



Région Grand Est  
Direction de la Compétitivité et de la Connaissance  
Service Enseignement Supérieur, Recherche et Transfert  
de Technologie  
Contact : [infrarecherche@grandest.fr](mailto:infrarecherche@grandest.fr)

## APPEL A PROJETS 2021

### SOUTIEN AUX PROJETS ET INFRASTRUCTURES DE RECHERCHE DE POINTE

Réserver à l'administration

**N° de dossier :**

(attribué par la Région Grand Est)

**Vos références dossier :** indiquer l'acronyme du pôle scientifique

Les dossiers doivent être envoyés par votre Établissement avant le 30/04/2021 à 18h00 en version électronique à [infrarecherche@grandest.fr](mailto:infrarecherche@grandest.fr), sous forme de fichiers Word/Excel et PDF nommés de la façon suivante : « **nom du projet-infrarecherche2021** ».

Après cette date, les dossiers ne seront pas pris en considération.

**Le dossier doit être rédigé en police ARIAL 10. Merci de respecter les indications concernant le nombre de page pour chaque item à remplir.**

**Titre / acronyme du programme :** **MoIAI4Cryo** : Modeling and Artificial Intelligence applied to Cryo-EM 3D structures to fight COVID and AMR

**Résumé scientifique du programme en français** (550 mots maximum) :

Les défis à relever par les pouvoirs publics en santé publique sont de plus en plus inquiétants, notamment ceux liés à la propagation de pathogènes. D'une part, l'infection par le SARS-Cov-2 qui s'est propagé dans le monde entier depuis fin 2019, créant une pandémie sans précédent sous le nom de maladie à coronavirus de 2019 ou COVID-19. Même si différents types de vaccins semblent actuellement efficaces sur les souches actuelles du virus, il paraît indispensable de mieux connaître les caractéristiques d'infection des virus à ARN afin de pouvoir lutter plus efficacement contre d'éventuelles pandémies futures. D'autre part, l'utilisation massive des antibiotiques en santé publique et en agro-alimentaire depuis des décennies, a généré l'émergence de nombreuses résistances aux antimicrobiens (AMR) chez les bactéries, posant aujourd'hui des problèmes notamment en milieu hospitalier avec l'émergence de bactéries multi-résistantes. Ces résistances se transmettent très efficacement entre bactéries par des mécanismes de transfert horizontaux de gènes.

Ce projet vise principalement à acquérir un équipement scientifique de pointe : un **cryo-microscope électronique**. Il permettra de mieux comprendre les mécanismes moléculaires biologiques à fort impact en santé publique. Cet équipement permettra **d'imager avec précisions la structure des virus**, des composants d'une cellule, ou encore des **macro-complexes moléculaires clés** réputés fragiles au plus près de leurs conditions naturelles à l'aide de techniques cryogéniques. Il permettra d'établir des **preuves de concept** en recherche fondamentale afin d'identifier des molécules d'intérêt pour de **nouvelles stratégies anti-infectieuses**. **Le premier défi** de ce projet sera de **décrypter l'interaction entre les protéines du coronavirus et la machinerie d'épissage des ARN messagers cellulaires** (le spliceosome) de l'hôte qui modifie la capacité de la cellule infectée à produire ses protéines pour assurer finalement une réplication efficace du virus. Par exemple, la méthylation de la coiffe située à l'extrémité 5' des ARNm cellulaires est un marqueur épigénétique du soi. Son absence déclenche la réponse immunitaire innée. Le coronavirus est capable de synthétiser une coiffe chimiquement identique lui permettant d'échapper à une réponse cellulaire antivirale. **Le second défi** concerne la conjugaison bactérienne. Nous chercherons à comprendre la structure de l'assemblage moléculaire du pore de conjugaison permettant le transfert d'ADN entre une bactérie donneuse et une bactérie receveuse (ADN codant potentiellement des AMR), notamment chez des bactéries à paroi de type Gram-positif, comme les streptocoques ou entérocoques, souvent responsables d'infections multi-résistantes. Le pore de conjugaison étant un important complexe multiprotéique membranaire, la résolution de sa structure 3D nécessitera l'emploi de la technique de **cryo-microscopie électronique** (Cryo-EM). Ces données permettront par la suite d'envisager des stratégies d'inhibition de ces transferts conjuguatifs afin de limiter les diffusions d'AMR.

Pour l'ensemble de ses défis, il est également nécessaire d'aborder la dynamique et la structure 3D de ces cibles thérapeutiques clés par des approches d'**intelligence artificielle** (IA) lors de l'analyse d'images en **cryo-EM**, de la classification 3D des structures et d'analyse de ces structures par **modélisation et calculs de dynamique moléculaire**.

Ce projet s'inscrit dans le cadre d'une recherche translationnelle pluridisciplinaire en **informatique, IA, chimie, biologie et santé**. La force du présent consortium est qu'il contribue à une **synergie inédite** entre 2 équipes expérimentales et 2 équipes de modélisation reconnues au niveau international et au sein du territoire lorrain de la région Grand-Est dans le domaine de la **biologie structurale intégrative** vers un **pôle unique de compétences** à partir d'expertises complémentaires dans la **biologie et biochimie structurales, le développement et l'application de méthodes d'IA, et de modélisation biomoléculaire**. Vu le contexte sanitaire actuel, un engagement particulier a été pris afin de lutter contre la COVID-19.

**Mots clés liés au programme** (6 mots maximum) :

Cryo-microscopie électronique  
Intelligence Artificielle  
Modélisation  
Macro-complexes protéine-ARN/ADN/protéine  
Protéines membranaires  
*Drug-design*

**Territoire** (Alsace, Champagne-Ardenne, Lorraine) : **Lorraine**

**Unité de recherche coordinatrice** : (acronyme + intitulé) : UMR 7365 CNRS-UL Ingénierie Moléculaire et Physiopathologie Articulaire (IMoPA, Dir : J-Y Jouzeau); Équipe « ARN, RNP, Structure-Fonction – Maturation » (**partenaire 1**)

**Responsable du programme** : (nom + prénom + fonction) : **Xavier Manival, Directeur de Recherche CNRS, responsable du groupe de biologie structurale intégrative de l'équipe.**

**Organisme gestionnaire du programme** :

**Université de Lorraine**  
34 cours Léopold  
CS 25233  
54052 NANCY CEDEX

**SECTEUR/CHAMP DISCIPLINAIRE**

**Principal : BIOLOGIE/SANTE**  
**Secondaire : CHIMIE/NUMERIQUE**

**Indiquez si votre programme correspond aux thèmes décrits (cochez la case concernée)**

**Lien avec un des trois enjeux de transition (SRESRI) :**

Les programmes doivent s'inscrire en cohérence avec les trois enjeux de transition écologique et environnementale, numérique et industrielle et dans les thématiques d'excellence déjà identifiées (santé, chimie, matériaux, ...) fédératrices couvrant l'ensemble des domaines de la recherche et susceptibles de générer des innovations.

- Transition numérique ;
- Transition industrielle ;
- Transition écologique et environnementale.

**Lien avec les domaines de la S3 :**

Les recherches menées doivent pouvoir alimenter les connaissances dans les domaines identifiés de la stratégie de spécialisation intelligente (S3) dont les perspectives de transfert et d'innovation sont avérées.

- Technologies et équipements pour la transition industrielle ;
- Recyclage et fonctionnalisation des matériaux ;
- Biotechnologies médicales ;
- Outils numériques pour la santé ;

- Dispositifs médicaux ;
- Molécules et matériaux biosourcés ;
- Outils et systèmes pour la gestion durable et intelligente des ressources naturelles ;
- Systèmes énergétiques et leur performance.

## LE PROGRAMME :

- **s'inscrit-t-il dans le champ de la « bioéconomie »**  Oui  Non

« La bioéconomie englobe l'ensemble des activités de production et transformation de la biomasse, qu'elle soit agricole, forestière, aquacole ou biodéchets à des fins de production alimentaire, de matériaux biosourcés, de molécules d'intérêt et d'énergie.

<https://www.bioeconomie-grandest.fr/>

**Si Oui, précisez le champ :**

- BIORESSOURCES**  
Biointrants, biocontrôle, agriculture de précision, outils numériques, agro-machinisme, services écosystémiques...
- CHIMIE DU VEGETAL ET BIOTECHNOLOGIES INDUSTRIELLES**  
Cosmétique, peintures, détergence, tensioactifs, arômes et parfums...
- MATERIAUX BIOSOURCÉS**  
Textile, BTP, composites, emballages, sport et loisirs, mobilier et design ...
- BIOÉNERGIES**  
Biogaz, biocarburants, biocombustibles, biochar...
- ALIMENTATION HUMAINE ET ANIMALE**  
Protéines végétales, ingrédients fonctionnels, texturants, arômes et colorants...
- PROCÉDÉS ET TECHNOLOGIES**  
Technologies clés, procédés, catalyse enzymatique...
- AUTRES :**

**Commenter votre sélection :**

- **relève-t-il des thématiques de recherche en Intelligence Artificielle (IA)?**  Oui  Non

**Si oui, précisez le champ :**

- INTÉGRATION DE L'IA**  
Vision artificielle, robotique, traitement des langues naturelles, systèmes multi-agents, interfaces homme machine, science des données, architectures et composants matériels pour l'IA...
- APPLICATION DE L'IA**  
Activités de recherche interdisciplinaire avec des spécialistes des domaines d'applications de l'IA (ex : sécurité et défense ; transport et mobilité ; santé ; environnement)
- CŒUR DE L'IA**  
Algorithmique, apprentissage machine, évolution artificielle, optimisation spécifique pour l'IA, raisonnement symbolique, lien avec les sciences cognitives

**Commenter votre sélection :**

Les approches de simulations moléculaires tout comme la cryo-microscopie électronique génèrent des ensembles massifs de données structurales complexes dont l'analyse est révolutionnée par l'application des techniques d'Intelligence Artificielle. Nos travaux s'appuient sur le développement de méthodes basées sur les réseaux de neurones profonds (*Deep Learning*) permettant de rationaliser et de prédire la dynamique de systèmes biomoléculaires complexes en combinant données expérimentales et théoriques, et traiter les images de cryo-EM.

• **relève-t-il des thématiques en lien avec d'autres politiques régionales telles que**

- Stratégie Hydrogène vert  
(<https://www.grandest.fr/hydrogene-le-nouveau-challenge-de-la-transition-energetique-en-grand-est/>)
- Plan Industrie du Futur  
(<https://www.grandest.fr/wp-content/uploads/2017/07/2178-4Pages-Industrie-du-Futur.pdf>)
- Feuille de route Santé  
(<https://www.grandest.fr/wp-content/uploads/2020/12/feuille-de-route-sante-web-bd.pdf>)

**Si Oui précisez :**

**Axe 2 : l'innovation et la recherche en santé : moteurs et vecteurs de transition – Ambition 1 : mettre en place une démarche d'accélération des innovations en santé structurée autour de trois priorités : Industrie, Académie, Intelligence Artificielle**

**Axe 4 : Soutien aux projets territoriaux de santé dans les programmes européens – Objectifs 1 et 2 : intégrer le volet REACT-EU de la région Grand Est et s'inscrire dans le programme FEDER-FSE 2021-2027 dans le cadre des mesures d'urgence sur la COVID-19.**

---

**Commentaires (le cas échéant, précisez si le programme est en lien avec une thématique émergente) :**

**Oui (voir ci-dessus)**

**Le projet a-t-il un lien avec une contractualisation précédente ou en cours (CPER 201'-2020, CTSC 2018-2020, CRSD...)?**

- Oui ;
- Non ;

Si oui, précisez :

## A. GESTIONNAIRE ET PARTENAIRE DU PROGRAMME

### I. Identification de l'organisme gestionnaire du programme

**Organisme gestionnaire : Université de Lorraine**

**Nature/statut juridique :** Établissement Public à Caractère Scientifique, Culturel et Professionnel, créé sous la forme de grand établissement

**Représentant légal** (nom, prénom, coordonnées) **et fonction :** **M. Pierre MUTZENHARDT** - Président de l'Université de Lorraine

**Adresse :** Université de Lorraine – 34 cours Léopold – BP 25233 – 54052 NANCY Cedex

**N° SIRET (ou SIREN le cas échéant) :** 130 015 506 00012 (130 015 506)

**Contacts Direction de la Recherche et de la Valorisation (DRV) – Université de Lorraine :**

**Contact administratif :**

Nom - Prénom : Mme Clotilde DINÉ      Courriel : [clotilde.dine@univ-lorraine.fr](mailto:clotilde.dine@univ-lorraine.fr)  
Tél. : 03.72.74.04.53

**Contacts financiers :**

Nom - Prénom : Mme Laura Thirion      Courriel : [laura.thirion@univ-lorraine.fr](mailto:laura.thirion@univ-lorraine.fr)  
Tél. : 03.72.74.04.83

Nom - Prénom : M. Mathieu Marchal      Courriel : [mathieu.marchal@univ-lorraine.fr](mailto:mathieu.marchal@univ-lorraine.fr)  
Tél. : 03.72.74.04.34

### II. Présentation de l'unité de recherche impliquée dans le programme

**Identité de l'unité de recherche P1** (acronyme + intitulé) : **UMR 7365 CNRS-UL Ingénierie Moléculaire et Physiopathologie Articulaire (IMoPA)**

**Discipline(s) :** Sciences de la vie – Biologie moléculaire et structurale - Santé

**Rattachement universitaire ou organisme de recherche :**

**Université de Lorraine**  
34 cours Léopold  
CS 25233  
54052 NANCY Cedex

**Adresse** (précisez Unité, Laboratoire, Département, Service, ...) : **Campus Biologie-Santé, Biopôle de l'Université de Lorraine**

Rue : **9 Avenue de la forêt de Haye**

Code postal : **54 505**      Ville : **Vandœuvre-lès-Nancy**

Tél. secrétariat : **03.72.74.65.80**      Courriel secrétariat : [imopa-secretariat@univ-lorraine.fr](mailto:imopa-secretariat@univ-lorraine.fr)

Site web : <http://www.imopa.cnrs.fr/>

**Nom/Prénom du directeur :** **JOUZEAU Jean-Yves**

Courriel : [jean-yves.jouzeau@univ-lorraine.fr](mailto:jean-yves.jouzeau@univ-lorraine.fr)

## **Intitulé de l'équipe** (acronyme + intitulé) : **Équipe « ARN, RNP, Structure-Fonction-Maturation**

Responsable de l'équipe : **Bruno Charpentier (PR1, UL) / Iouri Motorine (PRCE1, UL)**

Tel. : **03 72 74 66 27/03 72 74 66 29** Courriel : **bruno.charpentier@univ-lorraine.fr/yuri.motorin@univ-lorraine.fr**

### **Présentation de l'unité de recherche (P1) :**

#### **Description de l'unité de recherche et de l'équipe qui hébergera le projet. (300 mots maximum)**

IMoPA regroupe 6 équipes conduisant des recherches à l'échelle moléculaire, cellulaire ou de l'organisme entier. Les thèmes principaux sont : (1) les mécanismes régulant la biologie de l'ARN et des RNPs, incluant les aspects de leurs biogenèses et de leurs devenir (2) la régulation et le rôle pathologique des enzymes synthétisant les glycosaminoglycane dans l'arthrose, les maladies rares et le cancer; (3) les relations structure-fonction des enzymes et des protéines apparentées au biomédical et/ou à un intérêt biotechnologique; (4) les mécanismes liant les altérations phénotypiques des chondrocytes et la minéralisation pathologique; (5) la recherche des cibles pharmacologiques et des méthodes de vectorisation pour contrôler l'inflammation articulaire; et (6) l'ingénierie cellulaire et tissulaire à partir de cellules souches mésenchymateuses.

