

Étude *in silico* d'une mutation : application à la mutation fonctionnelle G12D du gène *KRAS*, retrouvée dans les cancers bronchiques non à petites cellules

In silico study of a functional mutation associated with non-small cell lung cancer: G12D mutation of exon 2 in KRAS gene

Noria Bouras^{1,2}
Abdelkader Bousahba³
Ahlam Megaïz³
Florence de Fraipont⁴
Tewfik Sahraoui¹
Fatima Zohra El Kebir¹

¹ Laboratoire de biologie du développement et de la différenciation (LBDD), Université Oran 1 Ahmed Ben Bella, Oran, Algérie

² Département de génétique moléculaire appliquée (GMA), Université des sciences et de la technologie d'Oran (USTO-MB), Algérie

³ Service d'oncologie médicale, Centre hospitalo-universitaire, CHU d'Oran, Algérie

⁴ Unité médicale de génétique moléculaire : maladies héréditaires et oncologie, Centre hospitalo-universitaire, CHU de Grenoble Alpes (CHUGA), France

Résumé. La biologie a connu un essor fulgurant durant le XX^e siècle et a été profondément bouleversée au cours de la dernière décennie par le passage à l'ère post-génomique, la généralisation de la biologie à haut débit et l'avènement d'une discipline relativement jeune, la bioinformatique. Cette dernière, qui englobe la collecte, l'organisation et l'analyse de données biologiques au moyen d'outils informatiques, est rapidement devenue indissociable des études relatives à la compréhension du génome. Les différentes mutations qui peuvent affecter nos gènes sont responsables d'un changement de la séquence protéique et sont susceptibles d'affecter par exemple la stabilité de la protéine, son adressage intracellulaire, sa maturation, son assemblage dans une structure multimérique, les sites essentiels à son activité enzymatique ou à l'interaction avec des ligands. Ainsi, un certain nombre de développements bioinformatiques ont permis de mettre en place des outils de prédiction *in silico* de la structure d'une protéine et de l'impact des mutations sur la structure des protéines. Dans notre étude, nous avons exploré *in silico* trois logiciels analytiques de prédiction et de modélisation, en choisissant comme modèle d'étude la mutation G12D qui affecte l'oncogène *KRAS* (V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog), retrouvée fréquemment chez les patients atteints de cancers bronchopulmonaires non à petites cellules originaires de l'ouest algérien. Ce travail nous a permis d'intégrer ces outils bioinformatiques au sein de notre laboratoire de biologie du développement et de la différenciation de l'université Oran 1 Ahmed Benbella, en Algérie. Ainsi, nous avons pu conclure que, même si la mutation trouvée est donnée comme tolérée et sans effet délétère sur la protéine Ras entière, la conséquence de cette mutation faux-sens dépend principalement de la position de l'acide aminé muté dans la protéine et de ses propriétés chimiques. En effet, la protéine Ras mutée montre une forte affinité avec la molécule GTP.

Mots clés : *étude in silico, mutation fonctionnelle G12D, KRAS, cancer bronchique non à petites cellules*

Abstract. Biology flourished during the XXth century and was profoundly disrupted during the last decade because of the transition to the post-genomic era, the spread of high-throughput biology, and the advent of a relatively new discipline, namely bioinformatics. This latter, which encompasses the collection, organization and analysis of biological data using the computer tool, has quickly become inseparable from the studies related to the genome understanding. The consequences of the different mutations that may affect our genes

Article reçu le 23 juillet 2018,
accepté le 25 février 2019

Correspondance : N. Bouras
<noriabouras@yahoo.fr>

are responsible for a change in the protein sequence and are likely to affect, for example, the stability of the protein, its intracellular targeting, its maturation, its assembly in a multimeric structure, the essential sites for its enzymatic activity or for the interaction with ligands. Thus, a number of bioinformatic developments have made it possible to set up *in silico* prediction tools of the structure of a protein that is aiming at predicting the impact of local mutations on the structure of proteins. Throughout our study, we have been interested in exploring, through *in silico* bioinformatic study, three analytical, prediction and modeling, software, choosing as exemple the G12D mutation that affects the proto-oncogene *KRAS* found in numerous algerian patients with bronchopulmonary cancers cells (NSCLC). This study allowed us to integrate these bioinformatic tools into our laboratory of developmental biology and LBDD differentiation at the University of Oran 1 Ahmed Benbella, in Algeria. Thus, we have been able to conclude, even if the found mutation is predicted to be tolerated and has no deleterious effect on the entire Ras protein, that the consequence of this missense mutation depends mainly on the position in the protein and the chemical properties of the amino acid involved in the substitution and which shows a strong affinity with the GTP molecule.

