dd, J = 7.6 and 0.8 Hz, 1H, H6), 4.43-4.32 (m, 1H, α [Leu3]), 4.28-4.18 (m, 1H, α [Leu2]), 3.65-3.56 (m, 1H, α [Leu1]), 2.87 (s, 6H, NMe_2), 1.77-1.35 (br, 9H, β and γ, overlapped), 0.95-0.77 (m, 12H, δ [Leu2 and Leu3]), 0.65 (d, J = 6.6 Hz, 3H, δ [Leu1]), 0.24 (d, J = 6.6 Hz, 3H, δ [Leu1]) ppm.

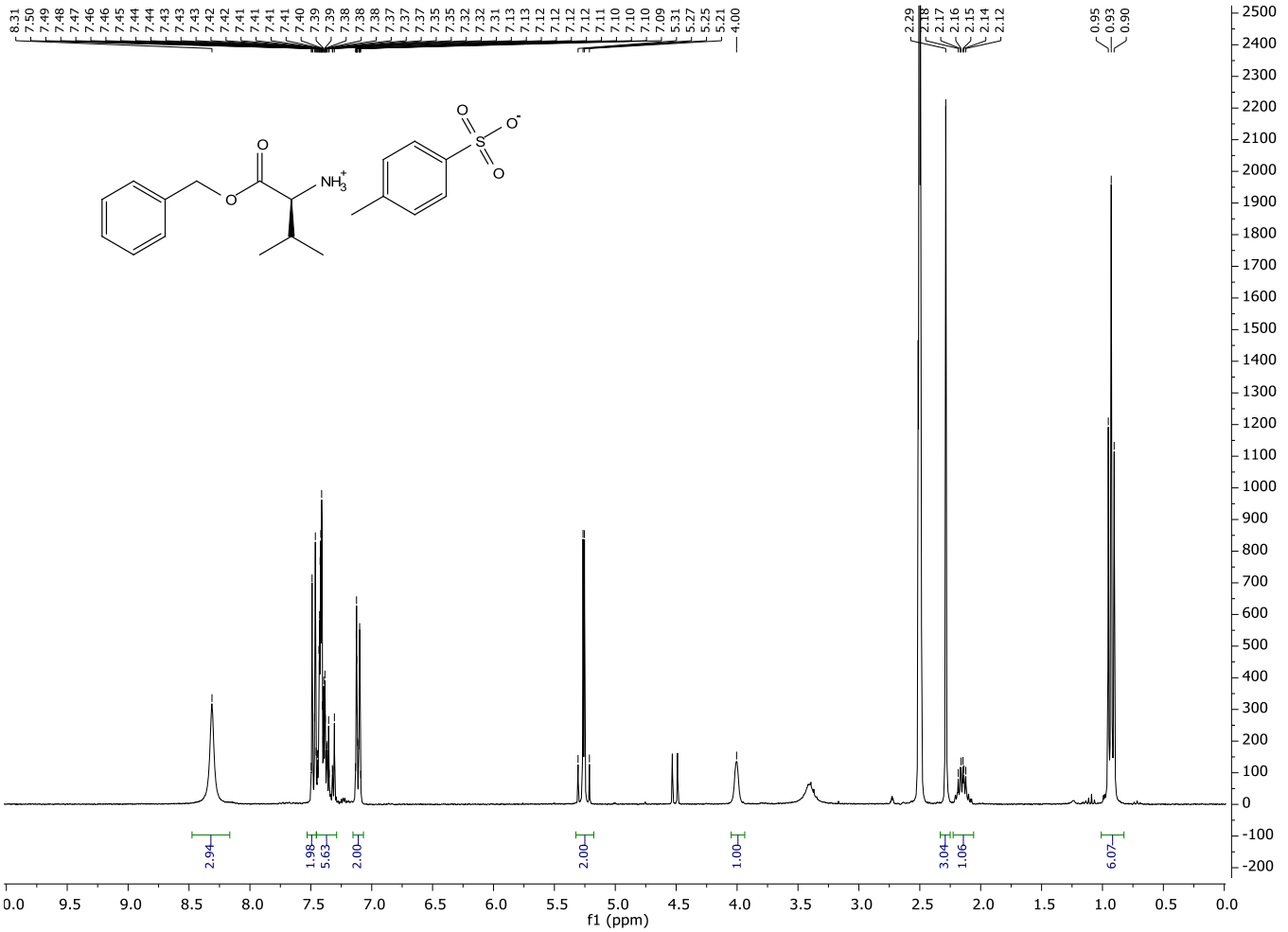
Dodecyl phosphate

Method of Lira and coll.[467] Dodecanol (930 mg, 5 mmol) and trichloroacetonitrile (1.2 mL, 12 mmol) were dissolved in anhydrous acetonitrile (5 mL) at 30°C. A suspension of tetrapropylammonium dihydrogen phosphate (prepared from 980 mg of phosphoric acid, 10 mmol, and appropriate amounts of a solution of tetrapropylammonium hydroxide followed by lyophilization) in anhydrous acetonitrile (25 mL) was added dropwise. The mixture was stirred for 3 hours at 30°C, then evaporated under vacuum. The oily residue was purified by flash chromatography on silica gel, eluting with CH_2Cl_2/MeOH/H_2O 80:20:1 to 65:35:4 to give the desired tetrapropylammonium hydrogen dodecyl phosphate as a pale sticky oil (1.00 g, 44%). The compound was converted to a free acidic form by passing through a column of freshly activated Amberlyst IR120 acidic resin (percolating with water/methanol) then titrated with sodium hydroxide to the monosodium salt and lyophilized.

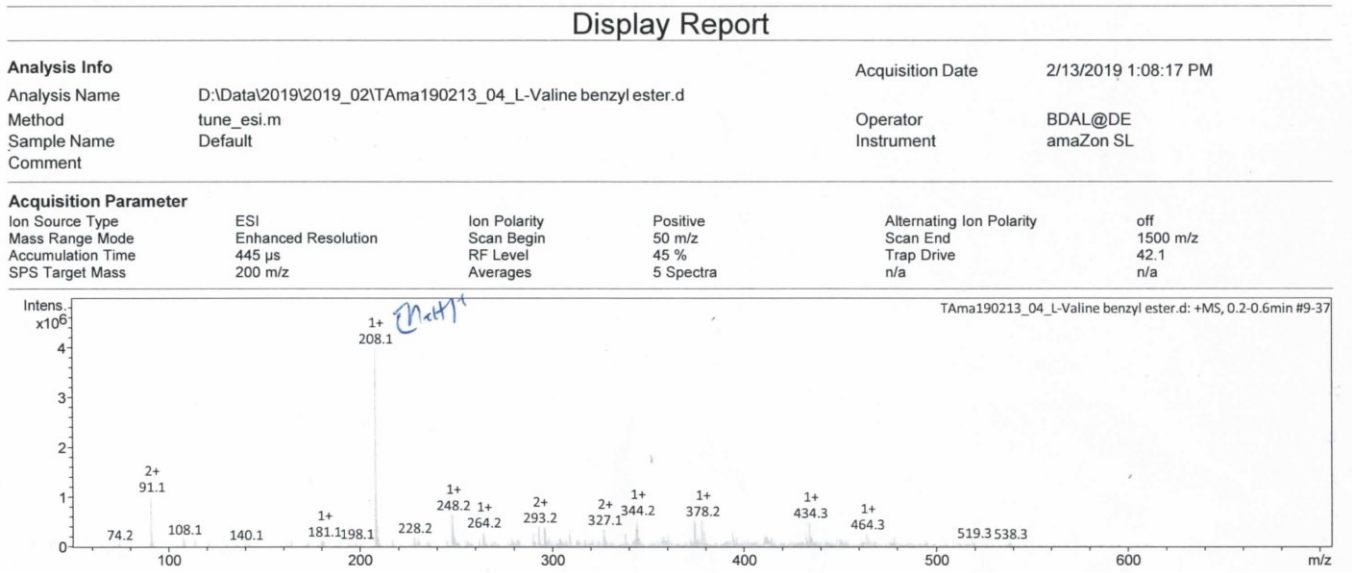
^{1}H NMR (MeOD, 300 MHz): δ = 3.84 (q, J = 6.5 Hz, 2H, CH_2OP), 3.34 (s, 1H, OH) 3.21-3.16 (m, 8H, 4×CH_2N from tetrapropylammonium), 1.77-1.67 (m, 8H, 4×CH_2 from tetrapropylammonium), 1.64-1.58 (m, 2H, CH_2), 1.41-1.28 (br, 18H, 9×CH_2 from the dodecyl chain), 1.00 (t, J = 7.36 Hz, 12H, 4×CH_3 from tetrapropylammonium), 0.89 (t, J = 6.8 Hz, 3H, CH_3) ppm.

^{31}P NMR (MeOD, 120 MHz): δ = 1.25 ppm.

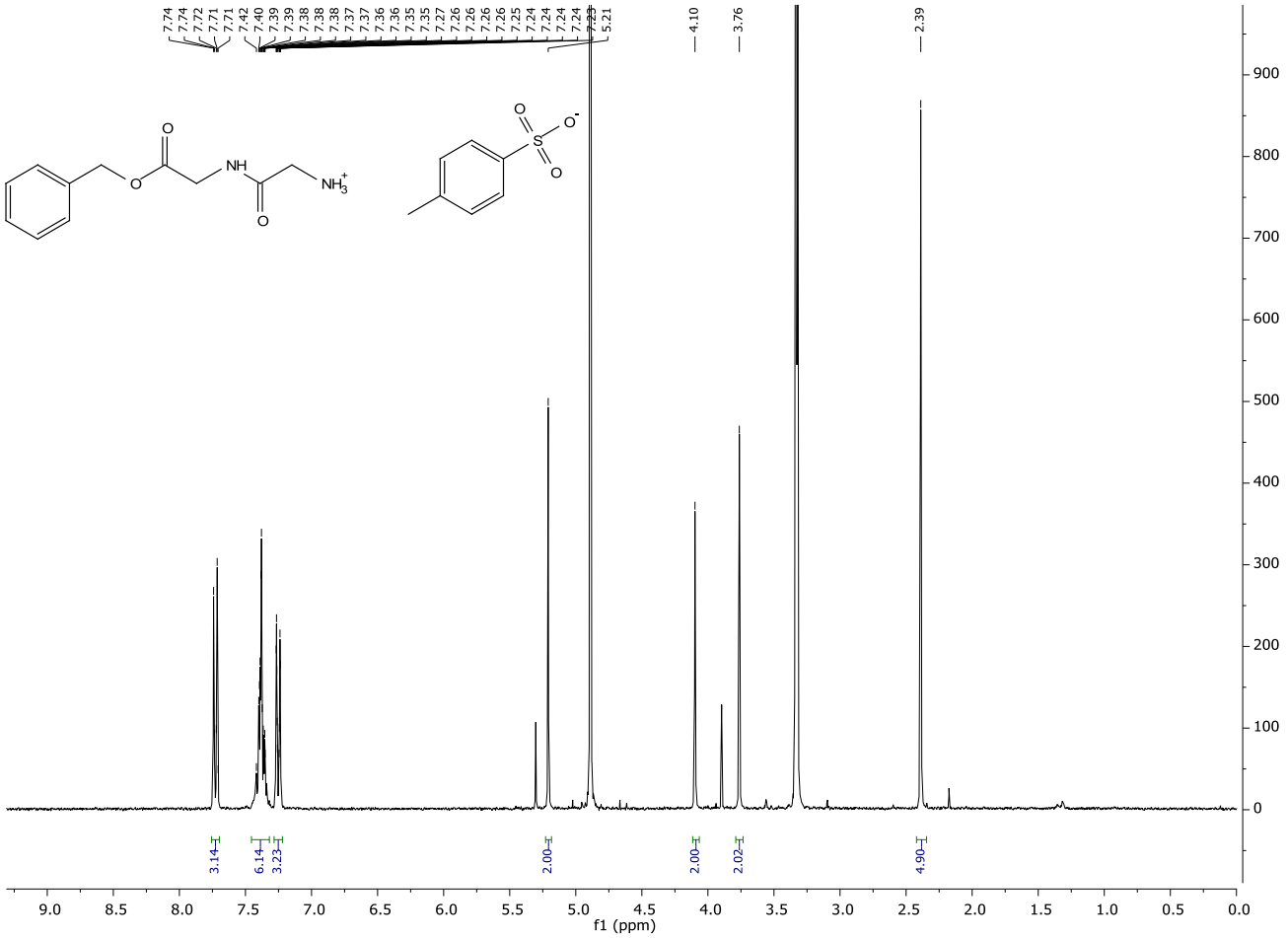
9. Analytical data for synthetic compounds



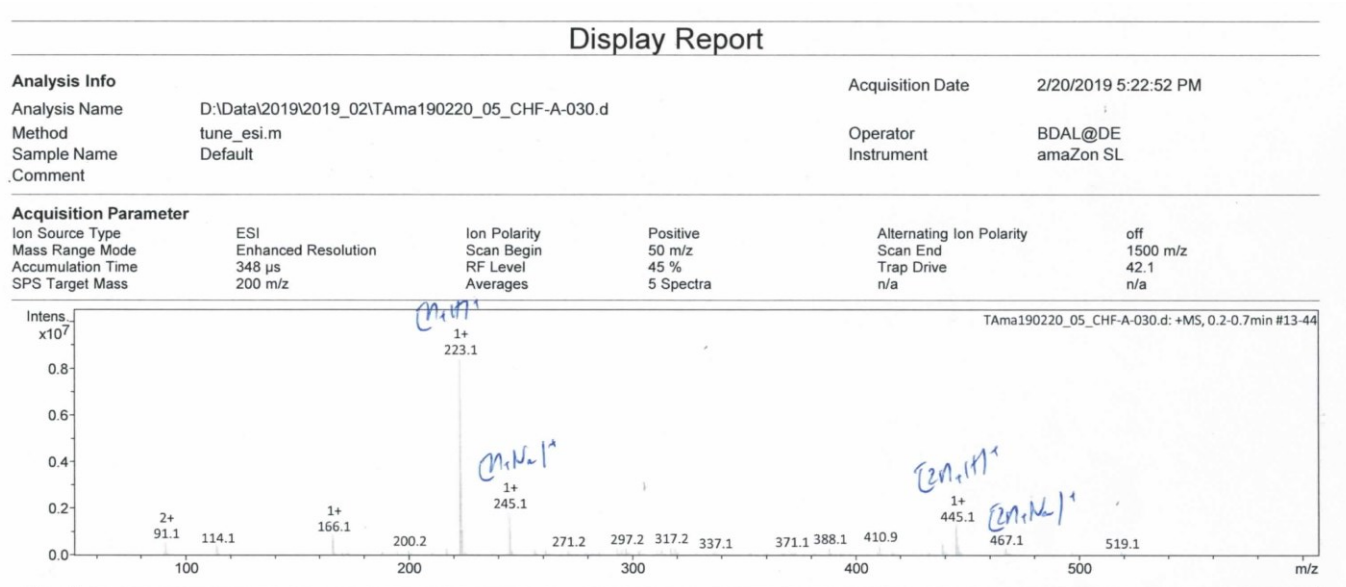
^{1}H NMR spectrum (DMSO-d_6, 300 MHz) of BnOVal.TsOH showing traces of benzyl alcohol.



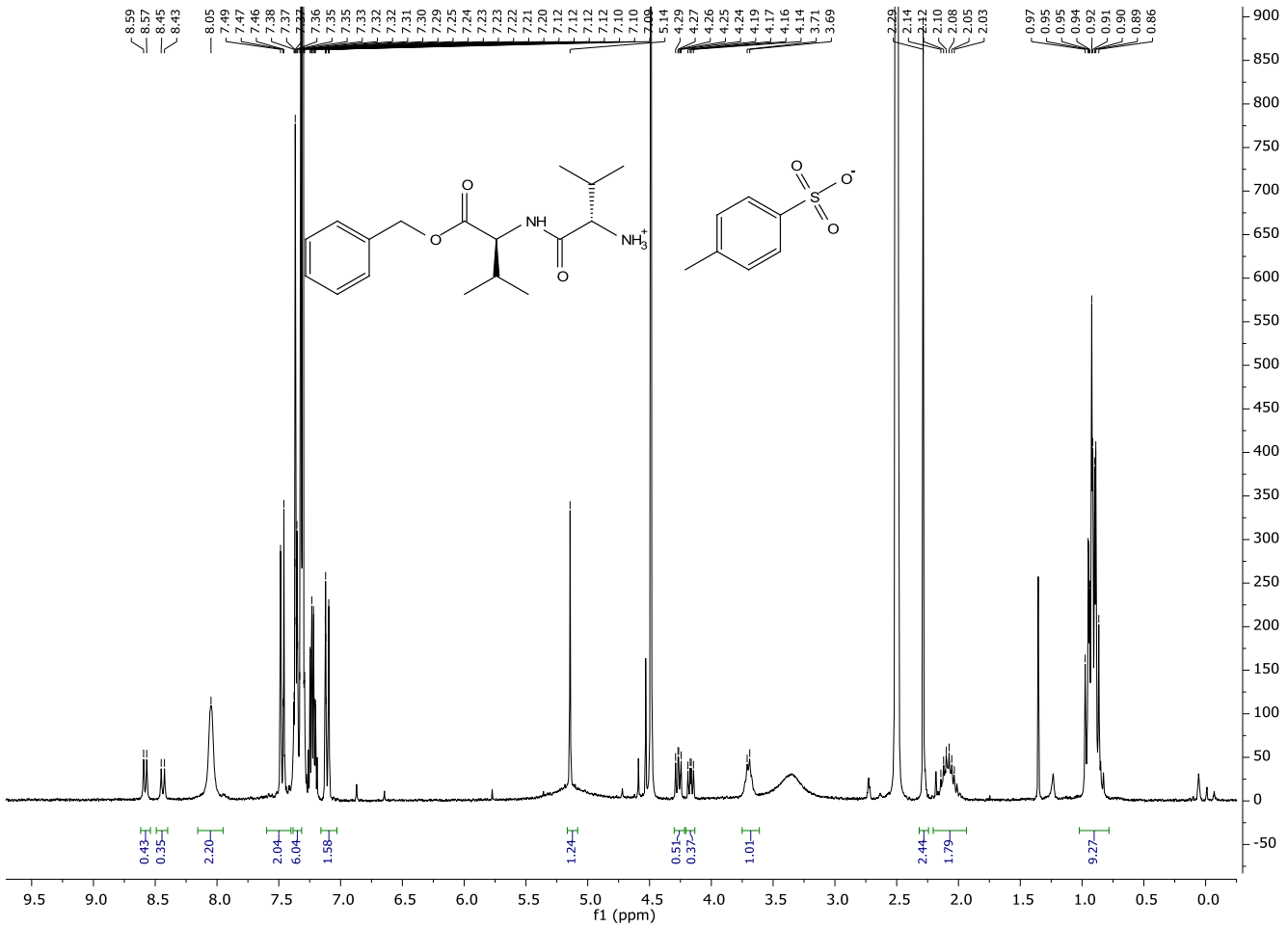
ESI-MS (+MS) of BnOVal.TsOH



[1]H NMR spectrum (MeOD, 300 MHz) of BnOGlyGly.TsOH showing a slight excess of TsOH and BnOH



ESI-MS (+MS) of BnOGlyGly.TsOH



^{1}H NMR spectrum (DMSO-d_6, 300 MHz) of the crude mixture containing compound BnOValVal.TsOH, Val$_2$ and excess benzyl alcohol.

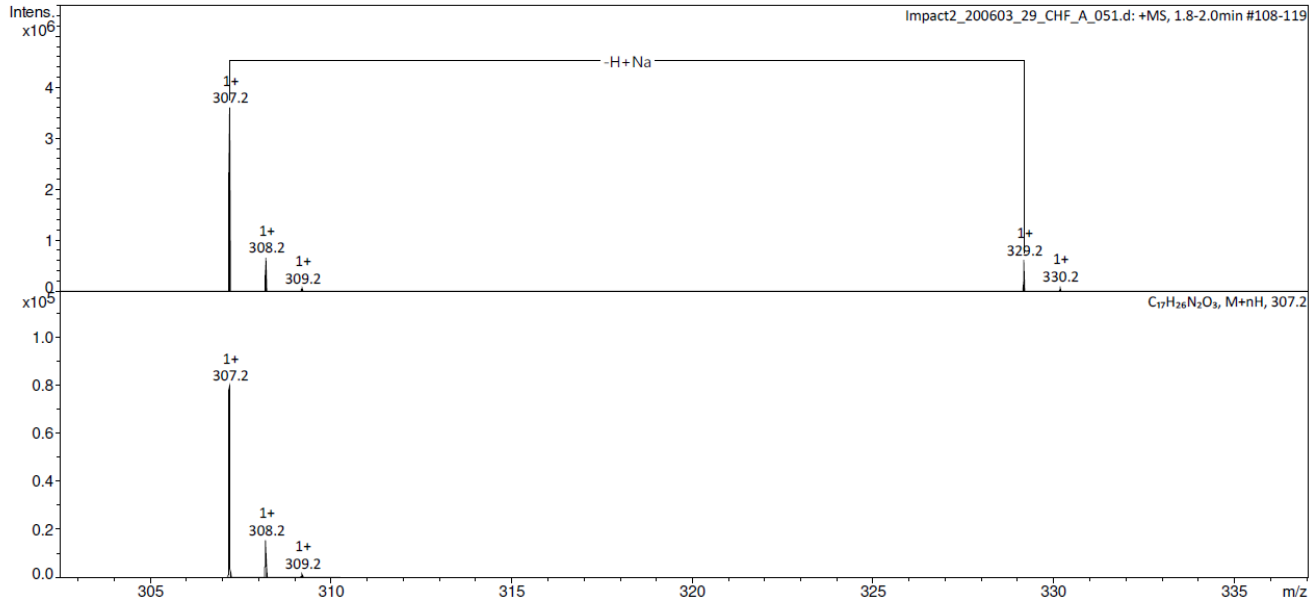
Display Report

Analysis Info

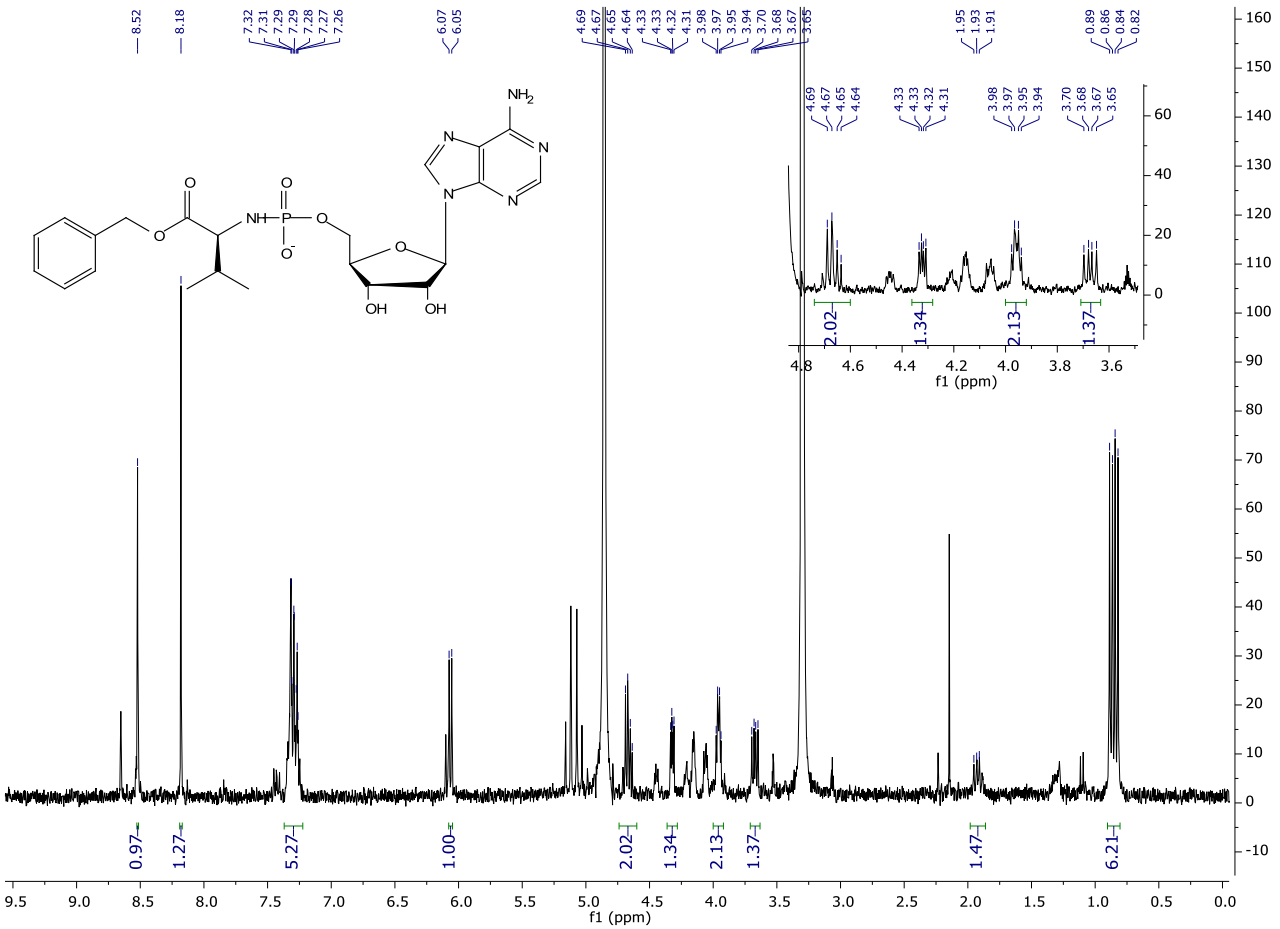
Analysis Name	D:\Data\Data\2020\2020_06\Impact2_200603_29_CHF_A_051.d	Acquisition Date	6/3/2020 5:07:53 PM
Method	Tune_pos_Standard.m	Operator	Demo User
Sample Name	TM	Instrument	impact II 1825265.10081
Comment			

Acquisition Parameter

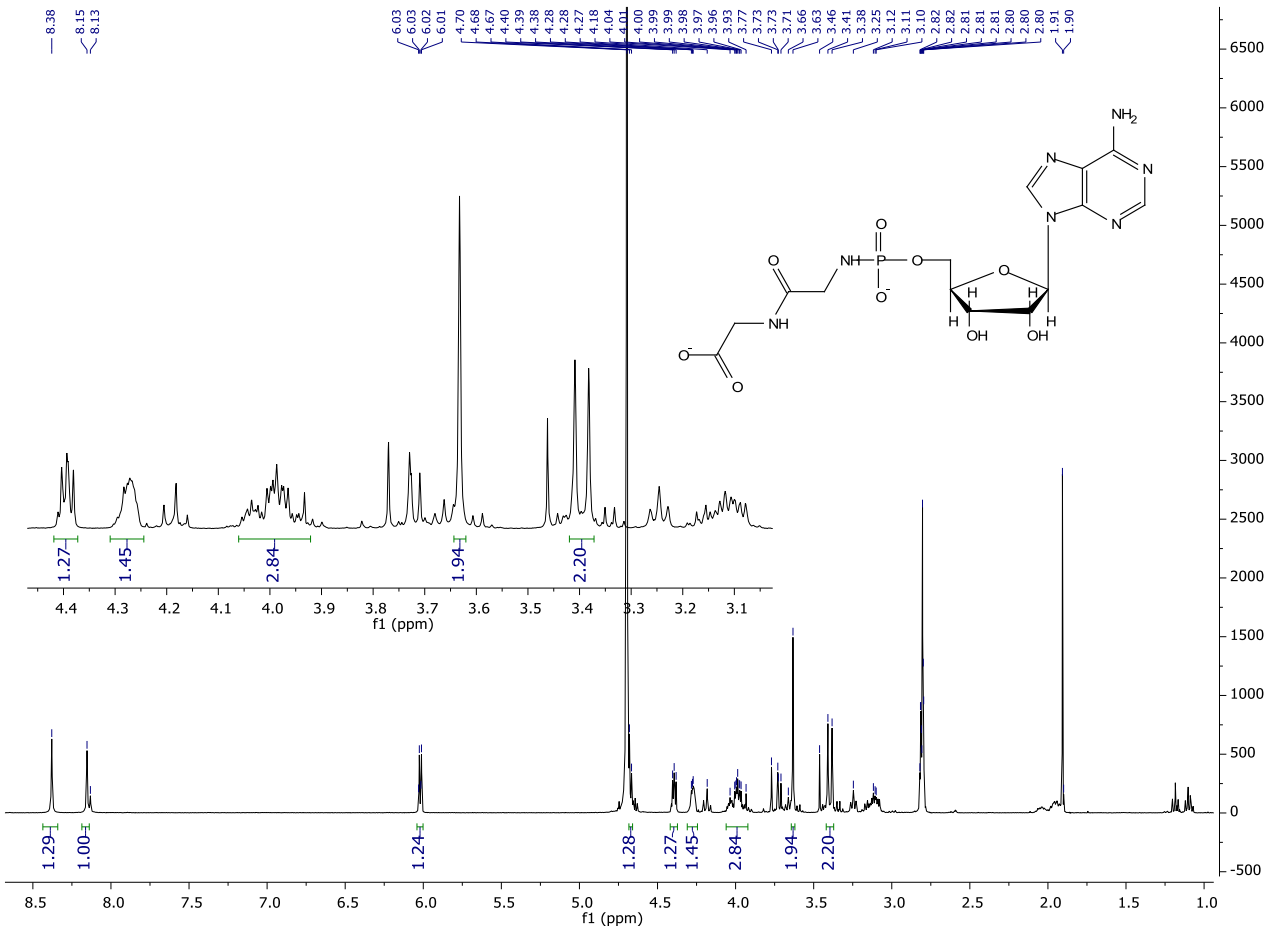
Source Type	ESI	Ion Polarity	Positive	Set Nebulizer	0.3 Bar
Focus	Active	Set Capillary	1500 V	Set Dry Heater	200 °C
Scan Begin	50 m/z	Set End Plate Offset	-500 V	Set Dry Gas	4.0 l/min
Scan End	1000 m/z	Set Charging Voltage	2000 V	Set Divert Valve	Source
		Set Corona	0 nA	Set APCI Heater	0 °C



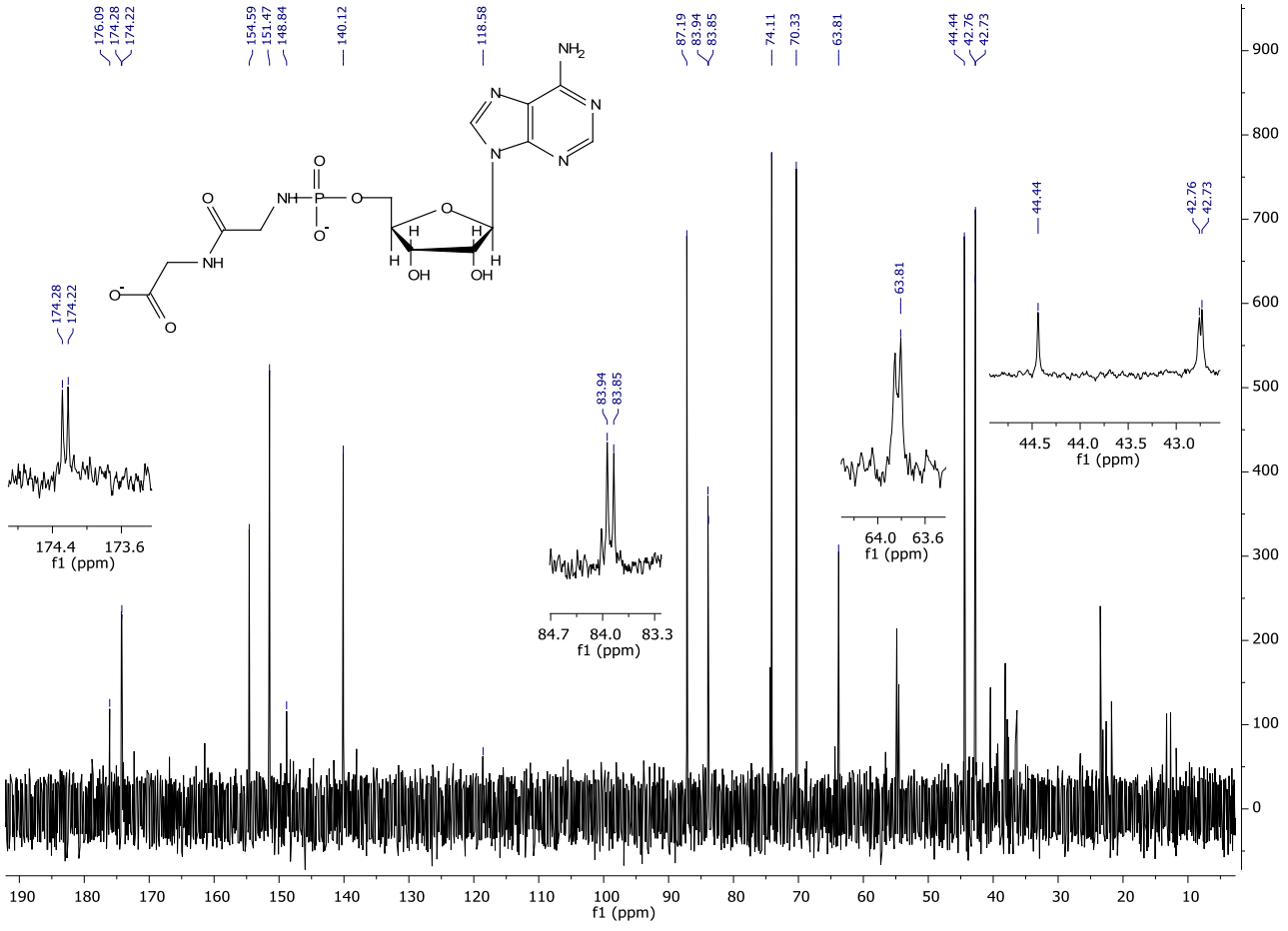
ESI-MS (+MS) of BnOValVal.TsOH



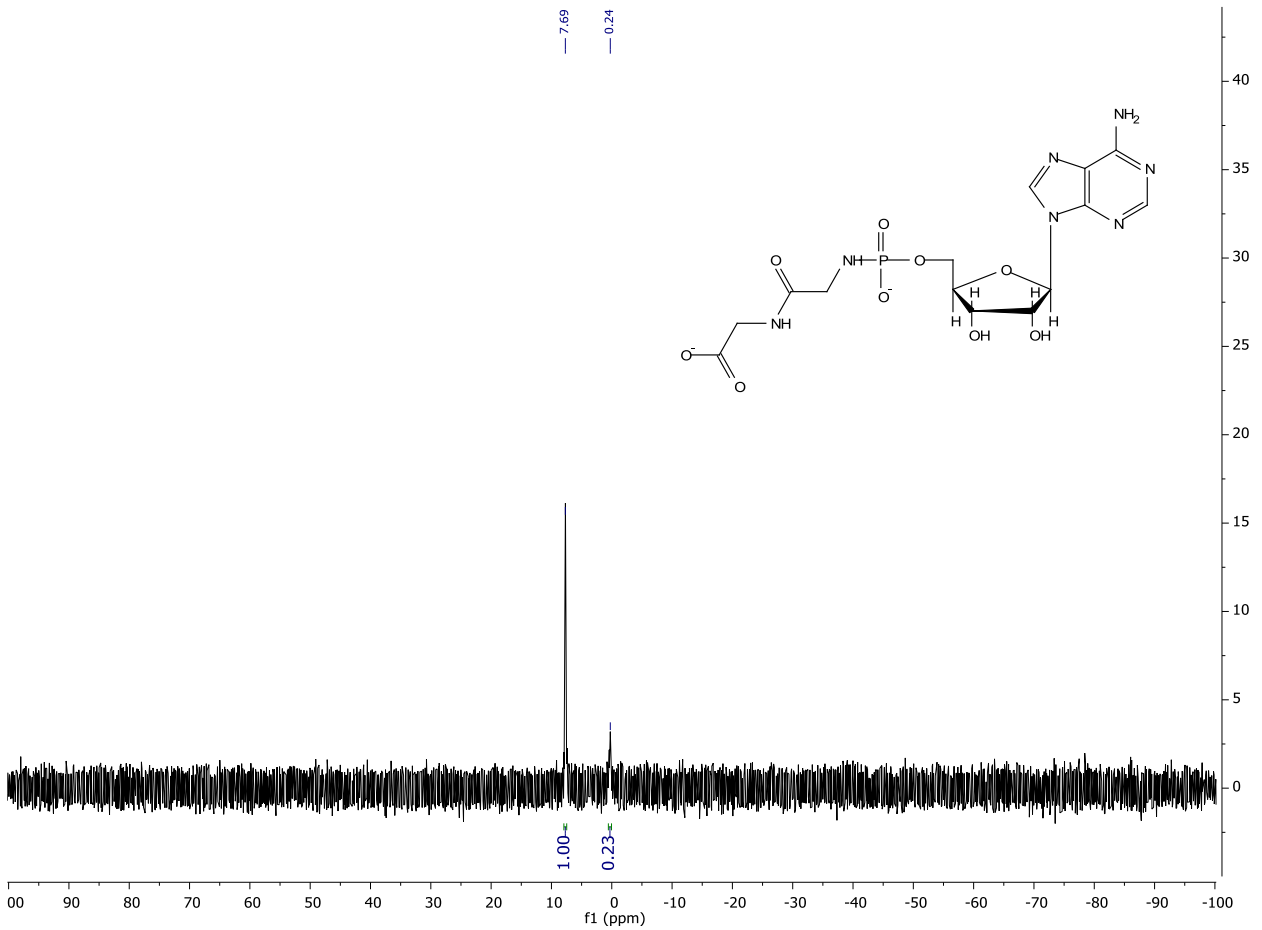
[1]H NMR spectrum (MeOD, 300 MHz) of impure BnOValpA obtained by the direct coupling method.