L'**équipe 1** étudie la biogenèse et la fonction de particules ribonucléoprotéiques (**RNPs**) ainsi que les fonctions des ARNs non codants et des complexes ARN-protéines, en situation normale ou pathologique. Ces travaux mettent en œuvre un large répertoire d'approches pluridisciplinaires incluant la biologie moléculaire et cellulaire, la génétique moléculaire, la protéomique, la transcriptomique, la biochimie, la physico-chimie et la biologie structurale. Ce projet s'adresse plus spécifiquement, au groupe de **Biologie Structurale Intégrative** dont X. Manival (DR2-CNRS ; 32 articles ; 600 citations ; h-index 14) est responsable. Il s'intéresse à comprendre le **mécanisme d'action des machineries d'assemblage** dans la biogenèse de RNPs spécialisées, la **snoRNP** à boîtes **C/D** et plus récemment, les **UsnRNPs** qui sont des composants des **spliceosomes**. X. Manival est également directeur scientifique de la plateforme RMN de l'UMS IBSLor et le coordinateur national du comité utilisateur de FRISBI (Infrastructure française pour la biologie structurale intégrée). Il possède une solide expérience en biochimie (co-expression, purification à grande échelle et caractérisation de grands complexes multiprotéique) et biologie structurale (RMN, CRX, SAXS). Au cours des six dernières années, il a déposé plus de 20 structures tridimensionnelles dans la PDB.

Son dernier travail en collaboration avec l'équipe 1 d'IMoPA et le consortium ANR *snoRNPASSEMBLY* porte sur les liens génétiques et physiques entre un facteur d'assemblage de la snoRNP C/D (Bcd1) et une des protéines chaperonnes contrôlant l'architecture de la chromatine (Rtt106). Il vient d'être publié dans la revue scientifique généraliste internationale *Nature Communications* (publication 10).

Notre travail s'appuie sur la plateforme de **Biophysique et Biologie Structurale** (S. Boschi-Muller) de l'UMS 2008 IBSLor (Y. Motorin). Historiquement, le groupe utilise les **rayons X** et la **RMN** pour résoudre les structures 3D de leurs complexes d'intérêt afin d'expliquer leur fonction biologique. Actuellement, nous collaborons avec C. Plisson-Chastang (LBME, Toulouse) pour l'analyse structurale par **cryo-EM** d'un complexe chaperon homologue au R2TP que nous avons découvert, le R2SP (publication 8).

**Activité et reconnaissance de l'unité de recherche (10 publications les plus significatives, distinctions, avis HCERES, Labex, Equipex, ...).**

**Evaluation de l'UMR 7365 CNRS-UL IMoPA pour la période 2011–2016** : L'évaluation par l'HCERES a conduit aux appréciations suivantes de l'unité par le comité d'experts internationaux (traduction littérale du rapport) : (i) qualité scientifique: très bonne à excellente; (ii) réputation et rayonnement: très bons; (iii) interaction avec le mode socio-économique et culturel: très bonne à excellente; (iv) organisation et vie de l'unité: très bonnes; (v) contribution à la formation par la recherche: très bonne; (vi) stratégie et projet à 5 ans: très bons à excellents.

La productivité globale de recherche de l'unité de recherche est excellente au niveau national et très bonne au niveau international avec 175 publications originales évaluées par des pairs, dont 71 dernier auteurs/correspondants. 70% des publications évaluées par des pairs sont dans des journaux en Q1 dont plusieurs dans des journaux à fort impact (IF > 10, voir ci-dessous) au niveau de spécialisation le plus haut. Le bilan inclut également 47 publications dans des revues prestigieuses telles que *Nat. Rev. Genetics* (IF=43), *J. Am. Chem. Soc.* (IF=14), *Nat. Chem. Biol.* (IF=12), *Nat. Commun.* (IF=12), *Nucleic Acids Res.* (IF=11), *Methods in Molecular Biol.* (IF=11), *Biomaterials* (IF=10), *Chem. Sci.* (IF=9,5), *Am. J. Human. Genet.* (IF=9), *J. Cell. Biol.* (IF=8,8), *Arthritis Rheum.* (IF=6,9), *Oncogene* (IF=6,8), *Antioxid. Redox Signal.* (IF=5,8) et *Free Rad. Biol. Med.* (IF=5,7), avec les responsables scientifiques comme auteurs principaux. Toutes les équipes contribuent à la forte production.

**Contribution à des initiatives d'excellence** : L'UMR 7365 IMoPA participe activement au projet I-SITE Lorraine Université d'Excellence. Deux projets Impact, « Biomolécules » et « Geenage », sont co-coordonnés par deux de ses membres (Kira J. Weissman pour Biomolécules et Jean-Yves Jouzeau pour Geenage).



L'UMR IMoPA est également partenaire du « FHU ARRIMAGE » coordonné par Jean-Louis Guéant et du « FHU CURE » (déposé janvier 2020) porté par Laurent Peyrin-Biroulet.

**Recommandation par l'HCERES pour l'axe RNP Structure-Fonction de l'équipe 1 pour la période 2017–2021** (traduction littérale du rapport) : l'axe 1 (**co-responsable X. Manival**) a démontré un **solide record de publications** au cours de cette période (13 articles publiés dans des revues internationales ou nationales avec un comité de lecture répertorié (ACL), 11 comme auteurs correspondant, 5 dans *Nucleic Acids Research*, 1 dans *Structure*). Cet axe est fortement encouragé à poursuivre ses efforts dans le futur. [...] **La cryo-microscopie électronique devrait également être considérée comme une future méthode de choix pour les grands complexes.**

**10 publications les plus significatives** (en gras les noms du laboratoire, en gras surligné les noms participant au projet) : l'équipe publie entre 10 et 15 articles par an dans des journaux à comité de lecture avec un facteur d'impact majoritairement supérieur à 10 tels que *Nature*, *Nucleic Acids Res.*, *Nat. Chem. Biol.*, *Nat. Commun.*, *Angew Chem. Int. Ed.*, *Structure*, ...

1) Watkins N.J., **Ségault V.**, **Charpentier B.**, Nottrott S., Fabrizio P., Bachi A., Wilm M., Rosbash M., **Branlant C.**, and Lührmann R. (2000). A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. *Cell* **103**: 457–466.

2) **Bizarro J.**, **Charron C.**, Boulon S., Westman B., Pradet-Balade B., Vandermoere F., **Chagot M. E.**, Hallais M., Ahmad Y., Leonhardt H., Lamond A., **Manival X.**, **Branlant C.**, **Charpentier B.**, Verheggen C.\* and Bertrand E.\* (2014). Proteomic and 3D structure analyses highlight the C/D box snoRNP assembly mechanism and its control. *J. Cell Biol.* **207**: 463–480.

3) Weldon C., **Behm-Ansmant I.**, Burley GA., Hurley LH., **Branlant C.**, Eperon IC.\* and Dominguez C.\* (2017). Identification of G-quadruplexes in long RNAs and in functional conditions using 7-deaza-Guanine substituted RNA *Nat. Chem. Biol.* **13**:18–20.

4) Eralles J., Marchand V., Panthu B., Gillot S., Belin S., Ghayad S. E., Garcia M., Laforets F., Marcel V., Baudin-Baillieu A., Bertin P., Coute Y., Adrait A., Meyer M., Therizols G., Yusupov M., Namy O., Ohlmann T., **Motorin Y.**, Catez F., Diaz J. J. (2017). Evidence for rRNA 2'-O-methylation plasticity: Control of intrinsic translational capabilities of human ribosomes, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 12943–12939.

5) Maurizy C., **Quinternet M.**, **Abel Y.**, Verheggen C., Santo P-E, Bourguet M., Paiva A-C-F, **Bragantini B.**, **Chagot M-E**, Robert M-C, Abeza C., **Fabre P.**, Fort P., Vandermoere F., Sousa P-M-F, Rain J-C, **Charpentier B.**, Cianféran S., Bandejas T-M, Pradet-Balade B., **Manival X.\*** and Bertrand E.\* (2018). The RPAP3-Cterminal domain identifies R2TP-like quaternary chaperones. *Nat. Commun.* **9**: 2093.

6) Marchand V., **Ayadi L.**, Ernst FGM, Hertler J., **Bourguignon-Igel V.**, **Galvanin A.**, Kotter A., Helm M., Lafontaine DLJ, **Motorin Y** (2018). AlkAniline-Seq: Profiling of m<sup>7</sup> G and m<sup>3</sup> C RNA Modifications at Single Nucleotide Resolution. *Angew Chem. Int. Ed.* **57**: 16785–16790.

7) Ringear M., Marchand V., Decroly E., **Motorin Y.**, Bannasser Y\* (2019). FTSJ3 is an RNA 2'-O-methyltransferase recruited by HIV to avoid innate immune sensing. *Nature* **565**: 500–504.

8) **Chagot, M.E.**, **Quinternet M.**, **Rothé B.**, **Charpentier B.**, Coutant J., **Manival X.** and Lebars I. (2019). The yeast C/D box snoRNA U14 adopts a "weak" K-turn like conformation recognized by the Snu13 core protein in solution. *Biochimie*, 164:70-82. *Biochimie article of the year 2019 award.*

9) **Abel Y.**, Paiva A-C.F., Bizarro J., **Chagot M-E**, Santo P-E, Robert M-C, **Quinternet M.**, F. Vandermoere, P- Sousa M.F., Fort P., **Charpentier B.**, **Manival X.**, Bandejas T-M., Bertrand E.\* and Verheggen C.\* (2020). \* auteur correspondant. NOPCHAP1 is a PAQosome cofactor that helps loading NOP58 on RUVBL1/2 during box C/D snoRNP biogenesis. *Nucleic Acids Res.*, doi: 10.1093/nar/gkaa1226.

10) **Bragantini B.\***, **Charron C.\***, Bourguet M.\*, **Paul A.\***, **Tiotiu D.**, **Rothé B.**, **Marty H.**, Terral G., Hessmann S., Decourty L., **Chagot M-E.**, Strub J-M, **Massenet S.**, Bertrand E., **Quinternet M.**, Saveanu C., Cianferani S., **Labialle S.\***, **Manival X.\*** and **Charpentier B.\*** (2021). \* Co premier auteur; \* auteur correspondant. The box C/D snoRNP assembly factor Bcd1 interacts with the histone chaperone Rtt106 controlling its association with RNA polymerase II. *Nat. Commun.* **12**: 1859.

### III. Présentation des unités de recherche partenaires

**Identité de l'unité de recherche P2** (acronyme + intitulé) : **UMR 1128 INRAE-UL Dynamique des Génomes et Adaptation Microbienne (DynAMic)**

**Discipline(s) : Sciences de la Vie – Biologie moléculaire et structurale – Santé**

**Rattachement universitaire ou organisme de recherche :**

**Université de Lorraine**

34 cours Léopold

CS 25233  
54052 NANCY Cedex

**Adresse** (précisez Unité, Laboratoire, Département, Service, ...): **Dynamique des Génomes et Adaptation Microbienne (DynAMic)**

Rue : **Campus des Aiguillettes, BP 70239**

Code postal : **54506**

Ville : **Vandœuvre-lès-Nancy**

Tél. secrétariat : **0372745149**

Courriel secrétariat : **layla.rateau@univ-lorraine.fr**

Site web : **http://dynamic.univ-lorraine.fr/**

**Nom/Prénom du directeur : Bertrand Aigle**

Courriel : **bertrand.aigle@univ-lorraine.fr**

**Intitulé de l'équipe** (acronyme + intitulé) : **Integrative and Conjugative Elements: Transfert and Adaptation (ICE-TeA)**

Responsable de l'équipe : **Nathalie Leblond-Bourget**

Tel. : **0372745146**

Courriel : **nathalie.leblond@univ-lorraine.fr**

### **Présentation de l'unité de recherche (P2) :**

#### **Description de l'unité de recherche et de l'équipe qui hébergera le programme (300 mots maximum)**

L'unité DynAMic s'intéresse à l'étude des mécanismes qui induisent la plasticité des génomes de bactéries Gram-positives. Elle se focalise sur deux modèles d'étude : les Streptocoques qui colonisent le tractus digestif des animaux, et les Streptomyces retrouvés dans les sols. Le laboratoire s'intéresse à l'impact des transferts horizontaux de gènes médiés par des éléments génétiques mobiles et aux phénomènes de recombinaison. Par ces mécanismes, les génomes de streptocoques peuvent échanger des gènes de virulence ou de résistance aux antibiotiques, tandis que les streptomyces brassent des clusters de gènes codant des métabolites secondaires, incluant des molécules antitumorales ou antimicrobiennes.