Key words: *in silico* study, functional mutation G12D, *KRAS*, non-small cell lung cancer

L'étude *in silico* (dont la traduction littérale en français est « en silice » un matériau principal retrouvé dans les puces informatiques de tous les ordinateurs) est un néologisme d'inspiration latine désignant une recherche ou un essai effectué au moyen de calculs complexes informatisés ou de modèles informatiques. Cette expression est surtout utilisée dans les domaines de la génomique associée à la bioinformatique (https://fr.wikipedia.org/wiki/In_silico). C'est un domaine qui prend de plus en plus d'importance avec le développement des analyses à grande échelle.

Dans ce cadre, une étude bioinformatique *in silico* est réalisée avec trois logiciels analytiques de prédiction et de modélisation afin de prédire l'impact de la mutation G12D chez des patients atteints de CBNPC au sein de la population de l'ouest algérien. Ce sont les logiciels SIFT, Aligne GVGD et Suiss PDP Viewver.

La mutation choisie est une mutation de la guanine en adénine (G-A) au niveau du codon 12 de l'exon 2 du gène *KRAS* induisant un changement de l'acide aminé glycine vers l'acide aspartique sur la chaîne polypeptidique de la protéine Ras. En effet, le gène *KRAS* est constitué de 6 exons et code pour une protéine de 189 acides aminés d'un poids moléculaire de 21 kDa, connue sous le nom de Ras qui a un rôle important dans la transmission de signaux intracellulaires provenant des récepteurs membranaires vers le noyau. Elle joue un rôle d'interrupteur au sein des voies de signalisation alternant entre deux états : un état actif où elle est liée au GTP (guanosine tri-phosphate), et un état inactif où elle est liée au GDP (guanosine diphosphate). Ainsi, l'activation des protéines Ras se produit

lors du remplacement du GDP par le GTP. En revanche, leur inactivation est provoquée par l'hydrolyse du GTP en GDP par des protéines de régulation telles que les GAP9 (*GTPase-activating proteins*), ainsi que par l'activité GTPase intrinsèque de la protéine Ras elle-même [1].

La mutation sur le gène *KRAS* est retrouvée dans 15 à 20 % de l'ensemble des cancers bronchiques dont 30 à 50 % sont des cancers broncho-pulmonaires non à petites cellules (CBNPC) [2]. Ces altérations du gène *KRAS* consistent en des mutations ponctuelles qui détériorent l'activité GTPasique intrinsèque de la protéine Ras. Ainsi, Ras s'accumule sous sa forme active et maintient l'activation constitutive des voies de signalisation qui va mener à une prolifération cellulaire incontrôlée [3]. Les mutations affectant le gène *KRAS* sont presque systématiquement localisées au niveau du codon 12. Néanmoins, ces mutations peuvent toucher d'autres codons tels que le 13, ou plus rarement le codon 61 [4].

L'étude *in silico* que nous présentons consiste à utiliser des logiciels analytiques de prédisposition en prenant comme modèle la mutation G12D. Celle-ci a été fréquemment retrouvée chez des patients atteints de CBNPC originaires de l'ouest algérien, dont les prélèvements ont été analysés par pyroséquençage dans le laboratoire de génétique moléculaire : maladies héréditaires et oncologie du CHU de Grenoble (GM : MHO, CHUGA, France) [5]. Cette mutation représente une transversion de la guanine vers l'adénosine au niveau du codon 12 de l'exon 2 du gène *KRAS* et qui est responsable de la substitution de l'acide aminé glycine vers l'acide aminé acide

aspartique au niveau de la chaîne polypeptidique de la protéine RAS.

Ce travail a permis d'une part, de prédire l'impact de la mutation dans notre population sur la protéine Ras perçue le plus souvent mutée dans les CBNPC et, d'autre part, d'étudier l'émergence des analyses numériques permettant ainsi l'acquisition rapide de données qu'elles soient génomiques et ou protéomiques pour la communauté biologique et scientifique.

Matériel et méthodes

Les logiciels que nous avons utilisés pour mener cette étude sont décrits ci-après.