^{1}H NMR spectrum (H_2O, 500 MHz) of Gly_2A. AMP, that could not be removed by preparative HPLC, is visible.



^{13}C NMR spectrum (H_2O, 125 MHz) of Gly_2A. AMP, that could not be removed by preparative HPLC, is visible.



^{31}P NMR spectrum (H_2O, 200 MHz) of Gly_2A. AMP, that could not be removed by preparative HPLC, is visible.

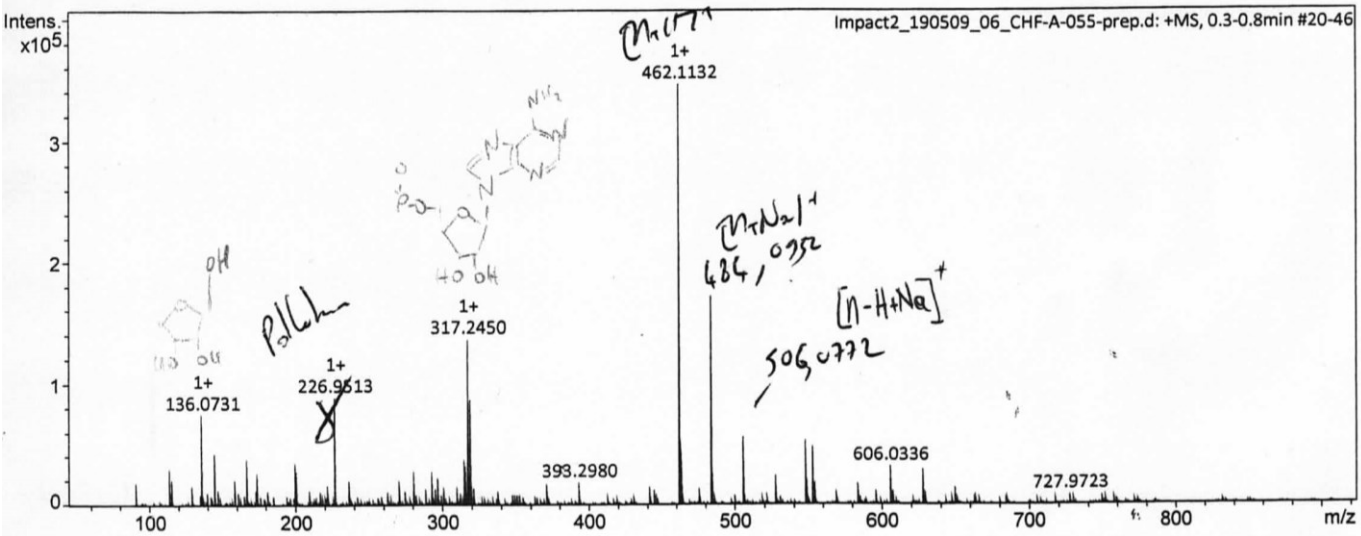
CENTRE COMMUN DE SPECTROMETRIE DE MASSE

Analysis Info

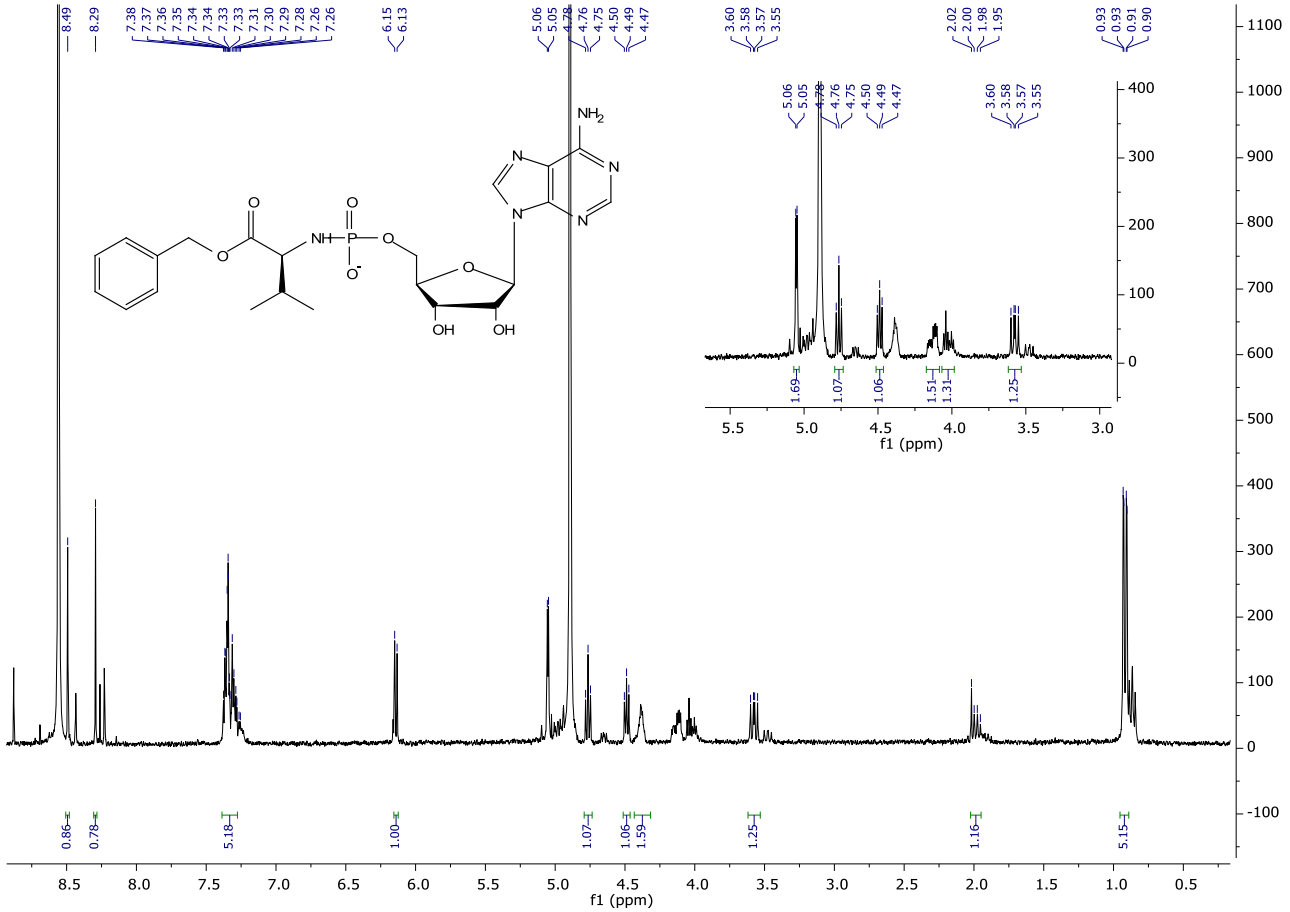
Analysis Name	Impact2_190509_06_CHF-A-055-prep.d		
Method	Tune_pos_Standard.m	Acquisition Date	5/9/2019 11:00:16 AM
Comment		Instrument / Ser#	impact II 1825265.1 0081

Acquisition Parameter

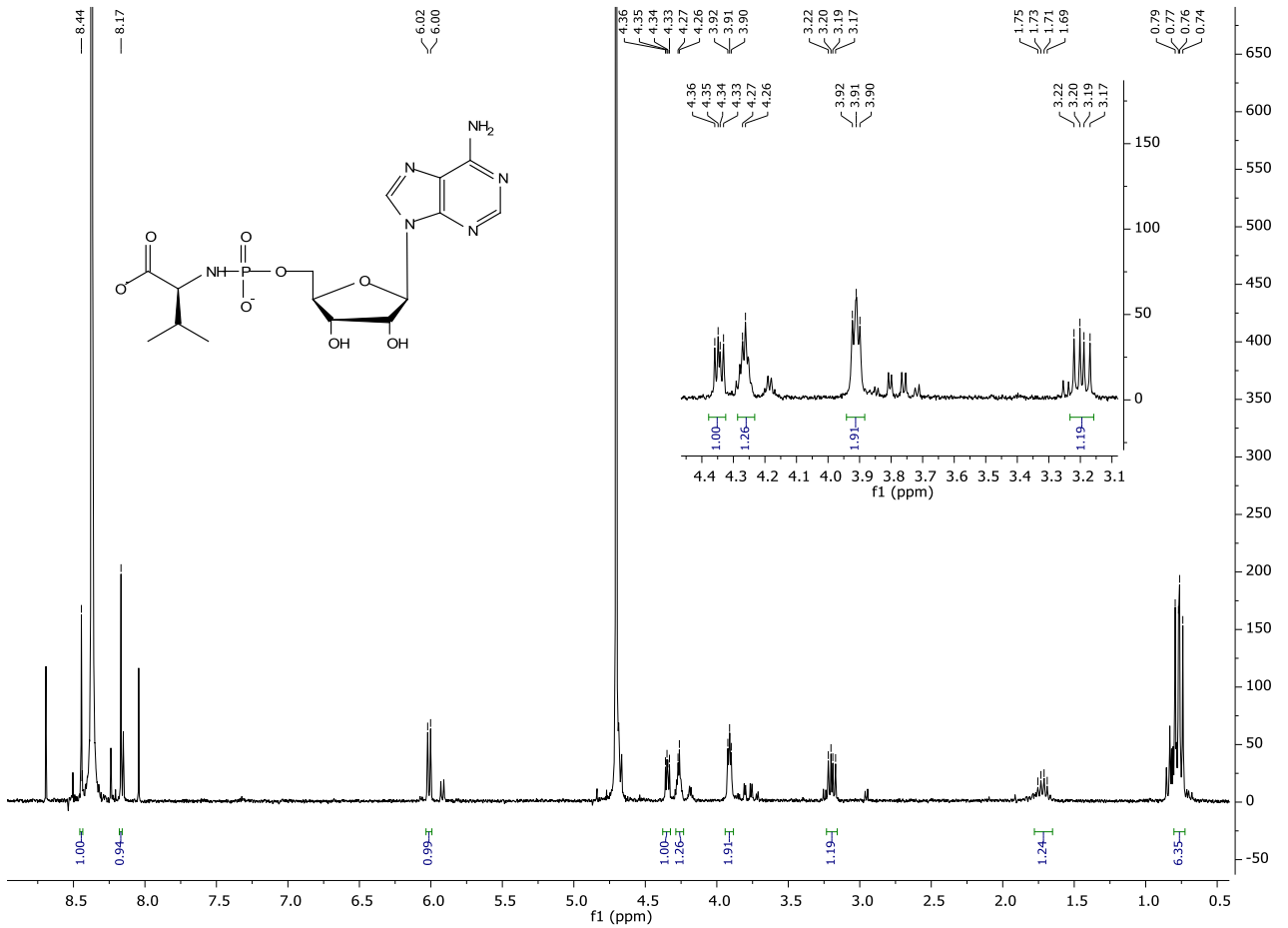
Source Type	ESI	Ion Polarity	Positive	Set Nebulizer	0.3 Bar
Focus	Active	Set Capillary	4500 V	Set Dry Heater	200 °C
Scan Begin	50 m/z	Set End Plate Offset	-500 V	Set Dry Gas	4.0 l/min
Scan End	1000 m/z	Set Collision Cell RF	750.0 Vpp	Set Divert Valve	Source



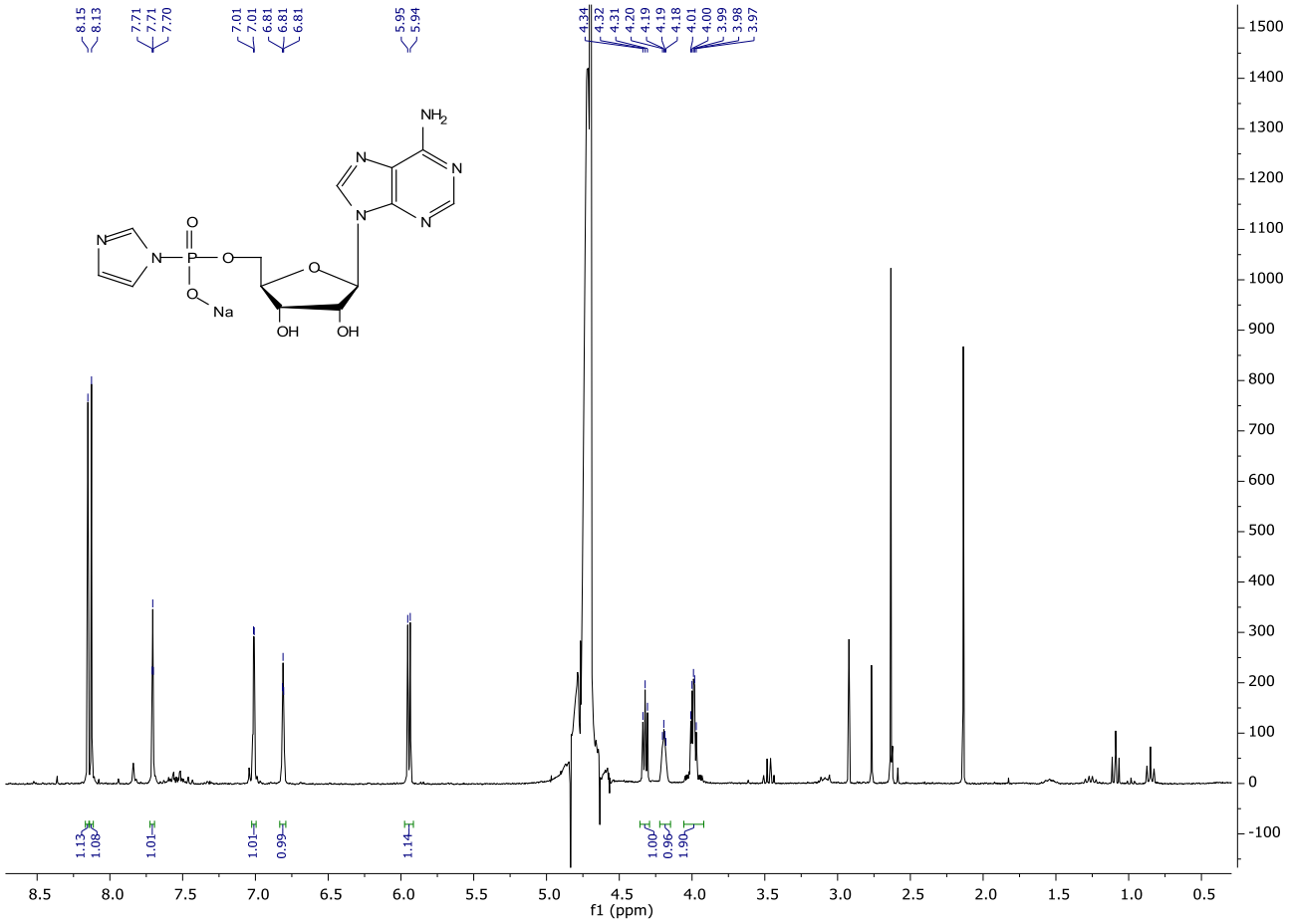
HRMS (+MS) of Gly_2A. The fragments visible at 136 (adenine) and 317 (dehydrated AMP) are typical from peptido-AMP.



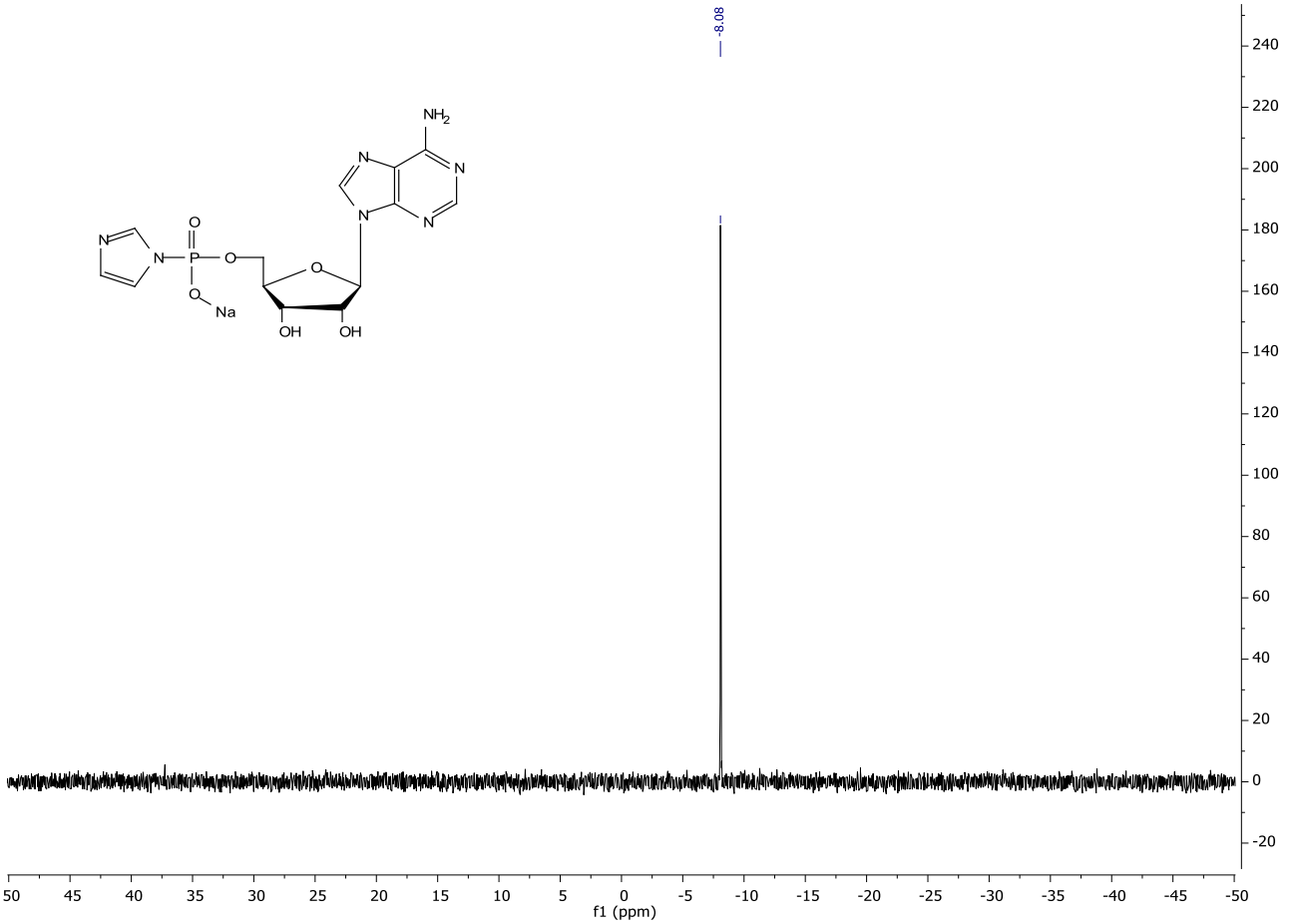
^{1}H NMR spectrum (D_2O, 300 MHz) of impure BnOValpA obtained by the benzotriazolide displacement method.



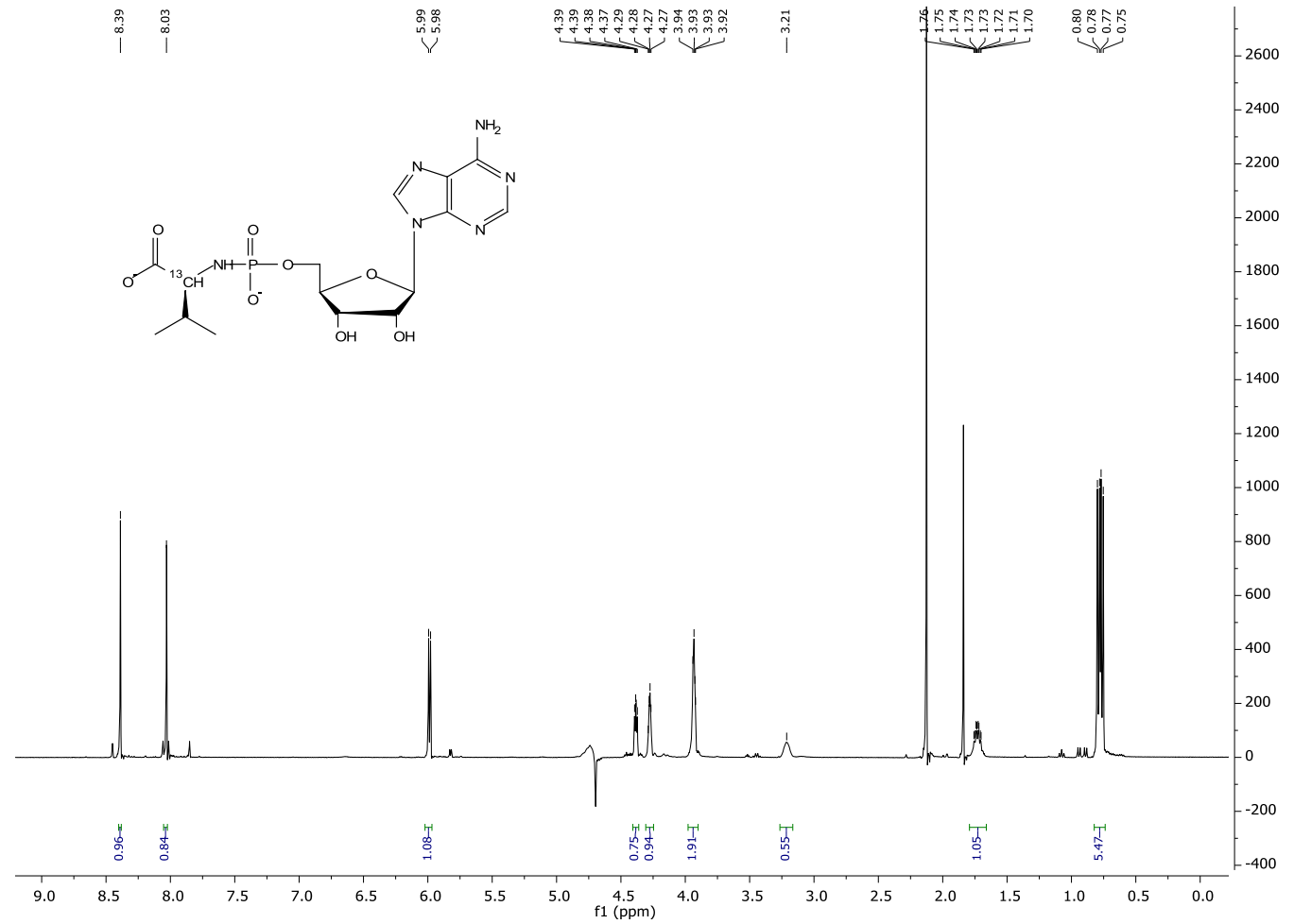
^{1}H NMR spectrum (D_2O, 300 MHz) of impure ValpA obtained by deprotection of BnOValpA.



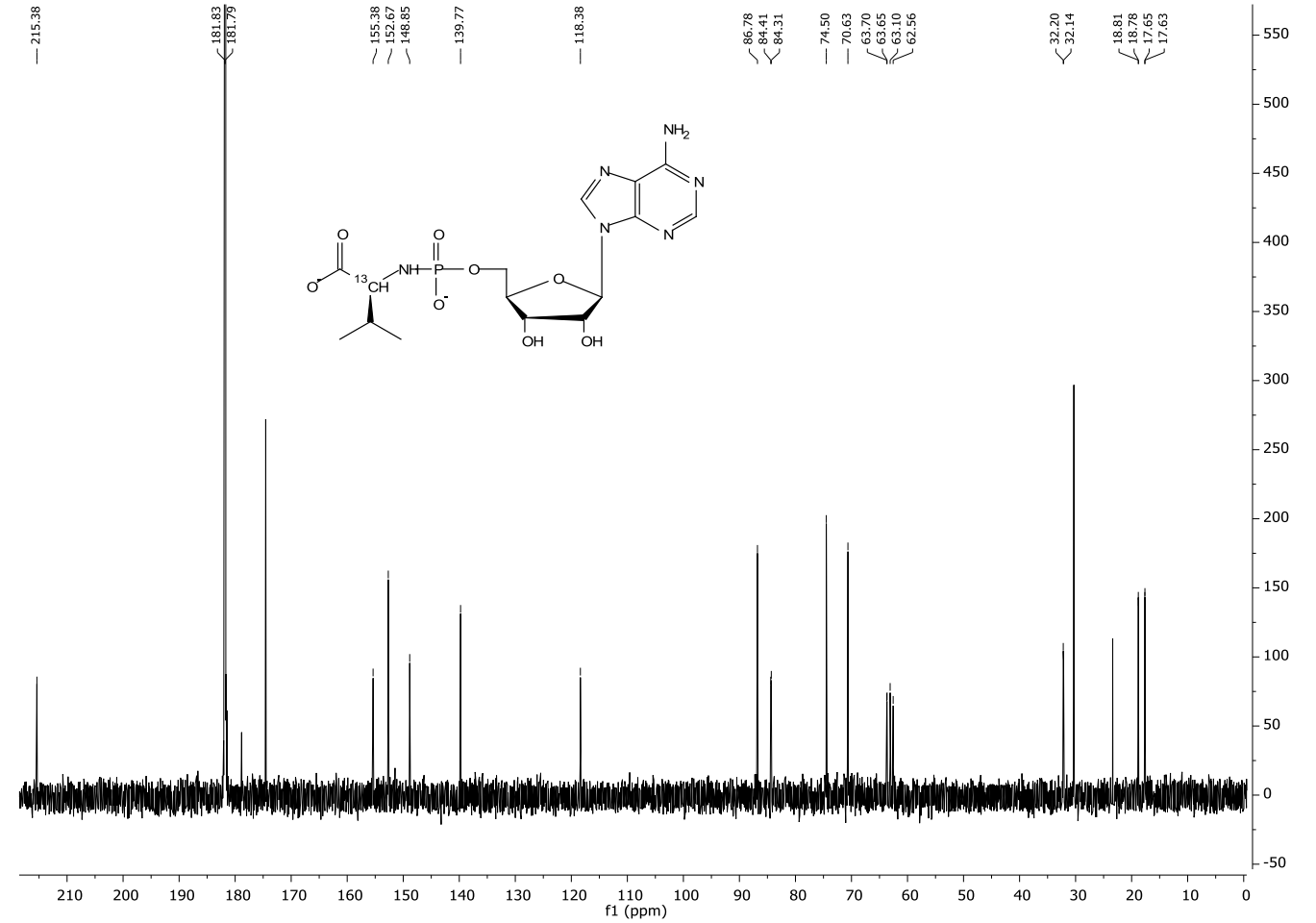
^{1}H NMR spectrum (D_2O, 300 MHz) of ImpA. Solvents from the synthesis that were not removed under high vacuum are visible.



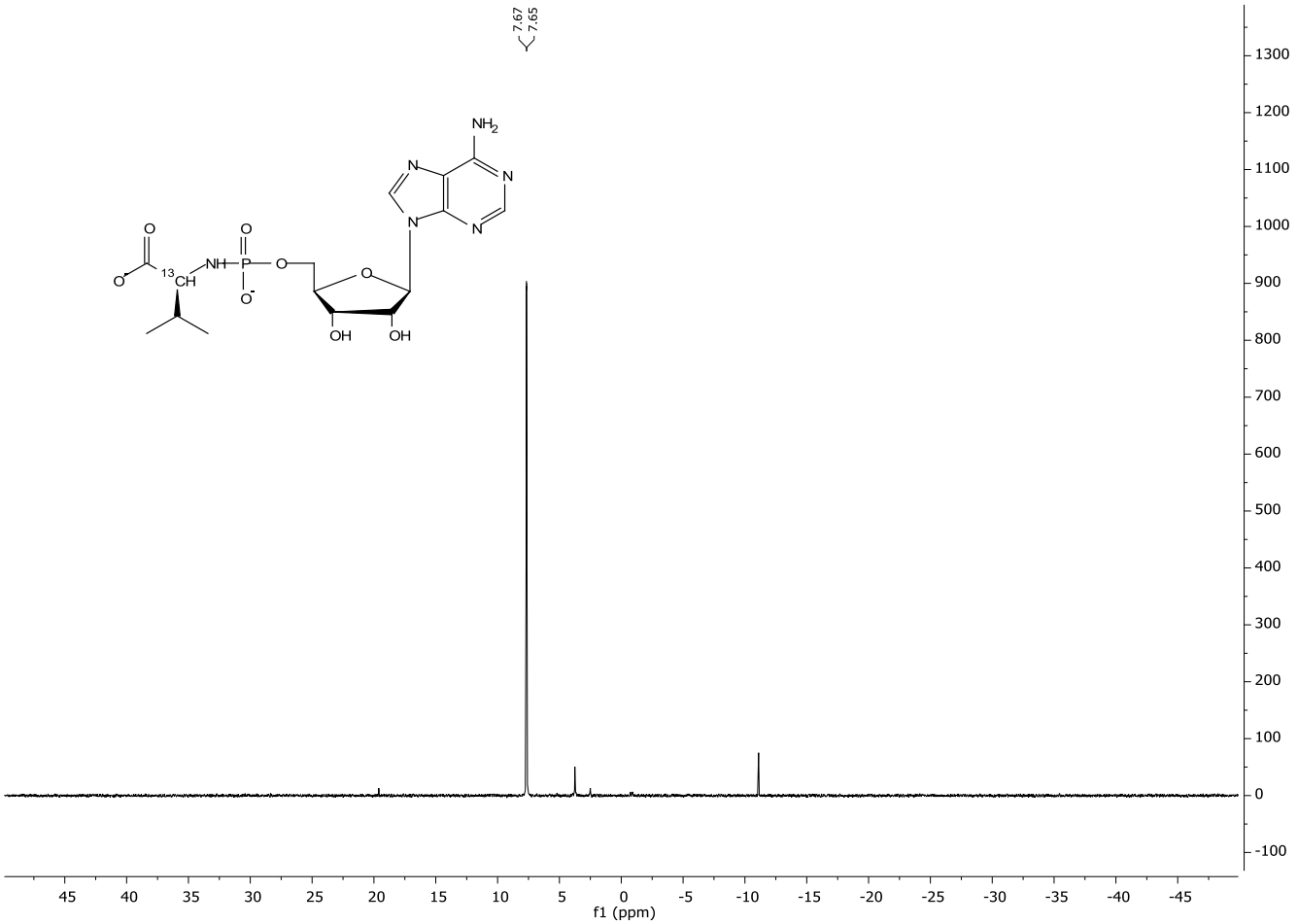
^{31}P NMR spectrum (H_2O, 125 MHz) of ImpA.