L'unité est composée de deux équipes, StrAda (Streptomyces) et ICE-TeA (Streptocoques). L'équipe ICE-TeA est pionnière au niveau international pour l'identification et l'étude des éléments conjugatifs et intégratifs (ICE), très répandus dans l'ensemble des génomes bactériens. Ces ICE codent souvent des gènes de résistance aux antibiotiques ou de virulence, et se transfèrent par conjugaison. L'équipe cherche à comprendre la contribution des ICE dans l'évolution des génomes de streptocoques, et à étudier leur biodiversité. Plus récemment, l'équipe ICE-TeA développe une nouvelle thématique visant à décrypter les mécanismes moléculaires permettant le recrutement de l'ADN de l'ICE et son transfert de la cellule donneuse à la cellule receveuse via un pore de conjugaison apparenté à un système de sécrétion de type IV. Pour cela, l'équipe utilise des approches biochimiques, biophysiques et structurales dans le but de caractériser ces complexes nucléo-protéiques impliqués dans le processus de transfert.

Nathalie Leblond-Bourget et Nicolas Soler (MCF recruté fin 2013), ont déjà coordonné des projets de type PEPS-CNRS, et des projets collaboratifs locaux (LUE-CITRAM-Mirabelle+ et FEDER) avec P4, visant à identifier des inhibiteurs de conjugaison en amont du pore. B. Douzi et N. Soler sont les initiateurs et organisateurs du réseau scientifique ProteinLorraine, créé en 2020, qui structure toutes les équipes lorraines travaillant sur les protéines. Activité et reconnaissance de l'unité de recherche (10 publications les plus significatives, distinctions, avis HCERES, Labex, Equipex, ...) :

#### **Activité et reconnaissance de l'unité de recherche (10 publications les plus significatives, distinctions, avis HCERES, Labex, Equipex, ...) :**

L'évaluation HCERES de l'unité DynAMic pour la période 2011-2016 spécifiait que « la recherche réalisée est globalement excellente. La quantité et la qualité de la production scientifique sont très bonnes. De plus, plusieurs actions sont entreprises afin d'exploiter les résultats de la recherche fondamentale, incluant des brevets de molécules et le développement de partenariats avec des industriels des secteurs pharmaceutique ou du bois. »

#### **10 publications les plus significatives :**

1- Cappelletti J, **Mohamad Ali A**, **Leblond-Bourget N**, Mathiot S, Dhalleine T, **Payot S**, Savko M, Didierjean C, Favier F, **Douzi B** (2021). Structural and Biochemical Analysis of OrfG: The VirB8-like Component of the Conjugative Type IV Secretion System of ICE St3 From *Streptococcus thermophilus*. *Front Mol Biosci*, 10.3389/fmolb.2021.642606.



- 2- Xu X, Risoul V, Byrne D, Champ S, **Douzi B**, Latifi A. (2020). HetL, HetR and PatS form a reaction-diffusion system to control pattern formation in the cyanobacterium nostoc PCC 7120. *Elife*. Aug 7;9:e59190. doi: 10.7554/eLife.59190.
- 3- **Lao J, Guédon G, Lacroix T, Charron-Bourgoin F, Libante V, Loux V, Chiapello H, Payot S, Leblond-Bourget N.** (2020). Abundance, Diversity and Role of ICEs and IMEs in the Adaptation of *Streptococcus salivarius* to the Environment. *Genes (Basel)*. doi: 10.3390/genes11090999.
- 4- **Libante V, Nombre Y, Coluzzi C, Staub J, Guédon G, Gottschalk M, Teatero S, Fittipaldi N, Leblond-Bourget N, Payot S.** (2019). Chromosomal Conjugative and Mobilizable Elements in *Streptococcus suis*: Major Actors in the Spreading of Antimicrobial Resistance and Bacteriocin Synthesis Genes. *Pathogens*. doi: 10.3390/pathogens9010022
- 5- **Soler N, Robert E, Chauvot de Beauchêne I, Monteiro P, Libante V, Maigret B, Staub J, Ritchie DW, Guédon G, Payot S, Devignes M-D, and Leblond-Bourget N** (2019). Characterization of a relaxase belonging to the MOBT family, a widespread family in Firmicutes mediating the transfer of ICEs. *Mob DNA*, 10:18.
- 6- **Tidjani AR, Lorenzi JN, Toussaint M, van Dijk E, Naquin D, Lespinet O, Bontemps C, Leblond P.** (2019). Massive gene flux drives genome diversity between sympatric *Streptomyces conspecifics*. *mBio*. doi: 10.1128/mBio.01533-19.
- 7- Germain E, Guiraud, P Byrne D, **Douzi B**, Djendli M and Maisonneuve E (2019). YtfK activates the stringent response by triggering the alarmone synthetase SpoT in *Escherichia coli*. *Nature Communications*, 2019, 10 (1).
- 8- Nguyen VS, **Douzi B**, Durand E, Roussel A, Cascales E, Cambillau C (2018). Towards a complete structural deciphering of Type VI secretion system. *Curr Opin Struct Biol*. doi: 10.1016/j.sbi.2018.01.007. Epub 2018 Feb 2. Review.
- 9- **Douzi B**, Logger L, Spinelli S, Blangy S, Cambillau C, Cascales E (2018). Structure-Function Analysis of the C-Terminal Domain of the Type VI Secretion TssB Tail Sheath Subunit. *J Mol Biol*. 2836(17)30565-X.
- 10- Bellanger X, **Payot S, Leblond-Bourget N, Guédon G.** (2014). Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. *FEMS Microbiol Rev*.

**Identité de l'unité de recherche P3** (acronyme + intitulé) : **UMR 7019 CNRS-UL Laboratoire de Physique et Chimie Théoriques (Dir : K. Dragi, co-dir : F. Dehez)**

**Discipline(s)** : Physique - Chimie

**Rattachement universitaire ou organisme de recherche :**

**Université de Lorraine**

34 cours Léopold

CS 25233

54052 NANCY Cedex

**Adresse** (précisez Unité, Laboratoire, Département, Service, ...) : **Faculté des Sciences et Techniques, UMR 7019, Laboratoire de Physique et Chimie Théoriques**

Rue : **Boulevard des Aiguillettes**

Code postal : **54 506**

Ville : **Vandœuvre-lès-Nancy**

Tél. secrétariat : **03 72 74 52 72**

Courriel secrétariat : [Severine.Bonenberger@univ-lorraine.fr](mailto:Severine.Bonenberger@univ-lorraine.fr)

Site web : <http://lpct.univ-lorraine.fr>

**Nom/Prénom du directeur/co-directeur** : **Dragi Karevski / Francois Dehez**

Courriel : [Dragi.Karevski@univ-lorraine.fr](mailto:Dragi.Karevski@univ-lorraine.fr) / [Francois.Dehez@univ-lorraine.fr](mailto:Francois.Dehez@univ-lorraine.fr)

**Intitulé de l'équipe** (acronyme + intitulé) : **Équipe « Biophysique et Biochimie »**

Responsable de l'équipe : **Antonio Monari**

Tel. : **03-72-74-52-78**

Courriel : [Antonio.Monari@univ-lorraine.fr](mailto:Antonio.Monari@univ-lorraine.fr)

**Présentation de l'unité de recherche (P3) :**

**Description de l'unité de recherche et de l'équipe qui hébergera le projet. (300 mots maximum)**

L'unité **LPCT** (UMR 7019 CNRS-UL) regroupe **5 équipes** (axes) œuvrant dans différents domaines de la Physique et de la Chimie théoriques : (i) **Biochimie et Biophysique**, (ii) Interaction Rayonnement-Matière, (iii) Etat liquide, (iv) Etat solide, (v) Dynamique et Symétrie. Nos travaux s'organisent autour de la mise au point de méthodes et de codes numériques permettant de décrire la dynamique équilibre et hors équilibre de systèmes complexes. Nous appliquons ces approches à des domaines très variés de la physique (optique quantique, ...) et de la chimie (solvatation en milieux non-conventionnels utilisés dans la Chimie Verte, ...).

Une partie de notre activité concerne la **modélisation moléculaire de phénomènes biochimiques et biophysiques** contrôlant des processus cellulaires clés. A partir de données issues de la biologie structurale, nous utilisons la **mécanique quantique** et la **dynamique moléculaire** pour **prédire** et **rationaliser** le **(dys)fonctionnement d'assemblages biomoléculaires complexes** et leurs interactions avec des principes actifs.

L'équipe **Biochimie-Biophysique (P3)** développe et utilise des approches **théoriques multi-échelles** permettant d'étudier des phénomènes allant de la **croissance de tissus** à la **dynamique fonctionnelle d'assemblages protéiques**. Nous nous intéressons aux cibles membranaires (canaux, récepteurs, transporteurs et enzymes) et aux pathologies qui leurs sont associées, ainsi qu'aux processus d'endommagement et de réparation de l'ADN. Notre équipe est reconnue pour son expertise dans le domaine des simulations moléculaires et des méthodes d'échantillonnages d'évènements rares. Pour extraire les informations pertinentes de la masse de données générée par les **simulations de dynamique moléculaire**, nous développons des approches d'**intelligence artificielle (deep learning)**. Notre équipe est impliquée dans le développement d'un des codes de dynamique moléculaire les plus utilisés au monde (NAMD, > 50 000 utilisateurs) et dans plusieurs collaborations autour de données structurales provenant de la **cryo-microscopie électronique** (canaux pentamérique et protéases, Nury, Schoehn, Gutsche, Grenoble ; ATPases, Chiu, Stanford ; Complexes de la chaîne respiratoire, Sazanov, Vienna). Notre équipe prend part à la recherche sur la COVID par la mise au point de modèles permettant de comprendre l'interaction des protéines du SRAS-Cov-2 avec des inhibiteurs ou avec de l'ARN et de rationaliser les conséquences des mutations observées dans les différents variants (articles 3, 5 et 9 de P3). Nous nous intéressons de manière générale à la problématique de l'infection virale (HIV, HCV) (articles 1 et 7 de P3). Nous nous impliquons également sur la problématique de la résistance antibactérienne au travers de l'étude de la régulation allostériques et de l'inhibition de protéases bactériennes (projet ANR en cours) (article 8 de P3).

**Activité et reconnaissance de l'unité de recherche (10 publications les plus significatives, distinctions, avis HCERES, Labex, Equipex, ...)** :

**Evaluation de l'UMR 7019 CNRS-UL LPCT pour la période 2013–2017** : le LPCT est un laboratoire créé à la dernière contractualisation. Il résulte de la fusion de 4 équipes issues de laboratoires distincts : Physique Statistique (Institut Jean Lamour, UMR 7198), Théorie-Modélisation-Simulation (Laboratoire Structure et Réactivité des Systèmes Moléculaires Complexes, UMR 7565), Modélisation Quantique (Laboratoire de Cristallographie, Résonance Magnétique et Modélisation, UMR 7036) et Biophysique et Physique Statistique (Laboratoire de Chimie et Physique - Approche Multi-échelle des Milieux Complexes, EA 4632).

L'évaluation de l'unité par l'HCERES a conduit aux appréciations suivantes : i) qualité scientifique, excellente à exceptionnelle ; ii) réputation et rayonnement, excellents ; iii) interaction avec le mode socio-économique et culturel, excellente ; iv) organisation et vie de l'unité, pas d'appréciation (nouvelle unité) ; v) contribution à la formation par la recherche, excellente ; vi) stratégie et projet à 5 ans, excellents.

La production scientifique de l'ensemble des membres de la future unité est abondante, 514 ACL pour 36 personnes (25 ETP) sur 5 ans, soit en moyenne 4,1 ACL/an/ETP. Dans l'ensemble, les publications sont dans des journaux représentatifs des thématiques et des disciplines de chaque équipe (52 *J. Phys. Chem. A, B, C et Lett.*, 20 *Phys. Rev. B*, 19 *Phys. Rev. A*, 16 *J. Chem. Theory Comput.*, 12 *Eur. Phys. Lett.*) et un nombre notable de publications sont dans des journaux à haut facteur d'impact et dans des journaux pluridisciplinaires (7 *PNAS*, 4 *JACS*, 4 *Chem. Eur. J.*, 3 *Acc. Chem. Res.*, 1 *ACS Nano*, 2 *Nat. Commun.*, 1 *Angew. Chem. Int. Ed.*, 1 *Small*). Certains articles publiés dans la période évaluée ont atteint un taux de citations élevé. La qualité de la production s'exprime également par la réalisation de codes numériques dont certains sont inclus dans les grands programmes commerciaux, un critère de qualité.

Les quatre équipes constituantes de ce projet sont toutes de grande qualité scientifique dans le domaine de la physique et de la chimie théorique. Les thématiques abordées sont très variées puisqu'elles incluent l'étude de la matière molle, notamment en biophysique, l'étude des matériaux, l'étude de la matière condensée et l'étude de systèmes réactifs allant de la physique à la biologie en passant par la chimie. **Toutes les équipes** ont une **production** ou une **activité excellente**, voire **exceptionnelle**. La publication dans les journaux à haut facteur d'impact de leur discipline est significative. Ces équipes ont également un rayonnement et une attractivité remarquables, qui se traduisent par de très bonnes performances sur les critères habituels (conférences invitées, organisations de congrès, visiteurs, échanges, doctorants d'origine étrangères, cotutelles, etc.), mais aussi par la mise en place de laboratoires internationaux (LEA et LIA).

**Contribution à des initiatives d'excellence** : outre son implication dans deux laboratoires internationaux du CNRS, dont 1 dirigé par un membre du laboratoire, l'UMR 7019 LPCT participe activement au projet I-SITE Lorraine Université d'Excellence au travers notamment de deux projets LUE Mirabelle+ (*Anticancer Iron*

Made+ et Conception d'Inhibiteurs du Transfert de Résistances aux agents Anti-Microbiens), 2 thèses LUE Doctorat et un poste de professeur invité Prof@Lorraine (Pr. Benoit Roux, Université de Chicago).

**Recommandation par l'HCERES pour l'axe Biochimie-Biophysique le prochain CQ 2018–2022** : cet axe bénéficie d'un long historique de contributions notables en biophysique structurale, allant des apports méthodologiques (calculs d'énergie libre) au traitement de phénomènes complexes et d'intérêt médical (électroporation, par exemple). Les activités de cet axe créent un levier intéressant au niveau local (Institut Jean Barriol, CPER IT2MP, ISITE LUE). La diversité dans les niveaux de modélisation (spatial, temporel, détail de la représentation...) et dans les outils employés (Amber, Gromacs, NAMD – pour ne nommer que la partie dynamique moléculaire classique) représente un défi important pour combiner différentes approches originales, et les solutions concrètes pour y parvenir (aussi bien d'un point de vue théorique que numérique) restent à mettre en place.