SIFT (*sorting intolerant from tolerant*) est un logiciel gratuit, téléchargeable à partir du site : <http://sift-dna.org>. Il est exclusivement basé sur des données de conservation entre séquences homologues et sur la construction d'une matrice de substitution spécifique des positions dans la séquence. Il utilise pour cela des informations de l'alignement multiple pour prédire si une substitution sera tolérée ou non à chaque position de la séquence (PSSM pour *position-specific substitution matrice*). Ainsi, SIFT calcule des probabilités

normalisées sous forme de SCORE pour toutes les substitutions possibles à chaque position. Ce score calculé varie entre 0 et 1. En conséquence, les valeurs comprises entre 0 et 0,05 sont considérées comme délétères. En revanche, les probabilités supérieures à ce seuil prédéfini seront considérées comme tolérées [6].

Aligne GVGD (*grantham variation grantham deviation*) est un programme gratuit, téléchargeable à partir du site : http://agvgd.iarc.fr/agvgd_input.php. Il détermine les caractéristiques biophysiques des acides aminés et des protéines selon des alignements multiples de séquences pour prédire l'impact des mutations sur la protéine, ainsi que leur conservation au sein de différentes espèces. Ce logiciel s'appuie sur deux scores différents : GV (variation biochimique à chaque position d'alignement entre différentes espèces) et GD (différence de propriétés physicochimiques entre l'acide aminé sauvage et muté). Ces derniers sont répartis en 7 classes : C0, C15, C25, C35, C45, C55 et C65. Le score C0 prédit que la substitution est considérée comme neutre. Les scores C15 et C25 prédisent que les substitutions sont considérées comme intermédiaires et que les changements de structure et de fonction sont incertains. En revanche, les scores C35 et plus, prédisent que les substitutions sont délétères [7].

Predict Not Tolerated	Position	Seq Rep	Predict Tolerated
y wvtsr qpn l ki h g f e d c a	1M	1.00	M
y wvsr qpnml ki h g f e d c a	2T	1.00	T
y wvtsr qpn ml ki h g f d c a	3E	1.00	E
w v t s r q p n m l ki h g f e d c a	4Y	1.00	Y
y wvtsr qpn ml i h g f e d c a	5K	1.00	K
y wvtsr qpn m ki h g f e d c a	6L	1.00	L
y wtsr qpnml ki h g f e d c a	7V	1.00	V
y wtsr qpnml ki h g f e d c a	8V	1.00	V
y wtsr qpnml ki h g f e d c a	9V	1.00	V
y wvtsr qpn ml ki h f e d c a	10G	1.00	G
y wvtsr qpn ml ki h g f e d c	11A	1.00	A
	12G	1.00	q c r i n e d k l t p V s a G
y wvtsr qpn ml ki h f e d c a	13G	1.00	G
y wtsr qpnml ki h g f e d c a	14G	1.00	V
y wvtsr qpn ml ki h f e d c a	15G	1.00	G
y wvtsr qpn ml i h g f e d c a	16G	1.00	K
y wvtr qpnml ki h g f e d c a	17S	1.00	S
y wvtsr qpn ml ki h g f e d c	18A	1.00	A

Figure 1. Résultats de l'analyse par le logiciel SIFT de la mutation G12D : prédiction pour la position des acides aminés 1 à 18. Code couleur des acides aminés : noir : non polaire, vert : polaire non chargé, rouge : basique, bleu : acide. Les lettres majuscules indiquent les acides aminés apparaissant dans l'alignement, les lettres minuscules résultent de la prédiction.

SPDBV (*Swiss-PdbViewer*) est un logiciel dont l'application est téléchargeable gratuitement sur le lien : <https://spdbv.vital-it.ch/>. Il s'agit d'un programme graphique moléculaire interactif pour visualisation et analyse des structures de protéines et d'acides nucléiques. Ces structures sont modélisées en 3D.

Résultats et discussion

Analyse par SIFT

Le résultat du logiciel SIFT est représenté sur la *figure 1*. L'analyse est représentée sous formes de colonnes. La colonne de gauche représente les acides aminés à une position donnée qui sont prédits intolérants. À l'inverse, celle de droite représente les acides aminés qui sont prédits tolérants.