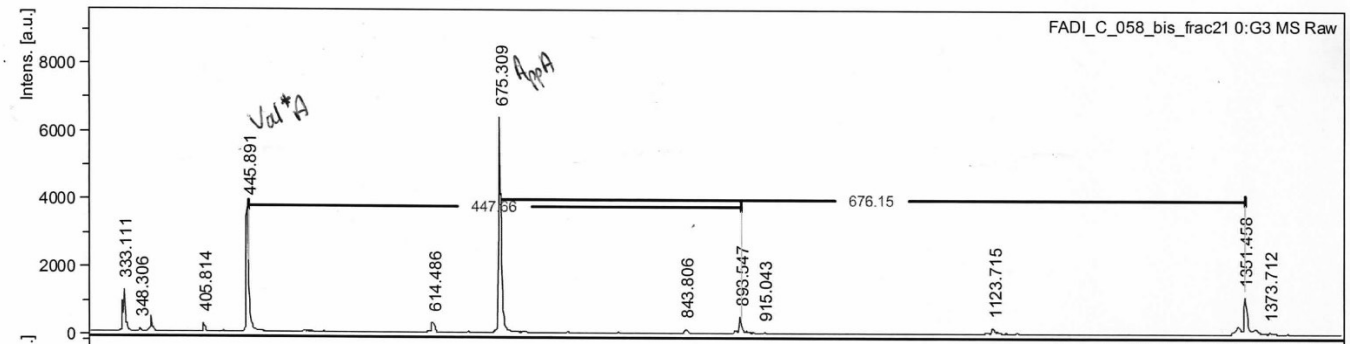
*^{1}H NMR spectrum (H_2O, 400 MHz) of Val*A. The presaturation sequence impacted the integration of the neighboring peaks.*



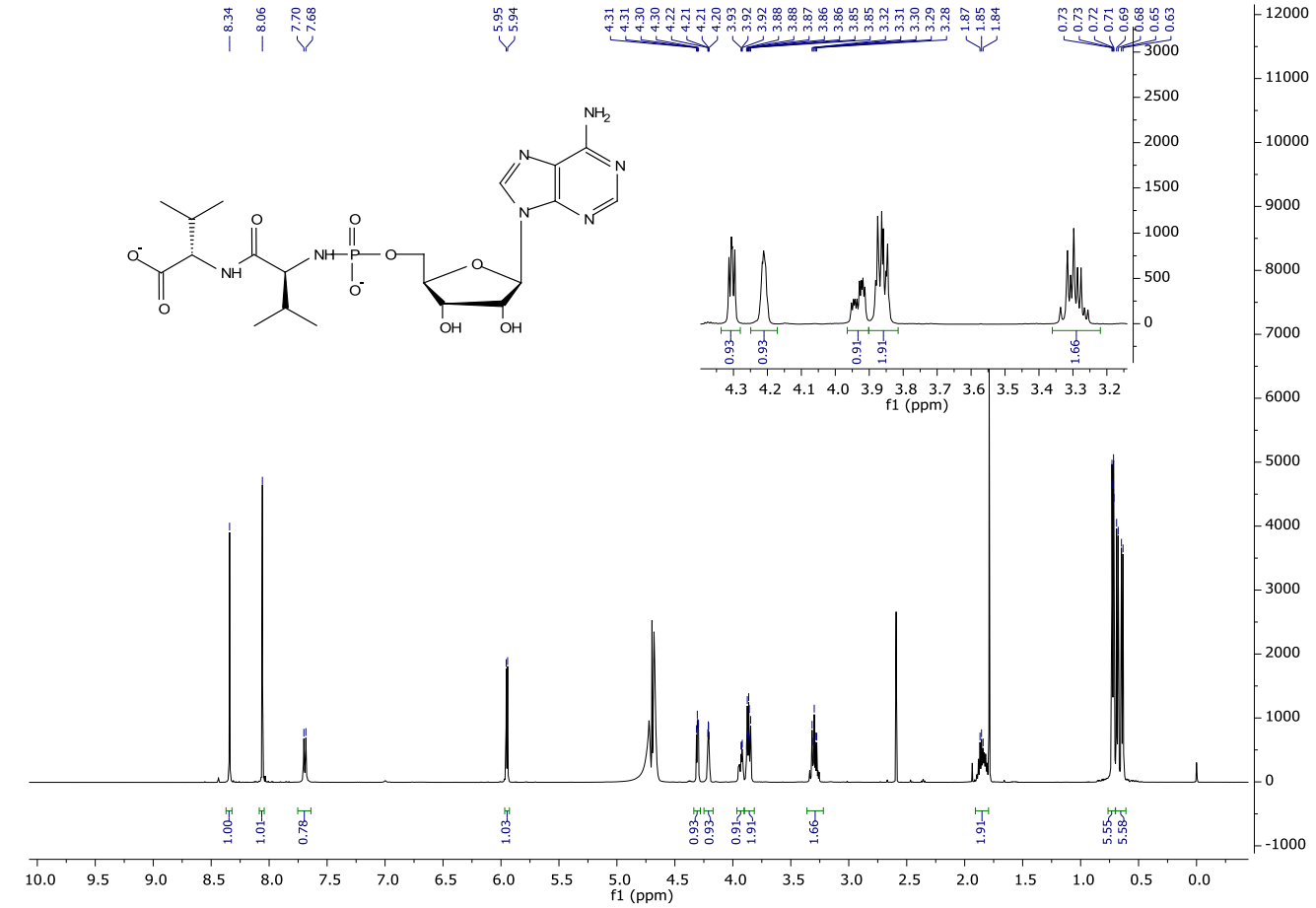
*^{13}C NMR spectrum (H_2O, 100 MHz) of Val*A.*



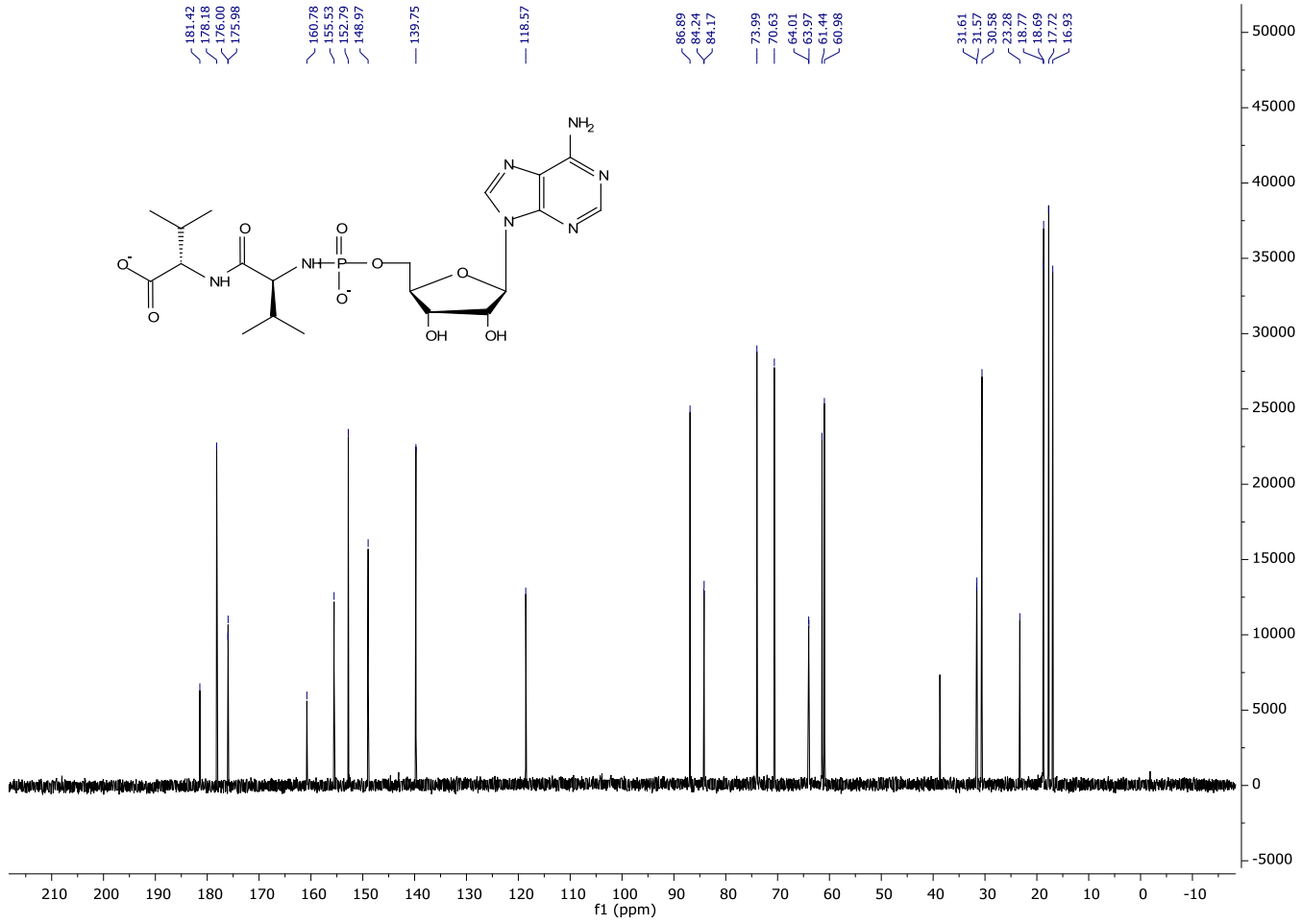
*^{31}P NMR spectrum (H_2O, 160 MHz) of Val*A.*



*MALDI-MS (THAP/citrate matrix, -MS linear mode, 0.5-2 kDa method) spectrum of Val*A. Although present in ca. 5%, AppA responds much better than ValA.*

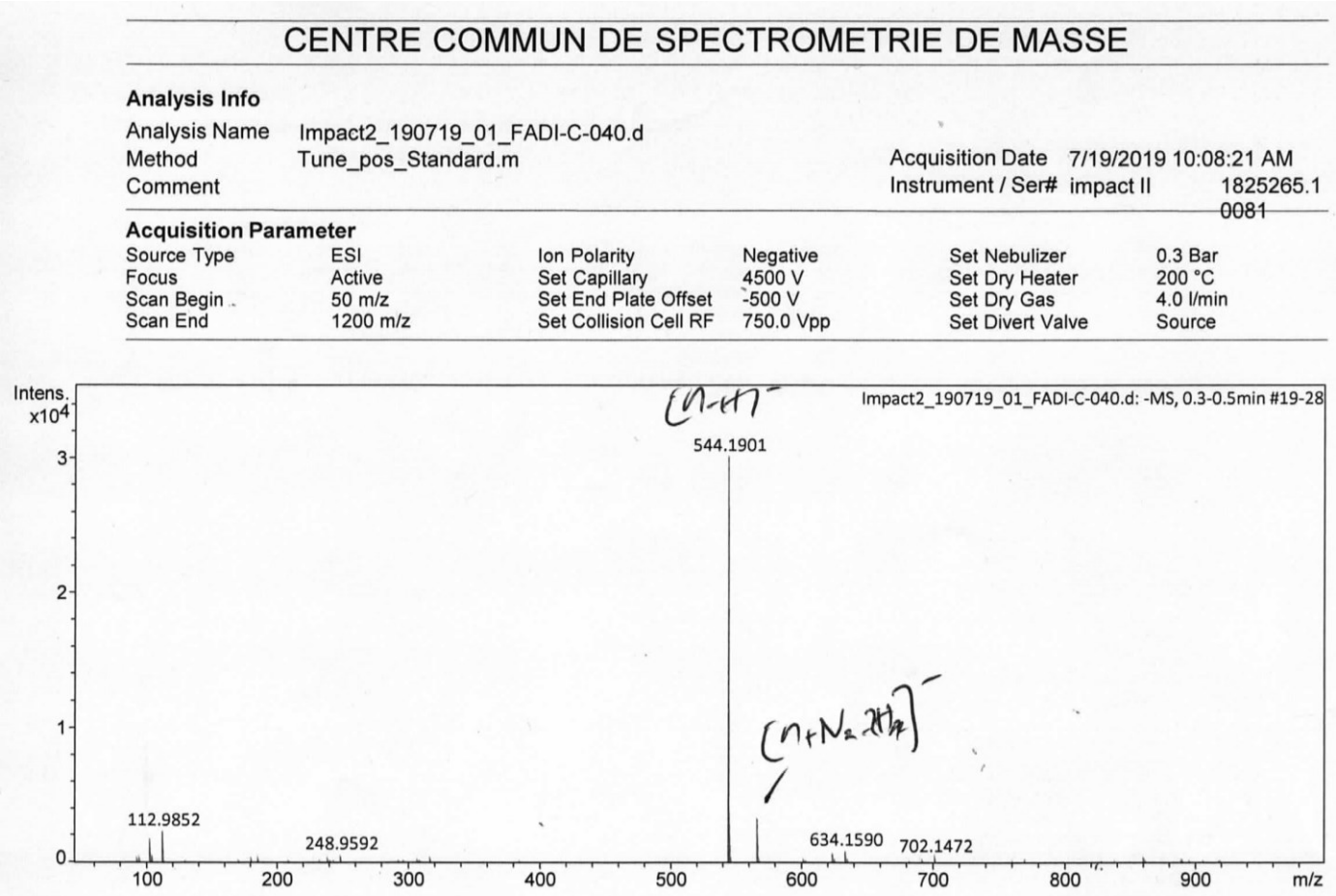


[1]H NMR spectrum (H_2O, 500 MHz) of Val_2A

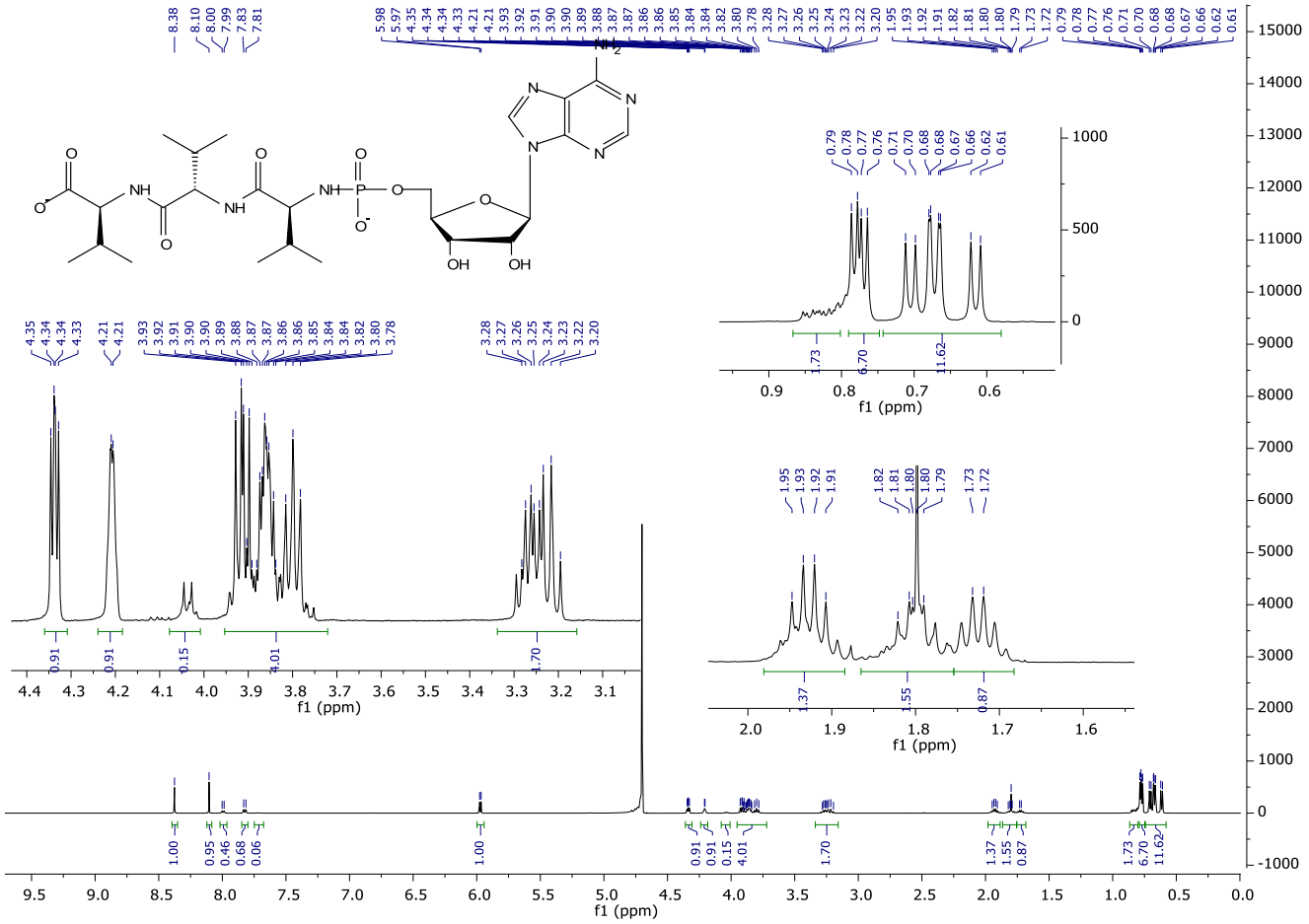


[13]C NMR spectrum (H_2O, 125 MHz) of Val_2A

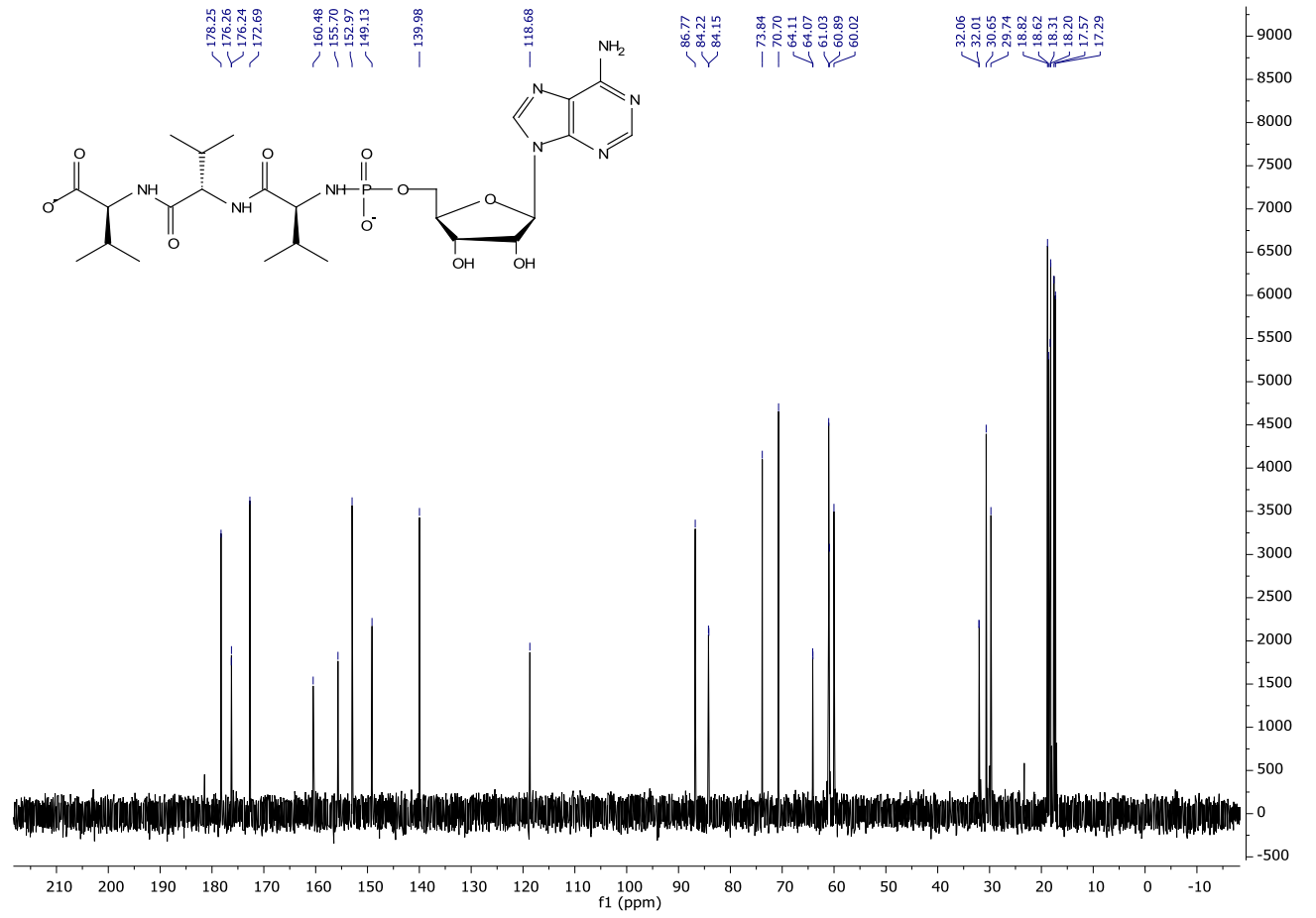
^{31}P NMR spectrum (H_2O, 200 MHz) of Val_2A.



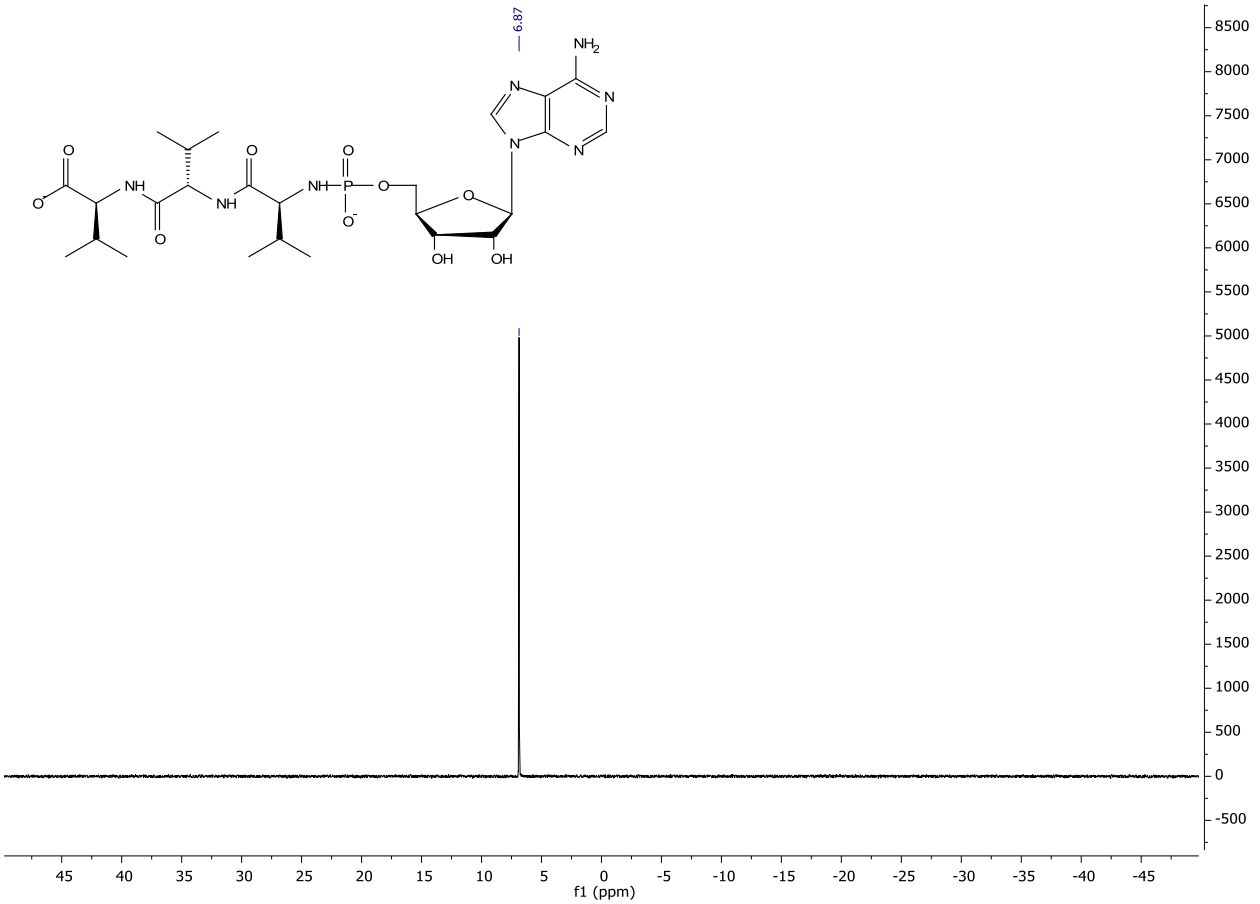
HRMS (-MS) of Val_2A.



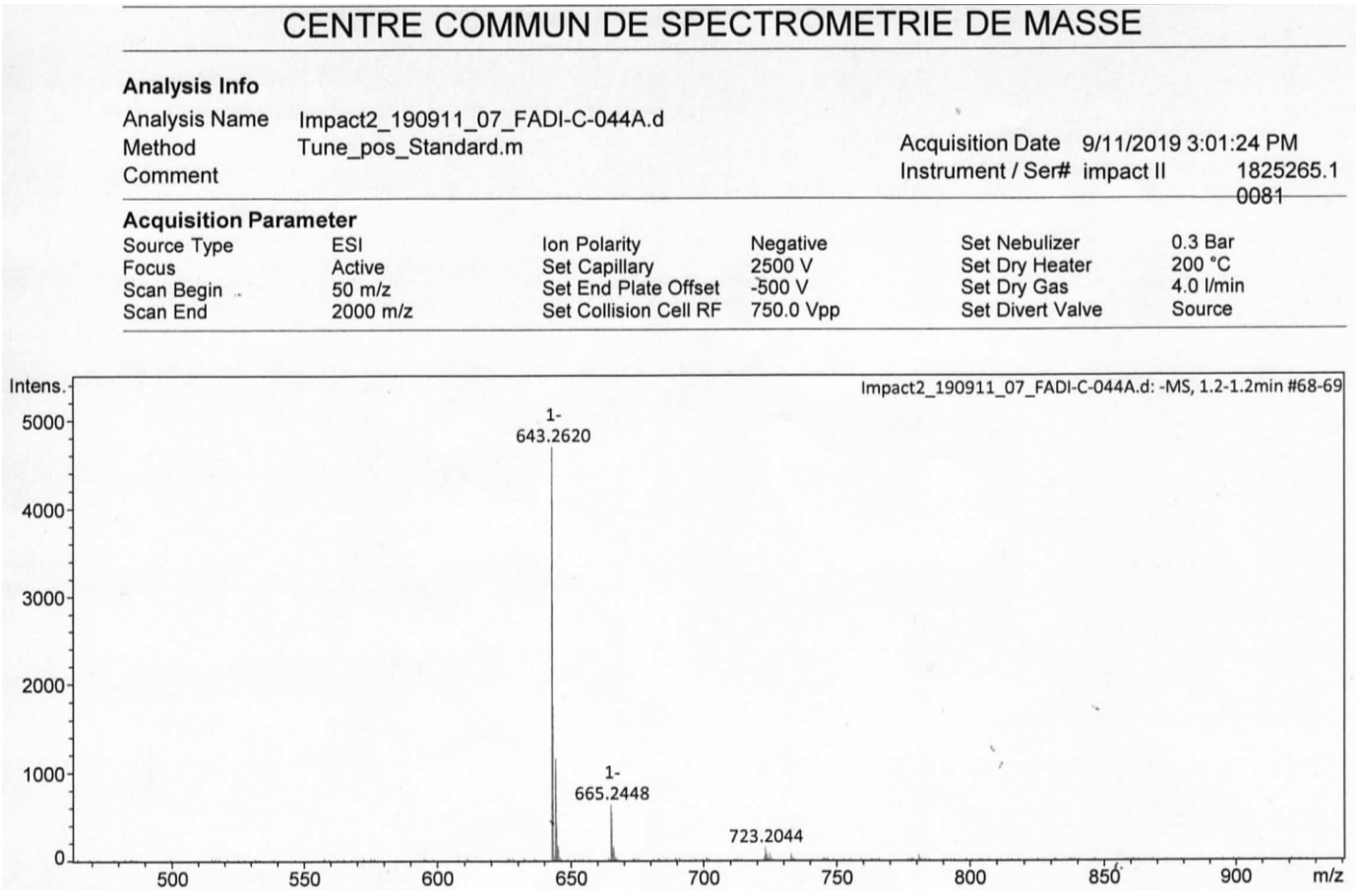
^{1}H NMR spectrum (H_2O, 500 MHz) of Val_3A



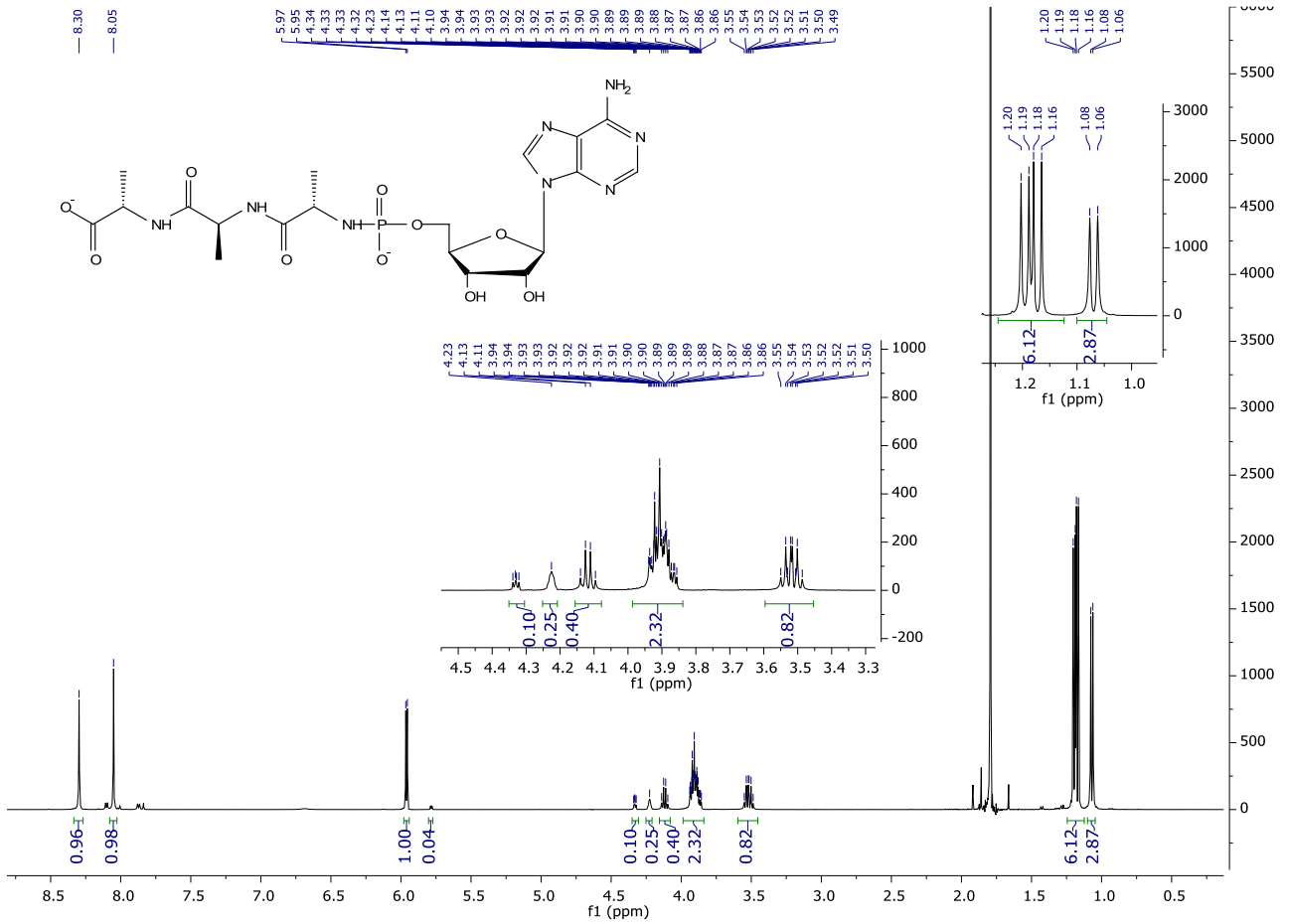
^{13}C NMR spectrum (H_2O, 125 MHz) of Val_3A



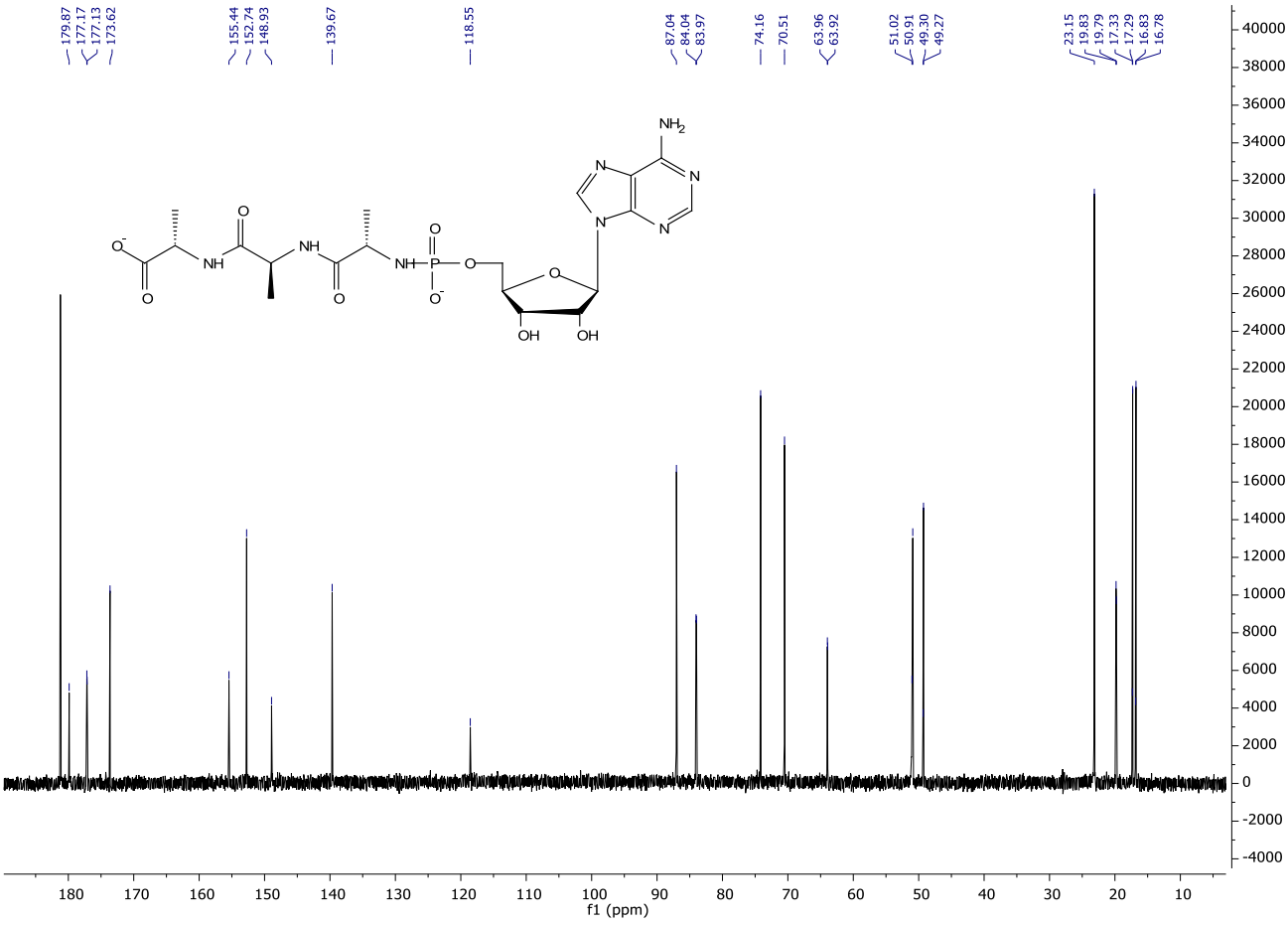
^{31}P NMR spectrum (H_2O, 200 MHz) of Val_3A.