**10 publications les plus significatives** : pour la période 2018–2019, environ 60 articles ont été publiés dans des revues internationales avec un comité de lecture, parmi lesquelles *Nature* ( $\times 2$ ,  $IF = 41,6$ ), *Cell* ( $IF = 36,2$ ), *Cell Host Microbe* ( $IF = 17,9$ ), *Chem. Rev.* ( $IF = 52,6$ ), *Acc. Chem. Res.* ( $IF = 20,9$ ), *J. Am. Chem. Soc.* ( $IF = 14,3$ ), *ACS Nano* ( $IF = 13,7$ ), *Nat. Commun.* ( $\times 2$ ,  $IF = 12,3$ ), *Sci. Adv.* ( $IF = 11,5$ ), *Nat. Struct. Mol. Biol.* ( $IF = 12,1$ ), *Annu. Rev. Biophys.* ( $IF = 11,7$ ).

1) B. P. Oestriker; J. H. Bolivar; M. Hensen; J. K. Claridge; **C. Chipot**; **F. Dehez**; N. Holzmann; N. Zitzmann; J. R. Schnell (2018). Re-evaluating the p7 viroporin structure. *Nature* **562**: E8–E18.

2) L. Polovinkin; G. Hassaine; J. Perot; E. Neumann; A. A. Jensen; S. N. Lefebvre; P.-J. Corringer; J. Neyton; **C. Chipot**; **F. Dehez**; G. Schoehn; H. Nury (2018). Conformational transitions of the serotonin 5-HT<sub>3</sub> receptor. *Nature* **563**: 275–279.

3) **Hognon, C., Miclot, T.,** García-Iriepa, C., Francés-Monerris, A., Grandemange, S., Terenzi, A., Marazzi, M., Barone, G. and **Monari, A.** (2020). Role of RNA guanine quadruplexes in favoring the dimerization of SARS unique domain in coronaviruses. *J. Phys. Chem. Lett.*, **11**: 5661-5667.

(4) **M.-E. Chagot**; **R. Dos Santos Morais**; **S. Dermouche**; **D. Lefebvre**; **X. Manival**; **C. Chipot**; **F. Dehez**; **M. Quinternet** (2019). Binding properties of the quaternary assembly protein SPAG1. *Biochem. J.* **476**: 1679–1694.

(5) Francés-Monerris, A., **Hognon, C., Miclot, T.,** García-Iriepa, C., Iriepa, I., Terenzi, A., Grandemange, S., Barone, G., Marazzi, M. and **Monari, A.** (2020). Molecular Basis of SARS-CoV-2 Infection and Rational Design of Potential Antiviral Agents: Modeling and Simulation Approaches. *J. Prot. Res.*, **19**:4291-4315.

(6) A. Singharoy; C. Maffeo; K.H. Delgado-Magnero; D. J. K. Swainsbury; M. Sener; U. Kleinekathofer; B. Isralewitz; I. Teo; D. Chandler; J. W. Vant; J. E. Stone; J. Phillips; T.V. Pogorelov; M. I. Mallus; **C. Chipot**; Z. Luthey-Schulten; P. Tieleman; C. N. Hunter; E. Tajkhorshid; A. Aksimentiev; K. Schulten (2019) Atoms to Phenotypes: Molecular Design Principles of Cellular Energy Metabolism. *Cell* **179**: 1098–1111.

(7) D. Pinto; C. Fenwick; C. Caillat; C. Silacci; S. Guseva; **F. Dehez**; **C. Chipot**; S. Barbieri; A. Minola; D. Jarrossay; G. D. Tomaras; X. Shen; A. Riva; M. Tarkowski; O. Schwartz; T. Bruel; J. Dufloo; M. S. Seaman; D. C. Montefiori; A. Lanzavecchia; D. Corti; G. Pantaleo; W. Weissenhorn (2019). Structural Basis for Broad HIV-1 Neutralization by a Novel MPER-Specific Human Broadly Neutralizing Antibody. *Cell Host Microbe* **26**: 623–637.

(8) J. Felix; K. Weinhaupl; **C. Chipot**; **F. Dehez**; A. Hessel; D. F. Gauto; C. Morlot; O. Abian; I. Gutsche; A. Velazquez-Campoy; P. Schanda; H. Fraga (2019) Mechanism of the allosteric activation of the ClpP protease machinery by substrates and active-site inhibitors. *Sci. Adv.* **5**: eaaw3818.

9) García-Iriepa, C., **Hognon, C.,** Francés-Monerris, A., Iriepa, I., **Miclot, T.,** Barone, G., **Monari, A.** and Marazzi, M. (2020). Thermodynamics of the Interaction between the Spike Protein of Severe Acute Respiratory Syndrome Coronavirus-2 and the Receptor of Human Angiotensin-Converting Enzyme 2. Effects of Possible Ligands. *Phys. Chem. Lett.*, **11**:9272-9281.

10) Lin, Y. C.; **Chipot, C.**; Scheuring, S. (2019) Annexin-A5 stabilizes membrane defects by inducing lipid phase transition. *Nat. Commun.* **11**: 230.

**Identité de l'unité de recherche P4** (acronyme + intitulé) : **UMR 7503 CNRS – UL – INRIA : Laboratoire Lorraine de Recherches en Informatique et ses Applications (LORIA)**

**Discipline(s) : Informatique**

**Rattachement universitaire ou organisme de recherche :**

**Université de Lorraine**

34 cours Léopold  
CS 25233  
54052 NANCY Cedex

**Adresse** (précisez Unité, Laboratoire, Département, Service, ...) : **Campus des aiguillettes**

Rue : **615 rue du Jardin Botanique**

Code postal : **54 500**

Ville : **Vandœuvre-lès-Nancy**

Tél. secrétariat : **03 83 59 20 00**

Courriel secrétariat : **josiane.reffort@loria.fr**

Site web : **www.loria.fr**

**Nom/Prénom du directeur : Jean-Yves Marion**

Courriel : **jean-yves.marion@loria.fr**

**Intitulé de l'équipe** (acronyme + intitulé) : « **Computational Algorithms for Protein Structures and Interactions** » (**CAPSID**)

Responsable de l'équipe : **Marie-Dominique Devignes**

Tel. : **03.83.59.20.65**

Courriel : **marie-dominique.devignes@loria.fr**

**Présentation de l'unité de recherche (P4) :****Description de l'unité de recherche et de l'équipe qui hébergera le projet. (300 mots maximum)**

Le LORIA regroupe 28 équipes INRIA/CNRS. Le département  **Systèmes complexes, intelligence artificielle et robotique** étudie les interactions homogènes et hétérogènes dans les systèmes complexes multi-échelles en biologie computationnelle, sciences cognitives, neurosciences et robotique. L'équipe **CAPSID** y est dédiée à la bio-informatique structurale et la modélisation moléculaire. Pluridisciplinaire, composée de 5 chercheurs permanents (informaticiens, biologistes, chimiste), elle développe des algorithmes et logiciels pour étudier les biomachines multi-moléculaires d'un point de vue structural. Les activités de l'équipe portent sur deux thèmes principaux :

(i) la **modélisation 3D** des interactions protéine – protéine/ARN (docking et simulations de dynamique moléculaire), avec un intérêt particulier pour les problèmes que pose la flexibilité moléculaire dans la modélisation des complexes ;

(ii) la classification des interactions protéiques par **extraction de connaissances** dans des bases de données. Un objet d'application particulier de l'axe (i) est le « *RNA-recognition motif* », présent dans la plupart des protéines liant les ARN flexibles simple-brin, et objet du projet H2020-ITN-RNAct de l'équipe.

L'équipe applique ses méthodes à des cibles biologiques d'intérêt, via de nombreux projets en collaborations avec le CHRU Nancy (CPER, ANR...), des équipes de biologie locales (FEDER, PEPS) et des partenaires académiques et industriels internationaux (**ITN-H2020**). Elle participe à des projets en intelligence artificielle appliquée à la santé (IMPACT, FAPs) et à la création de plateformes de modélisation, aux échelles locale (CPER\_IT2MP **SMEC**), nationale (plateforme **MBI**, membre de l'*Institut Français de Bioinformatique*) et internationale (membre du groupe *nucleic acids* du noeud **3D-bioinfo** de ELIXIR).

I. Chauvot de Beauchêne (**P4**), CRCN recrutée fin 2016, a déjà coordonné et participé à de nombreux projets individuel (PEPS-CNRS) et collaboratifs locaux (LUE-FEDER avec **P2**, PEPS-LUE avec **P1**, PEPS-LUE avec l'unité **P1-P3**) et international (ITN-H2020 pour le design des *RNA recognition motifs*), et a encadré ou encadre 2 post-doctorats et 4 doctorants.

**Activité et reconnaissance de l'unité de recherche (10 publications les plus significatives, distinctions, avis HCERES, Labex, Equipex, ...) :**

L'avis **HCERES 2011–2016**<sup>1</sup> stipule que « L'**activité scientifique** des membres de l'unité, de très bon niveau, peut souvent être qualifiée d'**excellente**, avec quelques pépites de premier plan international. [...] La visibilité de l'unité est attestée par sa participation à de nombreux projets : une trentaine de projets ANR, une petite vingtaine de projets du PIA, une quinzaine de projets européens... Certaines équipes ont un **rayonnement international exceptionnel**, attesté par l'obtention globale de **dix bourses ERC** obtenues sur la période, résultat que le comité d'experts juge remarquable. [...] La volonté de l'unité de développer des recherches « longitudinales » et d'accompagner l'ensemble d'une thématique, de ses recherches les plus fondamentales, jusqu'à ses possibles impacts économiques et sociétaux, porte ses fruits. [...] Les activités

1 <https://www.loria.fr/wp-content/uploads/2016/04/Evaluation-Loria-2011-2016-1.pdf>

pluridisciplinaires du LORIA sont aussi très visibles en région et contribuent à faire du LORIA un **acteur régional incontournable**. [...] Le LORIA a obtenu de très bons résultats en termes de **valorisation économique**, avec la création de 9 startups, sur la période, et des relations actives avec des entreprises à tous les niveaux internationaux »

### **Rayonnement**

Le LORIA est membre de la **Fédération Charles Hermite** qui regroupe les trois principaux laboratoires de recherche en mathématiques et STIC (science et technologies de l'information et de la communication) de Lorraine. L'équipe CAPSID est un nœud de l'**Institut Français de Bioinformatique**, un nœud de l'**INI-CRCT** du réseau *French Clinical Research Investigation Network*.

Les travaux de l'équipe CAPSID rayonnent grâce à la **distribution de nombreux logiciels** et serveurs très largement utilisés dans le monde : **gEMtools** (Cryo-EM fitting), **NAfragDB** (bibliothèque structurale d'ARN), **HEX** (docking par FFT), **SAM** (docking multimers), **EROS-DOCK** (docking gros grain), **KBDOCK** (docking template-based), **Kpax** (alignement structural), **lib3Dmol**, QRMSDmap, ...

### **10 publications les plus significatives :**

- 1) S. Z. Alborzi, **D. Ritchie**, **M.-D. Devignes**. (2018) Computational Discovery of Direct Associations between GO terms and Protein Domains. *BMC Bioinformatics* **19**: 413.
- 2) S. Z. Alborzi, **M.-D. Devignes**, **D. W. Ritchie**. (2017) ECDomainMiner: discovering hidden associations between enzyme commission numbers and Pfam domains. *BMC Bioinformatics* **18**: 107.
- 3) **D. W. Ritchie**, S. Grudin. Spherical polar Fourier assembly of protein complexes with arbitrary point group symmetry, *J. of Applied Crystallography*, 2016.
- 4) A. W. Ghoorah, **M.-D. Devignes**, **M. Smail-Tabbone**, **D. Ritchie** (2014). KBDOCK 2013: A spatial classification of 3D protein domain family interactions, *Nucleic Acids Res.* 42: 389-95.

Participant **I. Chauvot de Beauchêne** :

- 5) ME Ruiz Echarte, **DW Ritchie**, **I Chauvot de Beauchêne** (2020). Using Restraints in EROS-Dock Improves Model Quality in Pairwise and Multicomponent Protein Docking. *Proteins* 88(8):1121-1128.
- 6) ME Ruiz Echarte, **I Chauvot de Beauchêne**, **DW Ritchie** (2019). EROS-DOCK: protein-protein docking using exhaustive branch-and-bound rotational search. *Bioinformatics* 35: 5003-5010.
- 7) S.J. de Vries, **I. Chauvot de Beauchêne**, C.E. Schindler, M. Zacharias (2016). Cryo-EM Data Are Superior to Contact and Interface Information in Integrative Modeling. *Biophys J.* 110: 785-97.
- 8) SA Samsonov, M Zacharias, **I Chauvot de Beauchêne** (2018). Modeling large protein-glycosaminoglycan complexes using a fragment-based approach. *J. of comput. Chem.* 40: 1429-1439.
- 9) **I Chauvot de Beauchêne**, SJ de Vries, M. Zacharias (2016). Fragment-based modelling of single stranded RNA bound to RNA recognition motif containing proteins. *Nucleic Acids Res.* 44: 4565-80.
- 10) **I Chauvot de Beauchêne**, SJ de Vries, M. Zacharias (2016). Binding site identification and flexible docking of single stranded RNA to proteins using a fragment-based approach. *PloS Computational Biology* 12: e1004697.

## B. DESCRIPTION DU PROGRAMME DE RECHERCHE

**Intitulé :** MolAI4Cryo : Modeling and Artificial Intelligence applied to Cryo-EM 3D structures to fight COVID and AMR

**Dates de début et de fin de programme :** du 01/06/2022 au 31/05/2024 (2 ans)

**Localisation :** Université de Lorraine (UL, Nancy)

(1 à 2 pages)

### Contexte du programme, Présentation technique

#### Contexte, enjeux économiques, sociétaux, environnementaux, ...

Les défis à relever par les pouvoirs publics en santé publique sont nombreux comme enrayer les maladies infectieuses, élargir l'accès aux médicaments, tirer parti des nouvelles technologies, donner plus d'importance à la santé dans le débat climatique tout en allant vers une médecine adaptée à chaque citoyen et une maîtrise des coûts. Un des défis de santé de plus en plus inquiétants sera de se préparer aux prochaines épidémies en luttant contre la propagation de nouveaux pathogènes.

