



HAL
open science

Recalage de nuages de points 3D pour la comparaison de molécules pharmacologiques

Dominique Douguet, Frédéric Payan

► To cite this version:

Dominique Douguet, Frédéric Payan. Recalage de nuages de points 3D pour la comparaison de molécules pharmacologiques. CORESA (COMpression et REprésentation des Signaux Audiovisuels), Nov 2021, Sophia Antipolis, France. hal-03414972

HAL Id: hal-03414972

<https://hal.science/hal-03414972v1>

Submitted on 4 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Recalage de nuages de points 3D pour la comparaison de molécules pharmacologiques

Dominique Douguet
UCA - IPMC - Inserm

Frédéric Payan
UCA - I3S - CNRS

Résumé : *Le calcul de la similarité entre molécules est un sujet très étudié dans le domaine de la pharmacologie et de la modélisation moléculaire. Jusqu'à présent, bien que présentant de nombreux avantages, les représentations 3D des surfaces moléculaires ont été assez peu employées pour trouver des similarités entre molécules. Dans cette étude, nous montrons qu'il est possible d'appliquer des techniques de recalage 3D usuellement utilisées dans d'autres domaines (reconstruction 3D, vision par ordinateur, etc.) pour identifier des similarités entre molécules. Pour cela nous avons représenté ces dernières à l'aide de nuage de points 3D colorés, la couleur étant liée aux propriétés physico-chimiques des atomes composant les molécules. Ensuite, nous les avons alignés en appliquant successivement un recalage global, et un recalage local spécifique aux nuages colorés. En se basant sur des mesures de fitness et de distance géométrique, nous montrons que notre approche est particulièrement efficace pour superposer des molécules de taille semblable mais aussi lorsqu'elles sont de tailles différentes.*

Mots-clés : Nuage de points 3D coloré, recalage, surface moléculaire, descripteurs FPFH.

1 Introduction

Lorsqu'elle est prédite, la ressemblance entre structures moléculaires peut être calculée à partir d'une représentation 1D, 2D ou 3D de celles-ci. La représentation 3D des molécules a un avantage certain. En effet, une similarité géométrique 3D entre deux molécules peut, par exemple, induire des effets biologiques similaires, même si les molécules concernées ont une représentation 2D très différente comme dans le cas de deux molécules de deux familles structurales différentes. Il existe de nombreuses méthodes de calcul de la similarité exploitant la représentation 1D ou 2D, mais très peu utilisant des informations sur la forme géométrique 3D [1, 5].

Le but de notre projet actuel est de développer une méthode originale d'alignement de structures moléculaires pour la recherche de similarité, en s'inspirant des techniques de recalage de points 3D issues de la vision par ordinateur. Le recalage est en effet utilisé depuis plusieurs décennies pour fusionner dans un même repère global des ensembles de points 3D décrivant partiellement la surface d'un même objet ou d'une même scène, mais représentés dans des repères locaux (car capturés selon plusieurs points de vue ou à différents instants).

Nous sommes convaincus que ce genre de méthodes peut être utilisé dans le cadre de la recherche de la similarité moléculaire.

2 Travail proposé

A cette fin, nous développons SENSEAAS (SENSitive Surface As A Shape) [2], un outil original qui combine des méthodes récentes dédiées au recalage 3D, initialement développées pour la fusion de nuages de points 3D issus de scanners, de type LiDAR par exemple. Considérons deux molécules que l'on souhaite superposer pour rechercher des similarités. L'idée générale est d'aligner des représentations surfaciques de ces deux molécules, pour ensuite calculer un score de ressemblance. La figure ci-dessous décrit le principe général de SENSEAAS.

- **Génération d'un nuage de points 3D coloré de la surface moléculaire** de chaque molécule (Figure 1b). La couleur est liée aux propriétés physico-chimiques de l'atome sous-jacent à la surface.
- **Alignement global des 2 surfaces** (Figure 1c) en exploitant uniquement la géométrie des points. Pour cela, nous utilisons les descripteurs locaux FPFH [8] pour caractériser des points d'intérêt sur les surfaces. Les descripteurs de chaque surface sont ensuite appariés avec RANSAC [3], de telle sorte que les surfaces une fois recalées minimisent une distance géométrique.
- **Raffinement de l'alignement** précédent en prenant en compte la géométrie des points 3D mais aussi la couleur associée, liée aux propriétés physico-chimiques (Figure 1d). Pour cela, la technique proposée dans [7] est utilisée. Celle-ci prend en compte la couleur des points lors du recalage de nuages pour améliorer la superposition finale. Ainsi nous obtenons une superposition des surfaces (Figure 1e), que l'on peut reporter sur les graphes 3D des molécules associées (Figure 1f).