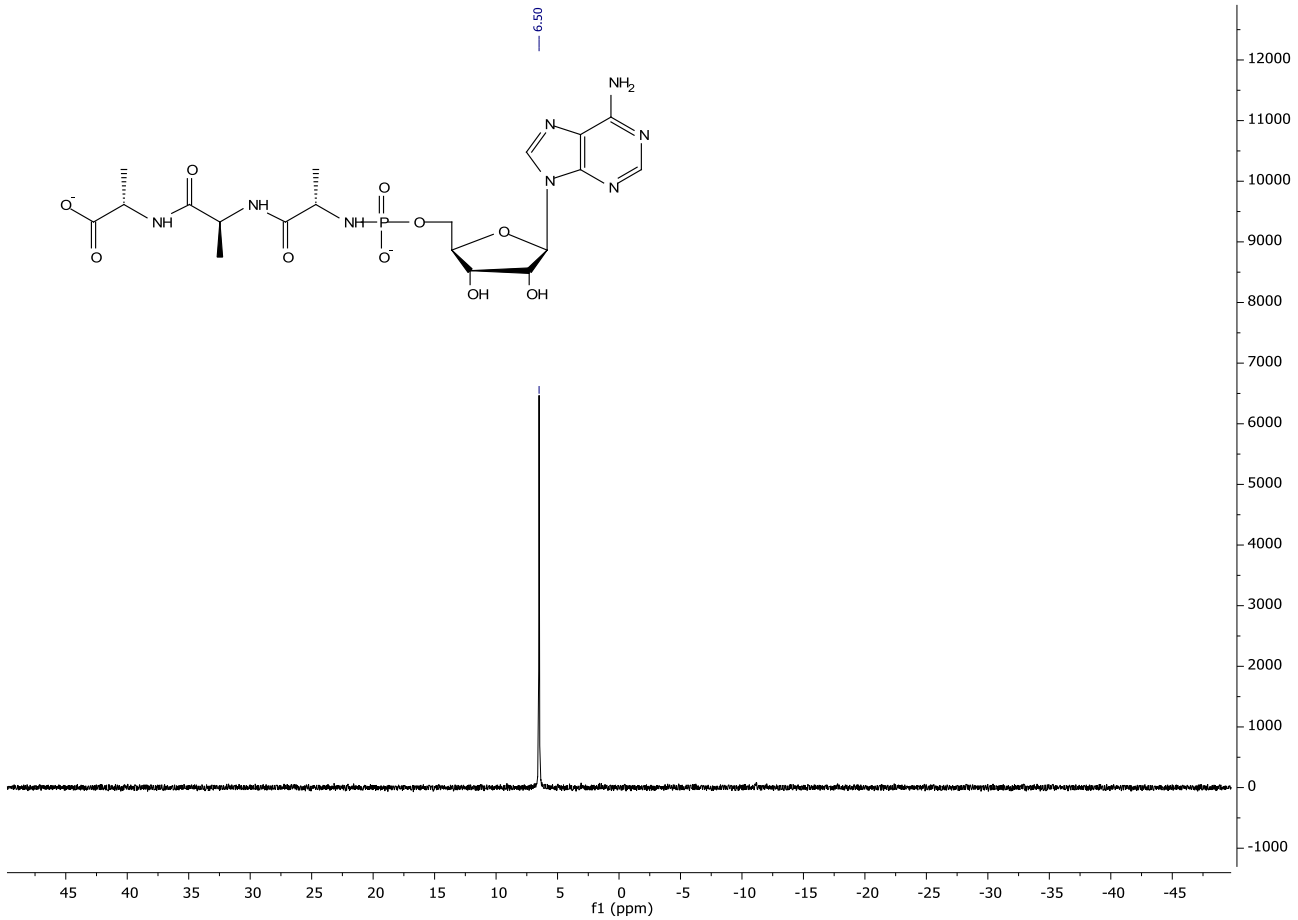
HRMS (-MS) of Val_3A



^{1}H NMR spectrum (D_2O, 500 MHz) of Ala_3A



^{13}C NMR spectrum (D_2O, 125 MHz) of Ala_3A



^{31}P NMR spectrum (D_2O, 200 MHz) of Ala_3A

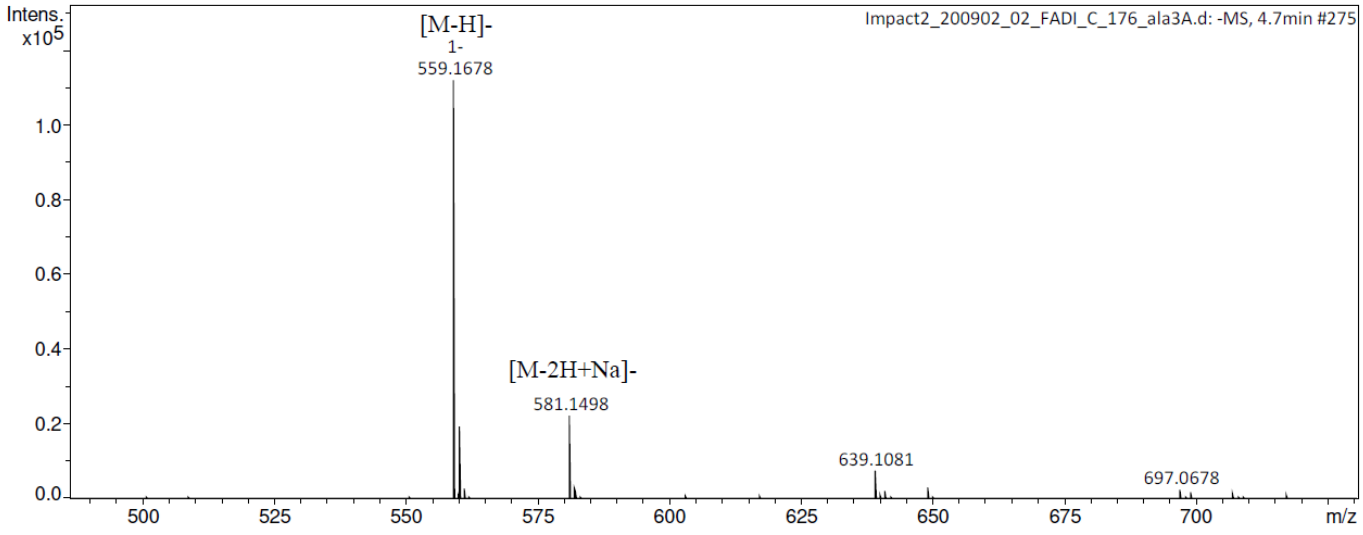
CENTRE COMMUN DE SPECTROMETRIE DE MASSE

Analysis Info

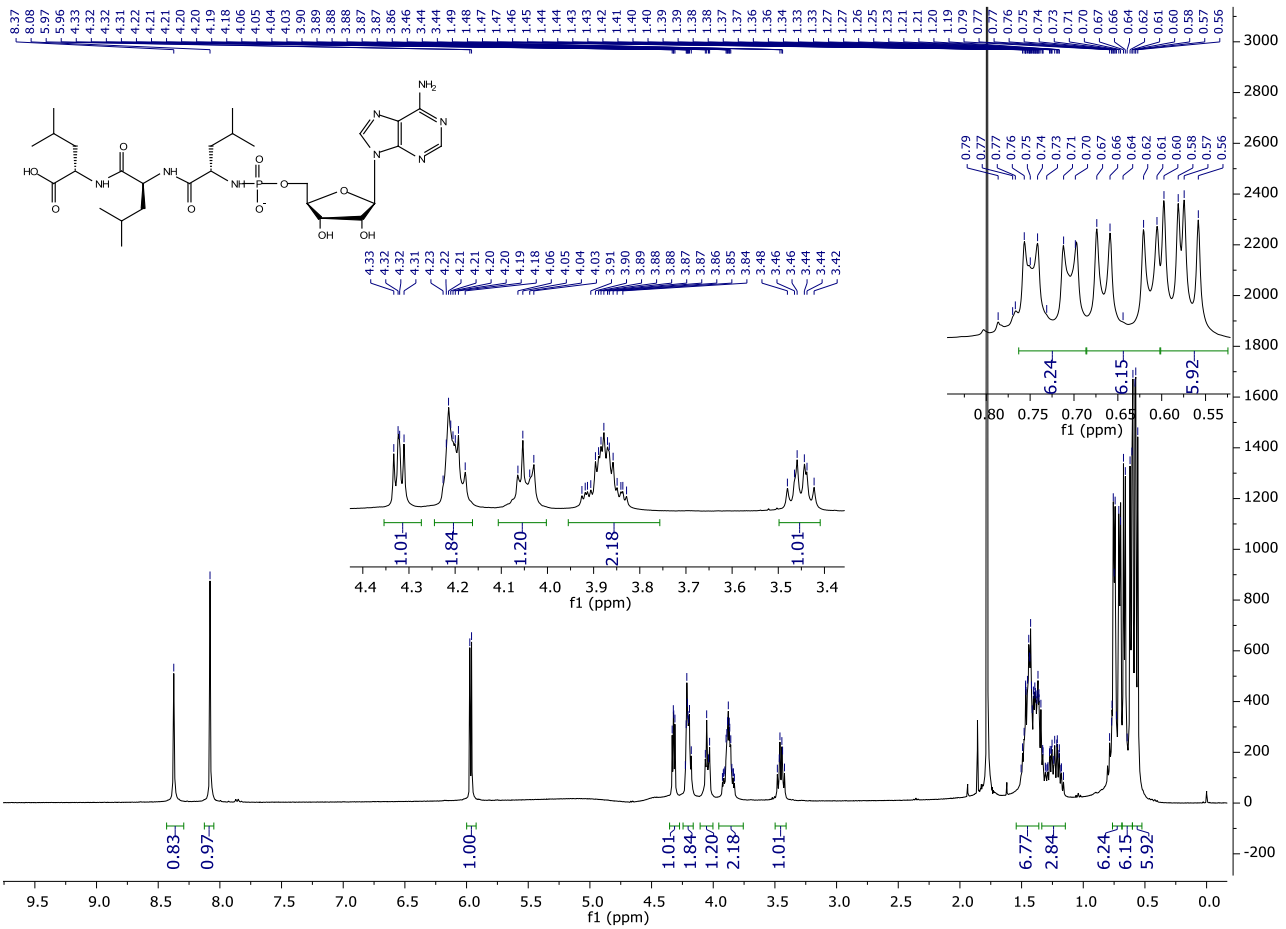
Analysis Name	Impact2_200902_02_FADI_C_176_ala3A.d		
Method	Tune_pos_Standard.m	Acquisition Date	9/2/2020 3:02:31 PM
Comment		Instrument / Ser#	impact II 1825265.10081

Acquisition Parameter

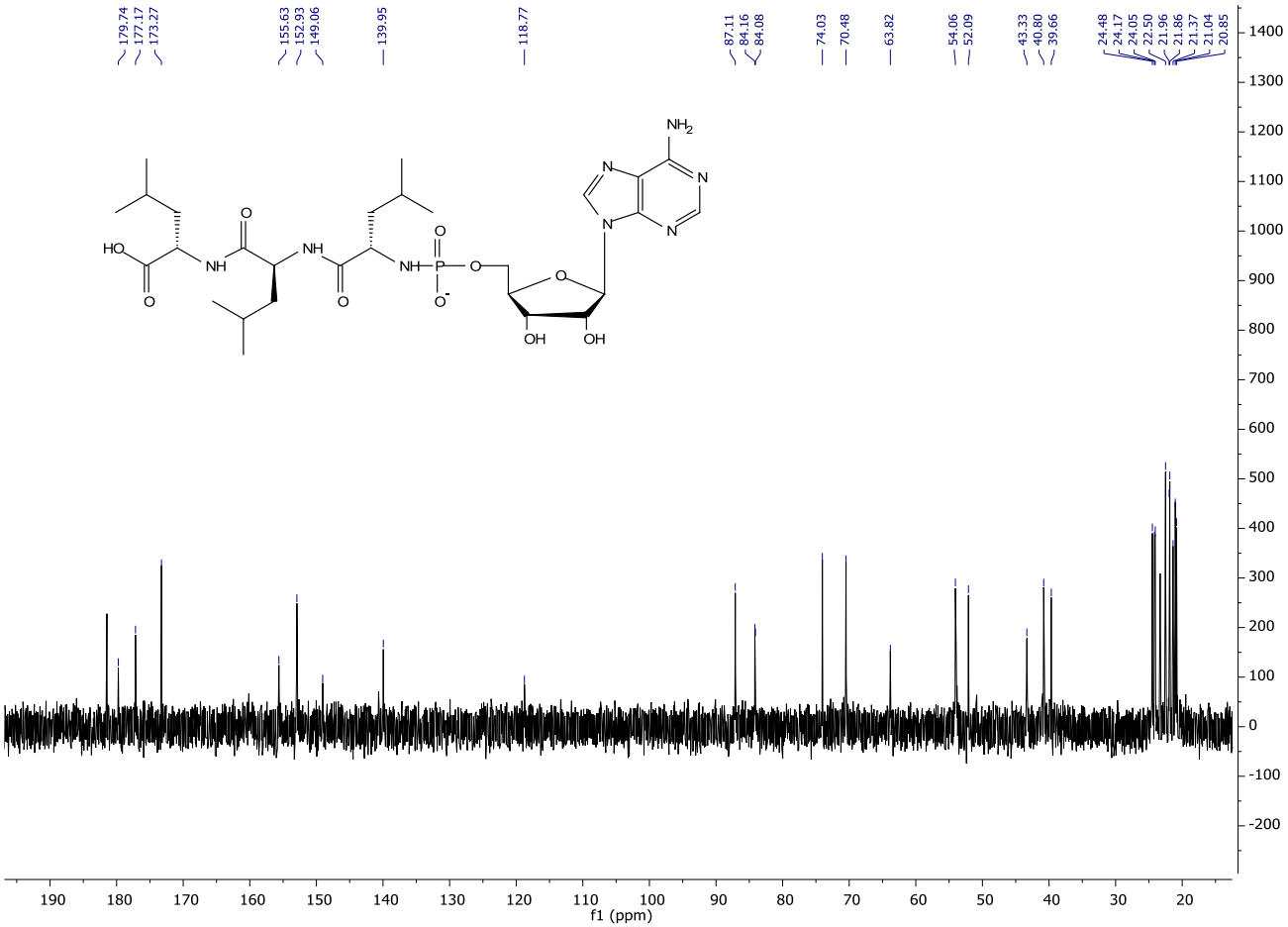
Source Type	ESI	Ion Polarity	Negative	Set Nebulizer	0.3 Bar
Focus	Active	Set Capillary	2500 V	Set Dry Heater	200 °C
Scan Begin	50 m/z	Set End Plate Offset	-500 V	Set Dry Gas	4.0 l/min
Scan End	1200 m/z	Set Collision Cell RF	750.0 Vpp	Set Divert Valve	Source



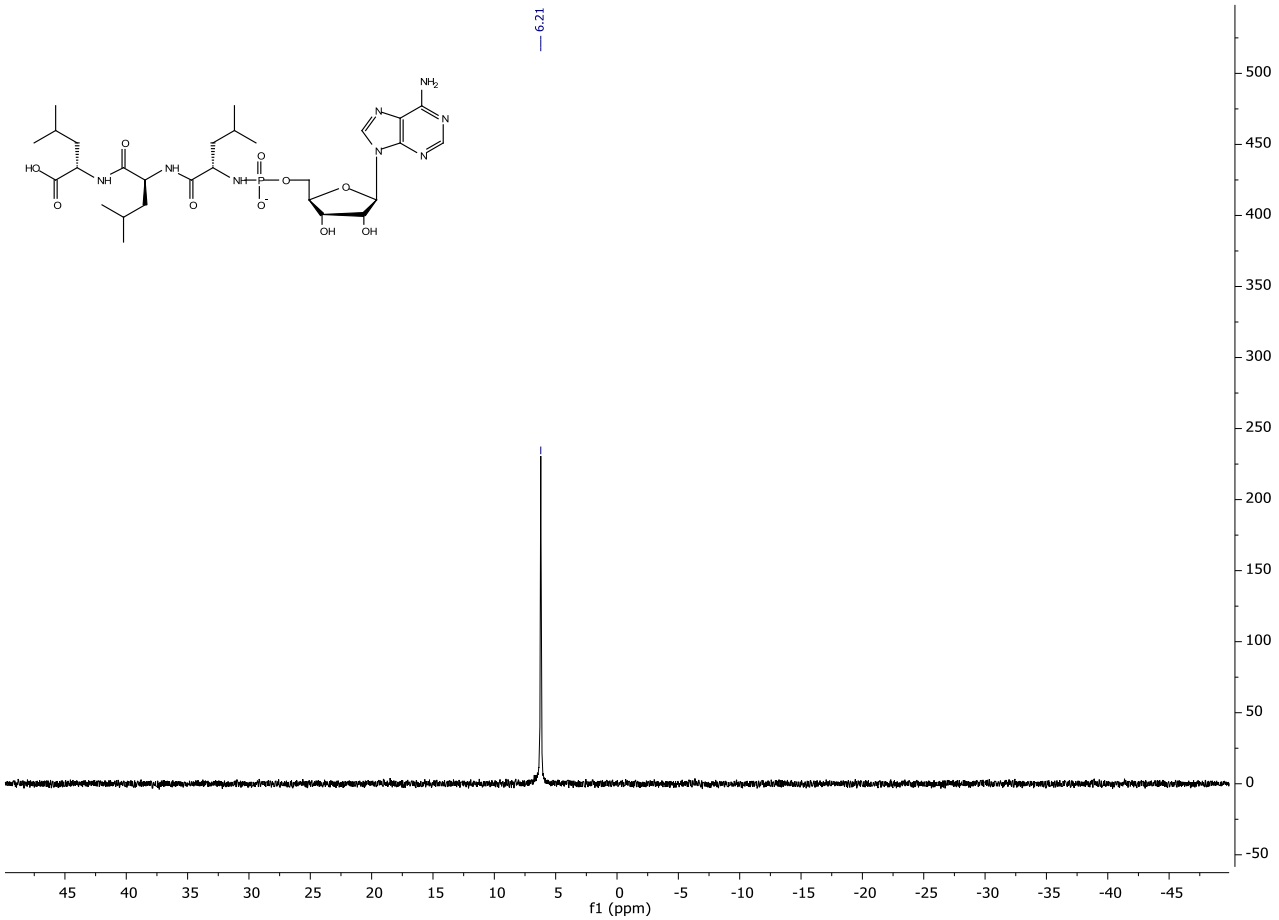
HRMS (-MS) of Ala_3A



^{1}H NMR spectrum (D_2O, 400 MHz) of Leu_3A



^{13}C NMR spectrum (D_2O, 100 MHz) of Leu_3A



^{31}P NMR spectrum (D_2O, 160 MHz) of Leu_3A.

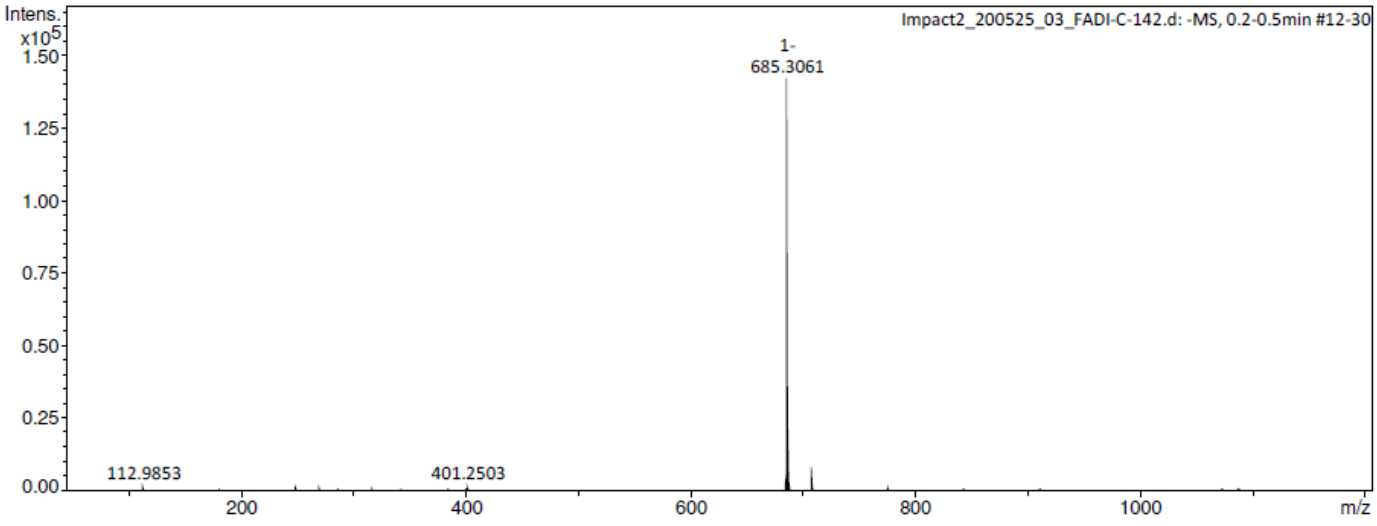
CENTRE COMMUN DE SPECTROMETRIE DE MASSE

Analysis Info

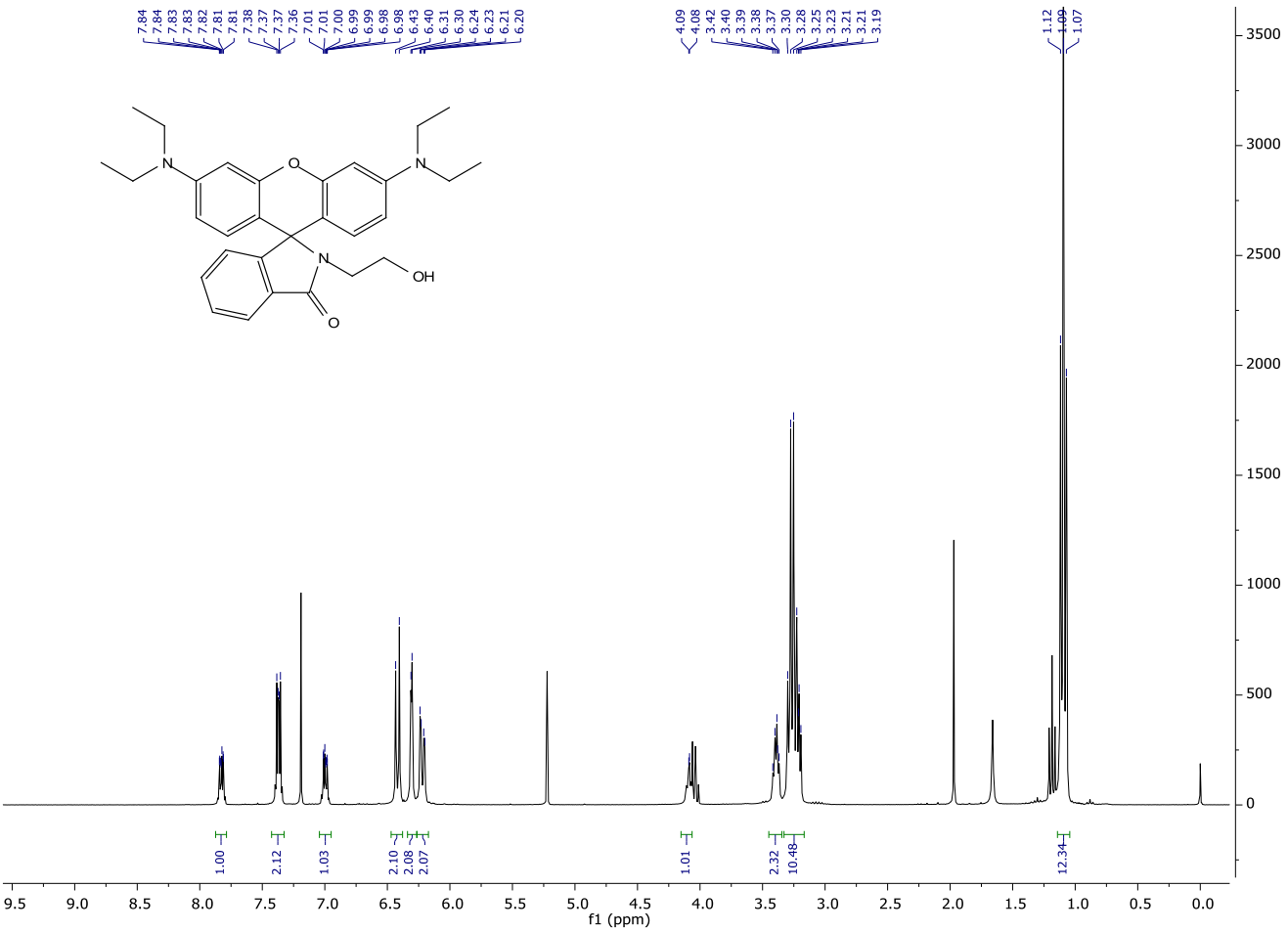
Analysis Name	Impact2_200525_03_FADI-C-142.d		
Method	Tune_pos_Standard.m	Acquisition Date	5/25/2020 10:56:04 AM
Comment		Instrument / Ser#	impact II 1825265.1 0081

Acquisition Parameter

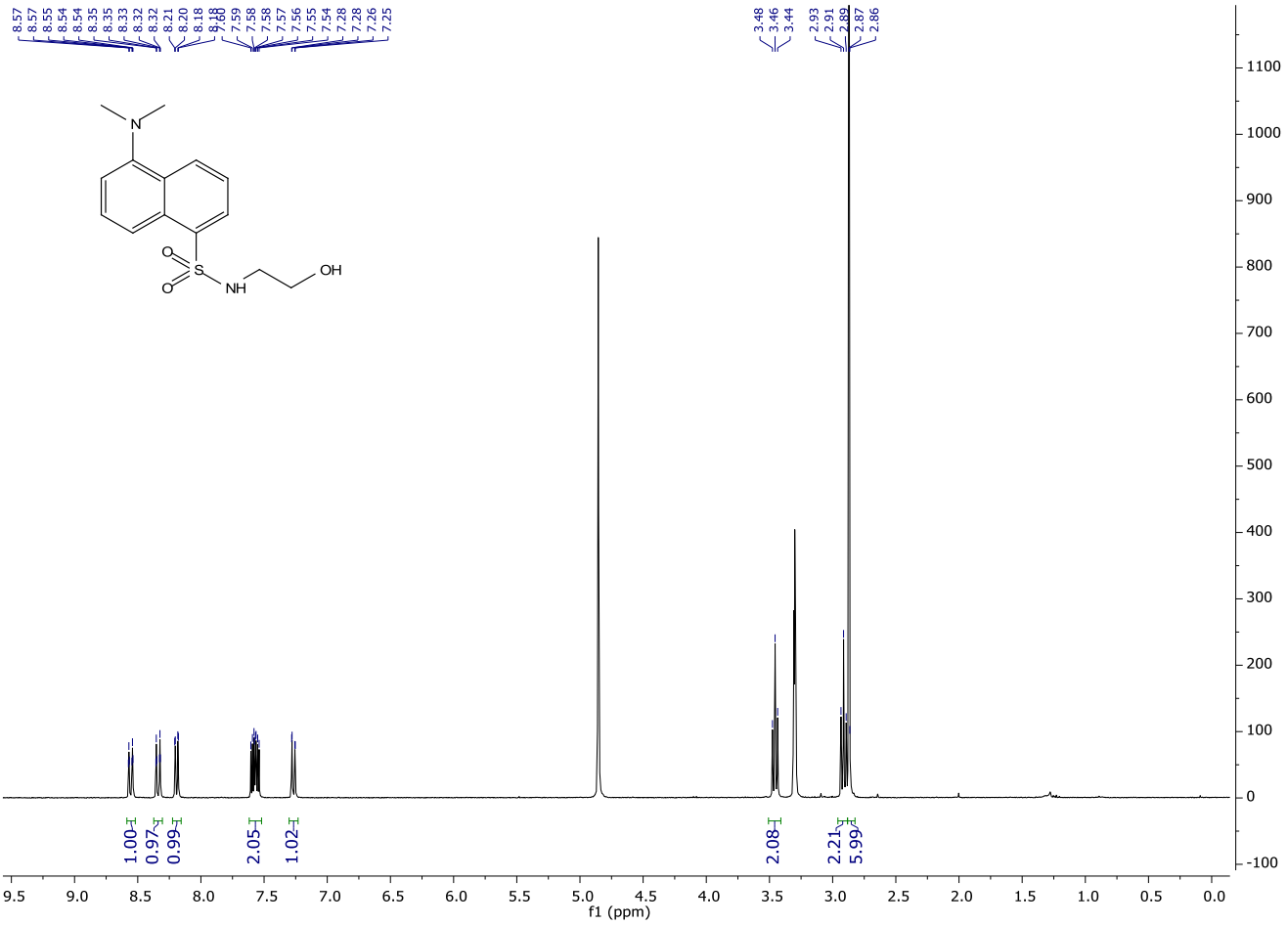
Source Type	ESI	Ion Polarity	Negative	Set Nebulizer	0.3 Bar
Focus	Active	Set Capillary	2500 V	Set Dry Heater	200 °C
Scan Begin	50 m/z	Set End Plate Offset	-500 V	Set Dry Gas	4.0 l/min
Scan End	1200 m/z	Set Collision Cell RF	750.0 Vpp	Set Divert Valve	Source



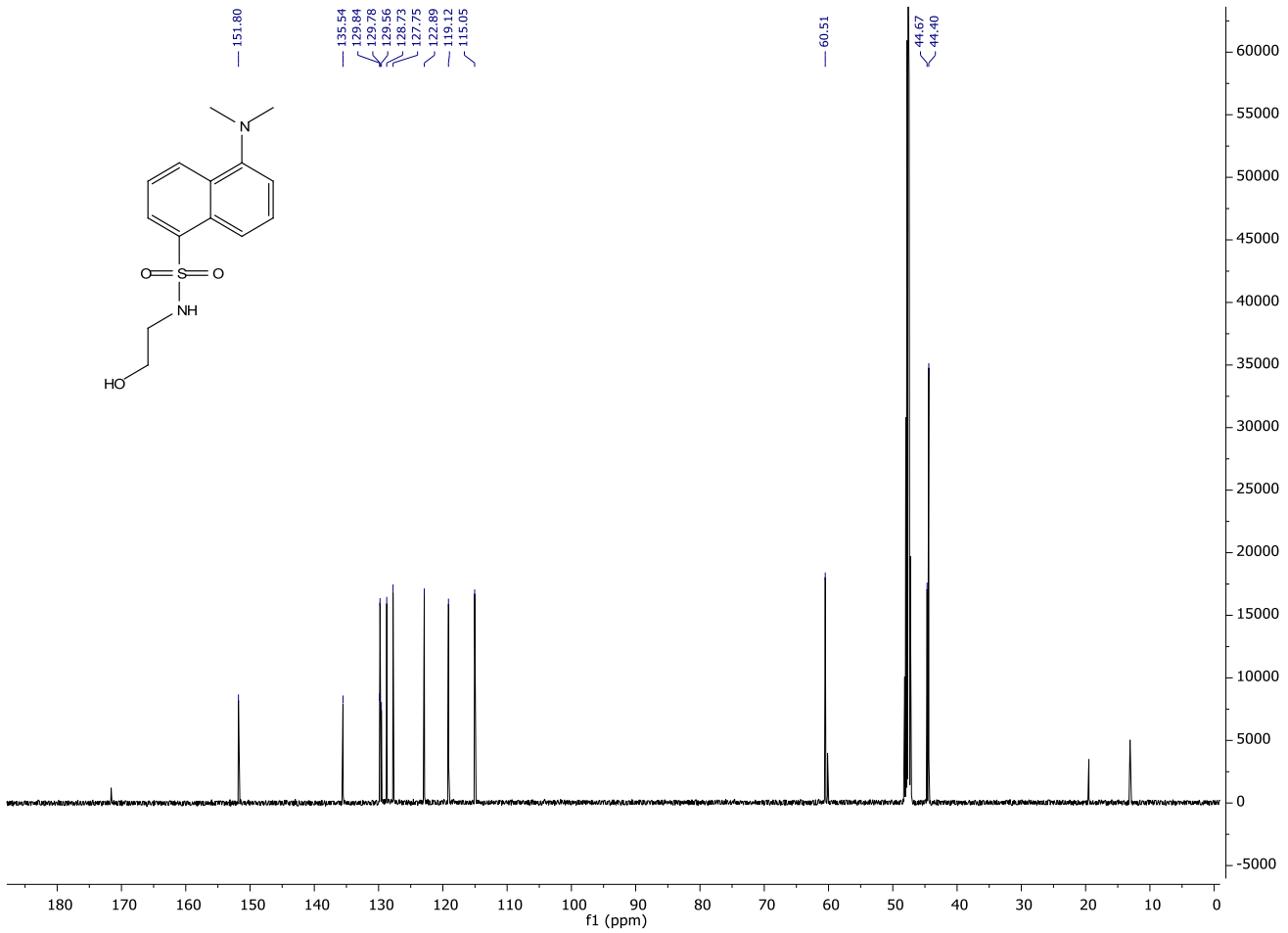
HRMS (-MS) of Leu_3A



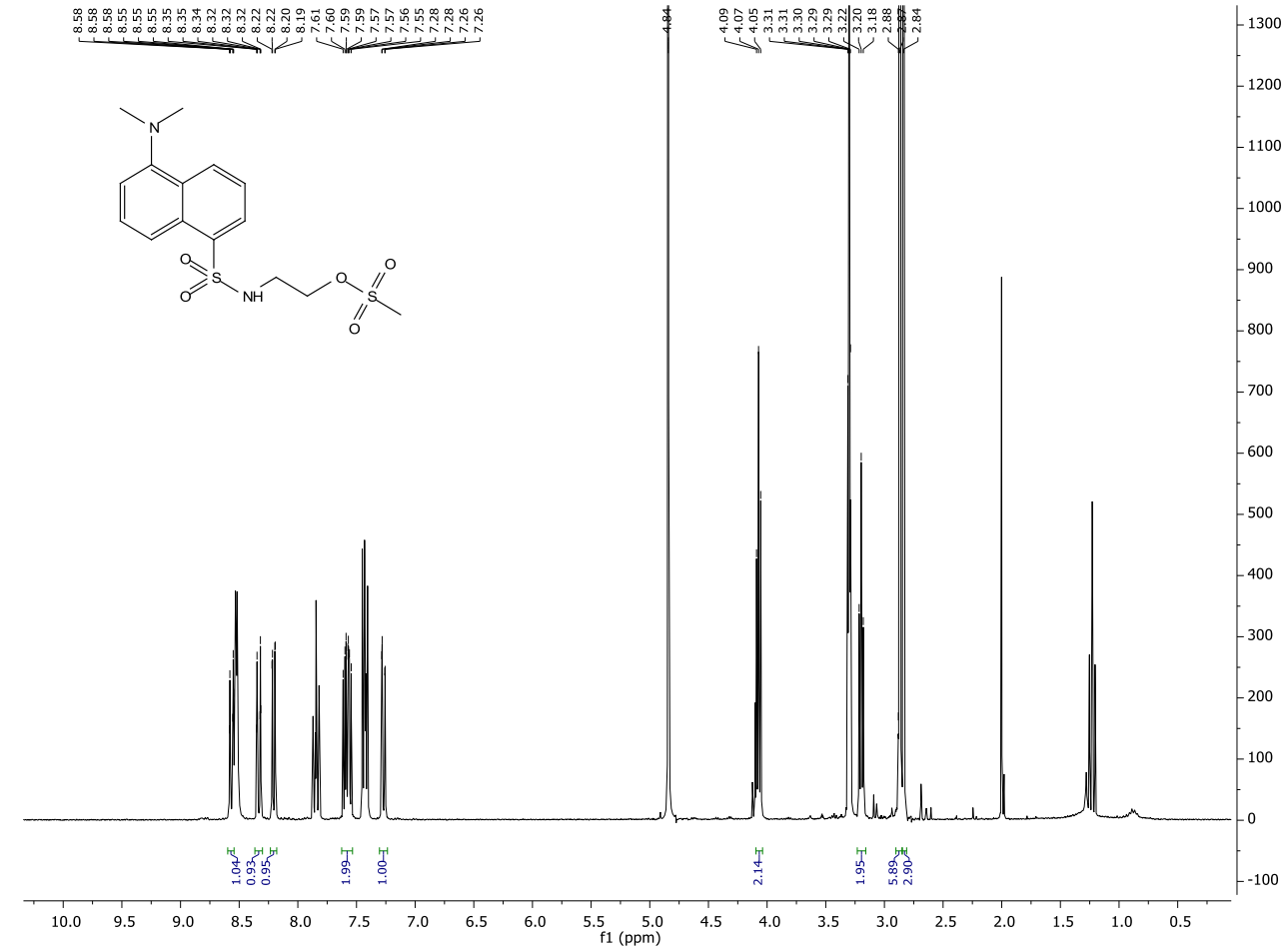
^{1}H NMR spectrum ($CDCl_3$, 300 MHz) of known acylated rhodamine



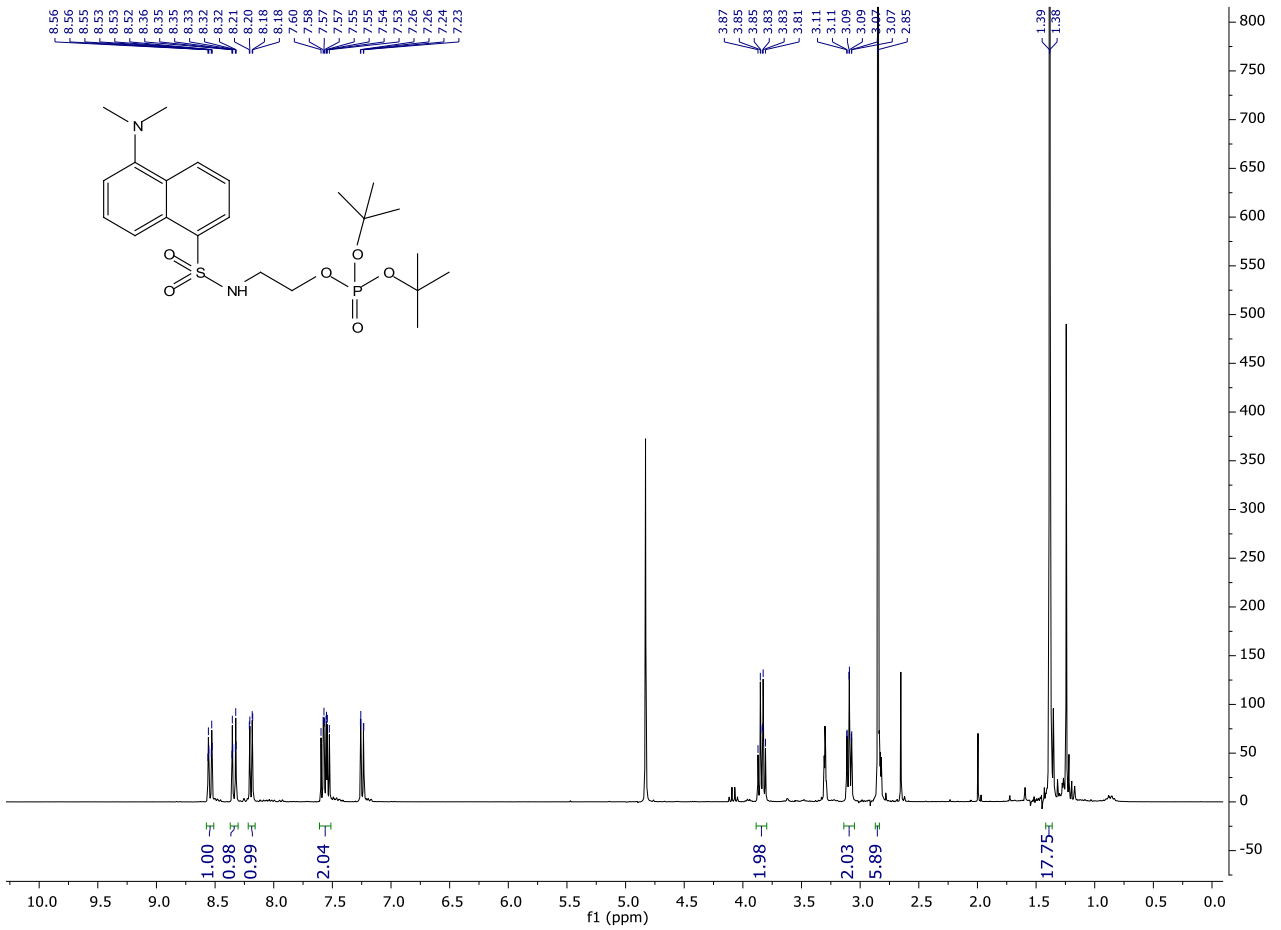
*^{1}H NMR spectrum (MeOD, 500 MHz) of **(29)***



