Programme de la journée MASIM 2019

5 Novembre 2019 @ IRIF, Univ. Paris 7

9h00 – 9h15	Frédéric Cazals (Inria Sophia) & Yann Ponty (CNRS/Ecole Polytechnique) *Introduction journée, présentation des actions MASIM*
9h15 – 9h40	**Hommage à Dave Ritchie (CAPSID, Inria/LORIA)** *Dave à Nancy* par Bernard Maigret (CAPSID) *Quelques contributions de Dave* par Sergei Grudinin (CNRS / Inria Grenoble)
9h40 – 10h00	Marie-Elisa Ruiz-Echartea (CAPSID) *EROS-DOCK : Exhaustive Rotational Search for pairwise and multi-body protein docking*
10h00 – 10h30	**Pause Café**
10h30 – 11h15	**Keynote by Michael Nilges (Institut Pasteur)** *Integrative structural biology, a Bayesian view*
11h15 – 11h35	Guillaume Postic (IFB) *Integration of MS-based proteomics and structural data for probing protein interaction networks*
11h35 – 11h55	Chloé QUIGNOT (CEA Saclay) *Taking evolutionary information to the atomic level in protein docking*
11h55 – 12h15	Didier Devaurs (Univ. Grenoble Alpes) *Studying Protein Structure through Hydrogen Exchange and Coarse-grained Conformational Sampling*
12h15 – 14h00	**Déjeuner + Session posters**
14h00 – 14h20	Frederic Cazals (Inria Sophia) *Multiscale analysis of structurally conserved motifs within flexible alignments*
14h20 – 14h40	Elodie Laine (Sorbonne Université) *Evolutionary decomposition and structural characterization of functionally distinct protein isoforms*
14h40 – 15h00	Benjamin Bouvier (CNRS / Univ. de Picardie Jules Verne) *Curvature as a collective coordinate in enhanced sampling membrane simulations*
15h00 – 15h20	Juan Cortés (LAAS-CNRS) *Protein loops with multiple meta-stable conformationsh a challenge for sampling and scoring methods*
15h20 – 16h00	**Pause Café**
16h00 – 16h45	**Keynote by Eric Westhof (IBMC, Univ. de Strasbourg)** *From RNA Architecture to Automatic RNA Modeling : RNA Puzzles*
16h45 – 17h05	Yann Ponty (CNRS/Ecole Polytechnique) *Integrative Probing Analysis of Nucleic Acids Empowered by Multiple Accessibility Profiles*
17h05 – 17h25	Audrey Legendre (IBISC, Univ Evry, Université Paris-Saclay) *Prédiction de structures secondaires de complexes d'ARN*
17h25 – 17h45	Sergei Grudinin (CNRS / Inria Grenoble) *Novel methods for integrative structural bioinformatics*
17h45 – 18h05	Vaitea Opuu (Ecole Polytechnique) *Computational design of proteins and enzymes*

Marie-Elisa Ruiz-Echartea

University of Lorraine, CNRS, Inria, LORIA, 54000 Nancy, France

EROS-DOCK : Exhaustive Rotational Search for pairwise and multi-body protein docking

Collaborateurs : Isaure Chauvot de Beauchêne, David W. Ritchie

Résumé : Protein-protein docking algorithms aim to predict the 3D structure of a complex using the structures of the individual proteins. For binary complexes, this involves searching and scoring in a six-dimensional space. Many docking algorithms use FFT techniques to exhaustively cover the search space and to accelerate the scoring calculation. However, the results often depend on the initial protein orientations with respect to the Fourier sampling grid. Furthermore, Fourier-transforming a physics-base force field can involve a serious loss of precision.

Here, we present a novel docking algorithm called EROS-DOCK (Exhaustive Rotational Search based Docking) [1] to rigidly dock two proteins using a series of exhaustive 3D rotational searches in which non-clashing orientations are scored using ATTRACT coarse-grained force field (ff) [2]. Initial positions are defined by putting each attractive pair of surface pseudo-atoms at their optimal distance in the ff. Thus, EROS-DOCK retains the exhaustive nature of FFT-based search algorithms while using a sensitive physics-based scoring function. Rather than calculating an $O(N \times M)$ interaction energy explicitly at every grid point, we use a quaternion π-ball$\check{\mathrm{T}}$ to represent the space of all possible 3D Euler angle rotations [3], and we recursively sub-divide the π-ball in order to cover the rotational space systematically, from each initial position. An associated tree-like data structure allows rotations that give steric clashes to be pruned efficiently using a "branch-and-bound" technique. To our knowledge, this is the first time that such a branch-and-bound pruning technique has been applied to the rigid-body protein docking problem.