Il s'agit d'une part, de l'infection par le coronavirus 2 du syndrome respiratoire aigu sévère (SARS-Cov-2) qui s'est propagé dans le monde entier depuis fin 2019, créant la pandémie sans précédent de COVID-19. Même si différents types de vaccins, en priorité ceux basés sur l'ARNm, semblent efficaces sur les souches actuelles du virus, il paraît indispensable de mieux connaître les caractéristiques d'infection, de réplication et d'évolution des virus à ARN afin de pouvoir lutter plus efficacement contre d'éventuelles pandémies virales futures. Il s'agit d'autre part, de l'utilisation massive des antibiotiques en santé publique et en agro-alimentaire depuis des décennies, qui a généré l'émergence de nombreuses résistances aux antimicrobiens (AMR) chez les bactéries, posant aujourd'hui des problèmes notamment en milieu hospitalier avec l'émergence de bactéries multi-résistantes. Ces résistances se transmettent désormais très efficacement entre bactéries par des mécanismes de transfert horizontaux de gènes. Ce projet vise principalement à acquérir un équipement scientifique de pointe (un cryo-microscope électronique) permettant notamment de mieux comprendre les mécanismes moléculaires biologiques à fort impact en santé publique.

La biologie structurale, historiquement basée sur les structures tridimensionnelles provenant des techniques de cristallographie aux rayons-X, puis de Résonance Magnétique Nucléaire, a connu un **bouleversement récent** sans précédent avec l'avènement de la **cryo-microscopie électronique** (cryo-EM) à haute résolution (Prix Nobel de chimie en 2017). La cryo-EM a connu ces 10 dernières années une formidable révolution technologique grâce aux développements de nouvelles générations de cryo-microscopes électroniques et de caméras à détection directe. L'observation à une échelle sub-nanométrique du vivant directement dans les cellules est à porter de main. C'est l'**opportunité inédite de visualiser des structures tridimensionnelles (3D) d'objets biologiques sans limite de taille, en solution et avec une résolution atomique**. Cette technique permet de visualiser un virus, un composant cellulaire mais également des macro-complexes moléculaires sous forme hydratée-congelée dans leurs états natifs et d'étudier leurs organisations structurales au plus proche de leurs conditions cellulaires natives par traitement d'images et reconstruction 3D. La cryo-EM est devenue indispensable aux chercheurs en biologie structurale dite intégrative pour caractériser à l'échelle atomique les différents processus cellulaires clés.

La **cryo-EM** implique un traitement conséquent d'images avec des quantités de données de l'ordre de plusieurs centaines de milliers de particules. Ceci nécessite des développements de logiciels basés sur l'**Intelligence Artificielle** (IA), par exemple pour la reconnaissance et la sélection de particules dans les images, l'attribution angulaire et la reconstruction 3D afin d'obtenir des cartes à une résolution d'environ 3 Å permettant de construire des modèles atomiques. Le **positionnement des molécules** dans des **cartes de densité atomique expérimentale** issues de la cryo-EM peuvent être obtenues par **arrimage data-driven** (ou *fitting*) de **structures 3D de chaque composant** (obtenues par RMN, cristallographie ou modélisation), via des approches gros grains.

La **dynamique moléculaire** est un outil puissant pour affiner ces modèles gros grains et ainsi **comprendre et prédire l'évolution de systèmes biomoléculaires complexes** à partir des données issues de la biologie structurale comme la cryo-EM. Toutefois, le coût informatique des simulations numériques associées limite les temps caractéristiques accessibles à des durées bien inférieures à celles des processus biologiques. Depuis quelques années, l'**IA** opère une véritable révolution dans notre domaine<sup>1</sup> ouvrant la voie à la résolution de problèmes jugés insolubles comme la **prédiction du repliement des protéines** (*AlphaFold*<sup>2</sup>). Les nouveaux résultats d'**AlphaFold** apportent la connaissance de la structure statique moyenne pour de nouvelles protéines, mais ne donne pas directement accès à leurs propriétés dynamiques. Nos laboratoires sont très impliqués dans la mise au point d'approche de type **Deep Learning** pour prédire la dynamique conformationnelle des protéines et l'effet de mutations (e.g. variants) ou de médicament sur celle-

ci grâce aux architectures de **calculs hybrides GPU/CPU**. A l'instar des trajectoires de dynamique moléculaire, les cartes des densités obtenues par cryo-EM contiennent également des informations sur la dynamique des systèmes. Ces informations sont toutefois "cachées" et restent aujourd'hui essentiellement inaccessibles. Récemment, des algorithmes couplant aux données expérimentales des approches de *Deep Learning* et de dynamique moléculaire se sont montrés particulièrement prometteur pour **révéler la dynamique de machineries biologiques complexes à des niveaux de détails inédits**<sup>3</sup>. Cette approche interdisciplinaire est au cœur des développements méthodologiques que nous mettrons en œuvre dans ce projet.

Ce programme de recherche va s'atteler à **déchiffrer l'architecture de complexes macromoléculaires clés dans certaines pathogénicités**, et permettra d'ouvrir des perspectives de contrôle de ces pathogènes. En cohérence avec la propagation des antibiorésistances, ces connaissances structurales tridimensionnelles pourront servir de base pour proposer de nouvelles pistes pour limiter les transferts d'AMR par conjugaison par la mise au point d'inhibiteurs de conjugaison. En complément, B. Douzi (CR-INRAE) est porteur au sein du **P2** d'une **ANR JCJC (ConjT4SS-Game)** en préparation pour la 2<sup>ème</sup> phase. Cette ANR vise à caractériser les protéines impliquées dans la formation du canal responsable du passage de l'ADN à travers la paroi des bactéries et qui permet notamment d'établir le contact entre la bactérie donneuse et la receveuse. En cohérence avec l'**urgence sanitaire de la COVID-19**, une demande de financement **ANR PRC CoVHijacksSplicing** portée par B. Charpentier (PR1, UL co-responsable de l'équipe 1 et co-directeur de **P1**) en collaboration avec I. Lebars (CRCN-CNRS, IBMC, Strasbourg) et E. Decroly (DR-CNRS, AFMB, Marseille) a été déposée pour la 2<sup>ème</sup> étape. Elle porte sur l'étude de la modulation de l'épissage au cours de l'infection par le SRAS-CoV-2 et des mécanismes moléculaires permettant le détournement du spliceosome de la cellule hôte par le virus et pourrait servir de preuve de concept pour l'élaboration de nouveau antiviraux.

Analyser les relations **structure, dynamique et fonction** des étapes d'assemblage de cette machinerie d'épissage en interaction avec certaines protéines virales ou un pore de conjugaison représente un vrai défi. Notamment au regard de leur variabilité structurale lors de la transition d'un état mature et fonctionnel à un état immature et non fonctionnel, permettra d'identifier leurs **dysfonctionnements** ou **détournements**, et de proposer des molécules à potentiel thérapeutique par *drug-design*. Ainsi, la dynamique et la structure de ces complexes seront abordées par le **développement et l'application d'approches innovantes en IA** dans des méthodes d'analyse d'images en **cryo-EM** (sélection de particules uniques), de classification 3D des structures et de l'analyse de ces structures par **modélisation d'amarrage et calculs de dynamique moléculaire** (calculs de trajectoires moléculaires).

Concrètement, il existe un fort besoin national pour le développement de la cryo-EM mais également en Lorraine avec un recensement se situant autour de 5 équipes supplémentaires par rapport aux partenaires inclus dans ce programme. C'est donc un **investissement urgent et primordial** pour la Lorraine afin que ses chercheurs puissent continuer à mener une recherche d'excellence en publiant leurs résultats dans des journaux de très forte notoriété scientifique.

Ce consortium crée une **nouvelle synergie inédite** entre **2 équipes expérimentales** et **2 équipes théoriques** reconnues au niveau international et réparties sur le territoire lorrain de la région Grand-Est avec une plus-value importante dans le décryptage de **machineries moléculaires biologiques complexes**. Les deux premières équipes bénéficieront du transfert des développements méthodologiques des deux dernières. Le consortium rassemble X. Manival (**P1**), responsable du groupe de Biologie Structurale Intégrative, N. Soler et B. Douzi (**P2**) de l'équipe ICE-TeA, F. Dehez (**P3**) co-directeur du LCPT et I. Chauvot de Beauchêne (**P4**), responsable du groupe interactions ARN/protéine de l'équipe CAPSID.

#### **Lien avec la stratégie de recherche de l'établissement**

Ce projet est en parti en lien avec un des domaines identifiés de la stratégie de spécialisation intelligente (**S3**), la **transition numérique**. Les perspectives de transfert et d'innovation sont avérées, à savoir dans les biotechnologies médicales et dans les outils numériques pour la santé. Il relève des thématiques de recherche en **IA** ainsi que de la politique régionale « **feuille de route Santé** » au niveau des axes 2 et 4 (voir page 4 du document). Il sera un vrai accélérateur d'innovations en santé notamment dans le cadre des **réponses thérapeutiques pour limiter la COVID-19** sur le territoire.

#### **Présentation générale et biologique du projet :**

De manière générale, les virus, parasites stricts, ont évolué pour détourner des voies cellulaires clés de l'hôte et échapper à sa réponse immunitaire innée en modulant la fonction de ses protéines et de ses voies de signalisation. Des publications récentes sur l'infection par le SARS-CoV-2 montre qu'elle affecte fortement l'expression des composants du spliceosome en conduisant à l'accumulation d'ARNm cellulaires non maturés<sup>4,5</sup>. L'inhibition de l'épissage par le pladiénolide B, un macrolide produit par *Streptomyces platensis* ou par la délétion du gène codant pour le facteur d'épissage RBM28 empêche la réplication du SARS-CoV-2<sup>5</sup>. Cela révèle l'**épissage des ARNm cellulaires** comme une voie indirecte mais essentielle pour la réplication du SARS-CoV-2 et donc comme une **cible thérapeutique potentielle**. Des interactions entre le spliceosome de l'hôte et les protéines virales ont déjà été identifiés<sup>6</sup>. Il s'agit notamment de l'interaction entre la protéine



virale **nsp16**, une **2'-O-méthyltransférase** impliquée dans la formation de la coiffe de l'ARN viral, et les **snRNAs U1 et U2** de la particule d'épissage<sup>4</sup> afin de perturber l'épissage cellulaire global. Ces deux ARNs font partis des 5 petits ARNs nucléaires riches en Uracile (UsnRNPs U1, U2, U4, U5 et U6) qui s'assemblent sur les introns des pré-ARNm et composent le spliceosome dont l'activité permet l'excision des introns et la ligature des exons de l'ARNm. **Nsp16 pourrait être un acteur majeur du remodelage de l'épitranscriptome** lors de l'infection mais également avoir un **spectre fonctionnel plus large** que son activité enzymatique 2'-O-MTase. Cependant, les bases moléculaires et structurales de ces interactions et leurs mécanismes directs ou indirects associés restent encore à caractériser.

De leur côté, les micro-organismes pathogènes sont capables d'infecter les cellules humaines, provoquant des affections pouvant parfois être mortelles. Les bactéries pathogènes **codent souvent des toxines ou autres facteurs de virulence qui, associées à de multiples AMR**, posent un problème majeur en santé publique. Ces AMR se propagent le plus souvent entre bactéries par transfert horizontal, dont la **conjugaison bactérienne** est un des mécanismes majeurs. La conjugaison nécessite un contact physique entre bactérie donneuse et receveuse, permettant l'établissement d'un pore de conjugaison par lequel l'ADN est transféré. Les bactéries sont divisées en deux grands types selon la structure de leur enveloppe. Les bactéries Gram-négatives possèdent deux membranes séparées par une fine couche de peptidoglycane, alors que les bactéries Gram-positives possèdent une seule membrane entourée d'une épaisse couche de peptidoglycane. Si l'architecture d'un pore de conjugaison de bactérie Gram- est connue depuis quelques années<sup>7</sup>, **celle d'un pore de conjugaison de bactérie Gram+ est à ce jour inconnue**. Ce projet vise donc en partie, à obtenir l'architecture structurale d'un pore de conjugaison de bactérie Gram+. Ces données pourront servir de base pour identifier des molécules permettant **de limiter ou inhiber les transferts conjuguatifs**, notamment dans le domaine hospitalier où les infections nosocomiales à germes Gram+ multirésistants sont fréquentes. Ces molécules pourront cibler l'assemblage du pore, ou bien le passage de l'ADN d'une cellule à une autre.

#### Etat de l'art :

Dans la cellule eucaryote, la formation de la **coiffe** à l'extrémité 5' des ARNm est une modification structurale essentielle mise en place dans le noyau et qui joue un rôle clé pour une traduction efficace de l'ARNm dans le cytoplasme. La guanosine (G<sub>0</sub>) présente dans la coiffe 5' est méthylée en position N7 et le **premier nucléotide (N<sub>1</sub>)** de l'ARNm est également **méthylé** mais cette fois-ci **en position 2'OH** de son ribose. Cette dernière marque épitranscriptomique est considérée comme un « **auto-marqueur** » d'ARNm cellulaire. Ainsi l'**absence de coiffe** ou une coiffe chimiquement incomplète permet la **reconnaissance des ARN étrangers** comme « non-soi » et induit la production d'interféron de type I déclenchant la **réponse immunitaire innée antivirale**<sup>8-10</sup>. Par conséquent, les virus, incluant ceux se répliquant uniquement dans le cytoplasme, ont mis au point des mécanismes pour synthétiser une coiffe indiscernable de celle des ARNm cellulaires. Les coronavirus expriment pour cela **nsp14**, la N7-MTase et **nsp16** (en association avec nsp10), la 2'-O-MTase convertissant la séquence G<sub>0</sub>pppN<sub>1</sub> de la coiffe en m<sup>7</sup>G<sub>0</sub>pppN<sub>1</sub>m afin de produire efficacement leurs protéines virales et d'échapper à la réponse immunitaire innée<sup>8,9,11</sup>. La résolution récente de plusieurs structures 3D de l'hétérodimère nsp10/nsp16 du SARS-CoV-2 a révélé comment nsp10 remodele le site de liaison SAM et étend la surface d'interaction avec les 5 premiers nucléotides des ARN viraux naissants<sup>9,12</sup>. De plus, il a été récemment démontré que la **nsp16 du SARS-CoV-2 interagissait avec les UsnRNAs U1 et U2 de l'hôte**, au niveau des séquences de reconnaissance respectivement des sites d'épissage à la bordure entre un exon et un intron (site 5' d'épissage) et du point de branchement à l'intérieur de l'intron, **perturbant l'épissage cellulaire global**<sup>4</sup>. Les déterminants moléculaires assurant la spécificité d'interaction de nsp16 en absence de nsp10 avec les snARNs, le mécanisme par lequel la protéine nsp16 est adressée au noyau et quelles sont les conséquences sur le programme d'épissage des pré-ARNm ne sont pas connus actuellement. Des études fonctionnelles et structurales sur ces différents sujets sont donc nécessaires.