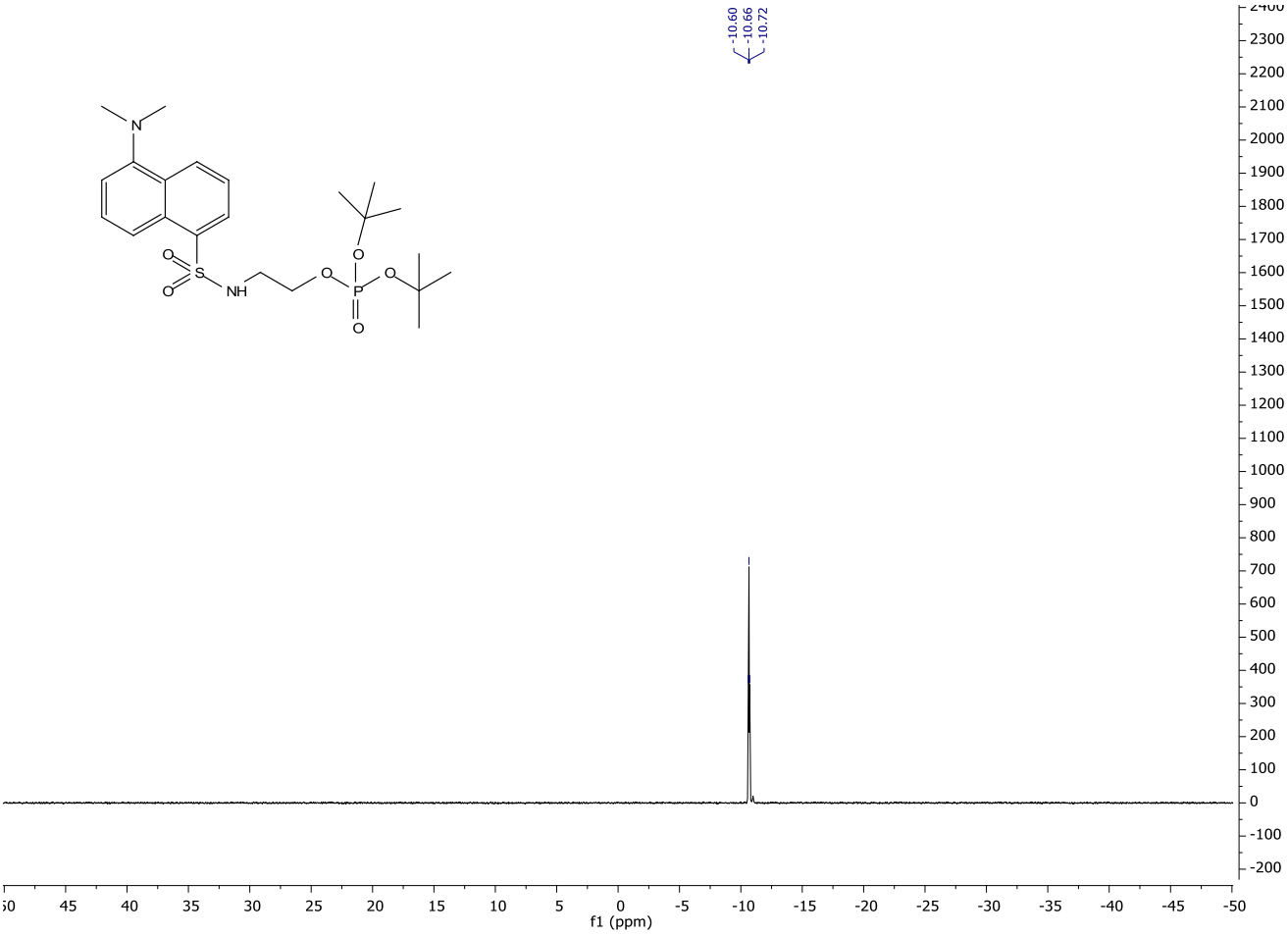
^{13}C NMR spectrum (MeOD, 125 MHz) of compound ***(29)***



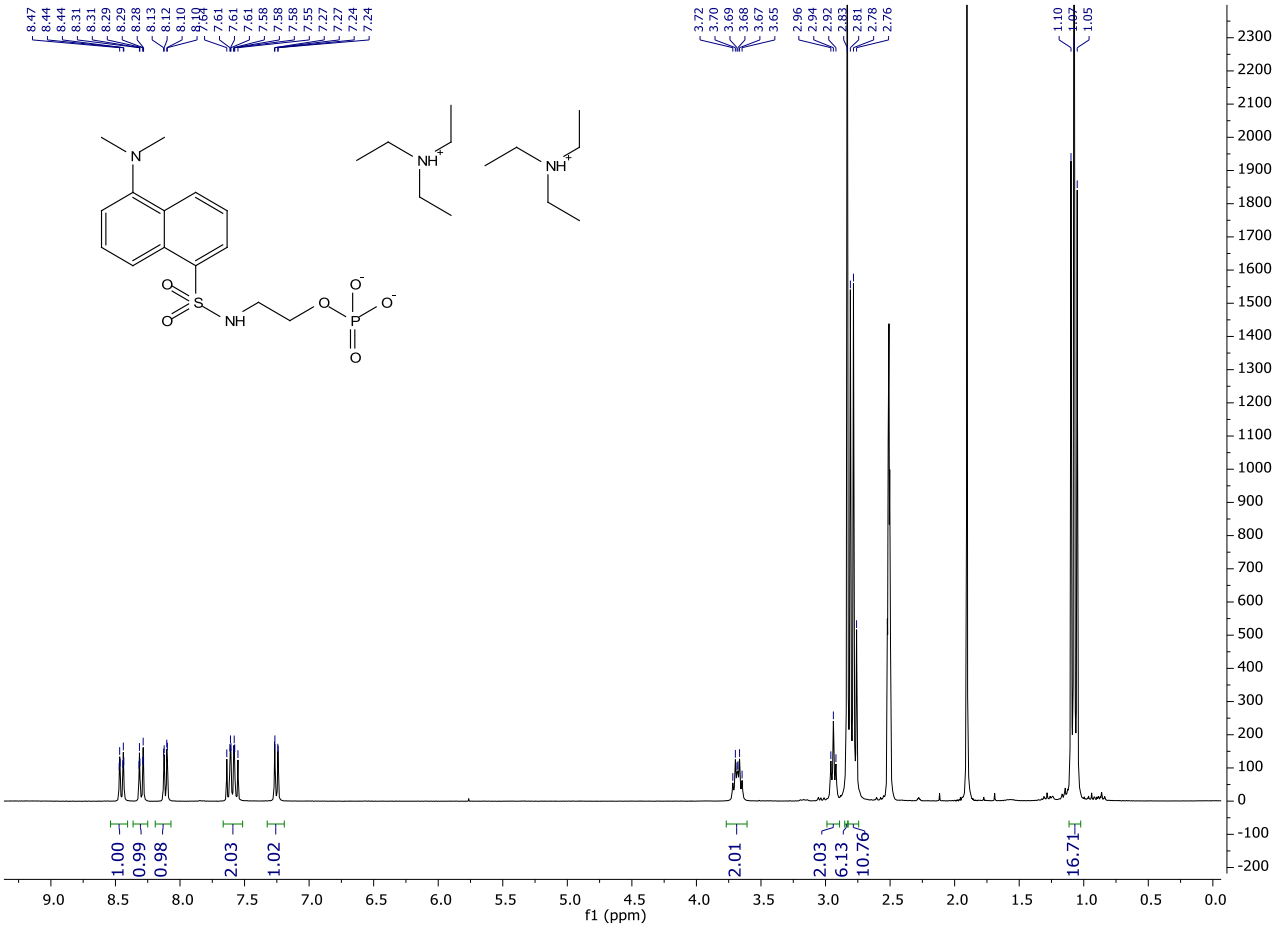
^{1}H NMR (MeOD, 300 MHz) of ***(30)*** *showing traces of EtOAc and pyridine that were deduced from the yield.*



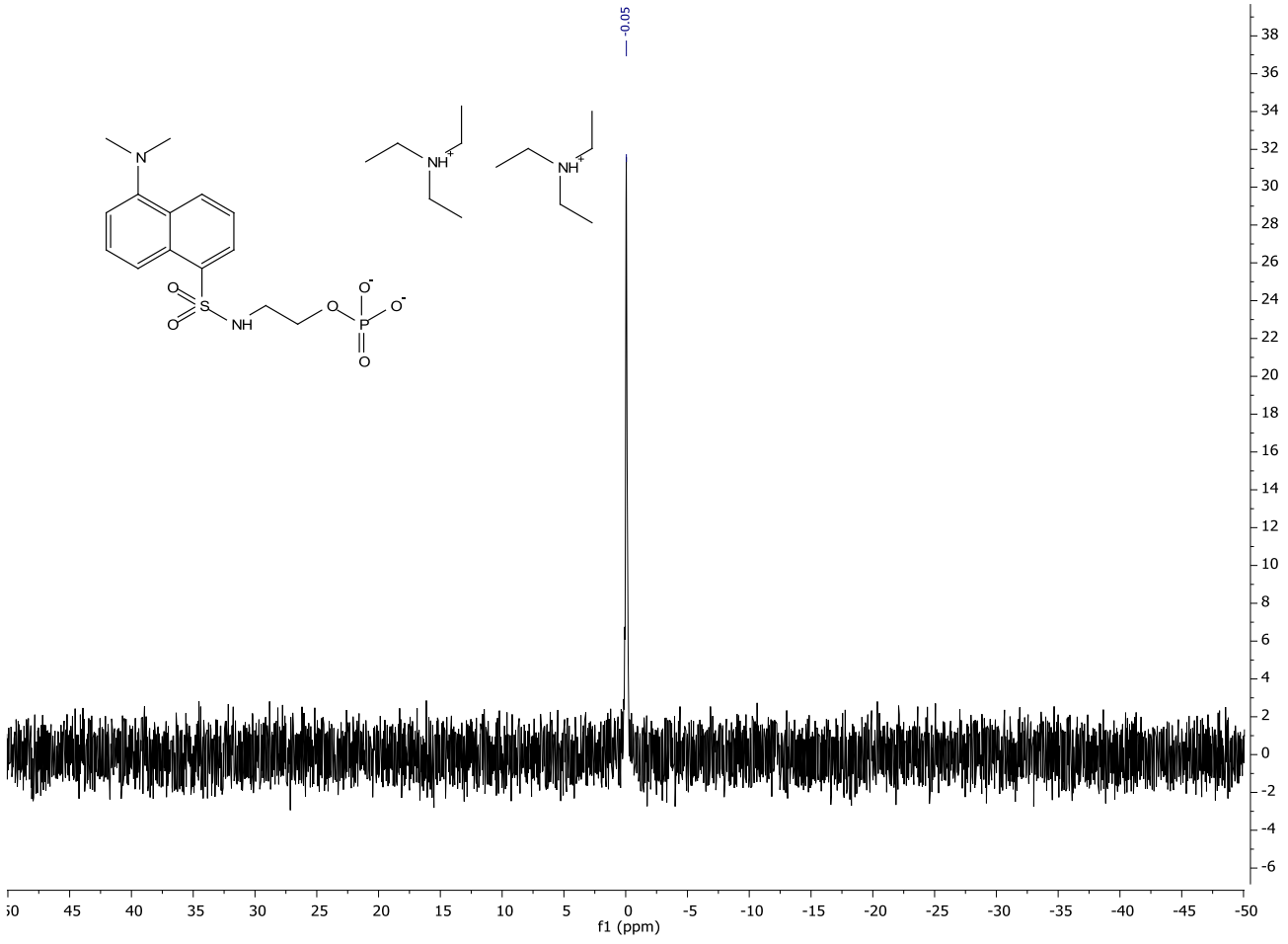
[1]H NMR (MeOD, 300 MHz) of ***(31)***



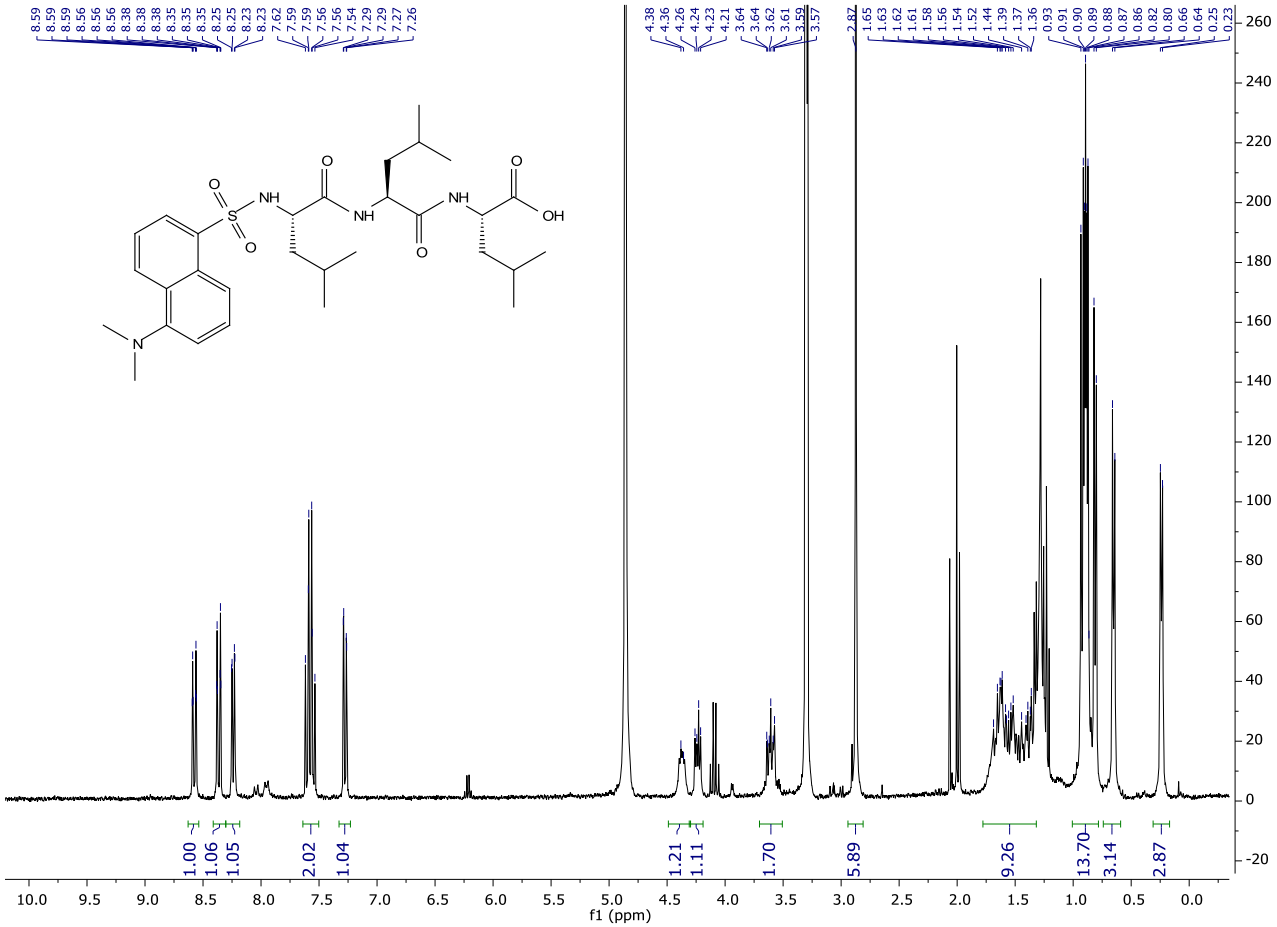
[31]P NMR (MeOD, 120 MHz) of ***(31)***



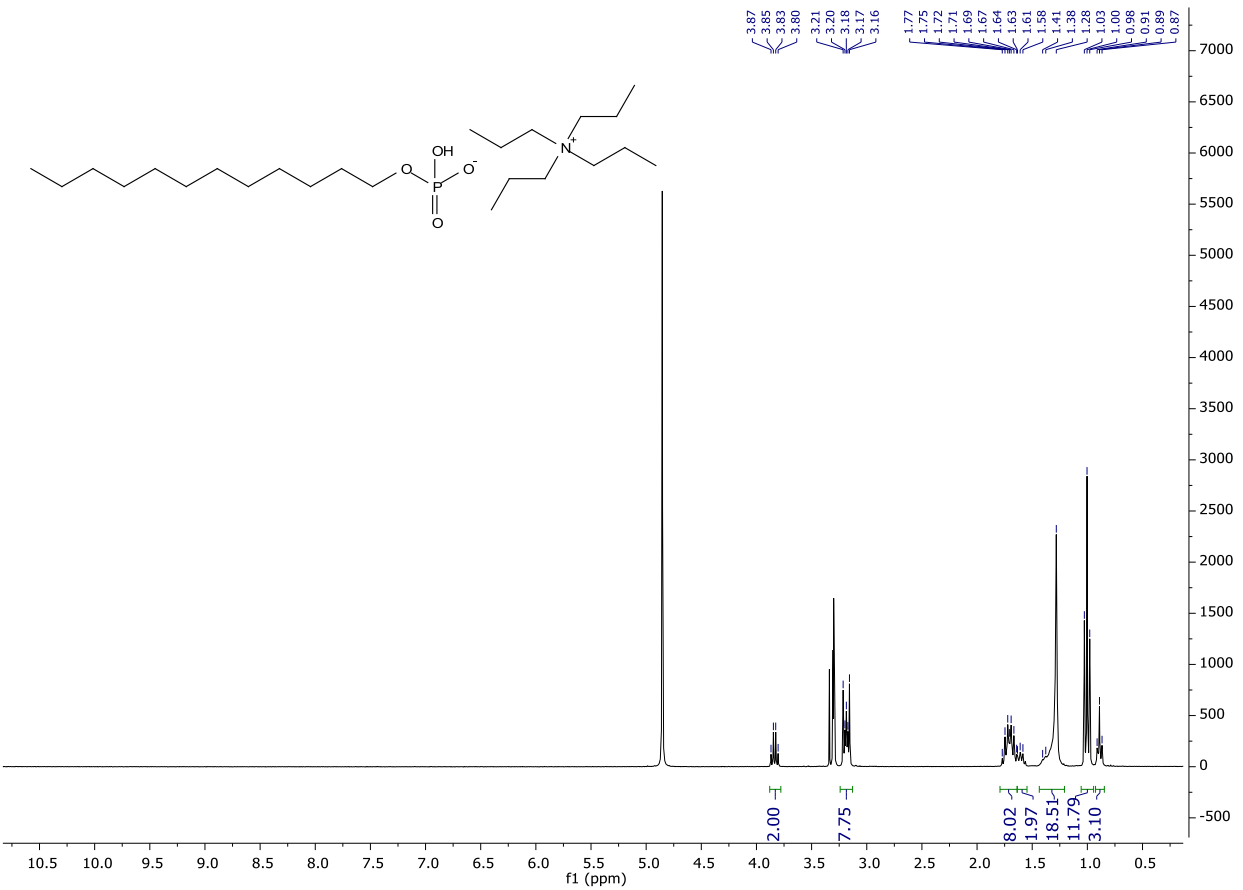
[1]H NMR (DMSO-d6, 300 MHz) of ***(32)***



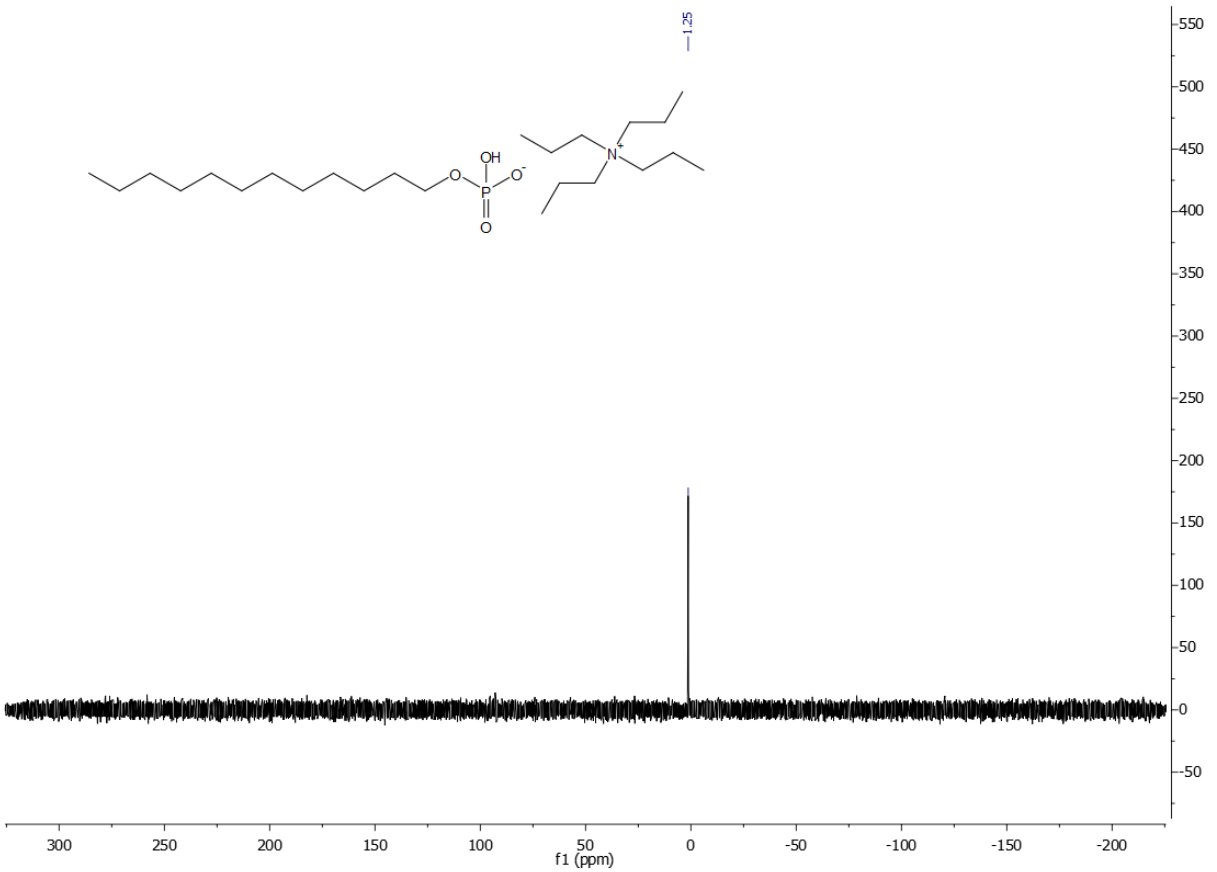
[31]P NMR (DMSO-d6, 120 MHz) of ***(32)***



^{1}H NMR spectrum (MeOD, 300 MHz) of dansyl Leu$_3$



^{1}H NMR spectrum (MeOD, 300 MHz) of the mono tetrapropylammonium salt of dodecyl phosphate



^{31}P NMR spectrum (MeOD, 120 MHz) of dodecyl phosphate

10. Biographical sketch and publications list

Dimitri Fayolle studied physical sciences in Preparative School at the Lycée Joffre (Montpellier, France) then at the University of Montpellier, where he obtained a bachelor's degree in physics and chemistry in 2015. In 2017, he received a master's degree in organic synthesis from the University of Lyon (France) after a 6-month internship in the laboratory of Prof. Peter Strazewski and Dr Michele Fiore dedicated to the synthesis of novel membranophobic clickable fluorophores. During his Master studies, he received a grant from the European Cooperation in Sciences and Technology (COST) action CM1304 to visit the laboratory of Pasquale Stano at the University of Rome (Italy) for a short research stay.

In 2017, he started his PhD studies under the direction of Prof. Peter Strazewski and co-supervision of Dr Michele Fiore, as part of an international collaborative project funded by the Volkswagen Stiftung. He studied the oligomerization and conjugation of biomolecules of prebiotic interest, and in particular the importance of lipids and amphiphiles in such processes. This work placed a particular emphasis on the development of analytical methods to study complex and heterogenous chemical systems, including heteronuclear and diffusion NMR and liquid chromatography, as well as phosphate chemistry. In 2019, he obtained a mobility grant from the Deutscher Akademischer Austauschdienst (DAAD) and spent one month in the laboratory of Prof. Clemens Richert (Stuttgart, Germany) to pursue research related to his PhD program.

Throughout his master and doctoral studies, he participated to 7 local and international congresses in France and in Europe and published 3 research articles as a first author, one as secondary author, with two others currently under revision, including a review article.

Peer-reviewed publications

1. Ruiz-Acosta, J., Altamura, E., Chieffo, C., **Fayolle, D.**, Lopez, A., Soulère, L., Oger, P., Popowycz, F., Buchet, R. & Fiore, M. The preparation and the evolutionary role of phospholipid esters and ethers (review article). **Under revision**

2. Lopez, A., **Fayolle, D.,** L., Fiore, M. & Strazewski, P., Chemical analysis of lipid boundaries after consecutive self-reproductions of supported giant vesicles. *iScience*, **accepted**

3. Altamura, E., Comte, A., D'Onofrio, A., Roussillon, C., **Fayolle, D.**, Buchet, R., Mavelli, F., Stano, P., Fiore, M. & Strazewski, P. Racemic phospholipids for origins of life studies (invited contribution to a special issue). *Symmetry*. **2020**, 12 (7), 1108

4. **Fayolle, D.**, Berthet, N., Doumèche, B., Renaudet, O., Strazewski, P. & Fiore, M. Towards the synthesis of outer membrane vesicle models with micromolar affinity to wheat germ agglutinin from alkyl dithioglycosides. *Beilstein J. Org. Chem*. **15**, 937-946 (2019).

5. **Fayolle, D.**, Fiore, M., Stano, P. & Strazewski, P. Rapid purification of giant lipid vesicles by microfiltration. *PLoS ONE 13***:** e0192975 (2018)

6. **Fayolle, D.**, Altamura, E., D'Onofrio, A., Madanamothoo, W. J., Fenet, B., Mavelli, F., Buchet, R., Stano, P., Fiore, M. & Strazewski, P. Crude phosphorylation mixtures containing racemic lipid amphiphiles self-assemble to give stable primitive compartments. *Sci. Rep.* **7**, 18106 (2017)

Oral communications and posters

1. **Fayolle, D.**, Fiore, M. & Strazewski, P. Formation of amphiphilic peptide-RNA conjugates in membranous environments. **Oral communication**, *ReCOB 18*, **2020**, Aussois (Modane), France. Postponed to December 2020 due to the epidemic situation.

2. **Fayolle, D.**, Fiore, M. & Strazewski, P. Co-evolution of peptides, RNA and lipids. **Poster**, *Volkswagen Stiftung Status Symposium on "Life?"*, **2019**, Hannover, Germany

3. **Fayolle, D.**, Fiore, M. & Strazewski, P. The formation of peptido-RNA in the presence of giant vesicles. **Oral communication**, *Annual seminar of the ICBMS*, **2019**, Villeurbanne, France. **Best oral communication prize**

4. **Fayolle, D.**, Fiore, M. & Strazewski, P. Abiotic, enzyme-free formation of peptide-RNA conjugates in water in the presence of lipid aggregates. **Oral communication**, *International Congress of the French Chemical Society*, **2018**, Montpellier, France

5. **Fayolle, D.**, Fiore, M., Stano, P. & Strazewski, P. Rapid purification of giant lipid vesicles by microfiltration. **Poster**, "*Systems Chemistry 2017" symposium of COST Action CM 1034*, **2017**, Sopron, Hungary

6. **Fayolle, D.**, Fiore, M., Stano, P. & Strazewski, P. Purification of giant lipid vesicles by microfiltration. **Oral communication**, *Working group 3 seminar, COST Action CM 1034*, **2017**, Warsaw, Poland

7. **Fayolle, D.**, Fiore, M., Dhenain, A., Lacôte, E., Larini, P. & Strazewski, P. "Tetrazine-Click-ARN". **Poster**, *Lyon Institute for Chemistry project restitution*, **2017**, Villeurbanne, France

6. References

1. Newton, I. S. *Philosophiae naturalis principia mathematica.* (S. Pepys, 1686).

2. Wöhler, F. Ueber künstliche Bildung des Harnstoffs. *Ann. Phys.* **88**, 253–256 (1828).

3. Ludlow, R. F. & Otto, S. Systems chemistry. *Chem. Soc. Rev.* **37**, 101–108 (2007).

4. Kolb, V. M. Defining life. Multiple perspectives. in *Handbook of Astrobiology* (ed. Kolb, V. M.) 57–64 (CRC Press, 2018).

5. Miller, S. L. & Urey, H. C. Organic compound synthesis on the primitive Earth. *Science* **130**, 245–251 (1959).

6. Boutlerow, M. A. Formation synthétique d'une substance sucrée. *Compte-Rendus Séances Académie Sci.* **53**, 145 (1861).

7. Islam, S. & Powner, M. W. Prebiotic systems chemistry: complexity overcoming clutter. *Chem* **2**, 470–501 (2017).

8. Richert, C. Prebiotic chemistry and human intervention. *Nat. Commun.* **9**, (2018).

9. Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D. & Sutherland, J. D. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.* **7**, 301–307 (2015).

10. Ruiz-Mirazo, K., Briones, C. & de la Escosura, A. Prebiotic systems chemistry: new perspectives for the origins of life. *Chem. Rev.* **114**, 285–366 (2014).

11. Szathmáry, E. & Maynard Smith, J. From replicators to reproducers: the first major transitions leading to life. *J. Theor. Biol.* **187**, 555–571 (1997).

12. Blain, J. C. & Szostak, J. W. Progress toward synthetic cells. *Annu. Rev. Biochem.* **83**, 615–640 (2014).

13. Strazewski, P. Omne vivum ex vivo … omne? How to feed an inanimate evolvable chemical system so as to let it self-evolve into increased complexity and life-like behaviour. *Isr. J. Chem.* **55**, 851–864 (2015).

14. Szostak, J. W., Bartel, D. P. & Luisi, P. L. Synthesizing life. *Nature* **409**, 387–390 (2001).

15. Damer, B. & Deamer, D. The hot spring hypothesis for an origin of life. *Astrobiology* **20**, 429–452 (2020).

16. Ragazzon, G. & Prins, L. J. Energy consumption in chemical fuel-driven self-assembly. *Nat. Nanotechnol.* **13**, 882–889 (2018).

17. Strazewski, P. The beginning of systems chemistry. *Life* **9**, 11 (2019).

18. Strazewski, P. The essence of systems chemistry. *Life* **9**, 60 (2019).

19. Ashkenasy, G., Hermans, T. M., Otto, S. & Taylor, A. F. Systems chemistry. *Chem. Soc. Rev.* **46**, 2543–2554 (2017).

20. Strazewski, P. Adding to Hans Kuhn's thesis on the emergence of the genetic apparatus: Of the Darwinian advantage to be neither too soluble, nor too insoluble, neither too solid, nor completely liquid. *Colloids Surf. B-Biointerfaces* **74**, 419–425 (2009).

21. Fiore, M. & Strazewski, P. Prebiotic lipidic amphiphiles and condensing agents on the early Earth. *Life* **6**, 17 (2016).

22. Biron, J.-P., Parkes, A. L., Pascal, R. & Sutherland, J. D. Expeditious, potentially primordial, aminoacylation of nucleotides. *Angew. Chem. Int. Ed.* **44**, 6731–6734 (2005).

23. Terenzi, S., Biała, E., Nguyen-Trung, N. Q. & Strazewski, P. Amphiphilic 3′-peptidyl-RNA conjugates. *Angew. Chem. Int. Ed.* **42**, 2909–2912 (2003).

24. Strazewski, P. How did translation occur? *Orig. Life Evol. Biospheres* **37**, 399–401 (2007).

25. Lazar, A. N., Coleman, A. W., Terenzi, S. & Strazewski, P. Observation of the formation of supported bilayers by amphiphilic peptidyl-RNA. *Chem. Commun.* 63–65 (2006) doi:10.1039/B509971B.

26. Isaad, A. L. C. *et al.* A hydrophobic disordered peptide spontaneously anchors a covalently bound RNA hairpin to giant lipidic vesicles. *Org Biomol Chem* **12**, 6363–6373 (2014).

27. Jauker, M., Griesser, H. & Richert, C. Spontaneous Formation of RNA Strands, Peptidyl RNA, and Cofactors. *Angew. Chem. Int. Ed.* **54**, 14564–14569 (2015).

28. Griesser, H. *et al.* Ribonucleotides and RNA promote peptide chain growth. *Angew. Chem. Int. Ed.* **56**, 1219–1223 (2017).

29. Liu, Z., Ajram, G., Rossi, J.-C. & Pascal, R. The chemical likelihood of ribonucleotide-α-amino acid copolymers as players for early stages of evolution. *J. Mol. Evol.* (2019) doi:10.1007/s00239-019-9887-7.

30. Vlassov, A. How was membrane permeability produced in an RNA world? *Orig. Life Evol. Biosphere J. Int. Soc. Study Orig. Life* **35**, 135–149 (2005).

31. Griesser, H., Bechthold, M., Tremmel, P., Kervio, E. & Richert, C. Amino acid-specific, ribonucleotide-promoted peptide formation in the absence of enzymes. *Angew. Chem. Int. Ed.* **56**, 1224–1228 (2017).

32. Tremmel, P., Griesser, H., Steiner, U. E. & Richert, C. How small heterocycles make a reaction network of amino acids and nucleotides efficient in water. *Angew. Chem. Int. Ed.* **58**, 13087–13092 (2019).

33. Motsch, S., Tremmel, P. & Richert, C. Regioselective formation of RNA strands in the absence of magnesium ions. *Nucleic Acids Res.* **48**, 1097–1107 (2020).

34. Rothschild, L. J. & Mancinelli, R. L. Life in extreme environments. *Nature* **409**, 10 (2001).

35. Oró, J. Synthesis of adenine from ammonium cyanide. *Biochem. Biophys. Res. Commun.* **2**, 407–412 (1960).

36. Sanchez, R. A., Ferris, J. P. & Orgel, L. E. Cyanoacetylene in prebiotic synthesis. *Science* **154**, 784–785 (1966).

37. Fox, S. W. & Harada, K. Synthesis of uracil under conditions of a thermal model of prebiological chemistry. *Science* **133**, 1923–1924 (1961).

38. Saladino, R., Crestini, C., Costanzo, G., Negri, R. & Di Mauro, E. A possible prebiotic synthesis of purine, adenine, cytosine, and 4(3*H*)-pyrimidinone from formamide: implications for the origin of life. *Bioorg. Med. Chem.* **9**, 1249–1253 (2001).

39. Pitsch, S., Eschenmoser, A., Gedulin, B., Hui, S. & Arrhenius, G. Mineral induced formation of sugar phosphates. *Orig. Life Evol. Biospheres* **25**, 297–334 (1995).

40. Kim, H.-J. *et al.* Synthesis of carbohydrates in mineral-guided prebiotic cycles. *J. Am. Chem. Soc.* **133**, 9457–9468 (2011).

41. Ricardo, A., Carrigan, M. A., Olcott, A. N. & Benner, S. A. Borate minerals stabilize ribose. *Science* **303**, 196–196 (2004).

42. Cafferty, B. J., Fialho, D. M., Khanam, J., Krishnamurthy, R. & Hud, N. V. Spontaneous formation and base pairing of plausible prebiotic nucleotides in water. *Nat. Commun.* **7**, (2016).

43. Bean, H. D. *et al.* Formation of a ß-pyrimidine nucleoside by a free pyrimidine base and ribose in a plausible prebiotic reaction. *J. Am. Chem. Soc.* **129**, 9556–9557 (2007).

44. Miller, S. L. A production of amino acids under possible primitive Earth conditions. *Science* **117**, 528–529 (1953).