3 Résultats

SENSEAAS reproduit 89% des superpositions de molécules telles qu'elles sont observées expérimentalement dans le jeu de données AstraZeneca fourni par le CCDC. Ce jeu de données contient 1465 molécules différentes réparties en 121 groupes de molécules alignées car « capturées » dans le site actif d'une même protéine.

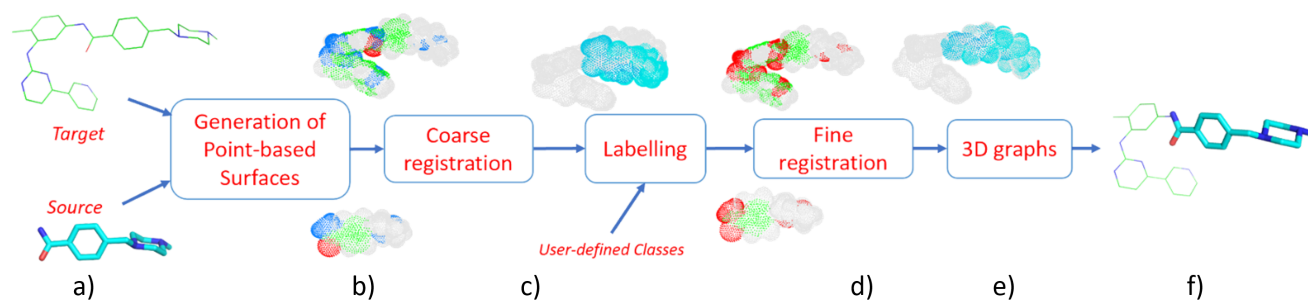


FIGURE 1 – principe général de notre algorithme SENSAAAS.

Une superposition est considérée comme reproduite si le RMSD, c'est-à-dire la différence moyenne entre les coordonnées prédites et expérimentales des atomes du graphe 3D, ne dépasse pas 2.0Å. Ces résultats sont comparables à ceux des deux autres méthodes d'alignement testées : 89% pour le programme ShaEP [9] et 87% pour le programme SHAFTS [6]. Cependant, nous avons montré que les alignements proposés par SENSAAAS sont plus précis puisqu'un plus grand nombre de molécules reproduites sont dans le premier intervalle de précision avec un $RMSD \leq 0.5\text{Å}$.

SENSAAAS est capable de recalculer parfaitement des molécules dont la structure moléculaire est identique ou très similaire comme une sous-structure (Figure 1) mais aussi lorsqu'elle est différente (2, *Bioisosteric matching*). L'alignement d'une sous-structure ou d'un fragment moléculaire sur une molécule de taille plus grande est ce que l'on appelle un sub-matching. C'est une propriété particulièrement intéressante que SENSAAAS permet, contrairement à d'autres méthodes de référence d'alignement par forme moléculaire (logiciels ROCS [4], SHAFTS [6], ou ShaEP [9], par exemple). En effet, une des problématiques à traiter en chimie médicinale est de pouvoir substituer une sous-structure par une autre sous-structure de famille chimique différente afin d'améliorer une propriété pharmacologique donnée et/ou d'acquies la propriété intellectuelle.

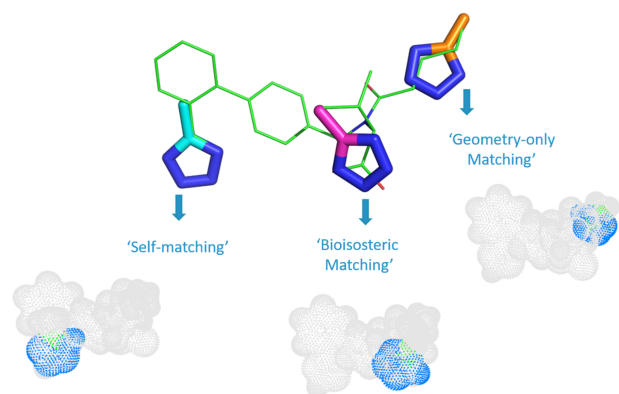


FIGURE 2 – Trois alignements possibles d'un fragment moléculaire (en cyan, magenta et orange) sur une structure moléculaire plus grande, le Valsartan représenté par sa structure en graphe 3D et coloré en vert.