**Aucun pore de bactérie Gram+** n'a à ce jour été étudié d'un point de vue de sa structuration générale. Ces pores de conjugaison correspondent à des **systèmes de sécrétion de type IV (T4SS)**. Si certaines protéines sont conservées entre les T4SS de bactéries Gram- et ceux de bactérie Gram+, d'autres sont spécifiques, et leur organisation tridimensionnelle peut différer. Récemment, nous avons réussi à caractériser par des approches biochimiques et structurales la protéine OrfG, un composant central dans l'assemblage du pore (article n°1 du P2). Par sa localisation membranaire et sa capacité à multimériser, cette protéine est le candidat principal parmi les composants du T4SS susceptible de former un **canal au niveau des parois bactériennes**, assurant ainsi le passage de l'ADN entre les deux bactéries. Des études structurales par microscopie électronique devraient nous renseigner davantage sur l'architecture globale de cette protéine et permettraient de mieux comprendre le processus de conjugaison.

#### Présentation scientifique et technique du programme :

**P1** est reconnu internationalement dans la biogenèse, l'assemblage et la maturation de macro-complexes protéine-ARN appelés **RiboNucléoProtéines (RNPs)**. Il s'agit notamment des petites RNPs nucléolaires (**snoRNPs**) et des petites RNPs nucléaires riches en U (**UsnRNPs**). Les snoRNPs à boîtes C/D catalysent l'addition d'un groupement méthyle en position 2'OH de riboses de l'ARNr. Le porteur du projet, X. Manival, possède une solide expérience en biochimie (stratégie de co-expressions chez *E.coli*) structurale

(RMN, CRX) et est responsable du groupe de Biologie Structurale Intégrative de l'équipe. N. Soler/B. Douzi (**P2**) sont des experts dans l'étude de complexes ADN-protéine et protéine/protéine par des approches biochimiques, biophysiques et structurales. Ils s'intéressent aux complexes macromoléculaires qui sont impliqués dans la conjugaison bactérienne. F. Dehez (**P3**) a animé l'axe Biochimie et Biophysique du LPCT avant de prendre la direction adjointe de ce laboratoire. Il est également membre du laboratoire international associé avec l'Université d'Illinois mondialement reconnue dans la modélisation et visualisation moléculaire. I. Chauvot de Beauchêne (**P4**) est responsable du groupe interactions protéine – protéine/ARN de l'équipe CAPSID du LORIA. Elle développe des méthodes innovantes pour la modélisation des assemblages macromoléculaires flexibles, avec un focus sur les motifs protéiques de reconnaissance de l'ARN, mais aussi leurs interactions avec d'autres protéines. Il est à noter que ce projet bénéficie du soutien de B. Klaholz (DR-CNRS, CBI/IGBMC, Strasbourg), responsable de l'équipe Grands complexes impliqués dans l'expression des gènes et responsable scientifique de la plateforme de biologie structurale intégrée de l'IGBMC. B. Klaholz est le coordinateur national de l'infrastructure FRISBI, centre européen de l'infrastructure Instruct-ERIC avec une expertise reconnue au niveau national et européen en Cryo-EM. X. Manival est le président du comité des utilisateurs de FRISBI.

**P1** a développé une méthode de co-expression chez *E. coli* pour palier au problème d'insolubilité de certaines protéines en interaction avec leur partenaire. Cette méthode permet également de produire directement *in vivo* chaque composant d'un macro-complexe. Les complexes protéines/ARN, nsp16/U1 et nps16/U2, seront quant à eux reconstitués *in vitro*. Nsp16 sera rajoutée à des complexes de pré-épissages purifiés à partir de cultures de cellules humaines contenant les snRNAs U1 et U2. Après caractérisations physico-chimiques par des techniques biophysiques (DLS, ITC, SEC-MALS, NMR) pour un contrôle de la qualité (pureté, homogénéité, stabilité et quantité), l'ensemble de ces différents échantillons sera déposé sur grilles afin d'optimiser les conditions d'observation (particules isolées, faible épaisseur de glace, ...) grâce à notre futur cryo-microscope. **P2** s'intéresse à l'étude des mécanismes moléculaires qui régissent le transfert de gènes par conjugaison. Dans le but de contrôler la dissémination de l'antibiorésistance, de nouvelles stratégies appellent à limiter ce transfert voir l'inhiber. Ceci ne pourra se faire qu'avec une connaissance fine des mécanismes moléculaires impliqués. **P2** développe pour cela des approches dédiées à la purification et caractérisation des complexes ADN-protéines et protéine/protéine, y compris pour des protéines membranaires (EMSA, SEC-MALS, SwitchSENSE®, ITC, Pull-Down, SPR). Les objets biologiques étudiés et les techniques utilisées par les deux partenaires expérimentaux sont majoritairement complémentaires.

Les puissantes méthodes développées par les équipes **P3** et **P4** du consortium serviront à analyser la structure et la dynamique des échantillons biologiques (macro-complexes) produits et purifiés par **P1** et **P2**. Il s'agit de la **dynamique moléculaire (P3)**, des **interactions de molécules et macromolécules par arrimage moléculaire (P4)**, de l'**arrimage à l'aide de données expérimentales à basse/moyenne résolution** (dont cryo-EM, **P4**), du **criblage virtuel et drug design (P3, P4)**, des **calculs d'énergies libres (P3)** et de l'**échantillonnage d'événement rares (P3)**. Grâce aux données structurales, à la robustesse des modèles et à la capacité d'échantillonnage des supercalculateurs actuels, la modélisation moléculaire permet d'accéder à des grandeurs thermodynamiques et cinétiques directement comparables aux données expérimentales. Prédire la formation de structures complexes résultant de l'assemblage de plusieurs macromolécules de structures connues, e.g. – protéines, ADN, ARN, lipides, médicaments –, voire inconnue e.g. – protéines désordonnées, ADN et ARN simple brin – reste à bien des égards une gageure. Il en va de même pour décrire la dynamique des processus biologiques, qui se déroule souvent sur des échelles de temps allant bien au-delà de la milliseconde, là où les simulations moléculaires standards se limitent à des échelles de temps de l'ordre de quelques dizaines de microsecondes. **L'objectif de ce projet est de développer des approches innovantes pour édifier un pont solide entre les modèles moléculaires et les données biochimiques et bio-structurales obtenues expérimentalement.** Les méthodes basées sur l'**IA (machine learning)** représentent une formidable opportunité pour développer des solutions nouvelles et ont d'ores et déjà montré leurs capacités à révolutionner notre domaine. Dans le cadre de ce projet, les 2 équipes de modélisation moléculaire vont **assembler leurs savoir-faire** pour établir une **stratégie méthodologique complète basée sur l'IA** allant de la **prédiction de structure d'assemblages biomoléculaires complexes en gros grain à l'étude de leur dynamique fonctionnelle à l'échelle atomique.**

**P3** va développer plusieurs approches permettant de modéliser des événements millisecondes avec des dynamiques moléculaires microsecondes. En particulier, nous allons utiliser les réseaux de neurones profonds (*deep learning*) pour extraire des trajectoires de dynamique moléculaire, les degrés de liberté pertinents (variables collectives, CVs) à même de décrire de manière minimale les transitions conformationnelles d'un système biomoléculaire donné. Pour cela, nous procéderons de manière itérative, en générant une trajectoire biaisée le long d'un jeu de CVs initial non-optimal. Cette trajectoire sera analysée par le réseau de neurones pour parfaire les CVs qui seront utilisés pour générer une nouvelle trajectoire jusqu'à convergence. Ces variables collectives seront ensuite utilisées pour déterminer un chemin de plus basse énergie connectant deux ou plusieurs conformations d'une protéine par une méthode de recherche de chemin (*String*) utilisant des essais de multiples trajectoires (*Swarm of Trajectories*). Le chemin ainsi déterminé

permettra d'établir le profil d'énergie libre de transition conformationnelle grâce à des simulations microsecondes en utilisant les approches de Force de Biais Adaptif développés dans l'équipe.

**P3** développe enfin depuis plusieurs années un ensemble de méthodes de calcul d'énergie libre (calorimètre computationnel) permettant d'échantillonner les transitions conformationnelles associées à la fonction des protéines, l'association protéine-protéine et protéine-ligand. L'ensemble de ces développements, réalisés dans les programmes **NAMD** et **VMD** dans le cadre du Laboratoire International Associé CNRS/UIUC (Université de l'Illinois à Urbana-Champaign, États-Unis), serviront de base aux développements prévus dans ce projet. Il est important de noter que NAMD dispose d'une interface permettant de **projeter** par *fit* flexible **des structure 3D de protéines dans des enveloppes de microscopie électronique** (MDFF).

**P4** développe des méthodes de modélisation des assemblages macromoléculaire (protéines/acides nucléiques/glycanes/petites molécules) par amarrage, en utilisant des modèles gros grains. Elle développe en particulier une **approche originale de modélisation des complexes protéine – ARN simple-brin** par amarrage « à l'aveugle » (*ab initio*) puis assemblage de fragments structuraux d'ARN, sur une structure rigide de protéine. Cette méthode d'amarrage combinatoire a permis de résoudre la difficulté que représente l'immense espace conformationnel des portions d'ARNs non structurés, qui sont majoritaires dans les interactions ARN – protéine (articles 9-10 de P4). Cette méthode a été adaptée avec succès à l'amarrage des glycosaminoglycanes, autres polymères hyper-flexibles (article 8 de P4). Au cours du présent projet, elle sera d'abord **étendue aux ARNs partiellement structurés**, par une combinaison de fragments simple et double brin à amarrer. Elle sera ensuite adaptée à l'amarrage des « boucles » (portions désordonnées) de protéines, qui sont le facteur limitant des méthodes d'amarrage protéine - protéine actuelles. Enfin, elle sera étoffée par la **prise en compte de la variabilité conformationnelle locale des protéines** majoritairement structurées, par une discrétisation des conformations de la surface protéique sur lesquelles amarrer les fragments d'ARN ou de boucles protéiques.

**P4** a aussi récemment développé et implémenté **EROS-DOCK**, une nouvelle approche d'amarrage rigide par exploration hiérarchique « *branch-and-bound* » de l'espace des rotations des protéines (articles 6-7). Cette méthode peut actuellement s'appuyer sur des données de biophysique ou biochimie expérimentale (mutagenèse, FRET, *cross-links*, ...) pour l'élagage de l'espace des solutions. Cette capacité sera étendue au cours du projet à l'**utilisation de carte de cryo-EM** en éliminant les sous-espaces de rotation en inadéquation avec leur densité atomique.

#### **Objectifs :**

L'objectif général du projet **MolAI4Cryo** est l'avancée des connaissances fondamentales dans les mécanismes d'**infections virale** (COVID) ou **bactérienne** (AMR) a des fins de **santé publique** en utilisant **des approches d'IA** en traitement d'images, reconnaissance d'objet, affinements multiparamétriques, reconstruction tridimensionnelle et modélisation pour leur étude par **cryo-EM**. Ce projet fournira une vue complète et détaillée (*i*) de l'effet de l'infection par le SRAS-CoV-2 sur la machinerie d'épissage de l'hôte et (*ii*) du phénomène de conjugaison d'une bactérie Gram-positif. Les objectifs plus spécifiques que nous aimerions aborder sont de **déchiffrer, à l'échelle atomique, via des structures tridimensionnelles obtenues en cryo-microscopie électronique, les mécanismes moléculaires** sous-jacents au détournement de l'assemblage et de la fonction de la particule d'épissage des cellules hôtes médié par la protéine virale nsp16 du SARS-CoV-2 d'un part et à l'assemblage du pore de conjugaison d'autre part. Il conduira à la mise en place d'une approche conjointe expérimentale ciblée et de modélisation dédiée pour augmenter la compréhension fine des relations structures/dynamiques/fonctions qui pourra être appliquée à l'avenir à d'autres systèmes biomoléculaires complexes. Les données structurales obtenues sur ces complexes macromoléculaires serviront de preuves de concept pour de futures **cibles anti-virales**.