45. Bada, J. L. New insights into prebiotic chemistry from Stanley Miller's spark discharge experiments. *Chem. Soc. Rev.* **42**, 2186–2196 (2013).

46. Strecker, A. Ueber einen neuen aus Aldehyd - Ammoniak und Blausäure entstehenden Körper. *Justus Liebigs Ann. Chem.* **91**, 349–351 (1854).

47. Fischer, E. Reduction von Säuren der Zuckergruppe. *Berichte Dtsch. Chem. Ges.* **22**, 2204–2205 (1889).

48. Kiliani, H. Ueber das Cyanhydrin der Lävulose. *Berichte Dtsch. Chem. Ges.* **18**, 3066–3072 (1885).

49. Ritson, D. & Sutherland, J. D. Prebiotic synthesis of simple sugars by photoredox systems chemistry. *Nat. Chem.* **4**, 895–899 (2012).

50. Ashe, K. *et al.* Selective prebiotic synthesis of phosphoroaminonitriles and aminothioamides in neutral water. *Commun. Chem.* **2**, 23 (2019).

51. Traube, W. Der Aufbau der Xanthinbasen aus der Cyanessigsäure. Synthese des Hypoxanthins und Adenins. *Justus Liebigs Ann. Chem.* **331**, 64–88 (1903).

52. Becker, S. *et al.* A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. *Science* **352**, 833–836 (2016).

53. Becker, S. *et al.* Wet-dry cycles enable the parallel origin of canonical and non-canonical nucleosides by continuous synthesis. *Nat. Commun.* **9**, 163 (2018).

54. Becker, S., Schneider, C., Crisp, A. & Carell, T. Non-canonical nucleosides and chemistry of the emergence of life. *Nat. Commun.* **9**, 5174 (2018).

55. Schneider, C. *et al.* Noncanonical RNA nucleosides as molecular fossils of an early earth-generation by prebiotic methylations and carbamoylations. *Angew. Chem. Int. Ed.* **57**, 5943–5946 (2018).

56. Becker, S. *et al.* Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides. *Science* **366**, 76–82 (2019).

57. Kim, H.-J. *et al.* Evaporite borate-containing mineral ensembles make phosphate available and regiospecifically phosphorylate ribonucleosides: borate as a multifaceted problem solver in prebiotic chemistry. *Angew. Chem. Int. Ed.* **55**, 15816–15820 (2016).

58. Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis: V. Synthesis and photoanomerization of pyrimidine nucleosides. *J. Mol. Biol.* **47**, 531–543 (1970).

59. Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–242 (2009).

60. Islam, S., Bučar, D.-K. & Powner, M. W. Prebiotic selection and assembly of proteinogenic amino acids and natural nucleotides from complex mixtures. *Nat. Chem.* **9**, 584–589 (2017).

61. Stairs, S. *et al.* Divergent prebiotic synthesis of pyrimidine and 8-oxo-purine ribonucleotides. *Nat. Commun.* **8**, 15270 (2017).

62. Xu, J., Green, N. J., Gibard, C., Krishnamurthy, R. & Sutherland, J. D. Prebiotic phosphorylation of 2-thiouridine provides either nucleotides or DNA building blocks via photoreduction. *Nat. Chem.* **11**, 457–462 (2019).

63. Teichert, J. S., Kruse, F. M. & Trapp, O. Direct prebiotic pathway to DNA nucleosides. *Angew. Chem. Int. Ed.* **58**, 9944–9947 (2019).

64. Carell, T. *et al.* Structure and function of noncanonical nucleobases. *Angew. Chem. Int. Ed.* **51**, 7110–7131 (2012).

65. Bhowmik, S. & Krishnamurthy, R. The role of sugar-backbone heterogeneity and chimeras in the simultaneous emergence of RNA and DNA. *Nat. Chem.* **11**, 1009–1018 (2019).

66. Krishnamurthy, R. *et al.* Pyranosyl-RNA: base pairing between homochiral oligonucleotide strands of opposite sense of chirality. *Angew. Chem. Int. Ed. Engl.* **35**, 1537–1541 (1996).

67. Schöning, K.-U., Scholtz, P., Guntha, X. W., Krishnamurthy, R. & Eschenmoser, A. Chemical etiology of nucleic acid structure: the alpha-threofuranosyl-(3'->2') oligonucleotide system. *Science* **290**, 1347–1351 (2000).

68. Egholm, M. *et al.* PNA hybridizes to complementary oligonucleotides obeying the Watson–Crick hydrogen-bonding rules. *Nature* **365**, 566–568 (1993).

69. Liu, B. *et al.* Spontaneous emergence of self-replicating molecules containing nucleobases and amino acids. *J. Am. Chem. Soc.* **142**, 4184–4192 (2020).

70. Okamura, H. *et al.* Proto-urea-RNA (Wöhler RNA) containing unusually stable urea nucleosides. *Angew. Chem. Int. Ed.* **58**, 18691–18696 (2019).

71. Hoshika, S. *et al.* Hachimoji DNA and RNA: a genetic system with eight building blocks. *Science* **363**, 884–887 (2019).

72. Canavelli, P., Islam, S. & Powner, M. W. Peptide ligation by chemoselective aminonitrile coupling in water. *Nature* **571**, 546–549 (2019).

73. Peretó, J. G., Velasco, A. M., Becerra, A. & Lazcano, A. Comparative biochemistry of CO_2 fixation and the evolution of autotrophy. *Int. Microbiol.* 3–10 (1999).

74. Smith, E. & Morowitz, H. J. Universality in intermediary metabolism. *Proc. Natl. Acad. Sci.* **101**, 13168–13173 (2004).

75. Martin, W. & Russell, M. J. On the origin of biochemistry at an alkaline hydrothermal vent. *Philos. Trans. R. Soc. B Biol. Sci.* **362**, 1887–1926 (2007).

76. Muchowska, K. B., Chevallot-Beroux, E. & Moran, J. Recreating ancient metabolic pathways before enzymes. *Bioorg. Med. Chem.* **27**, 2292–2297 (2019).

77. Varma, S. J., Muchowska, K. B., Chatelain, P. & Moran, J. Native iron reduces CO_2 to intermediates and end-products of the acetyl-CoA pathway. *Nat. Ecol. Evol.* **2**, 1019–1024 (2018).

78. Muchowska, K. B. *et al.* Metals promote sequences of the reverse Krebs cycle. *Nat. Ecol. Evol.* **1**, 1716–1721 (2017).

79. Lipshutz, B. H. *et al.* TPGS-750-M: a second-generation amphiphile for metal-catalyzed cross-couplings in water at room temperature. *J. Org. Chem.* **76**, 4379–4391 (2011).

80. Muchowska, K. B., Varma, S. J. & Moran, J. Synthesis and breakdown of universal metabolic precursors promoted by iron. *Nature* **569**, 104–107 (2019).

81. Orgel, L. E. The implausibility of metabolic cycles on the prebiotic earth. *PLOS Biol.* **6**, e18 (2008).

82. Wu, L.-F. & Sutherland, J. D. Provisioning the origin and early evolution of life. *Emerg. Top. Life Sci.* **3**, 459–468 (2019).

83. Ruiz-Acosta, J. *et al.* The preparation and the evolutionary role of phospholipid esters and ethers. *Chem. Rev.* **under revision**, (2020).

84. McCollom, T. M., Ritter, G. & Simoneit, B. R. T. Lipid synthesis under hydrothermal conditions by Fischer-Tropsch-type reactions. 14.

85. Rushdi, A. I. & Simoneit, B. R. T. Lipid formation by aqueous Fischer-Tropsch-type synthesis over a temperature range of 100 to 400°C. 16.

86. Simoneit, B. R. T. Prebiotic organic synthesis under hydrothermal conditions: an overview. *Adv. Space Res.* **33**, 88–94 (2004).

87. Klein, A. E. & Pilpel, N. Oxidation of *n*-alkanes photosensitized by 1-naphthol. *J. Chem. Soc. Faraday Trans. 1 Phys. Chem. Condens. Phases* **69**, 1729 (1973).

88. Dworkin, J. P., Deamer, D. W., Sandford, S. A. & Allamandola, L. J. Self-assembling amphiphilic molecules: Synthesis in simulated interstellar/precometary ices. *Proc. Natl. Acad. Sci.* **98**, 815–819 (2001).

89. Deamer, D. W. Boundary structures are formed by organic components of the Murchison carbonaceous chondrite. *Nature* **317**, 792–794 (1985).

90. Deamer, D. W. & Pashley, R. M. Amphiphilic components of the murchison carbonaceous chondrite: Surface properties and membrane formation. *Orig. Life Evol. Biospheres* **19**, 21–38 (1989).

91. Meierhenrich, U. J., Filippi, J.-J., Meinert, C., Vierling, P. & Dworkin, J. P. On the origin of primitive cells: from nutrient intake to elongation of encapsulated nucleotides. *Angew. Chem. Int. Ed.* **49**, 3738–3750 (2010).

92. Harvey, G. R., Mopper, K. & Degens, E. T. Synthesis of carbohydrates and lipids on kaolinite. *Chem. Geol.* **9**, 79–87 (1972).

93. Hargreaves, W. R., Mulvihill, S. J. & Deamer, D. W. Synthesis of phospholipids and membranes in prebiotic conditions. *Nature* **266**, 78–80 (1977).

94. Oró, J. Studies in experimental organic cosmochemistry. *Ann. N. Y. Acad. Sci.* **108**, 464–481 (1963).

95. Steinman, G., Lemmon, R. M. & Calvin, M. Cyanamide: a possible key compound in chemical evolution. *Proc. Natl. Acad. Sci.* **52**, 27–30 (1964).

96. Eichberg, J., Sherwood, E., Epps, D. E. & Orb, J. Cyanamide mediated syntheses under plausible primitive earth conditions. IV. The synthesis of acylglycerols. *J. Mol. Evol.* **10**, 221–230 (1977).

97. Epps, D. E., Sherwood, E., Eichberg, J. & Oró, J. Cyanamide mediated syntheses under plausible primitive earth conditions. V. The synthesis of phosphatidic acids. *J. Mol. Evol.* **11**, 279–292 (1978).

98. Epps, D. E., Nooner, D. W., Eichberg, J., Sherwood, E. & Oró, J. Cyanamide mediated synthesis under plausible primitive earth conditions. VI. The synthesis of glycerol and glycerophosphates. *J. Mol. Evol.* **14**, 235–241 (1979).

99. Fayolle, D. *et al.* Crude phosphorylation mixtures containing racemic lipid amphiphiles self-assemble to give stable primitive compartments. *Sci. Rep.* **7**, 18106 (2017).

100. Rao, M., Eichberg, J. & Oró, J. Synthesis of phosphatidylethanolamine under possible primitive earth conditions. *J. Mol. Evol.* **25**, 1–6 (1987).

101. Rao, M., Eichberg, J. & Oró, J. Synthesis of phosphatidylcholine under possible primitive Earth conditions. *J. Mol. Evol.* **18**, 196–202 (1982).

102. Fiore, M. The synthesis of mono-alkyl phosphates and their derivatives: an overview of their nature, preparation and use, including synthesis under plausible prebiotic conditions. *Org. Biomol. Chem.* **16**, 3068–3086 (2018).

103. Fiore, M. *et al.* Giant vesicles from rehydrated crude mixtures containing unexpected mixtures of amphiphiles formed under plausibly prebiotic conditions. *Org. Biomol. Chem.* **15**, 4231–4240 (2017).

104. Gibard, C., Bhowmik, S., Karki, M., Kim, E.-K. & Krishnamurthy, R. Phosphorylation, oligomerization and self-assembly in water under potential prebiotic conditions. *Nat. Chem.* **10**, 212–217 (2018).

105. Gibard, C. *et al.* Geochemical sources and availability of amidophosphates on the early earth. *Angew. Chem. Int. Ed.* **58**, 8151–8155 (2019).

106. Schwartz, A. W. Phosphorus in prebiotic chemistry - an update and a note on plausibility. in *Handbook of Astrobiology* (ed. Kolb, V. M.) 355–359 (CRC Press, 2018).

107. Fernández-García, C., Coggins, A. J. & Powner, M. W. A chemist's perspective on the role of phosphorus at the origins of life. *Life* **7**, 31 (2017).

108. Pasek, M. A., Gull, M. & Herschy, B. Phosphorylation on the early earth. *Chem. Geol.* **475**, 149–170 (2017).

109. Hulshof, J. & Ponnamperuma, C. Prebiotic condensation reactions in an aqueous medium: A review of condensing agents. *Orig. Life* **7**, 197–224 (1976).

110. Parker, E. T. *et al.* A plausible simultaneous synthesis of amino acids and simple peptides on the primordial earth. *Angew. Chem. Int. Ed.* **53**, 8132–8136 (2014).

111. Orgel, L. E. & Lohrmann, R. Prebiotic chemistry and nucleic acid replication. *Acc. Chem. Res.* **7**, 368–377 (1974).

112. Powner, M. W. & Sutherland, J. D. Prebiotic chemistry: a new modus operandi. *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 2870–2877 (2011).

113. Mariani, A., Russell, D. A., Javelle, T. & Sutherland, J. D. A light-releasable potentially prebiotic nucleotide activating agent. *J. Am. Chem. Soc.* **140**, 8657–8661 (2018).

114. Ni, F., Sun, S., Huang, C. & Zhao, Y. *N*-phosphorylation of amino acids by trimetaphosphate in aqueous solution—learning from prebiotic synthesis. *Green Chem.* **11**, 569 (2009).

115. Inoue, H., Baba, Y., Furukawa, T., Maeda, Y. & Tsuhako, M. Formation of dipeptide in the reaction of amino acids with cyclo-triphosphate. *Chem. Pharm. Bull. (Tokyo)* **41**, 1895–1899 (1993).

116. Chung, N. M., Lohrmann, R., Orgel, L. E. & Rabinowitz, J. The mechanism of the trimetaphosphate-induced peptide synthesis. *Tetrahedron* **27**, 1205–1210 (1971).

117. Quimby, O. T. & Flautt, T. J. Ammonolyse des Trimetaphosphats. *Z. Für Anorg. Allg. Chem.* **296**, 220–228 (1958).

118. Krishnamurthy, R., Arrhenius, G. & Eschenmoser, A. Formation of glycolaldehyde phosphate from glycolaldehyde in aqueous solution. *Orig. Life Evol. Biospheres* **29**, 333–354 (1999).

119. Krishnamurthy, R., Guntha, S. & Eschenmoser, A. Regioselective α-phosphorylation of aldoses in aqueous solution. *Angew. Chem. Int. Ed.* **39**, 2281–2285 (2000).

120. Mullen, L. B. & Sutherland, J. D. Formation of potentially prebiotic amphiphiles by reaction of ß-hydroxy-*n*-alkylamines with cyclotriphosphate. *Angew. Chem. Int. Ed.* **46**, 4166–4168 (2007).

121. Toparlak, Ö. D., Karki, M., Egas Ortuno, V., Krishnamurthy, R. & Mansy, S. S. Cyclophospholipids increase protocellular stability to metal ions. *Small* (2019) doi:10.1002/smll.201903381.

122. Kanavarioti, A., Monnard, P.-A. & Deamer, D. W. Eutectic phases in ice facilitate nonenzymatic nucleic acid synthesis. *Astrobiology* **1**, 271–281 (2001).

123. White, R. H. Hydrolytic stability of biomolecules at high temperatures and its implication for life at 250 °C. *Nature* **310**, 430–432 (1984).

124. Liu, Z., Rossi, J.-C. & Pascal, R. How prebiotic chemistry and early life chose phosphate. *Life* **9**, (2019).

125. Danger, G., Plasson, R. & Pascal, R. Pathways for the formation and evolution of peptides in prebiotic environments. *Chem. Soc. Rev.* **41**, 5416 (2012).

126. Ferris, J. P., Hill, A. R., Liu, R. & Orgel, L. E. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **381**, 59–61 (1996).

127. Coari, K. M., Martin, R. C., Jain, K. & McGown, L. B. Nucleotide selectivity in abiotic RNA polymerization reactions. *Orig. Life Evol. Biospheres* **47**, 305–321 (2017).

128. Wolfenden, R. Benchmark reaction rates, the stability of biological molecules in water, and the evolution of catalytic power in enzymes. *Annu. Rev. Biochem.* **80**, 645–667 (2011).

129. Minakuchi, N. *et al.* Versatile supramolecular gelators that can harden water, organic solvents and ionic liquids. *Langmuir* **28**, 9259–9266 (2012).

130. Levine, S. Synthesis of glycyl and alanyl chlorides. *J. Am. Chem. Soc.* **76**, 1382–1382 (1954).

131. Goodman, M. & McGahren, W. J. Mechanistic studies of peptide oxazolone racemization. *Tetrahedron* **23**, 2031–2050 (1967).

132. Fox, S. W. & Harada, K. Thermal copolymerization of amino acids to a product resembling protein. *Science* **128**, 1214–1214 (1958).

133. Lahav, N., White, D. & Chang, S. Peptide formation in the prebiotic era: thermal condensation of glycine in fluctuating clay environments. *Science* **201**, 67–69 (1978).

134. Schwendinger, M. G. *et al.* Salt induced peptide formation: on the selectivity of the copper induced peptide formation under possible prebiotic conditions. *Inorganica Chim. Acta* **228**, 207–214 (1995).

135. Rode, B. M., Son, H. L., Suwannachot, Y. & Bujdak, J. The combination of salt induced peptide formation reaction and clay catalysis: a way to higher peptides under primitive earth conditions. *Orig. Life Evol. Biospheres* **29**, 273–286 (1999).

136. Pedreira-Segade, U., Hao, J., Montagnac, G., Cardon, H. & Daniel, I. Spontaneous polymerization of glycine under hydrothermal conditions. *ACS Earth Space Chem.* **3**, 1669–1677 (2019).

137. Huber, C. & Wächtershäuser, G. Peptides by activation of amino acids with CO on (Ni,Fe)S surfaces: implications for the origin of life. *Science* **281**, 670–672 (1998).

138. Kawamura, K. Development of micro-flow hydrothermal monitoring systems and their applications to the origin of life study on earth. *Anal. Sci.* **27**, 675–675 (2011).

139. Goolcharran, C. & Borchardt, R. T. Kinetics of diketopiperazine formation using model peptides. *J. Pharm. Sci.* **87**, 283–288 (1998).

140. Nagayama, M., Takaoka, O., Inomata, K. & Yamagata, Y. Diketopiperazine-mediated peptide formation in aqueous solution. *Orig. Life Evol. Biosph.* **20**, 249–257 (1990).

141. Rodriguez-Garcia, M. *et al.* Formation of oligopeptides in high yield under simple programmable conditions. *Nat. Commun.* **6**, (2015).

142. Griffith, E. C. & Vaida, V. In situ observation of peptide bond formation at the water-air interface. *Proc. Natl. Acad. Sci.* **109**, 15697–15701 (2012).

143. Gorlero, M. *et al.* Ser-His catalyses the formation of peptides and PNAs. *FEBS Lett.* **583**, 153–156 (2009).

144. Weber, A. L. Prebiotic amino acid thioester synthesis: thiol-dependent amino acid synthesis from formose substrates (formaldehyde and glycolaldehyde) and ammonia. *Orig. Life Evol. Biospheres* **28**, 259–270 (1998).

145. Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. Synthesis of proteins by native chemical ligation. *Science* **266**, 776–779 (1994).

146. Forsythe, J. G. *et al.* Ester-mediated amide bond formation driven by wet-dry cycles: a possible path to polypeptides on the prebiotic earth. *Angew. Chem. Int. Ed.* **54**, 9871–9875 (2015).

147. Forsythe, J. G. *et al.* Surveying the sequence diversity of model prebiotic peptides by mass spectrometry. *Proc. Natl. Acad. Sci.* **114**, 7652–7659 (2017).

148. Carny, O. & Gazit, E. Creating prebiotic sanctuary: self-assembling supramolecular peptide structures bind and stabilize RNA. *Orig. Life Evol. Biospheres* **41**, 121–132 (2011).

149. Pascal, R. & Pross, A. Stability and its manifestation in the chemical and biological worlds. *Chem. Commun.* **51**, 16160–16165 (2015).

150. Hawker, J. R. & Oró, J. Cyanamide mediated syntheses of peptides containing histidine and hydrophobic amino acids. *J. Mol. Evol.* **17**, 285–294 (1981).

151. Halmann, M. Cyanamide-induced condensation reactions of glycine. *Arch. Biochem. Biophys.* **128**, 808–810 (1968).

152. Pascal, R., Boiteau, L. & Commeyras, A. From the prebiotic synthesis of α-amino acids towards a primitive translation apparatus for the synthesis of peptides. in *Prebiotic Chemistry* (ed. Walde, P.) 69–122 (Springer, 2005). doi:10.1007/b136707.

153. Commeyras, A. *et al.* Prebiotic synthesis of sequential peptides on the Hadean beach by a molecular engine working with nitrogen oxides as energy sources. *Polym. Int.* **51**, 661–665 (2002).

154. Ehler, K. W. & Orgel, L. E. N,N′-carbonyldiimidazole-induced peptide formation in aqueous solution. *Biochim. Biophys. Acta* **434**, 233–243 (1976).

155. Ehler, K. W., Girard, E. & Orgel, L. E. Reactions of polyfunctional amino acids with N,N′-carbonyldiimidazole in aqueous solution oligopeptide formation. *Biochim. Biophys. Acta* **491**, 253–264 (1977).

156. Danger, G., Boiteau, L., Cottet, H. & Pascal, R. The peptide formation mediated by cyanate revisited. *N*-carboxyanhydrides as accessible intermediates in the decomposition of *N*-carbamoylamino acids. *J. Am. Chem. Soc.* **128**, 7412–7413 (2006).

157. Leman, L., Orgel, L. & Ghadiri, M. R. Carbonyl sulfide-mediated prebiotic formation of peptides. *Science* **306**, 283–286 (2004).

158. Lamy, C., Lemoine, J., Bouchu, D., Goekjian, P. & Strazewski, P. Glutamate–glycine and histidine–glycine co-oligopeptides: batch co-oligomerization versus pulsed addition of N-carboxyanhydrides. *ChemBioChem* **9**, 710–713 (2008).

159. Blocher, M., Liu, D. & Luisi, P. L. Liposome-assisted selective polycondensation of α-amino acids and peptides: the case of charged liposomes. *Macromolecules* **33**, 5787–5796 (2000).

160. Chan, W. & White, P. *Fmoc solid phase peptide synthesis: a practical approach*. (Oxford University Press, 2000).

161. Coggins, A. J. & Powner, M. W. Prebiotic synthesis of phosphoenol pyruvate by α-phosphorylation-controlled triose glycolysis. *Nat. Chem.* **9**, 310–317 (2017).

162. Danger, G. *et al.* 5(4*H*)-oxazolones as intermediates in the carbodiimide- and cyanamide-promoted peptide activations in aqueous solution. *Angew. Chem. Int. Ed Engl.* **52**, 611–614 (2013).

163. Garcia, A. D. *et al.* The astrophysical formation of asymmetric molecules and the emergence of a chiral bias. *Life* **9**, 29 (2019).

164. Ajram, G., Rossi, J.-C., Boiteau, L. & Pascal, R. A reassessment of prebiotically relevant chemical agents for the activation of α-amino acids and peptides. *J. Syst. Chem.* **7**, 19–28 (2019).

165. Clark, J. C., Phillipps, G. H., Steer, M. R., Stephenson, L. & Cooksey, A. R. Resolution of esters of phenylglycine with (+)-tartaric acid. *J. Chem. Soc. Perkin 1* 471–474 (1976) doi:10.1039/P19760000471.

166. Brands, K. M. J. & Davies, A. J. Crystallization-induced diastereomer transformations. *Chem. Rev.* **106**, 2711–2733 (2006).

167. Noorduin, W. L. *et al.* Emergence of a single solid chiral state from a nearly racemic amino acid derivative. *J. Am. Chem. Soc.* **130**, 1158–1159 (2008).

168. Kaptein, B. *et al.* Attrition-enhanced deracemization of an amino acid derivative that forms an epitaxial racemic conglomerate. *Angew. Chem. Int. Ed.* **47**, 7226–7229 (2008).

169. Noorduin, W. L. *et al.* Complete deracemization by attrition-enhanced Ostwald ripening elucidated. *Angew. Chem. Int. Ed.* **47**, 6445–6447 (2008).

170. Lee, D. H., Granja, J. R., Martinez, J. A., Severin, K. & Ghadiri, M. R. A self-replicating peptide. *Nature* **382**, 525–528 (1996).

171. Severin, K., Lee, D. H., Martinez, J. A. & Ghadiri, M. R. Peptide self-replication via template-directed ligation. *Chem. – Eur. J.* **3**, 1017–1024 (1997).

172. Severin, K., Lee, D. H., Martinez, J. A., Vieth, M. & Ghadiri, M. R. Dynamic error correction in autocatalytic peptide networks. *Angew. Chem. Int. Ed.* **37**, 126–128 (1998).

173. Severin, K., Lee, D. H., Kennan, A. J. & Ghadiri, M. R. A synthetic peptide ligase. *Nature* **389**, 706–709 (1997).

174. Lee, D. H., Severin, K., Yokobayashi, Y. & Ghadiri, M. R. Emergence of symbiosis in peptide self-replication through a hypercyclic network. *Nature* **390**, 591–594 (1997).

175. Rochet, J.-C. & Lansbury, P. T., Jr. Amyloid fibrillogenesis: themes and variations. *Curr. Opin. Struct. Biol.* **10**, 60–68 (2000).

176. Reches, M. & Gazit, E. Casting metal nanowires within discrete self-assembled peptide nanotubes. *Science* **300**, 625–627 (2003).

177. Greenwald, J., Friedmann, M. P. & Riek, R. Amyloid aggregates arise from amino acid condensations under prebiotic conditions. *Angew. Chem. Int. Ed.* **55**, 11609–11613 (2016).

178. Rubinov, B., Wagner, N., Rapaport, H. & Ashkenasy, G. Self-replicating amphiphilic ß-sheet peptides. *Angew. Chem. Int. Ed.* **48**, 6683–6686 (2009).

179. Takahashi, Y., Ueno, A. & Mihara, H. Design of a peptide undergoing α–ß structural transition and amyloid fibrillogenesis by the introduction of a hydrophobic defect. *Chem. – Eur. J.* **4**, 2475–2484 (1998).

180. Takahashi, Y. & Mihara, H. Construction of a chemically and conformationally self-replicating system of amyloid-like fibrils. *Bioorg. Med. Chem.* **12**, 693–699 (2004).

181. Rubinstein, I., Eliash, R., Bolbach, G., Weissbuch, I. & Lahav, M. Racemic ß sheets in biochirogenesis. *Angew. Chem. Int. Ed.* **46**, 3710–3713 (2007).

182. Wieczorek, R., Adamala, K., Gasperi, T., Polticelli, F. & Stano, P. Small and random peptides: an unexplored reservoir of potentially functional primitive organocatalysts. the case of seryl-histidine. *Life* **7**, 19 (2017).

183. Engelhart, A. E., Powner, M. W. & Szostak, J. W. Functional RNAs exhibit tolerance for non-heritable 2′–5′ versus 3′–5′ backbone heterogeneity. *Nat. Chem.* **5**, 390–394 (2013).

184. Morávek, J. Formation of oligonucleotides during heating of a mixture of uridine 2′(3′) -phosphate and uridine. *Tetrahedron Lett.* **8**, 1707–1710 (1967).

185. Ogasawara, H. *et al.* Synthesizing oligomers from monomeric nucleotides in simulated hydrothermal environments. *Orig. Life Evol. Biospheres* **30**, 519–526 (2000).

186. Jacob, T. M. & Khorana, H. G. Studies on polynucleotides. XXX. A comparative study of reagents for the synthesis of the $C_{3'}$-$C_{5'}$ internucleotidic linkage. *J. Am. Chem. Soc.* **86**, 1630–1635 (1964).

187. Verlander, M. S., Lohrmann, R. & Orgel, L. E. Catalysts for the self-polymerization of adenosine cyclic 2′,3′-phosphate. *J. Mol. Evol.* **2**, 303–316 (1973).

188. Usher, D. A. & Yee, D. Geometry of the dry-state oligomerization of 2′,3′-cyclic phosphates. *J. Mol. Evol.* **13**, 287–293 (1979).

189. Costanzo, G. *et al.* Generation of RNA molecules by a base-catalysed click-like reaction. *ChemBioChem* **13**, 999–1008 (2012).

190. Morasch, M., Mast, C. B., Langer, J. K., Schilcher, P. & Braun, D. Dry polymerization of 3′,5′-cyclic GMP to long strands of RNA. *ChemBioChem* **15**, 879–883 (2014).

191. Costanzo, G., Pino, S., Botta, G., Saladino, R. & Di Mauro, E. May cyclic nucleotides be a source for abiotic RNA synthesis? *Orig. Life Evol. Biospheres* **41**, 559–562 (2011).

192. Da Silva, L., Maurel, M.-C. & Deamer, D. Salt-promoted synthesis of RNA-like molecules in simulated hydrothermal conditions. *J. Mol. Evol.* **80**, 86–97 (2015).

193. Monnard, P.-A., Kanavarioti, A. & Deamer, D. W. Eutectic phase polymerization of activated ribonucleotide mixtures yields quasi-equimolar incorporation of purine and pyrimidine nucleobases. *J. Am. Chem. Soc.* **125**, 13734–13740 (2003).

194. Attwater, J., Wochner, A., Pinheiro, V. B., Coulson, A. & Holliger, P. Ice as a protocellular medium for RNA replication. *Nat. Commun.* **1**, 76 (2010).

195. Attwater, J., Wochner, A. & Holliger, P. In-ice evolution of RNA polymerase ribozyme activity. *Nat. Chem.* **5**, 1011–1018 (2013).

196. Huang, W. & Ferris, J. P. One-step, regioselective synthesis of up to 50-mers of RNA oligomers by montmorillonite catalysis. *J. Am. Chem. Soc.* **128**, 8914–8919 (2006).

197. Inoue, T. & Orgel, L. E. A nonenzymatic RNA polymerase model. *Science* **219**, 859–862 (1983).

198. Li, L. *et al.* Enhanced nonenzymatic RNA copying with 2-aminoimidazole activated nucleotides. *J. Am. Chem. Soc.* **139**, 1810–1813 (2017).

199. Hagenbuch, P., Kervio, E., Hochgesand, A., Plutowski, U. & Richert, C. Chemical primer extension: efficiently determining single nucleotides in DNA. *Angew. Chem. Int. Ed.* **44**, 6588–6592 (2005).

200. Prabahar, K. J., Cole, T. D. & Ferris, J. P. Effect of phosphate activating group on oligonucleotide formation on montmorillonite: the regioselective formation of 3',5'-linked oligoadenylates. *J. Am. Chem. Soc.* **116**, 10914–10920 (1994).

201. Lohrmann, R. Formation of nucleoside 5′-phosphoramidates under potentially prebiological conditions. *J. Mol. Evol.* **10**, 137–154 (1977).

202. Lohrmann, R. Formation of nucleoside 5′-polyphosphates under potentially prebiological conditions. *J. Mol. Evol.* **8**, 197–210 (1976).

203. Yi, R., Hongo, Y. & Fahrenbach, A. C. Synthesis of imidazole-activated ribonucleotides using cyanogen chloride. *Chem. Commun.* **54**, 511–514 (2018).

204. Burcar, B. T., Jawed, M., Shah, H. & McGown, L. B. In situ imidazole activation of ribonucleotides for abiotic RNA oligomerization reactions. *Orig. Life Evol. Biospheres* **45**, 31–40 (2015).

205. Sulston, J., Lohrmann, R., Orgel, L. E. & Miles, H. T. Nonenzymatic synthesis of oligoadenylates on a polyuridylic acid template. *Proc. Natl. Acad. Sci.* **59**, 726–733 (1968).

206. Lohrmann, R., Bridson, P. K. & Orgel, L. E. Efficient metal-ion catalyzed template-directed oligonucleotide synthesis. *Science* **208**, 1464–1465 (1980).

207. Inoue, T. & Orgel, L. E. Substituent control of the poly(C)-directed oligomerization of guanosine 5'-phosphoroimidazolide. *J. Am. Chem. Soc.* **103**, 7666–7667 (1981).

208. Lohrmann, R. & Orgel, L. E. Template-directed synthesis of high molecular weight polynucleotide analogues. *Nature* **261**, 342–344 (1976).

209. Zielinski, W. S. & Orgel, L. E. Oligomerization of activated derivatives of 3′-amino-3′-deoxyguanosine on poly(C) and poly(dC) templates. *Nucleic Acids Res.* **13**, 2469–2484 (1985).

210. Zielinski, W. S. & Orgel, L. E. Oligoaminonucleoside phosphoramidates. Oligomerization of dimers of 3′-amino-3′-deoxy-nucleotides (GC and CG) in aqueous solution. *Nucleic Acids Res.* **15**, 1699–1715 (1987).

211. Zielinski, W. S. & Orgel, L. E. Autocatalytic synthesis of a tetranucleotide analogue. *Nature* **327**, 346–347 (1987).

212. Tohidi, M., Zielinski, W. S., Chen, C.-H. B. & Orgel, L. E. Oligomerization of 3′-amino-3′-deoxyguanosine-5′-phosphorimidazolidate on a d(CpCpCpCpC) template. *J. Mol. Evol.* **25**, 97–99 (1987).

213. Kiedrowski, G. von, Wlotzka, B., Helbing, J., Matzen, M. & Jordan, S. Parabolic growth of a self-replicating hexadeoxynucleotide bearing a 3'-5'-phosphoamidate linkage. *Angew. Chem. Int. Ed.* **30**, 423–426 (1991).

214. Sievers, D. & Kiedrowski, G. von. Self-replication of complementary nucleotide-based oligomers. *Nature* **369**, 221–224 (1994).

215. Zhang, S., Zhang, N., Blain, J. C. & Szostak, J. W. Synthesis of N3′-P5′-linked phosphoramidate DNA by nonenzymatic template-directed primer extension. *J. Am. Chem. Soc.* **135**, 924–932 (2013).

216. Ferris, J. P. & Ertem, G. Montmorillonite catalysis of RNA oligomer formation in aqueous solution. A model for the prebiotic formation of RNA. *J. Am. Chem. Soc.* **115**, 12270–12275 (1993).

217. Ferris, J. P. Montmorillonite-catalysed formation of RNA oligomers: the possible role of catalysis in the origins of life. *Philos. Trans. R. Soc. B Biol. Sci.* **361**, 1777–1786 (2006).

218. Ertem, G. & Ferris, J. P. Formation of RNA oligomers on montmorillonite: site of catalysis. *Orig. Life Evol. Biospheres* **28**, 485–499 (1998).

219. Kaddour, H. *et al.* Nonenzymatic RNA oligomerization at the mineral–water interface: an insight into the adsorption–polymerization relationship. *J. Phys. Chem. C* **122**, 29386–29397 (2018).

220. Miyakawa, S. & Ferris, J. P. Sequence- and regioselectivity in the montmorillonite-catalyzed synthesis of RNA. *J. Am. Chem. Soc.* **125**, 8202–8208 (2003).

221. Joshi, P. C., Aldersley, M. F. & Ferris, J. P. Progress in demonstrating homochiral selection in prebiotic RNA synthesis. *Adv. Space Res.* **51**, 772–779 (2013).

222. Franchi, M., Ferris, J. P. & Gallori, E. Cations as mediators of the adsorption of nucleic acids on clay surfaces in prebiotic environments. *Orig. Life Evol. Biospheres* **33**, 1–16 (2003).

223. Ertem, G. & Ferris, J. P. Template-directed synthesis using the heterogeneous templates produced by montmorillonite catalysis. a possible bridge between the prebiotic and RNA worlds. *J. Am. Chem. Soc.* **119**, 7197–7201 (1997).

224. Hud, N. V., Jain, S. S., Li, X. & Lynn, D. G. Addressing the problems of base pairing and strand cyclization in template-directed synthesis: a case for the utility and necessity of 'molecular midwives' and reversible backbone linkages for the origin of proto-RNA. *Chem. Biodivers.* **4**, 768–783 (2007).

225. Horowitz, E. D. *et al.* Intercalation as a means to suppress cyclization and promote polymerization of base-pairing oligonucleotides in a prebiotic world. *Proc. Natl. Acad. Sci.* **107**, 5288–5293 (2010).