Le logiciel SIFT a positionné la mutation G 12 D responsable de la substitution de l'acide aminé glycine vers l'acide aminé acide aspartique, au niveau de la colonne droite. Cela prédit que cette mutation est tolérante. En outre, le score calculé est de 0,08 (*figure 2*), ce qui indique que cette sub-

stitution n'altère pas la fonction de la protéine *Ras*. En effet, La protéine mutée aura dans ce cas la même longueur que la protéine sauvage car elle conservera le même nombre d'acides aminés. La mutation causale est tout à fait tolérée par la protéine qui continuera à assurer ses fonctions, contrairement aux autres mutations dites « non tolérées » qui peuvent avoir un effet délétère. La conséquence de cette mutation faux-sens dépend donc principalement de la position au sein de la protéine et des propriétés physico-chimiques de l'acide aminé impliqué dans la substitution. Selon les propriétés physicochimiques des acides aminés, la substitution G12D engendre le changement d'un résidu non polaire et hydrophobe (glycine) par un résidu acide (acide aspartique). Ce changement de groupe pourrait être à l'origine d'effets sévères de la pathologie comme le prédit le logiciel.

Analyse par Aligne GVGD

Les résultats par analyse du logiciel Aligne GVGD (*figure 3*) montrent que la valeur GV est de 109,55, ce qui prédit que la mutation siège au niveau d'une région non conservée entre les espèces analysées. Au cours de l'évolution, les acides aminés qui jouent un rôle important pour la structure ou la fonction d'une protéine ne sont pas ou peu modifiés et se retrouvent à la même place dans toutes les espèces et sont dits conservés. Cependant, les acides aminés moins 'vitaux' pour la fonction de la protéine peuvent être plus ou moins modifiés en fonction du moment auquel les espèces ont divergé. Ceci laisse supposer que la mutation

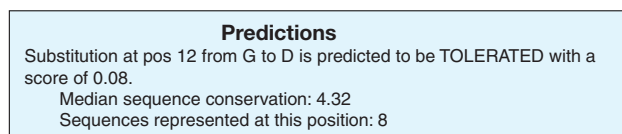


Figure 2. Résultat du score de mutation G12D par le logiciel SIFT.

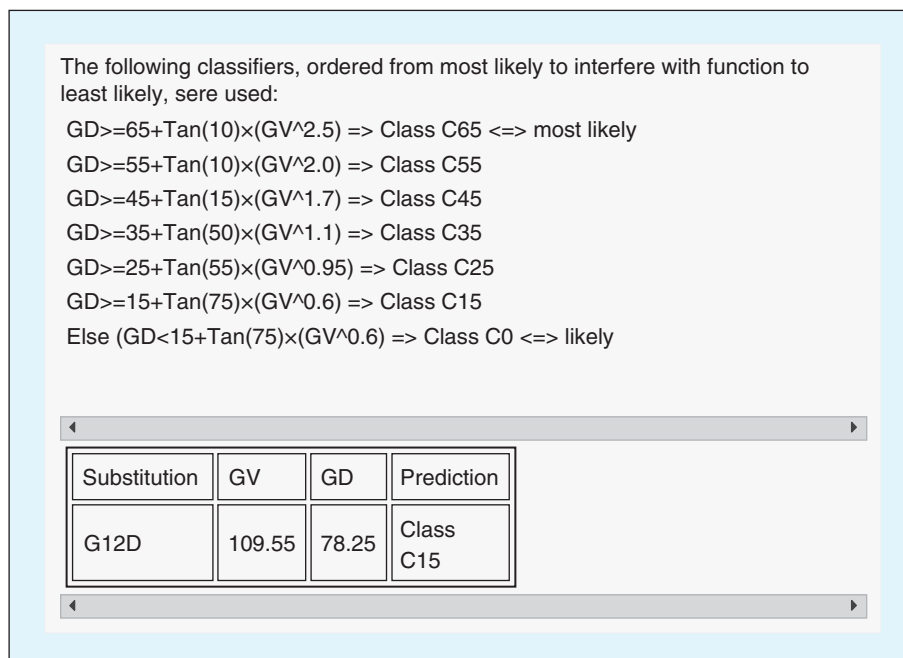


Figure 3. Résultat de l'analyse de la mutation G12D par le logiciel Align GVGD.