La Figure 2 présente un exemple de bioisostérie bien connue en chimie médicinale qui consiste à remplacer

un groupe acide carboxylique par un tétrazole. Dans cet exemple, SENSAAAS est capable de proposer des alignements locaux parmi lesquels le meilleur est une superposition du fragment tétrazole sur lui-même puisqu'il est présent dans la molécule target Valsartan (*self-matching*); le second est une superposition du fragment tétrazole sur l'acide carboxylique tel qu'on l'attendait (*bioisosteric matching*); le troisième est une superposition du fragment tétrazole sur une chaîne aliphatique mais qui est moins intéressante car seule la géométrie est similaire (*Geometry-only Matching*, pas de matching de points colorés).

Nos résultats montrent ainsi que SENSAAAS apporte une contribution pertinente aux méthodes d'alignement basées sur la forme, en particulier dans le domaine de l'optimisation de molécules où la propriété d'alignement local s'avère importante pour identifier des molécules prometteuses.

4 Conclusion et perspectives

Nous avons présenté SENSAAAS, un nouvel outil qui permet d'aligner des structures moléculaires à partir d'une représentation 3D des surfaces associées. Nous montrons que ces alignements, obtenus à partir de techniques de recalage de nuages de points très populaires dans d'autres domaines sont comparables à ceux des techniques de l'état de l'art, voire meilleurs si l'on évalue leur capacité à réaliser des sub-matching. Les perspectives sont nombreuses, tant au niveau de la méthodologie que des applications potentielles en drug discovery.

Le logiciel SENSAAAS est le fruit d'une volonté de créer des outils chimioinformatiques de haut niveau accessibles à d'autres chercheurs et ré-utilisables dans le développement d'autres outils, à l'instar de la suite logicielle RDKit (<http://rdkit.org>). L'accès et l'utilisation libre des bases de données et des outils sont courants dans le domaine de la bioinformatique structurale mais bien moins fréquent dans le domaine de la chimioinformatique où une propriété intellectuelle peut être associée aux petites molécules. **Afin de promouvoir de nouvelles méthodes utilisant des représentations surfaciques des molécules, nous avons créé un dépôt du code sur GitHub <https://github.com/SENSAAAS/sensaas> accompagné d'une documenta-**

tion détaillée et de tutoriels sur YouTube. Une version en démo est aussi accessible sur le serveur de l'IPMC : <https://chemoinfo.ipmc.cnrs.fr/SENSAAS>.

Références

- [1] Daniel Baum and Hans-Christian Hege. A point-matching based algorithm for 3d surface alignment of drug-sized molecules. In Michael R. Berthold, Robert C. Glen, and Ingrid Fischer, editors, *Computational Life Sciences II*, pages 183–193, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.
- [2] Dominique Douguet and Frédéric Payan. sensaas : Shape-based alignment by registration of colored point-based surfaces. *Molecular Informatics*, 39(8) :2000081, 2020.
- [3] Martin A. Fischler and Robert C. Bolles. Random sample consensus : A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6) :381–395, June 1981.
- [4] J. A. GRANT, M. A. GALLARDO, and B. T. PICKUP. A fast method of molecular shape comparison : A simple application of a gaussian description of molecular shape. *Journal of Computational Chemistry*, 17(14) :1653–1666, 1996.
- [5] Ashutosh Kumar and Kam Y. J. Zhang. Advances in the development of shape similarity methods and their application in drug discovery. *Frontiers in Chemistry*, 6 :315, 2018.
- [6] Xiaofeng Liu, Hualiang Jiang, and Honglin Li. Shafts : A hybrid approach for 3d molecular similarity calculation. 1. method and assessment of virtual screening. *Journal of Chemical Information and Modeling*, 51(9) :2372–2385, 2011. PMID : 21819157.
- [7] J. Park, Q. Zhou, and V. Koltun. Colored point cloud registration revisited. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 143–152, 2017.
- [8] R. B. Rusu, N. Blodow, and M. Beetz. Fast point feature histograms (fpfh) for 3d registration. In *2009 IEEE International Conference on Robotics and Automation*, pages 3212–3217, 2009.
- [9] Mikko J. Vainio, J. Santeri Puranen, and Mark S. Johnson. Shaep : Molecular overlay based on shape and electrostatic potential. *Journal of Chemical Information and Modeling*, 49(2) :492–502, 2009. PMID : 19196024.