226. Naylor, R. & Gilham, P. T. Studies on some interactions and reactions of oligonucleotides in aqueous solution. *Biochemistry* **5**, 2722–2728 (1966).

227. Kiedrowski, G. von. A self-replicating hexadeoxynucleotide. *Angew. Chem. Int. Ed.* **25**, 932–935 (1986).

228. Shabarova, Z. A. Chemical development in the design of oligonucleotide probes for binding to DNA and RNA. *Biochimie* **70**, 1323–1334 (1988).

229. Kozlov, I. A. & Orgel, L. E. Nonenzymatic template-directed synthesis of RNA from monomers. *Mol. Biol.* **34**, 781–789 (2000).

230. Orgel, L. E. Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.* **39**, 99–123 (2004).

231. Hänle, E. & Richert, C. Enzyme-free replication with two or four bases. *Angew. Chem. Int. Ed.* **57**, 8911–8915 (2018).

232. Fahrenbach, A. C. *et al.* Common and potentially prebiotic origin for precursors of nucleotide synthesis and activation. *J. Am. Chem. Soc.* **139**, 8780–8783 (2017).

233. Vogel, S. R., Deck, C. & Richert, C. Accelerating chemical replication steps of RNA involving activated ribonucleotides and downstream-binding elements. *Chem. Commun.* 4922–4924 (2005) doi:10.1039/B510775J.

234. Kervio, E., Sosson, M. & Richert, C. The effect of leaving groups on binding and reactivity in enzyme-free copying of DNA and RNA. *Nucleic Acids Res.* **44**, 5504–5514 (2016).

235. Tam, C. P. *et al.* Synthesis of a nonhydrolyzable nucleotide phosphoroimidazolide analogue that catalyzes nonenzymatic RNA primer extension. *J. Am. Chem. Soc.* **140**, 783–792 (2018).

236. Walton, T. & Szostak, J. W. A highly reactive imidazolium-bridged dinucleotide intermediate in nonenzymatic RNA primer extension. *J. Am. Chem. Soc.* **138**, 11996–12002 (2016).

237. Wu, T. & Orgel, L. E. Nonenzymic template-directed synthesis on hairpin oligonucleotides. 2. Templates containing cytidine and guanosine residues. *J. Am. Chem. Soc.* **114**, 5496–5501 (1992).

238. Sosson, M., Pfeffer, D. & Richert, C. Enzyme-free ligation of dimers and trimers to RNA primers. *Nucleic Acids Res.* **47**, 3836–3845 (2019).

239. Deck, C., Jauker, M. & Richert, C. Efficient enzyme-free copying of all four nucleobases templated by immobilized RNA. *Nat. Chem.* **3**, 603–608 (2011).

240. Kervio, E., Claasen, B., Steiner, U. E. & Richert, C. The strength of the template effect attracting nucleotides to naked DNA. *Nucleic Acids Res.* **42**, 7409–7420 (2014).

241. Jauker, M., Griesser, H. & Richert, C. Copying of RNA sequences without pre-activation. *Angew. Chem. Int. Ed.* **54**, 14559–14563 (2015).

242. Adamala, K. & Szostak, J. W. Nonenzymatic template-directed RNA synthesis inside model protocells. *Science* **342**, 1098–1100 (2013).

243. Sosson, M. & Richert, C. Enzyme-free genetic copying of DNA and RNA sequences. *Beilstein J. Org. Chem.* **14**, 603–617 (2018).

244. Schmidt, J. G., Nielsen, P. E. & Orgel, L. E. Information transfer from peptide nucleic acids to RNA by template- directed syntheses. *Nucleic Acids Res.* **25**, 4797–4802 (1997).

245. Koppitz, M., Nielsen, P. E. & Orgel, L. E. Formation of oligonucleotide−PNA−chimeras by template-directed ligation. *J. Am. Chem. Soc.* **120**, 4563–4569 (1998).

246. Yu, H., Zhang, S. & Chaput, J. C. Darwinian evolution of an alternative genetic system provides support for TNA as an RNA progenitor. *Nat. Chem.* **4**, 183–187 (2012).

247. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).

248. Weinger, J. S., Parnell, K. M., Dorner, S., Green, R. & Strobel, S. A. Substrate-assisted catalysis of peptide bond formation by the ribosome. *Nat. Struct. Mol. Biol.* **11**, 1101–1106 (2004).

249. Iannazzo, L. *et al.* Synthesis of 3′-fluoro-tRNA analogues for exploring non-ribosomal peptide synthesis in bacteria. *ChemBioChem* **16**, 477–486 (2015).

250. Wickramasinghe, N. S. M. D., Staves, M. P. & Lacey, J. C. Stereoselective, nonenzymic, intramolecular transfer of amino acids. *Biochemistry* **30**, 2768–2772 (1991).

251. Sievers, A., Beringer, M., Rodnina, M. V. & Wolfenden, R. The ribosome as an entropy trap. *Proc. Natl. Acad. Sci.* **101**, 7897–7901 (2004).

252. Szathmáry, E. The origin of the genetic code: amino acids as cofactors in an RNA world. *Trends Genet.* **15**, 223–229 (1999).

253. Benner, S. A., Kim, H.-J. & Biondi, E. Prebiotic chemistry that could not *not* have happened. *Life* **9**, 84 (2019).

254. Yarus, M. Amino acids as RNA ligands: a direct-RNA-template theory for the code's origin. *J. Mol. Evol.* **47**, 109–117 (1998).

255. Turk-MacLeod, R. M., Puthenvedu, D., Majerfeld, I. & Yarus, M. The plausibility of RNA-templated peptides: simultaneous RNA affinity for adjacent peptide side chains. *J. Mol. Evol.* **74**, 217–225 (2012).

256. Liu, Z., Rigger, L., Rossi, J.-C., Sutherland, J. D. & Pascal, R. Mixed anhydride intermediates in the reaction of 5(4*H*)-oxazolones with phosphate esters and nucleotides. *Chem. – Eur. J.* **22**, 14940–14949 (2016).

257. Biron, J.-P. & Pascal, R. Amino acid *N*-carboxyanhydrides: activated peptide monomers behaving as phosphate-activating agents in aqueous solution. *J. Am. Chem. Soc.* **126**, 9198–9199 (2004).

258. Liu, Z., Beaufils, D., Rossi, J.-C. & Pascal, R. Evolutionary importance of the intramolecular pathways of hydrolysis of phosphate ester mixed anhydrides with amino acids and peptides. *Sci. Rep.* **4**, 7440 (2015).

259. Liu, Z. *et al.* 5(4*H*)-oxazolones as effective aminoacylation reagents for the 3′-terminus of RNA. *Synlett* **28**, 73–77 (2017).

260. Darken, M. A. Puromycin inhibition of protein synthesis. *Pharmacol. Rev.* **16**, 223–243 (1964).

261. Chapuis, H. & Strazewski, P. Shorter puromycin analog synthesis by means of an efficient Staudinger-Vilarrasa coupling. *Tetrahedron* **62**, 12108–12115 (2006).

262. Saneyoshi, H., Michel, B. Y., Choi, Y., Strazewski, P. & Marquez, V. E. Synthesis of conformationally locked versions of puromycin analogues. *J. Org. Chem.* **73**, 9435–9438 (2008).

263. Krishnakumar, K. S., Goudedranche, S., Bouchu, D. & Strazewski, P. The shortest synthetic route to puromycin analogues using a modified Robins approach. *J. Org. Chem.* **76**, 2253–2256 (2011).

264. Moroder, H. *et al.* Non-hydrolyzable RNA–peptide conjugates: a powerful advance in the synthesis of mimics for 3′-peptidyl tRNA termini. *Angew. Chem. Int. Ed.* **48**, 4056–4060 (2009).

265. Steger, J. *et al.* Efficient access to nonhydrolyzable initiator tRNA based on the synthesis of 3′-azido-3′-deoxyadenosine RNA. *Angew. Chem. Int. Ed.* **49**, 7470–7472 (2010).

266. Steger, J. & Micura, R. Functionalized polystyrene supports for solid-phase synthesis of glycyl-, alanyl-, and isoleucyl-RNA conjugates as hydrolysis-resistant mimics of peptidyl-tRNAs. *Bioorg. Med. Chem.* **19**, 5167–5174 (2011).

267. Graber, D. *et al.* Reliable semi-synthesis of hydrolysis-resistant 3′-peptidyl-tRNA conjugates containing genuine tRNA modifications. *Nucleic Acids Res.* **38**, 6796–6802 (2010).

268. Chemama, M. *et al.* Stable analogues of aminoacyl-tRNA for inhibition of an essential step of bacterial cell-wall synthesis. *J. Am. Chem. Soc.* **129**, 12642–12643 (2007).

269. Chemama, M., Fonvielle, M., Arthur, M., Valéry, J.-M. & Etheve-Quelquejeu, M. Synthesis of stable aminoacyl-tRNA analogues containing triazole as a bioisoster of esters. *Chem. – Eur. J.* **15**, 1929–1938 (2009).

270. Fonvielle, M. *et al.* Efficient access to peptidyl-RNA conjugates for picomolar inhibition of non-ribosomal $FemX_{Wv}$ aminoacyl transferase. *Chem. - Eur. J.* **19**, 1357–1363 (2013).

271. Santarem, M. *et al.* Synthesis of 3′-triazoyl-dinucleotides as precursors of stable Phe-$tRNA^{Phe}$ and Leu-$tRNA^{Leu}$ analogues. *Bioorg. Med. Chem. Lett.* **24**, 3231–3233 (2014).

272. Fonvielle, M. *et al.* Electrophilic RNA for peptidyl-RNA synthesis and site-specific cross-linking with tRNA-binding enzymes. *Angew. Chem. Int. Ed.* **55**, 13553–13557 (2016).

273. Shim, J. L., Lohrmann, R. & Orgel, L. E. Poly(U)-directed transamidation between adenosine 5'-phosphorimidazolide and 5'-phosphoadenosine 2'(3')-glycine ester. *J. Am. Chem. Soc.* **96**, 5283–5284 (1974).

274. Zielinski, W. S. & Orgel, L. E. Polymerization of a monomeric guanosine derivative in a hydrogen-bonded aggregate. *J. Mol. Evol.* **29**, 367–369 (1989).

275. Li, L., Foley, J. P. & Helmy, R. Simultaneous separation of small interfering RNA and lipids using ion-pair reversed-phase liquid chromatography. *J. Chromatogr. A* **1601**, 145–154 (2019).

276. Heron, S., Maloumbi, M.-G., Dreux, M., Verette, E. & Tchapla, A. Method development for a quantitative analysis performed without any standard using an evaporative light-scattering detector. *J. Chromatogr. A* **1161**, 152–156 (2007).

277. El Haddad, M. *et al.* New approach for inorganic anion analysis by ion exchange chromatography coupled with evaporative light scattering detection. *J. Liq. Chromatogr. Relat. Technol.* **26**, 3117–3128 (2003).

278. Zagorevskii, D. V., Aldersley, M. F. & Ferris, J. P. MALDI analysis of oligonucleotides directly from montmorillonite. *J. Am. Soc. Mass Spectrom.* **17**, 1265–1270 (2006).

279. Burcar, B. T. *et al.* Potential pitfalls in MALDI-TOF MS analysis of abiotically synthesized RNA oligonucleotides. *Orig. Life Evol. Biospheres* **43**, 247–261 (2013).

280. Costanzo, G., Pino, S., Ciciriello, F. & Mauro, E. D. Generation of long RNA chains in water. *J. Biol. Chem.* **284**, 33206–33216 (2009).

281. Costanzo, G. *et al.* Non-enzymatic oligomerization of 3', 5' cyclic AMP. *PLOS ONE* **11**, e0165723 (2016).

282. Franco, A., Ascenso, J. R., Ilharco, L. & Da Silva, J. A. L. Synthesis of ribonucleotides from the corresponding ribonucleosides under plausible prebiotic conditions within self-assembled supramolecular structures. *New J. Chem.* **44**, 2206–2209 (2020).

283. Claridge, T. D. W. Chapter 12 - experimental methods. in *High-resolution NMR techniques in organic chemistry* (Elsevier, 2016).

284. Dal Poggetto, G., Castañar, L., Adams, R. W., Morris, G. A. & Nilsson, M. Dissect and divide: putting NMR spectra of mixtures under the knife. *J. Am. Chem. Soc.* **141**, 5766–5771 (2019).

285. Claridge, T. D. W. Chapter 10 - diffusion NMR spectroscopy. in *High-resolution NMR techniques in organic chemistry* (Elsevier, 2016).

286. Cohen, Y., Avram, L. & Frish, L. Diffusion NMR spectroscopy in supramolecular and combinatorial chemistry: an old parameter -- new insights. *Angew. Chem. Int. Ed.* **44**, 520–554 (2005).

287. Islam, S. *et al.* Detection of potential TNA and RNA nucleoside precursors in a prebiotic mixture by pure shift diffusion-ordered NMR spectroscopy. *Chem. - Eur. J.* **19**, 4586–4595 (2013).

288. Motsch, S., Pfeffer, D. & Richert, C. 2′/3′ regioselectivity of enzyme-free copying of RNA detected by NMR. *ChemBioChem* **21**, (2020).

289. Lombard, J., López-García, P. & Moreira, D. The early evolution of lipid membranes and the three domains of life. *Nat. Rev. Microbiol.* **10**, 507–515 (2012).

290. Walde, P., Cosentino, K., Engel, H. & Stano, P. Giant vesicles: preparations and applications. *ChemBioChem* **11**, 848–865 (2010).

291. Hargreaves, W. R. & Deamer, D. W. Liposomes from ionic, single-chain amphiphiles. *Biochemistry* **17**, 3759–3768 (1978).

292. Apel, C. L., Deamer, D. W. & Mautner, M. N. Self-assembled vesicles of monocarboxylic acids and alcohols: conditions for stability and for the encapsulation of biopolymers. *Biochim. Biophys. Acta BBA - Biomembr.* **1559**, 1–9 (2002).

293. Monnard, P.-A., Apel, C. L., Kanavarioti, A. & Deamer, D. W. Influence of ionic inorganic solutes on self-assembly and polymerization processes related to early forms of life: implications for a prebiotic aqueous medium. *Astrobiology* **2**, 139–152 (2002).

294. Monnard, P.-A. & Deamer, D. W. Preparation of vesicles from nonphospholipid amphiphiles. in *Methods in Enzymology* vol. 372 133–151 (Elsevier, 2003).

295. Maurer, S. E., Deamer, D. W., Boncella, J. M. & Monnard, P.-A. Chemical evolution of amphiphiles: glycerol monoacyl derivatives stabilize plausible prebiotic membranes. *Astrobiology* **9**, 979–987 (2009).

296. Chen, I. A., Salehi-Ashtiani, K. & Szostak, J. W. RNA catalysis in model protocell vesicles. *J. Am. Chem. Soc.* **127**, 13213–13219 (2005).

297. Black, R. A. *et al.* Nucleobases bind to and stabilize aggregates of a prebiotic amphiphile, providing a viable mechanism for the emergence of protocells. *Proc. Natl. Acad. Sci.* **110**, 13272–13276 (2013).

298. Cornell, C. E. *et al.* Prebiotic amino acids bind to and stabilize prebiotic fatty acid membranes. *Proc. Natl. Acad. Sci.* **116**, 17239–17244 (2019).

299. Jin, L., Kamat, N. P., Jena, S. & Szostak, J. W. Fatty acid/phospholipid blended membranes: a potential intermediate state in protocellular evolution. *Small* **14**, 1704077 (2018).

300. Hishida, M., Seto, H., Yamada, N. L. & Yoshikawa, K. Hydration process of multi-stacked phospholipid bilayers to form giant vesicles. *Chem. Phys. Lett.* **455**, 297–302 (2008).

301. Reeves, J. P. & Dowben, R. M. Formation and properties of thin-walled phospholipid vesicles. *J. Cell. Physiol.* **73**, 49–60 (1969).

302. Akashi, K., Miyata, H., Itoh, H. & Kinosita, K. Preparation of giant liposomes in physiological conditions and their characterization under an optical microscope. *Biophys. J.* **71**, 3242–3250 (1996).

303. Lira, R. B., Dimova, R. & Riske, K. A. Giant unilamellar vesicles formed by hybrid films of agarose and lipids display altered mechanical properties. *Biophys. J.* **107**, 1609–1619 (2014).

304. Pons, M., Foradada, M. & Estelrich, J. Liposomes obtained by the ethanol injection method. *Int. J. Pharm.* **95**, 51–56 (1993).

305. Domazou, A. S. & Luisi, P. L. Size distribution of spontaneously formed liposomes by the alcohol injection method. *J. Liposome Res.* **12**, 205–220 (2002).

306. Stano, P. *et al.* Novel camptothecin analogue (gimatecan)-containing liposomes prepared by the ethanol injection method. *J. Liposome Res.* **14**, 87–109 (2004).

307. Pautot, S., Frisken, B. J. & Weitz, D. A. Production of unilamellar vesicles using an inverted emulsion. *Langmuir* **19**, 2870–2879 (2003).

308. Angelova, M. & Dimitrov, D. S. A mechanism of liposome electroformation. in *Trends in Colloid and Interface Science II* (ed. Degiorgio, V.) vol. 76 59–67 (Steinkopff, 1988).

309. Traïkia, M., Warschawski, D. E., Recouvreur, M., Cartaud, J. & Devaux, P. F. Formation of unilamellar vesicles by repetitive freeze-thaw cycles: characterization by electron microscopy and ^{31}P-nuclear magnetic resonance. *Eur. Biophys. J.* **29**, 184–195 (2000).

310. Mui, B., Chow, L. & Hope, M. J. Extrusion technique to generate liposomes of defined size. in *Methods in Enzymology* vol. 367 3–14 (Academic Press, 2003).

311. Fanti, A., Gammuto, L., Mavelli, F., Stano, P. & Marangoni, R. Do protocells preferentially retain macromolecular solutes upon division/fragmentation? A study based on the extrusion of POPC giant vesicles. *Integr. Biol.* **10**, 6–17 (2018).

312. Lundahl, P., Zeng, C.-M., Lagerquist Hägglund, C., Gottschalk, I. & Greijer, E. Chromatographic approaches to liposomes, proteoliposomes and biomembrane vesicles. *J. Chromatogr. B. Biomed. Sci. App.* **722**, 103–120 (1999).

313. Tamba, Y., Terashima, H. & Yamazaki, M. A membrane filtering method for the purification of giant unilamellar vesicles. *Chem. Phys. Lipids* **164**, 351–358 (2011).

314. Karal, M. A. S. *et al.* Low cost non-electromechanical technique for the purification of giant unilamellar vesicles. *Eur. Biophys. J.* (2019) doi:10.1007/s00249-019-01363-6.

315. Fayolle, D., Fiore, M., Stano, P. & Strazewski, P. Rapid purification of giant lipid vesicles by microfiltration. *Plos One* **13**, e0192975 (2018).

316. Neice, A. Chapter 3 - methods and limitations of subwavelength imaging. in *Advances in Imaging and Electron Physics* (ed. Hawkes, P. W.) vol. 163 117–140 (Elsevier, 2010).

317. Pereira de Souza, T., Steiniger, F., Stano, P., Fahr, A. & Luisi, P. L. Spontaneous crowding of ribosomes and proteins inside vesicles: a possible mechanism for the origin of cell metabolism. *ChemBioChem* **12**, 2325–2330 (2011).

318. Stewart, J. C. Colorimetric determination of phospholipids with ammonium ferrothiocyanate. *Anal. Biochem.* **104**, 10–14 (1980).

319. Sotirhos, N., Herslöf, B. & Kenne, L. Quantitative analysis of phospholipids by 31P-NMR. *J. Lipid Res.* **27**, 386–392 (1986).

320. Fiore, M., Maniti, O., Girard-Egrot, A., Monnard, P.-A. & Strazewski, P. Glass microsphere-supported giant vesicles for the observation of self-reproduction of lipid boundaries. *Angew. Chem. Int. Ed.* **130**, 288–292 (2018).

321. Spitzer, J. & Poolman, B. The role of biomacromolecular crowding, ionic strength, and physicochemical gradients in the complexities of life's emergence. *Microbiol. Mol. Biol. Rev.* **73**, 371–388 (2009).

322. Spitzer, J. From water and ions to crowded biomacromolecules: in vivo structuring of a prokaryotic cell. *Microbiol. Mol. Biol. Rev.* **75**, 491–506 (2011).

323. de Souza, T. P., Fahr, A., Luisi, P. L. & Stano, P. Spontaneous encapsulation and concentration of biological macromolecules in liposomes: an intriguing phenomenon and its relevance in origins of life. *J. Mol. Evol.* **79**, 179–192 (2014).

324. Hanczyc, M. M., Fujikawa, S. M. & Szostak, J. W. Experimental models of primitive cellular compartments: encapsulation, growth, and division. *Science* **302**, 618–622 (2003).

325. Shimizu, Y. & Ueda, T. PURE technology. *Methods Mol. Biol. Clifton NJ* **607**, 11–21 (2010).

326. Stano, P., D'Aguanno, E., Bolz, J., Fahr, A. & Luisi, P. L. A remarkable self-organization process as the origin of primitive functional cells. *Angew. Chem. Int. Ed.* **52**, 13397–13400 (2013).

327. Stano, P. *et al.* Recent biophysical issues about the preparation of solute-filled lipid vesicles. *Mech. Adv. Mater. Struct.* **22**, 748–759 (2015).

328. Lohse, B., Bolinger, P.-Y. & Stamou, D. Encapsulation efficiency measured on single small unilamellar vesicles. *J. Am. Chem. Soc.* **130**, 14372–14373 (2008).

329. Budin, I. & Szostak, J. W. Physical effects underlying the transition from primitive to modern cell membranes. *Proc. Natl. Acad. Sci.* **108**, 5249–5254 (2011).

330. Chakrabarti, A. C. & Deamer, D. W. Permeability of lipid bilayers to amino acids and phosphate. *Biochim. Biophys. Acta BBA - Biomembr.* **1111**, 171–177 (1992).

331. Chakrabarti, A. C. & Deamer, D. W. Permeation of membranes by the neutral form of amino acids and peptides: Relevance to the origin of peptide translocation. *J. Mol. Evol.* **39**, 1–5 (1994).

332. Chakrabarti, A. C., Breaker, R. R., Joyce, G. F. & Deamer, D. W. Production of RNA by a polymerase protein encapsulated within phospholipid vesicles. *J. Mol. Evol.* **39**, 555–559 (1994).

333. Walde, P., Goto, A., Monnard, P.-A., Wessicken, M. & Luisi, P. L. Oparin's reactions revisited: enzymic synthesis of poly(adenylic acid) in micelles and self-reproducing vesicles. *J. Am. Chem. Soc.* **116**, 7541–7547 (1994).

334. Mansy, S. S. *et al.* Template-directed synthesis of a genetic polymer in a model protocell. *Nature* **454**, 122–125 (2008).

335. Wimley, W. C. & White, S. H. Experimentally determined hydrophobicity scales for proteins at membrane interfaces. *Nat. Struct. Biol.* **3**, 842–848 (1996).

336. Nagle, J. F. & Scott, H. L. Lateral compressibility of lipid mono- and bilayers. Theory of membrane permeability. *Biochim. Biophys. Acta BBA - Biomembr.* **513**, 236–243 (1978).

337. Holthuis, J. C. M. & Levine, T. P. Lipid traffic: floppy drives and a superhighway. *Nat. Rev. Mol. Cell Biol.* **6**, 209–220 (2005).

338. Nakano, M. *et al.* Flip-flop of phospholipids in vesicles: kinetic analysis with time-resolved small-angle neutron scattering. *J. Phys. Chem. B* **113**, 6745–6748 (2009).

339. Hamilton, J. A. Fatty acid transport: difficult or easy? *J. Lipid Res.* **39**, 467–481 (1998).

340. Luisi, P. L. & Varela, F. J. Self-replicating micelles — a chemical version of a minimal autopoietic system. *Orig. Life Evol. Biospheres* **19**, 633–643 (1989).

341. Bachmann, P. A., Walde, P., Luisi, P. L. & Lang, J. Self-replicating reverse micelles and chemical autopoiesis. *J. Am. Chem. Soc.* **112**, 8200–8201 (1990).

342. Walde, P., Wick, R., Fresta, M., Mangone, A. & Luisi, P. L. Autopoietic self-reproduction of fatty acid vesicles. *J. Am. Chem. Soc.* **116**, 11649–11654 (1994).

343. Wick, R., Walde, P. & Luisi, P. L. Light microscopic investigations of the autocatalytic self-reproduction of giant vesicles. *J. Am. Chem. Soc.* **117**, 1435–1436 (1995).

344. Blöchliger, E., Blocher, M., Walde, P. & Luisi, P. L. Matrix effect in the size distribution of fatty acid vesicles. *J. Phys. Chem. B* **102**, 10383–10390 (1998).

345. Lonchin, S., Luisi, P. L., Walde, P. & Robinson, B. H. A matrix effect in mixed phospholipid/fatty acid vesicle formation. *J. Phys. Chem. B* **103**, 10910–10916 (1999).

346. Hardy, M. D. *et al.* Self-reproducing catalyst drives repeated phospholipid synthesis and membrane growth. *Proc. Natl. Acad. Sci.* **112**, 8187–8192 (2015).

347. Brea, R., Bhattacharya, A. & Devaraj, N. Spontaneous phospholipid membrane formation by histidine ligation. *Synlett* **28**, 108–112 (2016).

348. Bhattacharya, A., Brea, R. J. & Devaraj, N. K. De novo vesicle formation and growth: an integrative approach to artificial cells. *Chem. Sci.* **8**, 7912–7922 (2017).

349. Malinin, V. S., Haque, Md. E. & Lentz, B. R. The rate of lipid transfer during fusion depends on the structure of fluorescent lipid probes: a new chain-labeled lipid transfer probe pair. *Biochemistry* **40**, 8292–8299 (2001).

350. Pereira de Souza, T. *et al.* New insights into the growth and transformation of vesicles: a free-flow electrophoresis study. *J. Phys. Chem. B* **119**, 12212–12223 (2015).

351. Oberholzer, T., Wick, R., Luisi, P. L. & Biebricher, C. K. Enzymatic RNA replication in self-reproducing vesicles: an approach to a minimal cell. *Biochem. Biophys. Res. Commun.* **207**, 250–257 (1995).

352. Chen, I. A., Roberts, R. W. & Szostak, J. W. The emergence of competition between model protocells. *Science* **305**, 1474–1476 (2004).

353. Chen, I. A. & Szostak, J. W. Membrane growth can generate a transmembrane pH gradient in fatty acid vesicles. *Proc. Natl. Acad. Sci.* **101**, 7965–7970 (2004).

354. Miele, Y. *et al.* Self-division of giant vesicles driven by an internal enzymatic reaction. *Chem. Sci.* **11**, 3228–3235 (2020).

355. Adamala, K. & Szostak, J. W. Competition between model protocells driven by an encapsulated catalyst. *Nat. Chem.* **5**, 495–501 (2013).

356. Banerjee, P., Pal, S., Kundu, N., Mondal, D. & Sarkar, N. A cell-penetrating peptide induces the self-reproduction of phospholipid vesicles: understanding the role of the bilayer rigidity. *Chem. Commun.* **54**, 11451–11454 (2018).

357. Dwars, T., Paetzold, E. & Oehme, G. Reactions in micellar systems. *Angew. Chem. Int. Ed.* **44**, 7174–7199 (2005).

358. Segré, D., Ben-Eli, D., Deamer, D. W. & Lancet, D. The lipid world. *Orig. Life Evol. Biospheres* **31**, 119–145 (2001).

359. Furuuchi, R., Imai, E.-I., Honda, H., Hatori, K. & Matsuno, K. Evolving lipid vesicles in prebiotic hydrothermal environments. *Orig. Life Evol. Biospheres* **35**, 333–343 (2005).

360. Toppozini, L., Dies, H., Deamer, D. W. & Rheinstädter, M. C. Adenosine monophosphate forms ordered arrays in multilamellar lipid matrices: insights into assembly of nucleic acid for primitive life. *PLOS ONE* **8**, e62810 (2013).

361. Olasagasti, F., Kim, H. J., Pourmand, N. & Deamer, D. W. Non-enzymatic transfer of sequence information under plausible prebiotic conditions. *Biochimie* **93**, 556–561 (2011).

362. Kamat, N. P., Tobé, S., Hill, I. T. & Szostak, J. W. Electrostatic localization of RNA to protocell membranes by cationic hydrophobic peptides. *Angew. Chem. Int. Ed.* **54**, 11735–11739 (2015).

363. Kheyfets, B. B. & Mukhin, S. I. Area per lipid in DPPC-cholesterol bilayers: analytical approach. in *Biophysical Society 59th Annual Meeting* (2015). doi:arXiv:1501.02727.

364. Grochmal, A., Prout, L., Makin-Taylor, R., Prohens, R. & Tomas, S. Modulation of reactivity in the cavity of liposomes promotes the formation of peptide bonds. *J. Am. Chem. Soc.* **137**, 12269–12275 (2015).

365. Ohkubo, K., Urabe, K., Usui, S. & Sagawa, T. The role of the membrane-assisted hydrophobic interaction between di-, tri-, or tetrapeptide catalysts and amino acid esters for the enhancement of stereoselective hydrolysis reactions. *Macromol. Rapid Commun.* **17**, 109–116 (1996).

366. Ishigami, T., Suga, K. & Umakoshi, H. Chiral recognition of L-amino acids on liposomes prepared with L -phospholipid. *ACS Appl. Mater. Interfaces* **7**, 21065–21072 (2015).

367. Hirose, M., Ishigami, T., Suga, K. & Umakoshi, H. Liposome membrane as a platform for the L-Pro-catalyzed Michael addition of *trans*-ß-nitrostyrene and acetone. *Langmuir* **31**, 12968–12974 (2015).

368. Guo, L. & Santschi, P. H. Ultrafiltration and its applications to sampling and characterisation of aquatic colloids. in *Environmental Colloids and Particles* (eds. Wilkinson, K. J. & Lead, J. R.) 159–221 (John Wiley & Sons, Ltd, 2007). doi:10.1002/9780470024539.ch4.

369. Wimley, W. C. & White, S. H. Quantitation of electrostatic and hydrophobic membrane interactions by equilibrium dialysis and reverse-phase HPLC. *Anal. Biochem.* **213**, 213–217 (1993).

370. Wimley, W. C., Creamer, T. P. & White, S. H. Solvation energies of amino acid side chains and backbone in a family of host−guest pentapeptides. *Biochemistry* **35**, 5109–5124 (1996).

371. Bishop, C. M., Walkenhorst, W. F. & Wimley, W. C. Folding of ß-sheets in membranes: specificity and promiscuity in peptide model systems. *J. Mol. Biol.* **309**, 975–988 (2001).

372. Hegedüs, Z. *et al.* Foldameric α/ß-peptide analogs of the ß-sheet-forming antiangiogenic anginex: structure and bioactivity. *J. Am. Chem. Soc.* **135**, 16578–16584 (2013).

373. Zhang, W. *et al.* Quantifying binding of ethylene oxide–propylene oxide block copolymers with lipid bilayers. *Langmuir* **33**, 12624–12634 (2017).

374. Brown, J. W. & Huestis, W. H. Structure and orientation of a bilayer-bound model tripeptide: a proton NMR study. *J. Phys. Chem.* **97**, 2967–2973 (1993).

375. Bonechi, C., Ristori, S., Martini, G., Martini, S. & Rossi, C. Study of bradykinin conformation in the presence of model membrane by nuclear magnetic resonance and molecular modelling. *Biochim. Biophys. Acta* **1788**, 708–716 (2009).

376. Claridge, T. D. W. Chapter 2 - introducing high-resolution NMR. in *High-resolution NMR techniques in organic chemistry* (Elsevier, 2016).

377. Bystrov, V. F., Dubrovina, N. I., Barsukov, L. I. & Bergelson, L. D. NMR differentiation of the internal and external phospholipid membrane surfaces using paramagnetic Mn^{2+} and Eu^{3+} ions. *Chem. Phys. Lipids* **6**, 343–350 (1971).

378. Marcotte, I., Separovic, F., Auger, M. & Gagné, S. M. A multidimensional 1H NMR investigation of the conformation of methionine-enkephalin in fast-tumbling bicelles. *Biophys. J.* **86**, 1587–1600 (2004).

379. Jia, L. *et al.* REDOR solid-state NMR as a probe of the membrane locations of membrane-associated peptides and proteins. *J. Magn. Reson.* **253**, 154–165 (2015).

380. Walde, P., Umakoshi, H., Stano, P. & Mavelli, F. Emergent properties arising from the assembly of amphiphiles. Artificial vesicle membranes as reaction promoters and regulators. *Chem Commun* **50**, 10177–10197 (2014).

381. Rajamani, S. *et al.* Lipid-assisted synthesis of RNA-like polymers from mononucleotides. *Orig. Life Evol. Biospheres* **38**, 57–74 (2008).

382. Mungi, C. V. & Rajamani, S. Characterization of RNA-like oligomers from lipid-assisted nonenzymatic synthesis: implications for origin of informational molecules on early earth. *Life* **5**, 65–84 (2015).

383. Blocher, M., Liu, D., Walde, P. & Luisi, P. L. Liposome-assisted selective polycondensation of α-amino acids and peptides. *Macromolecules* **32**, 7332–7334 (1999).

384. Blocher, M., Hitz, T. & Luisi, P. L. Stereoselectivity in the oligomerization of racemic tryptophan *N*-carboxyanhydride (NCA-Trp) as determined by isotope labeling and mass spectrometry. *Helv. Chim. Acta* **84**, 842–848 (2001).

385. Morigaki, K., Dallavalle, S., Walde, P., Colonna, S. & Luisi, P. L. Autopoietic self-reproduction of chiral fatty acid vesicles. *J. Am. Chem. Soc.* **119**, 292–301 (1997).

386. Murillo-Sánchez, S., Beaufils, D., González Mañas, J. M., Pascal, R. & Ruiz-Mirazo, K. Fatty acids' double role in the prebiotic formation of a hydrophobic dipeptide. *Chem. Sci.* **7**, 3406–3413 (2016).

387. Trost, B. M. Cycloadditions via TMM-Pd intermediates: new strategies for asymmetric induction and total synthesis. Presentation in SympCOMet 2018, ENSCP, Paris, France. (2018).

388. Mukaiyama, T. Oxidation-reduction condensation a new method for peptide synthesis. *Synth. Commun.* **2**, 243–265 (1972).

389. Knorre, D., Vlassov, V. & Zarytova, V. Reactive oligonucleotide derivatives and sequence-specific modification of nucleic acids. *Biochimie* **67**, 785–789 (1985).

390. Sumbatyan, N. V. *et al.* The solution synthesis of antisense oligonucleotide-peptide conjugates directly linked via phosphoramide bond by using a fragment coupling approach. *Nucleosides Nucleotides Nucleic Acids* **23**, 1911–1927 (2004).

391. Garipova, I. Y. & Silnikov, V. N. Synthesis of phosphamide conjugates of 5′ mononucleotides with carcinine and its analogs. *Russ. Chem. Bull. Int. Ed.* **51**, 1940–1944 (2002).

392. Zarytova, V. F., Sergeyev, D. S. & Godovikova, T. S. Synthesis of bleomycin A5 oligonucleotide derivatives and site-specific cleavage of the DNA target. *Bioconjug. Chem.* **4**, 189–193 (1993).

393. Grandas, A., Robles, J. & Pedroso, E. Phosphitylation of primary carboxamides. synthesis of peptide-oligonucleotide conjugates with acylphosphoramidate linkages. *Nucleosides Nucleotides Nucleic Acids* **14**, 825–828 (1995).

394. Cahard, D., McGuigan, C. & Balzarini, J. Aryloxy phosphoramidate triesters as Pro-Tides. *Mini Rev. Med. Chem.* **4**, 371–381 (2004).

395. Mehellou, Y., Balzarini, J. & McGuigan, C. Aryloxy phosphoramidate triesters: a technology for delivering monophosphorylated nucleosides and sugars into cells. *ChemMedChem* **4**, 1779–1791 (2009).

396. Lönnberg, T., Ora, M. & Lönnberg, H. Hydrolytic reactions of nucleoside phosphoramidates: kinetics and mechanisms. *Mini-Rev. Org. Chem.* **7**, 33–43 (2010).