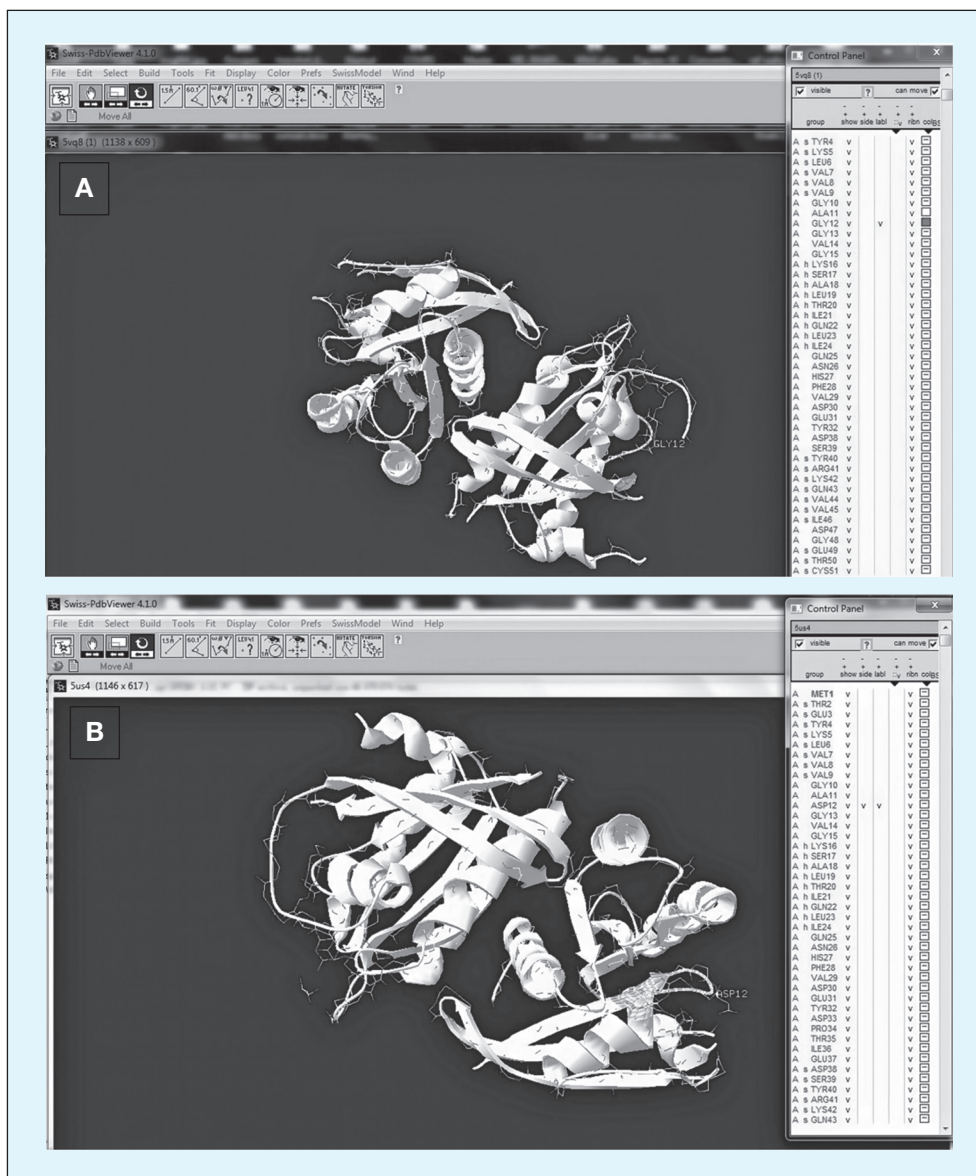


Figure 4. Structure 3D de la protéine Ras obtenue par le logiciel Swiss PDP Vieuver : **A** : Ras avec résidu glycine à la position 12 ; **B** : Ras avec résidu acide aspartique à la position 12.

G12D n'altère pas encore une fois la structure protéique retrouvée dans différentes espèces puisqu'elle continuera à exercer son rôle dans les voies de signalisation [8]. Concernant la valeur GD qui est de 78,25, ceci signifie que les propriétés physico-chimiques entre l'acide aminé glycine et l'acide aminé acide aspartique sont modérément conservées. En outre, cette mutation appartient à la classe C15 ce qui signifie que la substitution n'altère pas la fonction de la protéine Ras. En effet, il est bien connu que toutes les mutations qui affectent le gène *KRAS* entraînent l'expression d'une protéine Ras constitutionnellement active et donc liée au GTP d'une manière continue

sans pour autant altérer sa fonction. Cette activation permanente de Ras a pour conséquence d'activer la voie des MAP kinases (*mitogen-activated protein kinases*), qui stimule la prolifération cellulaire et qui peut même activer la voie PI3kinase (phosphatidylinositol-3-kinase) [1].