The EROS-DOCK algorithm was tested on 173 target complexes from the Protein Docking Benchmark (v4) [4], and results were compared with those of ATTRACT and ZDOCK [5]. Overall, EROS-DOCK was able to find local minima that were not explored by the ATTRACT gradient-driven atom-based search. After refinement by a short coarse-grained minimization, the EROS-DOCK results were generally better than those of ATTRACT and ZDOCK, according to the standard CAPRI criteria.

EROS-DOCK can use contact restraints as an additional pruning criteria. Our results show that using even just one residue-residue restraint in each interaction interface is sufficient to increase the number of cases with acceptable solutions within the top 10 from 51 to 121 out of 173 pairwise docking cases. We used EROS-DOCK with restraints to dock trimeric complexes by combinatorial assembly of pairwise solutions. We expected that all interfaces in a multi-body docking solution should be similar to at least one interface in each lists of pairwise docking solutions. Thus, we used a new fast technique to calculate the RMSD between pairs of transformation matrices [6], and an adaptation of the branch-and-bound rotational search algorithm to accelerate the search for low RMSD docking solutions. By test on a home-made benchmark of 11 three-body cases, 7 obtained at least one acceptable quality solution in the top 50 solutions.

References

[1] EROS-DOCK : Protein-protein docking using exhaustive branch-and-bound rotational search. M.-E. Ruiz-Echartea, I. Chauvo de Beauchene, D. W. Ritchie, Bioinformatics (2019), pii : btz434

[2] Binding site prediction and improved scoring during flexible protein-protein docking with ATTRACT. S Fiorucci and M Zacharias. Proteins (2010), 78 :3131-3139.

[3] Global Optimization through Rotation Space Search. RI Hartley and F Kahl. Int. J. Comp. Vis. (2009), 82 :64-79.

[4] Protein-protein docking benchmark version 4.0. H Hwang et al. Proteins (2010), 78 :3111-3114.

[5] Accelerating protein docking in ZDOCK using an advanced 3D convolution library. BG Pierce, Y Hourai and Z Weng. PLoS One (2011), 6(9) :e24657.

[6] Rapid determination of RMSDs corresponding to macromolecular rigid body motions. P. Popov, S. Grudinin. J. Comp. Chem. (2014), 35(12) :950-956.

Michael Nilges (Keynote)

Institut Pasteur, Paris

Integrative structural biology, a Bayesian view

Résumé : Bientôt disponible

Guillaume Postic

Institut Français de Bioinformatique (IFB), UMS 3601-CNRS, Universite Paris-Saclay, Orsay, France

Integration of MS-based proteomics and structural data for probing protein interaction networks

Collaborateurs : J. Andreani, R. Guerois, J. Marcoux, E. Mouton-Barbosa, Y. Vandenbrouck, S. Cianferani, G. Labesse, O. Schiltz, P. Tufféry

Résumé : Introduction
Mass spectrometry (MS) has become essential for characterizing molecular species and their interactions. Most of the time, proteomic studies stop at listing the interacting proteins, without performing the analysis of the identified sequences. This is a wasted opportunity when considering the fact that structural and evolutionary aspects provide a powerful analysis framework for biologists : e.g. for interpreting patients mutations that interfere with assemblies, setting up directed mutagenesis and functional dissection experiments, or virtual screening.

Methods : The MS2MODELS proteomics pipeline integrates structural biology to MS data, in order to enhance the analysis of the protein-protein interaction networks. The homology-based detection of relevant structures from the Protein Data Bank (PDB) is carried out with HHsearch. Annotations of homomultimeric complexes, as well as interaction data from BioGRID and the eukaryotic linear motifs (ELM) resource are also integrated into the analysis.