#### **Originalité et/ou caractère innovant**

L'originalité concerne en premier lieu la stratégie sous-jacente de cette demande qui se décompose en deux points. Le premier consiste à installer au niveau de la plateforme de Biophysique et Biologie Structurale (B2S, responsable S. Boschi-Muller, **P1**) de l'UMS Ingénierie-Biologie-Santé en Lorraine (IBSLor, <https://umsibslor.univ-lorraine.fr/fr>, responsable I. Motorin, **P1**) un automate de vitrification d'échantillons et un cryo-microscope de criblage (100 kV) pour respectivement préparer les grilles cryo-EM et effectuer leurs contrôles qualité afin d'optimiser les préparations pour l'enregistrement à haute résolution sur le microscope Titan Krios à Strasbourg (B. Klaholz, CBI/IGBMC). Actuellement, seuls 2 autres cryo-microscopes aussi puissants sont présents en France. Par conséquent, le temps d'acquisition sur ce type d'appareil est extrêmement précieux, étant très demandé et très coûteux. L'investissement dans ce type de microscope en Lorraine serait une première et permettrait également de lever ce verrou en évitant de saturer la plateforme de Strasbourg avec l'étude de grilles de qualité incertaine. Les contraintes techniques de ce nouvel instrument seront complémentaires avec celles des autres équipements de biologie structurale déjà existants à l'UMS, à savoir un spectromètre RMN 600 MHz équipé d'une cryosonde TCI et un diffractomètre de rayons X. En effet, la RMN s'adresse à des objets biologiques en solution en dessous de 50 kDa, la cristallographie impose une cristallisation de l'échantillon et la cryo-EM n'est pas adaptée à des objets en dessous de 100 kDa, tolère une certaine flexibilité de ce dernier et nécessite une quantité de matériels dix fois inférieure. Le deuxième concerne la mise au point d'un mur d'image composé d'un ensemble (>12) d'écrans 4K et 3D

capable d'afficher les données structurales issues des simulations de dynamique moléculaire et de Cryo-EM. Plus largement cet équipement pourra servir à toutes les problématiques de visualisations massive de données du site Lorrain. Ce mur d'image sera piloté par un ensemble de serveurs GPU dédiées permettant l'affichage haute définition simultanément des données sur tous les écrans permettant ainsi un travail collaboratif ([Mur d'image de l'IBPC à Paris](#)). Cette plateforme sera couplée à des stations de travail dédiées au sein des équipes associées. La partie traitement des données fera appel à quelques serveurs hybrides de développement intégrant des cœurs Tensor spécifiques qui seront mutualisés avec les infrastructures de calcul régionales (EXPLOR) et accueillis dans le *data center* à Nancy (**P3** et **P4**).

L'originalité de ce programme réside également dans la synergie de l'ensemble des compétences développées dans chacune des équipes pour proposer des outils d'ajustement structural aidant à la résolution fine de la structure d'assemblages moléculaires à l'aide de données issue de la cryo-EM et à la compréhension de la dynamique fonctionnelle de systèmes biologiques complexes. Ces travaux permettront non seulement de maintenir le leadership international de chacune des équipes dans leurs domaines, mais aussi de faire **émerger un pôle de compétences théorie/expérience rares et précieux** autour de la **résolution structurale et dynamique en biologie**.

Enfin, ce projet prend en compte la **machinerie d'épissage de l'ARN messager** comme une voie essentielle pour la réplication du SARS-CoV-2 et donc comme **cible potentielle** pour limiter la pandémie de la **COVID-19** et ses futurs variants plus contagieux. Jusqu'à maintenant cette possibilité n'a pas été explorée et donc représente un caractère innovant évident. Ce projet s'attaquera également à l'étude structurale du **pore de conjugaison des bactéries Gram+**, un complexe macromoléculaire essentiel pour la dissémination de l'antibiorésistance et l'émergence des bactéries multi-résistantes. Les données structurales collectées sur nous permettront de révéler de  **futures cibles moléculaires pour limiter les diffusions d'AMR** entre bactéries.

#### ***Dimension interrégionale, transfrontalière, européenne, internationale***

Ce projet a été initié afin de **générer des synergies fortes** entre les expertises complémentaires des équipes de modélisation (**P3** et **P4**) et expérimentales (**P1** et **P2**) au sein du territoire **lorrain** de la **Région Grand-Est**. Il bénéficie du soutien de B. Klaholz (DR-CNRS, CBI/IGBMC, Strasbourg), responsable de l'équipe Grands complexes impliqués dans l'expression des gènes et responsable scientifique de la plateforme de biologie structurale intégrée de l'IGBMC qui accueille un des cryo-microscopes les plus puissants au monde.

Ce consortium rassemble des responsables scientifiques avec une visibilité internationale et d'horizons divers interdisciplinaires mais complémentaires, incluant des biologistes structuralistes (**P1**, IMoPA et **P2**, DynAMic), un chimiste théoricien (**P3**, LPCT) et une bio-informaticienne (**P4**, LORIA). Plusieurs partenaires du consortium collaborent déjà ensemble : **P1** et **P3** (article n°4 en 2020 du P3), **P1** et **P4** (projet PEPS « interANRIL »). **P2** et **P4** (projet LUE-CITRAM, article n°5 de **P2**), **P2**, **P3** et **P4** (projet FEDER). **P4** est impliqué dans 3 plateformes structurales reconnues au sein de l'Institut Français de Bioinformatique (IFB), permettant de distribuer les outils spécifiques développés dans ce projet à cette communauté. Ce projet renforcera et mutualisera des collaborations internationales existantes des différents partenaires. D'une part pour l'émergence d'une **synergie transfrontalière** « très Grand-Est » : collaborations de **P4** avec la VUB en Belgique (ITN-H2020\_RNAct) et la TUM en Allemagne (publications 7-8-9-10), et collaboration de **P1** avec R. Lührmann (MPI, Göttingen, Allemagne, publication n°1 de P1) avec un Laboratoire Européen Associé (LEA *pre-mRNA splicing*). D'autre part à l'échelle mondiale : **P3** coordonne le Laboratoire International Associé SML entre le CNRS et l'université d'Illinois à Urbana-Champaign. Au travers de ce consortium et de son partenaire américain (TCBG au Beckmann Institut, ressource NIH), **P3** est impliqué dans les **développements des programmes** de DM de visualisation **parmi les plus utilisées au monde** (NAMD et VMD).

#### ***Avancées attendues, verrous techniques/scientifiques à lever***

La **préparation des grilles d'échantillons** compatible pour la cryo-EM à haute résolution reste une **étape limitante** dans l'étude des macro-complexes biologiques. Afin d'optimiser les échantillons à analyser, **P1** et **P2** souhaitent investir dans un **automate de cryo-préparation d'échantillons** permettant la congélation ultra-rapide pour préserver l'échantillon sous forme hydratée dans son état natif, et un **cryo-microscope électronique basse résolution de criblage** (100 keV) pour contrôler la qualité des grilles, et par conséquent optimiser le temps d'utilisation des cryo-microscopes électroniques à haute résolution nationaux dont principalement celui de Strasbourg (B. Klaholz, CBI/IGBMC). **P2** souhaite également pouvoir acquérir des images en coloration négative afin d'optimiser ses protocoles de purifications de protéines transmembranaires. La création d'un pôle d'analyses et une plateforme de cryo-EM en Lorraine est inédite et permettra à une dizaine d'équipes locales d'accéder à cette technique pour résoudre la structure 3D de leurs systèmes biologiques d'intérêt même avec une dynamique et une hétérogénéité conformationnelle notable (ce qui était difficilement possible avec les autres techniques de biologie structurale) avec des quantités d'échantillons (0,1 mg/mL) inférieures d'au moins un facteur 10 par rapport aux techniques historiquement précédentes de biologie structurale. Il permettra également de visualiser des particules notamment virales à une résolution subcellulaire. L'investissement dans l'achat d'un cryo-microscope électronique de criblage en

Lorraine se justifie également comme une antenne de la plateforme d'excellence en cryo-EM de Strasbourg, permettant en retour de renforcer son statut national.

**Suites attendues : applications, marchés possibles, start-up, nouvelle piste de recherche, nouveaux produits, nouvelles technologies, nouveaux investissements...**

Ce projet pourra avoir un impact sur la santé publique car l'ensemble de ses données pourront révéler de **nouvelles cibles antivirales** basées sur l'inhibition de l'interaction de nsp16 avec les composants de la machinerie d'épissage de l'ARNm, sur la correction de défaut d'épissages de certains pré-ARNm et donc des **opportunités thérapeutiques pour lutter contre la pandémie de la COVID-19**. La molécule d'ARN représente une des technologies d'avenir pour l'innovation médicale de cette décennie.

Concernant la **lutte contre la diffusion des AMR**, ces données pourront apporter des outils permettant d'utiliser à plus long terme certains antibiotiques de dernier recours, par exemple en couplant leur administration avec des **molécules inhibitrices des transferts de gènes par conjugaison** entre bactéries Gram-positives, ce qui limitera l'acquisition des AMR pour ces antibiotiques de la part de potentielles souches pathogènes. De telles approches sont déjà en cours de développement entre **P2** et **P4** pour des étapes clés de la conjugaison situées en amont du passage de l'ADN par le pore.

**P4** travaille en collaboration avec la **start-up Harmonic pharma** (co-fondé par l'équipe) de repositionnement moléculaire, *i.e.* l'utilisation pour de nouvelles indications thérapeutiques de médicaments déjà approuvés (ce qui assure des retombées économiques plus rapides qu'un nouveau médicament).

**Calendrier prévisionnel du programme : (1 page)**

**Phasage du projet** : Les équipements demandés sont essentiels pour le développement du projet et seront achetés au cours de la première année du projet.

Un co-financement pour l'achat de l'automate de cryo-préparation d'échantillons est prévu dans le cadre de la demande ANR PRC *CoVHijacksSplicing* de P1. Un deuxième co-financement, plus important, via notamment le 4<sup>ème</sup> Programme d'Investissement d'Avenir (PIA<sub>4</sub>) dans la recherche et l'innovation sera demandé pour l'acquisition du cryo-microscope de criblage. Les tutelles (Université de Lorraine, CNRS, INRAe) des partenaires directement concernés (P1 et P2) seront sollicités pour une contribution éventuelle.

L'investissement pour le stockage et le mur d'image se fera au sein des équipes concernées (P3 et P4) en concertation entre eux via des contrats et de l'autofinancement public.

**Nature des livrables attendus et dates prévisionnelles de livraison :**

La liste des livrables est la suivante en fonction du diagramme de Gantt.

L1 : rapport d'achat de l'automate de cryo-préparation

L2 : Ingénieur d'Etudes CDD (statutaire permanent, déjà financé)

L3 : Chercheur contractuel (à recruter, financement ANR)

L4 : Technicien CDD (statutaire permanent, déjà financé)

L5 : Chercheur contractuel (à recruter, non financé)

L6 : *kickoff*

L7 : rapport d'achat du mur d'images

L8 : rapport d'achat du cryo-microscope de criblage

L9-11 : rapports à mi-parcours des axes thématiques 1, 2 et 3 ?

L12-13 : réunion de travail annuelle

L14-17 : rapports finaux et publications des axes thématiques 1, 2 et 3

**Diagramme de Gantt :**

Axes	Coordination	Partenaires	0 mois	6 mois	12 mois	24 mois
Management	P1					
Equipements	P1, P3	Tous	L1	L7, L8		
Axe 1 : coronavirus/Epissage	P1	P1, P3, P4	L2, L3		L9	
Axe 2 : conjugaison bactérienne/AMR	P2	P2, P3, P4	L4		L10	
Axe 3 : Modélisation et IA	P3, P4	P3, P4	L5, L6		L11	
Réunions	P1	Tous	L6		L12	L13
Rapports Intermédiaires/Finaux	P1	Tous			L9-11	L14-17

(1 à 2 pages)

**Adéquation entre le programme et les moyens mis en œuvre pour sa réalisation :**

**Moyens d'accompagnement en termes d'infrastructures et de personnels pour le fonctionnement et la gestion des équipements :**

**Moyens humains (personnel permanent en ETP et en coûts) :** Chaque partenaire apportera les moyens humains complémentaires ci-dessous pour la réalisation du projet.

- **P1** ; 1 chercheur (DR-CNRS) et 1 ITA (IE-CNRS) statutaires pour environ 1 ETP/an pour un coût annuel de 100 k€. L'ingénieur d'étude du groupe (L2) a acquis la compétence pour préparer les grilles de cryo-EM dans l'équipe de D. Levy (Institut Curie, Paris) et cela sera sa priorité dès la réception de

l'équipement. L'équipe 1 a présenté un candidat aux concours 2021 de chargés de recherche du CNRS avec notamment des compétences cryo-EM. Il n'a pas été sélectionné cette année mais sera de nouveau présenté pour 2022. C'est une des priorités de recrutement pour le laboratoire IMoPA. En attendant, le chercheur contractuel recruté dans le cadre de l'ANR (L3) effectuera l'acquisition des images de cryo-EM et leur traitement.

- **P2** ; 4 chercheurs/chercheurs-enseignant/ITA statutaires dont environ 1,5 ETPT/an pour un coût annuel de 150 k€. L'équipe s'appuiera sur un de ses techniciens statutaires.

- **P3** ; 4 chercheurs/chercheurs-enseignant/ITA statutaires pour environ 1,5 ETP/an pour un coût annuel de 150 k€. Le chercheur contractuel à recruter (L5) aura en charge le développement d'algorithmes de *deep learning* permettant de décrire par des variables collectives réduites les changements conformationnels associés à la fonction des protéines.

- **P4** ; 5(7) chercheurs/ITA statutaires dont environ 1(3) ETPT/an pour un coût annuel de 100(240) k€. Le groupe compte recruter un chercheur CNRS et un chercheur INRIA en 2021 (sélectionnés pour audition). Le 1<sup>er</sup> candidat est expert en modélisation d'assemblages moléculaires (« *docking* »), en particulier membranaires, par approches de data-driven et basées sur l'apprentissage automatique. Le 2<sup>nd</sup> candidat est expert en *docking* basé sur des cartes de Cryo-EM (« *fitting* »), ainsi qu'en échantillonnage de conformations des molécules flexibles (ARN, boucles protéiques...).

**Moyens matériels, ...** : Chaque unité partenaire apportera des infrastructures complémentaires par l'utilisation de plateaux techniques/plateformes hautement performants qui seront utiles au bon déroulement du projet.

- **P1** : plateforme de Biophysique et Biologie Structurale (B2S, resp : S. Boschi-Muller) de l'UMS 2008 IBSLor (<https://umsibslor.univ-lorraine.fr/fr>) qui offre une combinaison d'équipements destinées à la caractérisation physicochimique des protéines, des ARNs et de leurs interactions. Elle donne accès à plusieurs équipements de biophysique (dichroïsme circulaire, microcalorimétrie, résonance plasmonique de surface, fluorescence) et de biologie structurale (spectromètre RMN 600 MHz équipé d'une cryo-sonde TCI et appareil à diffraction des rayons-X). Le plateau technique de purification d'IMoPA sera utilisé pour la purification de nos échantillons.

- **P2** : accès à la plateforme ASIA (Approches fonctionnelles et Structurales des InterActions cellulaires) du pôle A2F (<https://a2f.univ-lorraine.fr/asia/>). Cette plateforme offre un ensemble d'équipements complémentaire à ceux de la plateforme B2S, permettant de caractériser au niveau biophysique les macromolécules biologiques et leurs interactions (incluant ITC, SwitchSENSE, SEC-MALS). Resp. JM. Girardet. Par ailleurs, P2 est équipé pour la purification de protéines (AKTA, ultracentrifugation).