Analyse par Swiss PDP Vieuver

La figure 4 montre la structure 3D de la protéine Ras avec en position 12, le résidu sauvage glycine (figure 4A) et le résidu muté acide aspartique (figure 4B). Les résultats montrent que le résidu muté est situé à la surface d'un

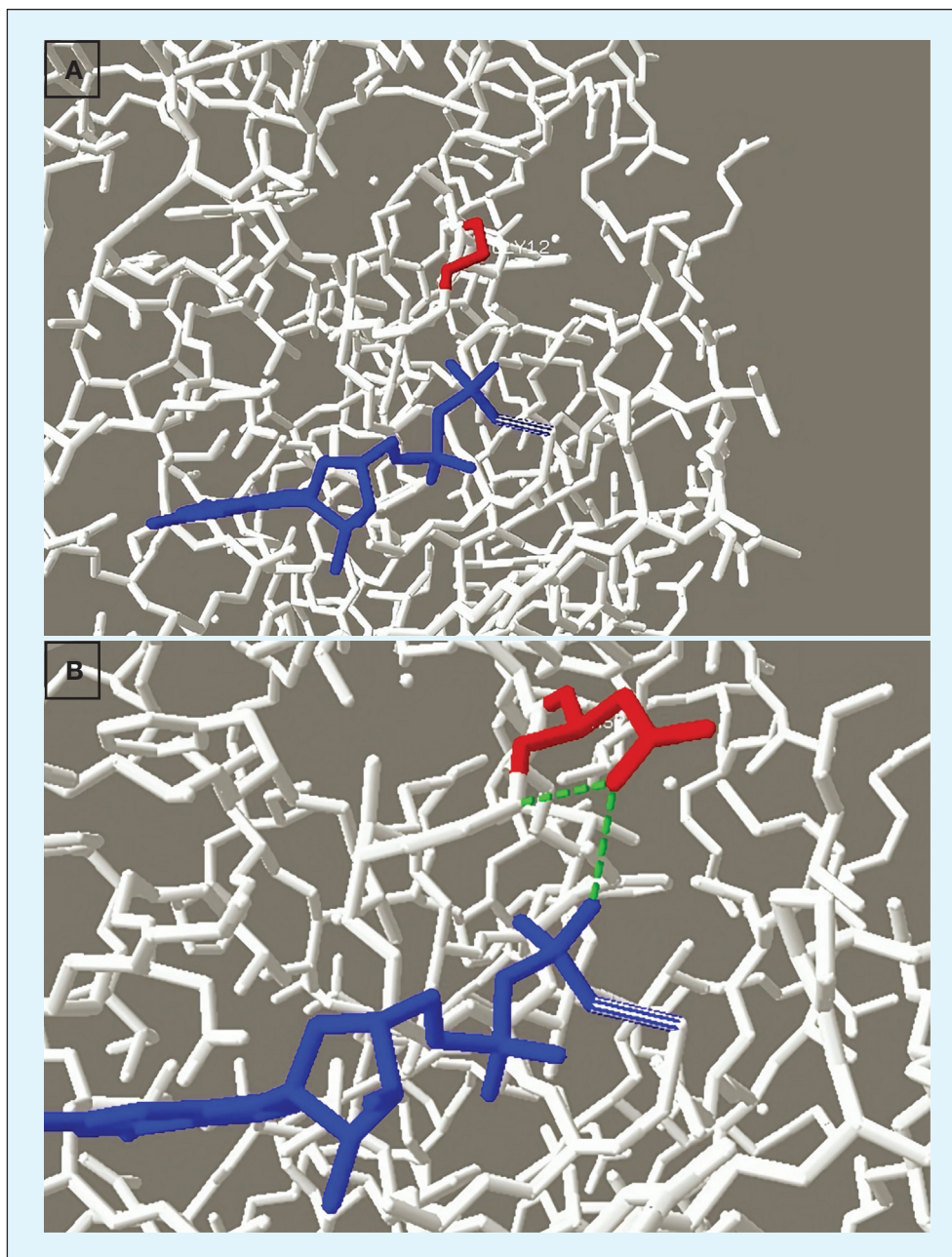


Figure 5. Structure 3D de l'acide aminé à la position 12 obtenu par le logiciel Swiss PDP Viewer : **A** : acide aminé sauvage glycine ; **B** : acide aminé muté acide aspartique.