Results : We have used MS2MODELS on several MS datasets containing up to hundreds of proteins. Thanks to the integration of structural information, the pipeline is able (i) to identify true positives in MS data by validating interactions within the input list of proteins, and (ii) to find additional partners that are either below the MS detection threshold (false negatives) or not detected at all. Moreover, MS2MODELS indicates the potential involvement of each input protein into a homomultimeric complex. The pipeline comes with an easy-to-use web interface. Thus, the protein-protein interaction networks can be conveniently visualized in a web browser. Although MS can detect protein complexes, it cannot identify the protein residues involved in the interactions. This is why MS2MODELS offers the possibility to visualize the 3D structure of each partner within the context of its complex. The structure of the latter is either experimental or predicted, depending on its availability in the PDB or the Swiss Model Repository, respectively.

Conclusions : The MS2MODELS project shows the interest of integrating protein structure data to the analysis of interactomes. In this way, MS2MODELS may benefit the community of biologists working on macromolecular interactions, with important applications such as the analysis of pathological dysfunctions related to altered molecular interactions or isoforms. Further development will focus on predicting the 3D structure of multiprotein complexes.

Chloé QUIGNOT

Laboratoire de Biologie Structurale et Radiobiologie - CEA Saclay

Taking evolutionary information to the atomic level in protein docking

Collaborateurs : Jessica ANDREANI, Raphaël GUEROIS

Résumé : Protein complexes are of fundamental importance in most biological processes and the structure of their interface can give us significant information to probe and understand the mechanisms behind these processes. As the experimental determination of 3D complex structures is not always possible, in silico predictions through molecular docking of these interfaces are very helpful to study how two proteins interact. Molecular docking is composed of two main steps, the generation of possible interface models (sampling step) followed by the scoring of these models in order to choose the most plausible ones. A key interest of our team is the use of co-evolutionary information to help predict the structure of a protein interface. It is the main driving force of the scoring function developed in the team, InterEvScore, which combines evolutionary information through homologous sequence alignments at residue level with a simple coarse-grained statistical potential [1]. InterEvScore has shown its value in many applications such as collaborations with experimentalists or the international CAPRI (Critical Assessment of Predicted Interactions) protein docking challenge [2] and is part of our docking server, InterEvDock2 [3]. Recent updates of InterEvDock2 now allow input proteins to be submitted as oligomeric structures or as sequences for which homology models are built automatically, and offer the possibility of filtering docking poses using constraints e.g. derived from biological data. We are currently working on a new evolutionary score that takes into account the atomic interface plasticity that occurs during evolution through explicit modelling of the structures of homologous proteins. Preliminary results on the large PPI4DOCK benchmark [4] are promising as we not only increase the general ranking of correct predictions but also, interestingly, improve the quality of the correct models found in the top-scored predictions.

[1] Andreani, Faure and Guerois. Bioinformatics, 2013. 29(14), 1742-1749.
http://dx.doi.org/10.1093/bioinformatics/btt260
[2] Yu, Andreani, Ochsenbein and Guerois. Proteins, 2017. 85(3) : 378-390.
http://doi.org/10.1002/prot.25180
[3] Quignot, Rey, Yu, Tuffery, Guerois and Andreani. Nucleic Acids Res, 2018. 46(W1) : W408-W416.
http://doi.org/10.1093/nar/gky377
http://bioserv.rpbs.univ-paris-diderot.fr/services/InterEvDock2/
[4] Yu and Guerois. Bioinformatics, 2016. 32(24) : 3760-3767.
http://doi.org/10.1093/bioinformatics/btw533

Didier Devaurs

Univ. Grenoble Alpes

Studying Protein Structure through Hydrogen Exchange and Coarse-grained Conformational Sampling