- **P3** : accès à la plateforme EXPLOR.

- **P4** : accès au cluster de calculs GRID5000 (15000 cores), plateforme SMEC (Simulation, Modélisation et Extraction de Connaissance), plateforme MBI-DS4H (Modélisation des Biomolécules et de leurs Interactions – Data Science for Health).

**Montant du programme correspondant aux dépenses d'investissement et de fonctionnement ; Pertinence et justification des coûts du programme**

Voir également le plan de financement joint.

Le coût du projet représente principalement le prix d'achat d'un cryo-microscope (1,5 Millions d'euros). Ce montant élevé est justifié par la présence d'une source d'électrons X-FEG extrêmement brillante, une station de chargement cryogénique très rapide, une automatisation optimisée quasi-complète de l'ensemble des étapes par apprentissage via l'IA y compris pour l'acquisition des images et leurs traitements. Ce cryo-microscope est facile d'utilisation permettant ainsi à des chercheurs non spécialistes en cryo-EM de pouvoir se former. Il est dédié pour l'analyse de particules uniques.

(1 page maximum)

**Modalité de pilotage et de coordination du programme**

**Aspects organisationnels du programme** : le projet sera organisé autour de 3 axes thématiques qui seront animés par un ou deux des partenaires. Le coordinateur s'assurera du respect des délais des livrables en cohérence avec le diagramme de Gantt. Il veillera à ce que la communication soit transparente entre tous les partenaires du consortium. Les résultats pourront être mis à disposition de l'ensemble des partenaires via un site internet sécurisé (type Mycore). Les logiciels développés par les équipes de simulations seront en accès libre. Le coordinateur, X. Manival (**P1**), maintiendra des contacts réguliers entre les différents partenaires pour faciliter l'avancement régulier et efficace du projet **MolAI4Cryo**. En plus des e-mails envoyés via une liste de diffusion, il organisera régulièrement des visioconférences et un workshop annuel en présentiel (avec invitation d'une personnalité experte d'un axe thématique) afin de discuter en profondeur des avancées scientifiques et des problèmes rencontrés. Un rapport final sera réalisé à la fin du contrat et transmis aux tutelles et aux financeurs. Nous veillerons à remercier les différents financeurs dont la région Grand-Est dans toutes les communications liées au projet et les réunions de présentation.

Pour chaque axe thématique, les principaux résultats scientifiques seront publiés dans des journaux généralistes ou spécialisés à forte visibilité mais également déposés dans l'archive pluridisciplinaire ouverte



et gratuite HAL (politique sur la science ouverte du ministère de la recherche). Les résultats seront également disséminés lors de communications orales dans divers congrès nationaux et internationaux reliés au projet.

**Qualification, rôle et implication des participants** : les équipes et unités ont été choisies de telles sortes qu'elles amèneront des systèmes biologiques et expertises complémentaires. Les **P3** et **P4** développeront des logiciels intégrant l'IA pour traiter le flux énorme des images cryo-EM des particules afin d'optimiser leur reconstruction 3D et pour étudier leurs interactions et dynamiques. Les **P1** et **P2** seront utilisateurs de ces derniers développements et utiliseront la plateforme de cryo-EM sur Nancy comme une satellite de celle de Strasbourg avec le soutien de B. Klaholz (CBI/IGBMC). Le présent consortium est tout à la fois complémentaire mais aussi très équilibré, présentant 2 partenaires expérimentaux et 2 partenaires computationnels.

**Modalité de pilotage et de coordination du projet** : IMoPA (**P1**, X. Manival) animera le réseau **MolAI4Cryo** avec un représentant pour chaque unité partenaire ; B. Douzi pour **P2** (DynAMic), F. Dehez pour **P3** (LPCT) et I. Chauvot de Beauchêne pour **P4** (LORIA). X. Manival et F. Dehez coordonneront l'achat et les appels d'offres des équipements en lien avec leurs tutelles respectives. Un comité directeur sera constitué avec les responsables de chaque partenaire pour discuter régulièrement des orientations à prendre au cours des deux années du projet. L'**axe scientifique 1** est coordonné par X. Manival, l'**axe 2** par B. Douzi, et l'**axe 3** par F. Dehez et I. Chauvot de Beauchêne. Nous organiserons une réunion de démarrage dans les premières semaines du début du projet, suivie d'une réunion par an (la dernière étant une réunion bilan, voir diagramme de Gantt).

(1 à 2 pages)

**Impacts et retombées socio-économiques :**

**Impact global du projet, potentiel de valorisation** : le projet **MolAI4Cryo** générera en particulier de nouveaux outils en IA pour le traitement d'images en cryo-EM et pour les calculs de DM qui permettront au consortium de créer une nouvelle synergie en Lorraine afin d'étudier la structure, dynamique et fonction des systèmes complexes et des machineries comprenant de multiples sous-unités. Ceci apportera des **connaissances fondamentales** essentielles sur la perturbation de l'assemblage et la structure-dynamique-fonction de la particule d'épissage en interaction avec la protéine virale nsp16 ou encore sur le complexe multi-protéique du pore de la conjugaison bactérienne. Le projet inclura des approches multidisciplinaires : biologie moléculaire, microbiologie, chimie, biochimie, biophysique computationnelle, biologie structurale, bioinformatique structurale, informatique, intelligence artificielle, traitement d'images, intégration multi-échelle.

Au niveau **biomédical**, l'information structurale obtenue par cryo-EM pourrait révéler de nouvelles pistes antivirales comme l'ARN constituant les particules d'épissage ou encore un inhibiteur des transferts de conjugaison et révolutionner notre capacité à lutter contre la pandémie de la COVID-19 et limiter les AMR.

Au niveau **biotechnologique**, **P4** a un partenariat étroit avec la start-up *Harmonic Pharma* pour le repositionnement de molécules actives, parmi lesquelles seront recherchés par criblage *in silico* des inhibiteurs des systèmes étudiés dans ce projet.

Au niveau **valorisation industrielle**, toutes les équipes sont directement en lien avec le service de valorisation de leurs tutelles et de la **SATT** (Société d'Accélération du Transfert de Technologies) Grand-Est **SAYENS** (Unistra et UL) qui facilitent le transfert technologique, le dépôt de brevets et des collaborations avec des entreprises de biotechnologie locales. Un doctorant de P1 vient d'être recruté comme chef de projet auprès de l'antenne en Lorraine ce qui facilitera la démarche.

**Perspectives du projet en matière de rayonnement, de propriété industrielle, d'enjeu compétitif ...** : le projet fournira une énorme quantité de données et d'informations qui seront stockées dans un entrepôt de données qui sera ouvert à la communauté après un embargo inhérent à la publication (respect de la politique de l'*open data*). Nos résultats seront rapidement publiés dans des journaux à haut facteur d'impact (grâce à l'IA et la cryo-EM qui sont des techniques très porteuses et prometteuses) et seront présentés dans des congrès, mettant ainsi en avant la visibilité nationale et internationale de la région Grand-Est. Les développements et *workflows* numériques produits pourront être mis à disposition de la communauté scientifique, voire selon les résultats obtenus, conduire à un serveur d'interrogation web.

X. Manival et F. Dehez doivent organiser le prochain *Rhine-Knee Regional Meeting* (existe depuis 1987, rotation entre la Suisse, l'Allemagne et la France) sur la biologie structurale à Nancy en 2022. L'ensemble des partenaires du réseau et leurs collaborateurs seront invités à ce congrès transfrontalier international.

(2 à 3 pages maximum)

**Descriptifs des équipements demandés dans le cadre du programme :**

- Description détaillée, caractère structurant et dimension mutualisée de l'équipement
- Qualité du modèle économique associé à l'activité de l'équipement
- Offres de service, organisation, ouverture aux communautés scientifiques académiques et industrielles intéressées



Trois types d'équipements essentiels à la réussite du projet sont demandés :

(1) Une première série d'équipements concerne la préparation des grilles pour la cryo-EM avec un **nettoyeur par plasma** (*plasma cleaner*, élimine les contaminations hydrocarbonées et rend les grilles hydrophiles), un **automate de vitrification d'échantillons** biologiques pour cryo-EM et un **porte-objet dédié aux cryo-transferts**. Ils sont indispensables pour ensuite acquérir des images sur le cryo-microscope de criblage. Ils seront installés sur la plateforme B2S (UMS 2008 IBSLor) dont les locaux sont dans le même bâtiment que l'unité coordinatrice (P1, IMoPA). Ils seront à disposition pour l'ensemble des équipes de Lorraine qui désire utiliser la cryo-EM. Ce type d'équipement n'existe pas à ce jour en Lorraine.

(2) Le deuxième équipement consistera en l'**infrastructure de pointe** principale, à savoir un **cryo-microscope électronique à transmission de criblage** (100 kV, pointe X-FEG haute brillance, passeur automatique d'échantillons, automatisation optimisée quasi-complète de l'ensemble des étapes par apprentissage via l'IA y compris pour l'acquisition des images et leurs traitements) compatible avec la plateforme haute résolution de Strasbourg. Il faudra qu'il puisse basculer rapidement à température ambiante pour faire de la coloration négative pour répondre à la demande spécifique de certaines équipes de l'université de Lorraine (dont **P2**) et donc accroître son caractère structurant. Il sera essentiel pour cribler les meilleures conditions de préparation des échantillons en vue d'une analyse par cryo-EM à haute résolution à Strasbourg (CBI). Ce microscope sera également installé sur la plateforme B2S de Nancy. Il sera essentiel aux partenaires du projet et au moins à 5 équipes de recherche en Lorraine (prospective effectuée par P1). Ce type d'équipement sera ouvert à la communauté scientifique et il n'existe pas à ce jour en Lorraine.

(3) Le troisième équipement correspond à une infrastructure numérique complète permettant de visualiser les données sur une architecture 3D collaborative (mur d'image piloté par des serveurs graphiques GPU) qui sera accessible à l'ensemble des partenaires.

Ces équipements sera localisé dans des plateformes hautement performantes qui seront ouvertes à l'ensemble des partenaires du réseau **MolAI4Cryo**, des laboratoires académiques et industriels régionaux (Grand-Est) et nationaux. Ces plateformes sont labellisées avec des démarches de qualité. A titre d'exemple, la plateforme **B2S** de Nancy est labellisée Star-LUE (3 étoiles sur trois possibles). Ce label est délivré par Lorraine-Université d'Excellence (LUE).

L'offre tarifaire à la journée sera différenciée en fonction des partenaires internes, externes, académiques ou industriels.

#### Références :

1. Noe, F., Olsson, S., Kohler, J. & Wu, H. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science* **365**, (2019).
2. Senior, A. W. *et al.* Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).
3. Matsumoto, S. *et al.* Extraction of protein dynamics information from cryo-EM maps using deep learning. *Nat. Mach. Intell.* **3**, 153–160 (2021).
4. Banerjee, A. K. *et al.* SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses. *Cell* **183**, 1325–1339.e21 (2020).
5. Bojkova, D. *et al.* Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature* **583**, 469–472 (2020).
6. Gordon, D. E. *et al.* Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science* **370**, (2020).
7. Low, H. H. *et al.* Structure of a type IV secretion system. *Nature* **508**, 550–553 (2014).
8. Park, A. & Iwasaki, A. Type I and Type III Interferons – Induction, Signaling, Evasion, and Application to Combat COVID-19. *Cell Host Microbe* **27**, 870–878 (2020).
9. Decroly, E., Ferron, F., Lescar, J. & Canard, B. Conventional and unconventional mechanisms for capping viral mRNA. *Nat. Rev. Microbiol.* **10**, 51–65 (2012).
10. Fung, T. S. & Liu, D. X. Human Coronavirus: Host-Pathogen Interaction. *Annu. Rev. Microbiol.* **73**, 529–557 (2019).
11. Ribero, M. S., Jouvenet, N., Dreux, M. & Nisole, S. Interplay between SARS-CoV-2 and the type I interferon response. *PLOS Pathog.* **16**, e1008737 (2020).
12. Chen, Y. *et al.* Biochemical and Structural Insights into the Mechanisms of SARS Coronavirus RNA Ribose 2'-O-Methylation by nsp16/nsp10 Protein Complex. *PLOS Pathog.* **7**, e1002294 (2011).

## C. BUDGET DU PROGRAMME

Coût total prévisionnel du programme :	3 279 500 €
dont Investissement :	1 999 500 €
dont Fonctionnement :	1 280 000 €

Régime TVA : X HT

Montant total de l'aide régionale sollicitée : 499 875 € en Investissement

### I. Les dépenses prévisionnelles

Se reporter à **l'Annexe « Plan de financement Infrastructure 2021 »** pour renseigner les dépenses prévisionnelles du programme, poste par poste, et par section de dépenses (investissement et fonctionnement).

### II. Echancier du programme

Les montants des dépenses prévisionnelles peuvent être ventilés par année dans la même annexe. **La durée maximale du programme est de 2 ans.**

### III. Les ressources

Se reporter à l'Annexe « Plan de financement Infrastructure 2021 » pour renseigner le tableau détaillé des ressources du programme par section de dépenses (investissement et fonctionnement).



**L'annexe « Plan de financement Infrastructure 2021 » est à produire et à compléter pour le programme dans sa globalité pour les sections d'investissement et de fonctionnement même si les dépenses éligibles pour la Région Grand Est concernent uniquement l'investissement.**

## D. APPRECIATION ET ENGAGEMENT DE L'ORGANISME GESTIONNAIRE

### I. Appréciation et engagement de l'organisme gestionnaire

Ce dossier doit avoir fait l'objet d'un avis du Conseil Scientifique ou de la Commission Recherche de l'établissement demandeur (pour les Universités) ou de l'autorité compétente (pour les autres établissements d'enseignement supérieur et de recherche).

### II. Engagement de l'organisme gestionnaire

Je soussigné, M. Pierre Mutzenhardt, Président de l'Université de Lorraine, sollicite une aide de la Région Grand Est de 500 000 € pour la réalisation du programme intitulé « ..... » évalué à ..... € HT.

Fait à ....., le .././...

Cachet et signature (représentant légal ou délégué)