domaine qui est important pour la liaison avec d'autres molécules et qui peut contenir un grand nombre de résidus à cette position. Ainsi, les résidus mutants pourraient perturber ces interactions intermoléculaires et par conséquent, affecter des assemblages de protéines qui sont susceptibles d'être biologiquement pertinentes. En effet, l'analyse par cristallographie des différentes formes (active ou inactive) de la protéine Ras a montré un domaine catalytique, identifié entre les résidus 1 et 171, incluant

la région impliquée dans la liaison des acides aminés codés par les codons 12 et 13 de l'exon 2 du gène. En revanche, la région comprise entre les résidus 32 et 40 change de structure et représente un élément essentiel pour toutes les interactions avec les molécules effectrices et les GAP. Enfin la région en C-terminal regroupant les quatre derniers acides aminés a été identifiée comme essentielle à l'ancrage de la protéine Ras dans la membrane plasmique [9].

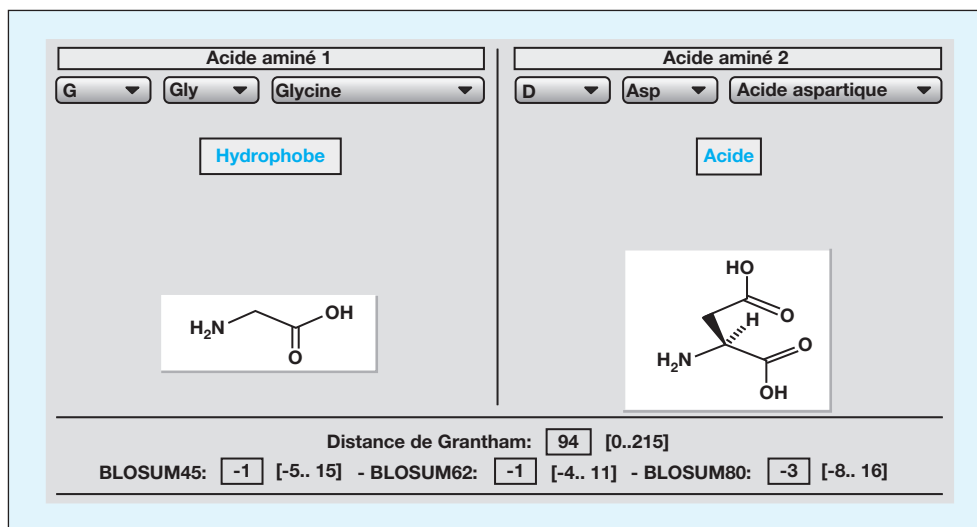


Figure 6. Comparaison de l'acide aminé glycine versus acide aspartique.

Dans la structure 3D, on peut voir que le résidu à la position 12 de la chaîne polypeptidique, coloré en rouge, a des interactions avec le ligand GTP, coloré en bleu (figure 5). En effet, la différence entre les propriétés de type sauvage et la mutation peut facilement entraîner un gain ou une perte d'interactions avec le ligand. Parce que la liaison du ligand est souvent importante pour la fonction de la protéine, cette fonction pourrait être perturbée par cette mutation. Dans notre étude, il est bien clair que le résidu mutant présenterait des liaisons hydrogène (colorées en vert) (figure 5B) contrairement au résidu sauvage (figure 5A), ce qui laisse prédire que l'interaction Ras GTP est renforcée quand *KRAS* est muté. Cela peut avoir un effet direct sur la protéine puisqu'il a été démontré que les mutations qui affectent le gène *KRAS* bloquent la protéine Ras en configuration active à travers des changements au niveau du domaine de liaison GTP [4].

En outre, l'analyse Swiss PDB décrit que le résidu à la position 12 fait partie d'un domaine interpro nommée petite GTPase qui fait partie des Superfamilles *Ras*. En effet, quand Ras est sous forme activée, sa fonction GTPasique est réduite, ce qui limite son action sur la régulation de la croissance cellulaire [10]. Ainsi, et étant donné que la mutation G12D engendre une protéine Ras constitutivement activée, cela semblerait modifier sa fonction GTPasique et par conséquent, réduire son rétrocontrôle négatif.