Collaborateurs : Dinler Antunes, Gregory Lizée, Lydia Kavraki

Résumé : Experimentally observing and/or computationally modeling large proteins and macromolecular complexes remain critical challenges for structural biology. Our work aims to address these challenges by combining experimental and computational techniques to overcome their respective shortcomings. At one end of the experimental spectrum, X-ray crystallography yields atomic-resolution structural models, but presents strong limitations in terms of cost and applicability. At the other end of the spectrum, hydrogen-exchange monitoring is cheap and easier to carry out, but cannot produce structural models because of its low resolution. To mitigate this issue, one side of our coupled approach consists of developing computational methods to complement such low-resolution experimental techniques. As these computational methods suffer from the curse of dimensionality when applied to large molecular systems, the other side of our coupled approach consists of guiding them with experimental data.
Here, we present three applications of our coupled approach combining hydrogen exchange on the experimental side and a robotics-inspired method for coarse-grained conformational sampling on the computational side. First, we argue that using coarse-grained conformational sampling of protein structure improves the fit between computationally-generated conformations and experimental hydrogen-exchange data. Second, we show that our approach allows analyzing the variability of a protein's native state described by crystallographic and hydrogen-exchange data. Finally, we explain how to obtain an atomic-resolution structural model of a protein state for which only hydrogen-exchange data is available.

Frederic Cazals

Inria

Multiscale analysis of structurally conserved motifs within flexible alignments

Collaborateurs : Romain Tetley

Résumé : This talk will present a generic framework to perform a multiscale structural analysis of two structures (homologous proteins, conformations) undergoing conformational changes. Practically, given a seed structural alignment, we identify structural motifs with a hierarchical structure, characterized by three unique properties. First, the hierarchical structure sheds light on the trade-off between size and flexibility. Second, motifs can be combined to perform an overall comparison of the input structures in terms of combined RMSD – an improvement over the classical least RMSD. Third, motifs can be used to seed iterative aligners, and to design hybrid sequence-structure profile HMM characterizing protein families.

From the methods standpoint, our framework is reminiscent from the bootstrap and combines concepts from rigidity analysis (distance difference matrices), graph theory, computational geometry (space filling diagrams), and topology (topological persistence).

On challenging cases (class II fusion proteins, flexible molecules) we illustrate the ability of our tools to localize conformational changes, shedding light of commonalities of structures which would otherwise appear as radically different. Our tools are available within the Structural Bioinformatics Library at https://sbl.inria.fr/doc/Structural_motifs-user-manual.html

We anticipate that they will be of interest to perform structural comparisons at large, and for remote homology detection.

Elodie Laine

Sorbonne Université

Evolutionary decomposition and structural characterization of functionally distinct protein isoforms

Collaborateurs : Diego Javier Zea, Antoine Labeeuw, Hugues Richard

Résumé : Alternative splicing (AS) is a regulatory process by which multiple protein isoforms are produced from a single gene. It has the potential to greatly expand the proteome and its deregulation is involved in several diseases, like cancer. Although the mechanisms of AS are well documented at the gene level, very little is known about the impact of AS on protein three-dimensional (3D) structures. Here, we present an automated tool for the structural annotation and 3D modelling of protein isoforms. Our approach is centred on orthologous exonic regions which we call s-exons. They represent the minimal evolutionary units describing the whole diversity of protein isoforms across a set of species. Decomposing the isoforms into s-exons provides a convenient way to describe and compare them, and to map alternative splicing events onto their predicted structures. We rely on comparative modelling and iteratively search templates until reaching the maximum coverage of the s-exons. We show how the presented method, combined with an evolutionary analysis, can be used to infer the functional outcomes of alternative splicing events on a set of about 50 genes with known functional isoforms. The tool is implemented as part of the PhyloSofS package, freely available to the community at https://github.com/PhyloSofS-Team/PhyloSofS.

Benjamin Bouvier

CNRS / Université de Picardie Jules Verne

Curvature as a collective coordinate in enhanced sampling membrane simulations

Résumé : The plasticity of membranes plays an important functional role in cells, cell components and micelles, where bending, budding and remodeling implement numerous recognition and communication processes. Comparatively, molecular simulation methods to induce, control and quantitatively characterize such deformations remain scarce. This work defines a novel collective coordinate associated with membrane bending, which strives to combine realism (by preserving the notion of local atomic curvatures) and low computational cost (allowing its evaluation at every time step of a molecular dynamics simulation). Enhanced sampling simulations along this conformational coordinate provide convenient access to the underlying bending free energy landscape.
To showcase the method's potential, I will briefly present its application to the remodeling of POPE bilayers and the study of the influence of the Pseudomonas quinolone signal on the budding of Gram-negative bacterial outer membranes.