397. McGuigan, C., Pathirana, R. N., Mahmood, N., Devine, K. G. & Hay, A. J. Aryl phosphate derivatives of AZT retain activity against HIV1 in cell lines which are resistant to the action of AZT. *Antiviral Res.* **17**, 311–321 (1992).

398. McGuigan, C. *et al.* Anti-cancer ProTides: tuning the activity of BVDU phosphoramidates related to thymectacin. *Bioorg. Med. Chem.* **13**, 3219–3227 (2005).

399. McGuigan, C. *et al.* Phosphoramidate ProTides of 2′- *c* -methylguanosine as highly potent inhibitors of hepatitis C virus. study of their in vitro and in vivo properties. *J. Med. Chem.* **53**, 4949–4957 (2010).

400. Pertusati, F. & McGuigan, C. Diastereoselective synthesis of P-chirogenic phosphoramidate prodrugs of nucleoside analogues (ProTides) via copper catalysed reaction. *Chem. Commun.* **51**, 8070–8073 (2015).

401. Ross, B. S., Ganapati Reddy, P., Zhang, H.-R., Rachakonda, S. & Sofia, M. J. Synthesis of diastereomerically pure nucleotide phosphoramidates. *J. Org. Chem.* **76**, 8311–8319 (2011).

402. Simmons, B., Liu, Z., Klapars, A., Bellomo, A. & Silverman, S. M. Mechanism-based solution to the protide synthesis problem: selective access to sofosbuvir, acelarin, and INX-08189. *Org. Lett.* **19**, 2218–2221 (2017).

403. Tran, K. *et al.* Development of a diastereoselective phosphorylation of a complex nucleoside via dynamic kinetic resolution. *J. Org. Chem.* **80**, 4994–5003 (2015).

404. Atherton, F. R., Openshaw, H. T. & Todd, A. R. 174. Studies on phosphorylation. Part II. The reaction of dialkyl phosphites with polyhalogen compounds in presence of bases. A new method for the phosphorylation of amines. *J. Chem. Soc. Resumed* 660 (1945) doi:10.1039/jr9450000660.

405. Le Corre, S. S., Berchel, M., Couthon-Gourvès, H., Haelters, J.-P. & Jaffrès, P.-A. Atherton–Todd reaction: mechanism, scope and applications. *Beilstein J. Org. Chem.* **10**, 1166–1196 (2014).

406. Gryaznov, S. M. & Sokolova, N. I. A new method for the synthesis of oligodeoxyribonucleotides containing internucleotide phosphoramidate bonds. *Tetrahedron Lett.* **31**, 3205–3208 (1990).

407. Chen, J.-K., Schultz, R. G., Lioyd, D. H. & Gryaznov, S. M. Synthesis of oligodeoxyribonucleotide N3′→P5′ phosphoramidates. *Nucleic Acids Res.* **23**, 2661–2668 (1995).

408. Kers, I., Stawiński, J. & Kraszewski, A. A new synthetic method for the preparation of nucleoside phosphoramidate analogues with the nitrogen atom in bridging positions of the phosphoramidate linkage. *Tetrahedron Lett.* **39**, 1219–1222 (1998).

409. Chen, W. Z. & Zhao, Y. F. The synthesis of amino acid methyl ester 5′-phosphoamidates of protected uridine. *Phosphorus Sulfur Silicon Relat. Elem.* **185**, 2054–2063 (2010).

410. Whalen, L. J., McEvoy, K. A. & Halcomb, R. L. Synthesis and evaluation of phosphoramidate amino acid-based inhibitors of sialyltransferases. *Bioorg. Med. Chem. Lett.* **13**, 301–304 (2003).

411. Winnick, T. & Scott, E. M. Phosphorylated amino acids. *Arch. Biochem.* **12**, 201–208 (1947).

412. Jampel, E., Wakselman, M. & Vilkas, M. Phosphorylation sélective de composés aminés en milieu aqueux. *Tetrahedron Lett.* 3533–3536 (1968).

413. Wu, L. Y. & Berkman, C. E. Synthesis of N-phosphoryl amino acids using bis(9-fluorenylmethyl)phosphite. *Tetrahedron Lett.* **46**, 5301–5303 (2005).

414. Saxon, E. & Bertozzi, C. R. Cell Surface Engineering by a Modified Staudinger Reaction. *Science* **287**, 2007–2010 (2000).

415. Serwa, R. *et al.* Chemoselective Staudinger-Phosphite Reaction of Azides for the Phosphorylation of Proteins. *Angew. Chem. Int. Ed.* **48**, 8234–8239 (2009).

416. Wilkening, I., Signore, G. del & Hackenberger, C. P. R. Synthesis of N,N-disubstituted phosphoramidates via a Lewis acid-catalyzed phosphorimidate rearrangement. *Chem. Commun.* 2932–2934 (2008) doi:10.1039/B802030B.

417. Wuts, P. G. M. & Greene, T. W. Protection for the carboxyl group. in *Greene's Protective Groups in Organic Synthesis* 533–646 (John Wiley & Sons, Ltd, 2006). doi:10.1002/9780470053485.ch5.

418. Ogawa, T., Hatayama, K., Maeda, H. & Kita, Y. Mild and facile cleavage of 2-cyanoethyl ester using sodium sulfide or tetrabutylammonium fluoride. Synthesis of 1,4-dihydropyridine monocarboxylic acids and unsymmetrical 1,4-dihydropyridine dicarboxylates. *Chem. Pharm. Bull. (Tokyo)* **42**, 1579–1589 (1994).

419. Kita, Y. *et al.* Protecting group for carboxyl function: mild and facile cleavage of 2-cyanoethyl ester under non-hydrolytic conditions. *Chem. Pharm. Bull. (Tokyo)* **42**, 147–150 (1994).

420. Hugel, H. M., Bhaskai, K. V. & Longmore, R. W. Cyanomethyl esters: useful protection for carboxylic acids. *Synth. Commun.* **22**, 693–697 (1992).

421. Peyrottes, S., Mestre, B., Burlina, F. & Gait, M. J. The synthesis of peptide-oligonucleotide conjugates by a frangment coupling approach. *Tetrahedron* **54**, 12513–12522 (1998).

422. Lohrmann, R. & Orgel, L. E. Prebiotic activation processes. *Nature* **244**, 418–420 (1973).

423. Kozhemyakin, L. A. & Kozhemyakin, A. L. Compounds comprising disulfide-containing peptides and nitrogenous bases, and medicinal uses thereof. Pat. No US2003/0073618 A1. (2003).

424. Rojas, S. J. A. *et al.* Polymerase-independent analysis of the sequence of polynucleotides. Pat. No WO2006063717 (A2).

425. Simon, E. S., Grabowski, S. & Whitesides, G. M. Convenient syntheses of cytidine 5'-triphosphate, guanosine 5'-triphosphate, and uridine 5'-triphosphate and their use in the preparation of UDP-glucose, UDP-glucuronic acid, and GDP-mannose. *J. Org. Chem.* **55**, 1834–1841 (1990).

426. Joshi, P. C., Dubey, K., Aldersley, M. F. & Sausville, M. Clay catalyzed RNA synthesis under Martian conditions: Application for Mars return samples. *Biochem. Biophys. Res. Commun.* **462**, 99–104 (2015).

427. Baidya, M., Brotzel, F. & Mayr, H. Nucleophilicities and Lewis basicities of imidazoles, benzimidazoles, and benzotriazoles. *Org. Biomol. Chem.* **8**, 1929 (2010).

428. Lenarcik, B. & Ojczenasz, P. The influence of the size and position of the alkyl groups in alkylimidazole molecules on their acid-base properties. *J. Heterocycl. Chem.* **39**, 287–290 (2002).

429. Sanghvi, Y. S. A status update of modified oligonucleotides for chemotherapeutics applications. *Curr. Protoc. Nucleic Acid Chem.* **46**, (2011).

430. Srivastava, S. C., Pandey, D., Srivastava, N. P. & Bajpai, S. P. RNA synthesis by reverse direction process: phosphoramidites and high purity RNAs and introduction of ligands, chromophores, and modifications at 3'-end. in *Current protocols in nucleic acid chemistry* vol. 45 3.20.1-3.20.39 (Wiley WCH, 2011).

431. Thieriet, N., Guibé, F. & Albericio, F. Solid-phase peptide synthesis in the reverse (N → C) direction. *Org. Lett.* **2**, 1815–1817 (2000).

432. Bollhagen, R., Schmiedberger, M., Barlos, K. & Grell, E. A new reagent for the cleavage of fully protected peptides synthesised on 2-chlorotrityl chloride resin. *J. Chem. Soc. Chem. Commun.* 2559 (1994) doi:10.1039/c39940002559.

433. Thermo Fisher Scientific. DNAPac PA-100 Columns. (2015).

434. Gilles, M. A., Hudson, A. Q. & Borders, C. L. Stability of water-soluble carbodiimides in aqueous solution. *Anal. Biochem.* **184**, 244–248 (1990).

435. Williams, A. & Ibrahim, I. T. A new mechanism involving cyclic tautomers for the reaction with nucleophiles of the water-soluble peptide coupling reagent 1-ethyl-3-(3'-(dimethylamino)propyl)carbodiimide (EDC). *J. Am. Chem. Soc.* **103**, 7090–7095 (1981).

436. Seel, F. & Klein, N. *N*-methylcarbamoylphosphate, i. Synthese / *N*-methylcarbamoyl phosphates, i. synthesis. *Z. Für Naturforschung B* **38**, 797–803 (1983).

437. Christie, S. M. H., Elmore, D. T., Kenner, G. W., Todd, A. R. & Weymouth, F. J. Nucleotides. Part XXII. Syntheses of P1P2-diadenosine-5' and P1P2-diuridine-5' pyrophosphates. *J. Chem. Soc.* 2947–2953 (1953) doi:10.1039/jr9530002947.

438. Lucas, L. H. & Larive, C. K. Measuring ligand-protein binding using NMR diffusion experiments. *Concepts Magn. Reson. Part A* **20A**, 24–41 (2004).

439. Pagès, G., Gilard, V., Martino, R. & Malet-Martino, M. Pulsed-field gradient nuclear magnetic resonance measurements (PFG NMR) for diffusion ordered spectroscopy (DOSY) mapping. *Analyst* **142**, 3771–3796 (2017).

440. Hwang, T. L. & Shaka, A. J. Water suppression that works. excitation sculpting using arbitrary wave-forms and pulsed-field gradients. *J. Magn. Reson. A* **112**, 275–279 (1995).

441. Franck, J. M., Scott, J. A. & Han, S. Nonlinear scaling of surface water diffusion with bulk water viscosity of crowded solutions. *J. Am. Chem. Soc.* **135**, 4175–4178 (2013).

442. Wu, X., Wu, Z. & Han, S. Chromogenic and fluorogenic detection of a nerve agent simulant with a rhodamine-deoxylactam based sensor. *Chem. Commun.* **47**, 11468 (2011).

443. Peng, T. & Yang, D. Construction of a library of rhodol fluorophores for developing new fluorescent probes. *Org. Lett.* **12**, 496–499 (2010).

444. Chen, X. *et al.* Fluorophore-labeled *S*-nitrosothiols. *J. Org. Chem.* **66**, 6064–6073 (2001).

445. Crane, C. M. *et al.* Fluorescent inhibitors for IspF, an enzyme in the non-mevalonate pathway for isoprenoid biosynthesis and a potential target for antimalarial therapy. *Angew. Chem. Int. Ed.* **45**, 1069–1074 (2006).

446. Takeuchi, T. HPLC of amino acids as dansyl and dabsyl derivatives. in *Quantification of amino acids and amines by chromatography* vol. 70 229–241 (Elsevier, 2005).

447. Dujols, F., Marty, L. & Mulliez, M. Phospholidines incorporating a ß N-sulfonylaminoalcohol moiety: first observed selectivity of phosphorus heterocycle aminolysis in the presence of water. *Org Biomol Chem* **3**, 227–232 (2005).

448. Gissot, A., Camplo, M., Grinstaff, M. W. & Barthélémy, P. Nucleoside, nucleotide and oligonucleotide based amphiphiles: a successful marriage of nucleic acids with lipids. *Org. Biomol. Chem.* **6**, 1324–1333 (2008).

449. Baillet, J., Desvergnes, V., Hamoud, A., Latxague, L. & Barthélémy, P. Lipid and nucleic acid chemistries: combining the best of both worlds to construct advanced biomaterials. *Adv. Mater.* **30**, 1705078 (2018).

450. Grayson, N. A. & Westkaemper, R. B. Stable analogs of acyl adenylates. Inhibition of acetyl- and acyl-CoA synthetase by adenosine 5′-alkylphosphates. *Life Sci.* **43**, 437–444 (1988).

451. Campins, N., Dieudonné, P., Grinstaff, M. W. & Barthélémy, P. Nanostructured assemblies from nucleotide-based amphiphiles. *New J. Chem.* **31**, 1928–1934 (2007).

452. Fernández-García, C. & Powner, M. Selective acylation of nucleosides, nucleotides, and glycerol-3-phosphocholine in water. *Synlett* **28**, 78–83 (2016).

453. Johansson, S. J. R. *et al.* A convenient protocol for the synthesis of fatty acid amides. *Synlett* **30**, 213–217 (2019).

454. Brachwitz, H., Bergmann, J., Thomas, Y., Wollny, T. & Langen, P. Synthesis and antiproliferative potency of 9-ß-D-arabinofuranosyl-2-fluoroadenine phospholipid adducts. *Bioorg. Med. Chem.* **7**, 1195–1200 (1999).

455. Ma, X.-B. & Zhao, Y.-F. Synthesis of N-phosphoryl di(or tri)-peptides through the active ester method. *Phosphorus Sulfur Silicon Relat. Elem.* **61**, 9–18 (1991).

456. Leveau, G., Kervio, E. & Richert, C. Solution-phase synthesis and characterization of fluorescent peptido RNA. (Confidential Master thesis, Universität Stuttgart, 2018).

457. Roseman, M. A. Hydrophobicity of the peptide C□O ··H□N hydrogen-bonded group. *J. Mol. Biol.* **201**, 621–623 (1988).

458. Jacobs, R. E. & White, S. H. The nature of the hydrophobic binding of small peptides at the bilayer interface: implications for the insertion of transbilayer helices. *Biochemistry* **28**, 3421–3437 (1989).

459. MacDonald, M. J., Lavis, L. D., Hilvert, D. & Gellman, S. H. Evaluation of the Ser-His dipeptide, a putative catalyst of amide and ester hydrolysis. *Org. Lett.* **18**, 3518–3521 (2016).

460. Wieczorek, R., Dörr, M., Chotera, A., Luisi, P. L. & Monnard, P.-A. Formation of RNA phosphodiester bond by histidine-containing dipeptides. *ChemBioChem* **14**, 217–223 (2013).

461. IUPAC-IUB joint commission on biochemical nomenclature (JCBN): abbreviations and symbols for the description of conformations of polynucleotide chains. *Pure Appl. Chem.* **56**, 595–624 (1983).

462. Cornish-Bowden, A. Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. *Nucleic Acids Res.* **13**, 3021–3030 (1985).

463. Eissler, S. *et al.* Substitution determination of Fmoc-substituted resins at different wavelengths. *J. Pept. Sci.* **23**, 757–762 (2017).

464. Furukawa, Y., Kim, H.-J., Hutter, D. & Benner, S. A. Abiotic regioselective phosphorylation of adenosine with borate in formamide. *Astrobiology* **15**, 259–267 (2015).

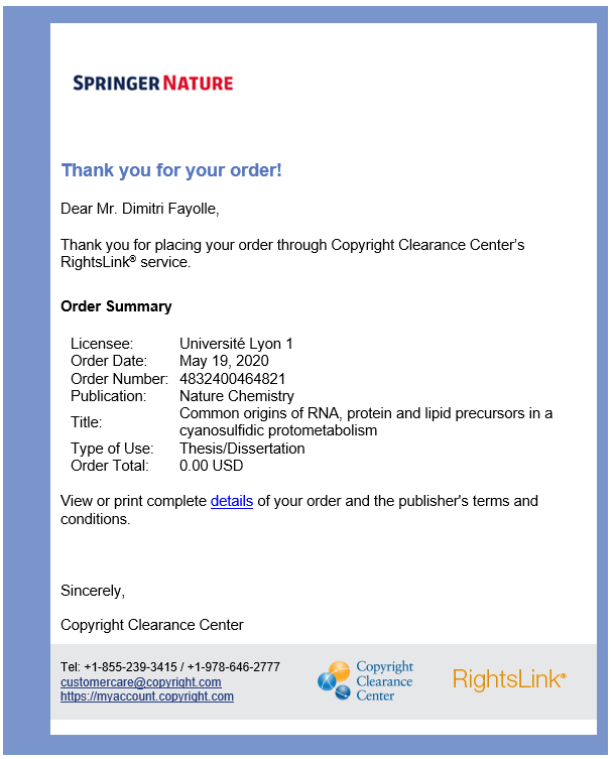
465. Brotsman, V. A. *et al.* Tightly bound double-caged [60]fullerene derivatives with enhanced solubility: structural features and application in solar cells. *Chem. – Asian J.* **12**, 1075–1086 (2017).

466. Huang, K. & Frey, P. A. Engineering human Fhit, a diadenosine triphosphate hydrolase, into an efficient dinucleoside polyphosphate synthase. *J. Am. Chem. Soc.* **126**, 9548–9549 (2004).

467. Lira, L. M., Vasilev, D., Pilli, R. A. & Wessjohann, L. A. One-pot synthesis of organophosphate monoesters from alcohols. *Tetrahedron Lett.* **54**, 1690–1692 (2013).

7. License agreement for copyrighted material

The figures reproduced in this work from external sources were obtained in accordance with copyright laws. Material under Creative Commons License was cited and reused as intended. Material under an Editor's copyright was requested following the Editor's procedures, as justified by the following License Agreements.



SPRINGER NATURE

Thank you for your order!

Dear Mr. Dimitri Fayolle,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Université Lyon 1
Order Date:	May 19, 2020
Order Number:	4832400464821
Publication:	Nature Chemistry
Title:	Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism
Type of Use:	Thesis/Dissertation
Order Total:	0.00 USD

View or print complete details of your order and the publisher's terms and conditions.

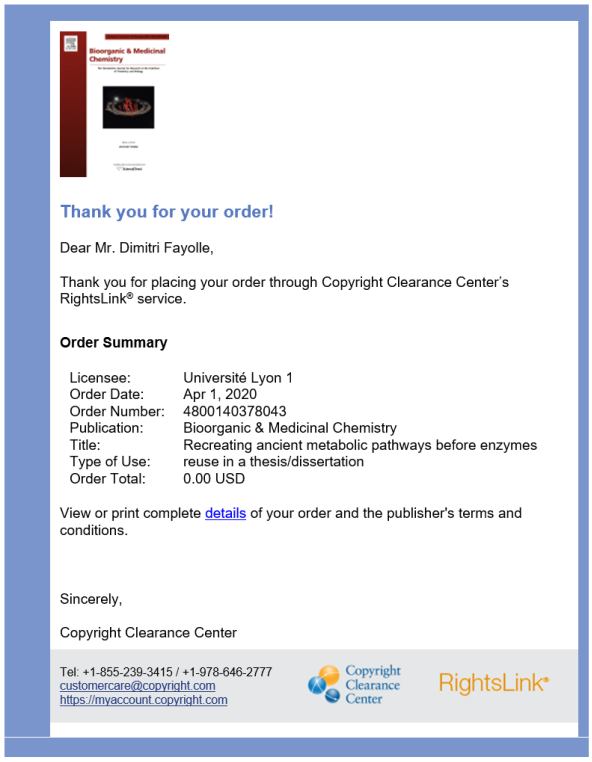
Sincerely,

Copyright Clearance Center

Tel: +1-855-239-3415 / +1-978-646-2777
customercare@copyright.com
https://myaccount.copyright.com

Copyright Clearance Center RightsLink®

Reproduction of Figure 1 from Patel and coll.[9] integrated in this document as Figure 2.4 on page 50.



Thank you for your order!

Dear Mr. Dimitri Fayolle,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Université Lyon 1
Order Date:	Apr 1, 2020
Order Number:	4800140378043
Publication:	Bioorganic & Medicinal Chemistry
Title:	Recreating ancient metabolic pathways before enzymes
Type of Use:	reuse in a thesis/dissertation
Order Total:	0.00 USD

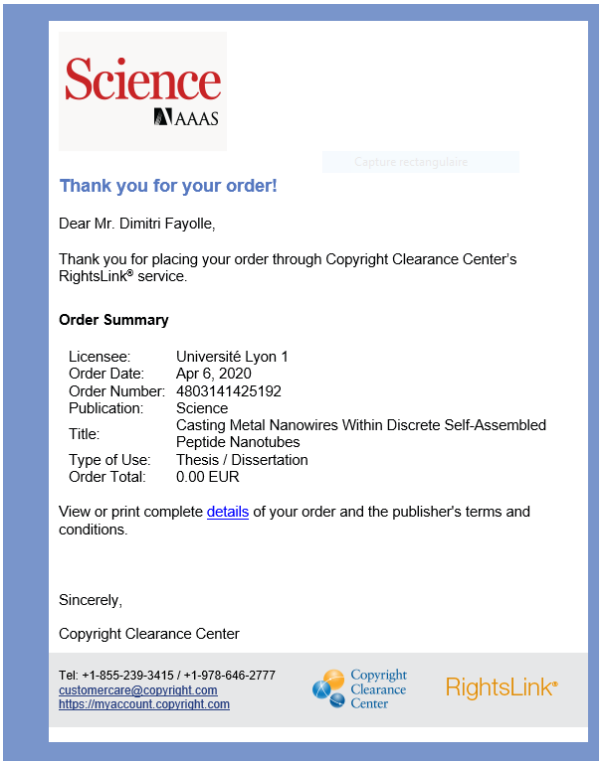
View or print complete details of your order and the publisher's terms and conditions.

Sincerely,

Copyright Clearance Center

Tel: +1-855-239-3415 / +1-978-646-2777
customercare@copyright.com
https://myaccount.copyright.com

Reproduction of Figure 4 from Muchowska and coll.[76] integrated in this document as Figure 2.5 on page 53.



Thank you for your order!

Dear Mr. Dimitri Fayolle,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Université Lyon 1
Order Date:	Apr 6, 2020
Order Number:	4803141425192
Publication:	Science
Title:	Casting Metal Nanowires Within Discrete Self-Assembled Peptide Nanotubes
Type of Use:	Thesis / Dissertation
Order Total:	0.00 EUR

View or print complete details of your order and the publisher's terms and conditions.

Sincerely,

Copyright Clearance Center

Tel: +1-855-239-3415 / +1-978-646-2777
customercare@copyright.com
https://myaccount.copyright.com

Reproduction of Figure 3 from Reches and Gazit[176] integrated in this document as Figure 2.8 on page 67

SPRINGER NATURE

Thank you for your order!

Dear Mr. Dimitri Fayolle,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Université Lyon 1
Order Date:	Apr 7, 2020
Order Number:	4803561451652
Publication:	Nature
Title:	A self-replicating peptide
Type of Use:	Thesis/Dissertation
Order Total:	0.00 EUR

View or print complete details of your order and the publisher's terms and conditions.

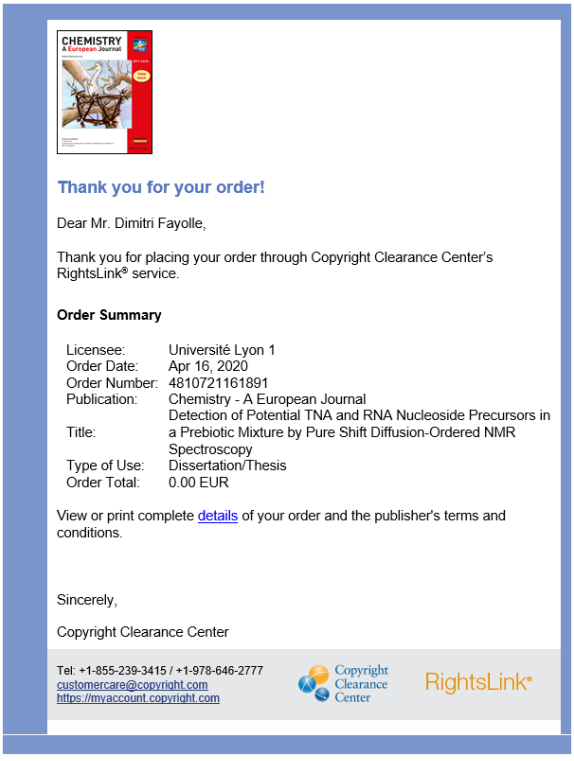
Sincerely,

Copyright Clearance Center

Tel: +1-855-239-3415 / +1-978-646-2777
customercare@copyright.com
https://myaccount.copyright.com

Copyright Clearance Center RightsLink®

Reproduction of Figure 1 from Lee et al[1] integrated in this document as Figure 2.7 on page 65



Thank you for your order!

Dear Mr. Dimitri Fayolle,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Université Lyon 1
Order Date:	Apr 16, 2020
Order Number:	4810721161891
Publication:	Chemistry - A European Journal
Title:	Detection of Potential TNA and RNA Nucleoside Precursors in a Prebiotic Mixture by Pure Shift Diffusion-Ordered NMR Spectroscopy
Type of Use:	Dissertation/Thesis
Order Total:	0.00 EUR

View or print complete details of your order and the publisher's terms and conditions.

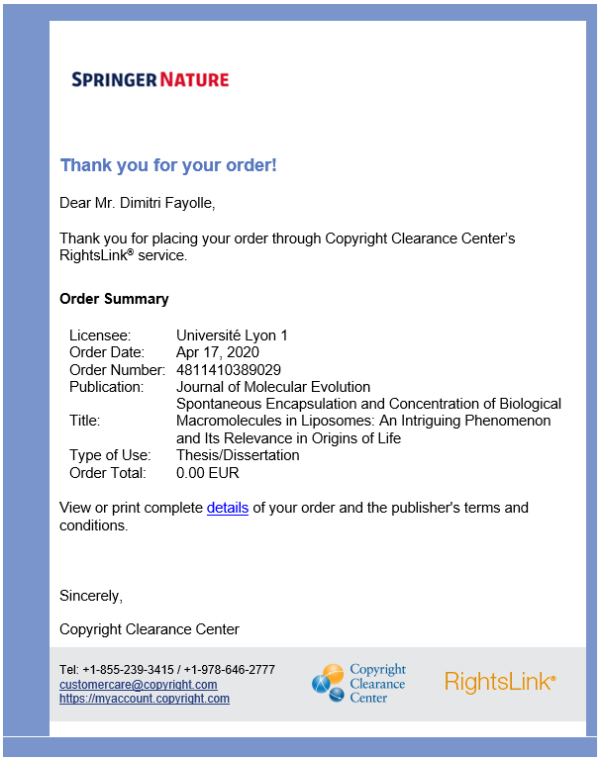
Sincerely,

Copyright Clearance Center

Tel: +1-855-239-3415 / +1-978-646-2777
customercare@copyright.com
https://myaccount.copyright.com

Copyright Clearance Center RightsLink®

Reproduction of Figure 6 from Islam and coll.[287] integrated in this document as Figure 2.17 on page 86



SPRINGER NATURE

Thank you for your order!

Dear Mr. Dimitri Fayolle,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Université Lyon 1
Order Date:	Apr 17, 2020
Order Number:	4811410389029
Publication:	Journal of Molecular Evolution
Title:	Spontaneous Encapsulation and Concentration of Biological Macromolecules in Liposomes: An Intriguing Phenomenon and Its Relevance in Origins of Life
Type of Use:	Thesis/Dissertation
Order Total:	0.00 EUR

View or print complete details of your order and the publisher's terms and conditions.

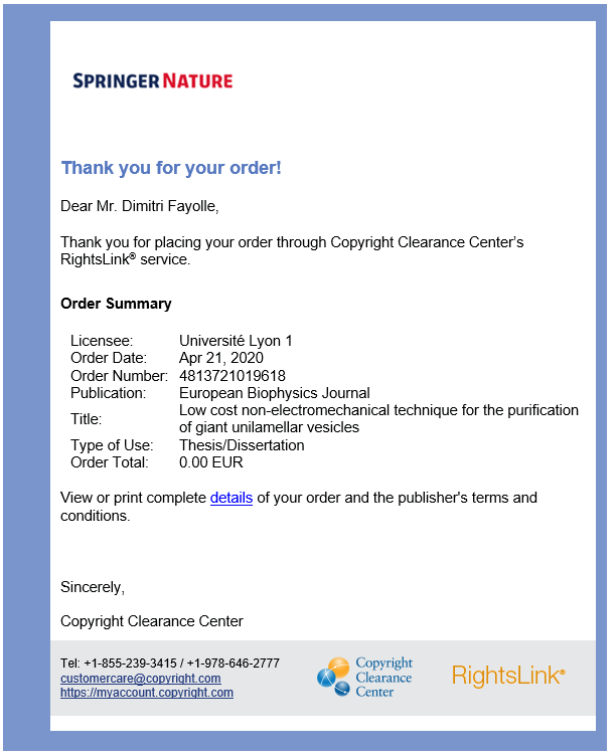
Sincerely,

Copyright Clearance Center

Tel: +1-855-239-3415 / +1-978-646-2777
customercare@copyright.com
https://myaccount.copyright.com

Copyright Clearance Center RightsLink®

Reproduction of Figures 2a, 3b and 6 from Stano and coll.[323] integrated in this document as Figure 2.23 on page 96 and Figure 2.24 on page 97.



SPRINGER NATURE

Thank you for your order!

Dear Mr. Dimitri Fayolle,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Université Lyon 1
Order Date:	Apr 21, 2020
Order Number:	4813721019618
Publication:	European Biophysics Journal
Title:	Low cost non-electromechanical technique for the purification of giant unilamellar vesicles
Type of Use:	Thesis/Dissertation
Order Total:	0.00 EUR

View or print complete details of your order and the publisher's terms and conditions.

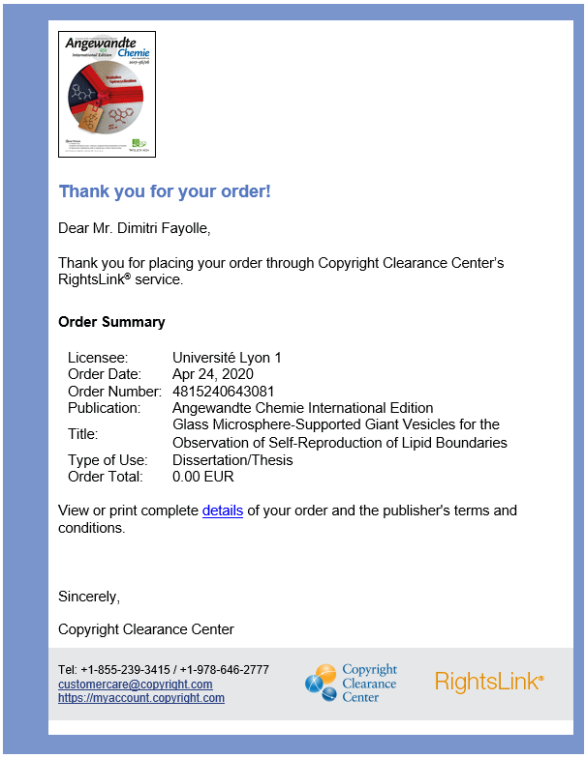
Sincerely,

Copyright Clearance Center

Tel: +1-855-239-3415 / +1-978-646-2777
customercare@copyright.com
https://myaccount.copyright.com

Copyright Clearance Center RightsLink®

Reproduction of Figure 2 from Karal and coll.[314] integrated in this document as Figure 2.20 on page 93.



Thank you for your order!

Dear Mr. Dimitri Fayolle,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Université Lyon 1
Order Date:	Apr 24, 2020
Order Number:	4815240643081
Publication:	Angewandte Chemie International Edition
Title:	Glass Microsphere-Supported Giant Vesicles for the Observation of Self-Reproduction of Lipid Boundaries
Type of Use:	Dissertation/Thesis
Order Total:	0.00 EUR

View or print complete details of your order and the publisher's terms and conditions.

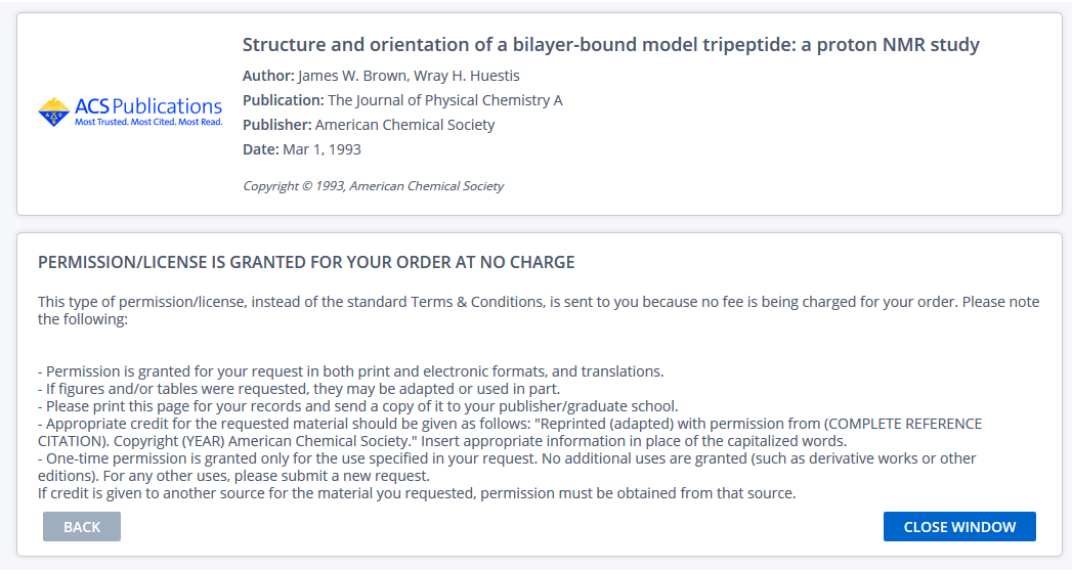
Sincerely,

Copyright Clearance Center

Tel: +1-855-239-3415 / +1-978-646-2777
customercare@copyright.com
https://myaccount.copyright.com

Copyright Clearance Center RightsLink®

Reproduction of Figure 2 from Fiore and coll.[320] integrated in this document as Figure 2.29 on page 101.



ACS Publications
Most Trusted. Most Cited. Most Read.

Structure and orientation of a bilayer-bound model tripeptide: a proton NMR study

Author: James W. Brown, Wray H. Huestis
Publication: The Journal of Physical Chemistry A
Publisher: American Chemical Society
Date: Mar 1, 1993

Copyright © 1993, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK CLOSE WINDOW

Reproduction of Figures 3a and 6 from Brown and Huestis[374] integrated in this document as Figure 2.32 on page 112.

Concomitant synthesis of peptides and oligonucleotides in the presence of lipid vesicles.

Translation of nucleic acids into proteins is a key mechanism to all known forms of life. It relies on a complex biological machinery, and chemists are struggling to mimic translation under conditions that may be representative of the emergence of early forms of life from a prebiotic world. Recently, it was suggested that peptido-RNA, a family of conjugates in which a peptide is bound to an oligonucleotide through a *N*→5' phosphoramidate, could provide a simple example of ribonucleotide-promoted peptide synthesis. In addition, cellular life is dependent on the formation and growth of compartments primarily based on self-assembled lipid membranes. In this work, we study the interaction of amphiphilic peptido-RNA, in which a lipophilic peptide is bound to a hydrophilic nucleotide, with membranes in three different ways. We found that amphiphilic peptido-RNA of random or defined sequences could be formed in the presence of lipid membranes but that these exerted no significant influence on the reaction networks at play. In contrast, peptido-RNA with short and medium-length membranophilic peptides could durably interact with lipid membranes, providing a localization mechanism. Finally, it is envisioned that, because phosphoramidates are easily cleaved, this localization could be exploited to allow the permeation of nucleotides and short nucleic acids through membranes without the assistance of enzymes. Our results contribute to shed light on a prebiotic peptide-RNA-lipid world.

Keywords: peptido-RNA; phosphoramidate; membranophilic peptides; giant vesicles.

Defended by Dimitri Fayolle

Université Claude Bernard Lyon 1, 9 September 2020