La figure 6 obtenue par le logiciel montre les différences structurelles des deux acides aminés sauvage et muté. Il est connu que ces deux acides aminés sont différents selon plusieurs caractéristiques dont les plus importantes sont :
 – les acides de type sauvage et mutant diffèrent par la taille. En effet le résidu mutant (acide aspartique) est plus grand que le résidu de type sauvage ;

– le résidu de type sauvage glycine est un acide aminé caractérisé par une chaîne latérale qui se réduit à un simple atome d'hydrogène et dont le carbone α n'est donc pas un centre stéréogène ;

– il y a une différence de charge entre le type sauvage et de l'acide aminé mutant puisque l'acide aminé sauvage est un acide aminé neutre et hydrophobe. Concernant le résidu mutant, l'acide aspartique, est un acide aminé caractérisé par une chaîne chargée négativement. Cela peut provoquer la répulsion entre ce résidu mutant et des résidus voisins (https://fr.wikipedia.org/wiki/Acide_amin%C3%A9). Toutes ces caractéristiques peuvent influencer sur l'interaction avec les ligands GDP et GTP.

Conclusion

Dans cette étude, nous avons démontré avec la mutation G12D du gène *KRAS*, très fréquente dans différents types de tumeurs, que l'utilisation d'une combinaison de logiciels disponibles gratuitement peut être intéressante pour caractériser les effets de cette mutation. En effet, nous avons confirmé que la mutation G12D n'a pas d'effet délétère sur la protéine Ras entière, mais que la conséquence de cette mutation faux-sens dépend principalement de la position de l'acide aminé dans la protéine et de ses propriétés chimiques. Nous avons décidé d'implanter cette étude *in silico* dans notre laboratoire de biologie du développement et de la différenciation de l'Université Oran 1 Ahmed Benbella. En effet, il nous semble intéressant de pouvoir effectuer la même démarche avec d'autres mutations pour d'autres gènes, comme par exemple les variants détectés lors du séquençage à haut débit (séquençage NGS) des tumeurs.

Parfois, pour ces nouveaux variants, peu d'informations sur leur pathogénicité sont disponibles dans les bases de données telles que la base COSMIC (*Catalogue of somatic mutation in cancer* : <https://cancer.sanger.ac.uk/cosmic>). L'étude *in silico* nous permettrait donc d'extrapoler ces effets au phénotype clinique des patients en fournissant les informations nécessaires pour lier un génotype à un phénotype.

Liens d'intérêts : les auteurs déclarent ne pas avoir de lien d'intérêts en rapport avec cet article.

Références

1. Lièvre A, Laurent-Puig P. La voie de signalisation RAS/MAPK. *Cancéro Dig* 2010 ; 2 : 38-42.
2. Riely GJ, Kris MG, Rosenbaum D, Marks J, Li A, Chitale DA, *et al.* Frequency and distinctive spectrum of *KRAS* mutations in never smokers with lung adenocarcinoma. *Clin Cancer Res* 2008 ; 14 : 5731-4.
3. Trahey M, McCormick F. A cytoplasmic protein stimulates normal *N-Ras* p21 GTPase, but does not affect oncogenic mutants. *Science* 1987 ; 238 : 542-5.
4. Mounawar M, Hainaut P. Biologie cellulaire et moléculaire des cancers broncho-pulmonaires. In : *Cancers broncho-pulmonaires : mise au point*. Montrouge : John Libbey Eurotext, 2006 : 145-60.
5. Dufort S, Richard MJ, de Fraipont F. Pyrosequencing method to detect *KRAS* mutation in formalin-fixed and paraffin-embedded tumor tissues. *Anal Biochem* 2009 ; 391 : 166-8.
6. Ng PC, Henikoff S. SIFT : predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003 ; 31 : 3812-4.
7. Tavtigian SV, Byrnes GB, Goldgar DE, Thomas A. Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. *Hum Mutat* 2008 ; 29 : 1342-54.
8. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* 1970 ; 48 : 443-53.
9. Tong MV, deVos L, Brunger AM, Yamaizumi AZ, Nishimura S, Kim SH. Molecular Switch for signal transduction : structural differences between active and inactive forms of protooncogenic *ras* proteins. *Science* 1990 ; 247 : 939-45.
10. DeClue JE, Papageorge A, Fletcher J, Diehl SR, Ratner N, Vass WC, *et al.* Abnormal regulation of mammalian p21ras contributes to malignant tumor growth in von Recklinghausen (type 1) neurofibromatosis. *Cell* 1992 ; 69 : 265-73.