Juan Cortés

LAAS-CNRS

Protein loops with multiple meta-stable conformations : a challenge for sampling and scoring methods

Collaborateurs : Amélie Barozet, Marc Bianciotto, Marc Vaisset, Thierry Siméon, Hervé Minoux

Résumé : Flexible regions in proteins, such as loops, cannot be represented by a single conformation. Instead, conformational ensembles are needed to provide a more global picture. In this context, identifying statistically meaningful conformations within an ensemble generated by loop sampling techniques remains an open problem. We have analyzed the performance of state-of-the-art methods to sample and score protein loop conformations on a set of 8 protein loops that are known to be flexible. The ability of each method to identify and select all of the known conformations is assessed, and the underlying energy landscapes are produced and projected to visualize the qualitative differences between the methods.
Our results show that statistical potentials provide considerable reliability despite their being designed to tradeoff accuracy for lower computational cost. On a large pool of loop models, they are capable of filtering out statistically improbable states while retaining those that resemble known (and thus likely) conformations. However, computationally expensive methods are still required for more precise assessment and structural refinement. The results also highlight the importance of employing several scaffolds for the protein, due to the high influence of small structural rearrangements in the rest of the protein over the modeled energy landscape for the loop.

Eric Westhof (Keynote)

Université de Strasbourg, Institut de biologie moléculaire et cellulaire du CNRS

From RNA Architecture to Automatic RNA Modeling : RNA Puzzles

Résumé : The key architectural elements of RNA structure will be described. In the past, RNA modeling achieved some success through the manual and computer-aided assembly of RNA fragments coupled with extensive sequence alignments and, sometimes, chemical probing data. The objectives nowadays aim at the automatic modeling of RNA sequences coupled or not with additional experimental data. In order to assess such models and approaches, independent evaluations tests have to be developed.

RNA-Puzzles is a CASP-like collective blind experiment for the evaluation of RNA three-dimensional structure prediction. The primary aims of RNA-Puzzles are to determine the capabilities and limitations of current methods of 3D RNA structure prediction based on sequence, to find whether and how progress has been made, and to illustrate whether there are specific bottlenecks that hold back the field. More than twenty RNA-Puzzles have been set up with automatic assessments of the agreements with X-ray structures. Several groups of modelers around the world participate in this collective effort. Difficulties and progress in RNA structure prediction will be reported.

Cruz, et al. RNA (2012), 18 :610625
Miao, et al. RNA (2015), 21 :1066-1084

Miao, et al. RNA (2017), 23 :118

Yann Ponty

CNRS/LIX, Ecole Polytechnique

Integrative Probing Analysis of Nucleic Acids Empowered by Multiple Accessibility Profiles

Collaborateurs : Afaf Saaidi, Bruno Sargueil and Delphine Allouche

Résumé : The manual production of reliable RNA structure models from chemical probing experiments, benefits from the integration of information derived from multiple protocols and reagents. However, the interpretation of multiple probing profiles remains a complex task, hindering the quality and reproducibility of modeling efforts.

In this talk we present an automated method called for the computational modeling of RNA structures from multiple probing reactivity profiles. Input profiles can result from experiments based on diverse protocols, reagents, or even a collection of mutants, and are jointly analyzed to predict the dominant conformations of an RNA.

Our method, called Integrative Probing Analysis of Nucleic Acids Empowered by Multiple Accessibility Profiles (IPANEMAP) combines statistical sampling, unsupervised learning, and multi-optimization, to produce secondary structure models that are both stable and well supported by experimental evidences. The analysis of several reactivity, both publicly available and specifically produced experimentally, demonstrates the good performances of IPANEMAP, even in a mono-probing setting. It confirms the potential of integrating multiple sources of probing data, leading to recommendations for the experimental design of informative probing assays.

Audrey Legendre

IBISC, Univ Evry, Université Paris-Saclay

Prédiction de structures secondaires de complexes d'ARN

Collaborateurs : Eric Angel, Fariza Tahi

Résumé : Les ARN peuvent interagir et former des complexes ayant des rôles variés dans la cellule, comme le ribosome ou le splicéosome. La prédiction de la structure secondaire de ces complexes est une premiere étape afin d'identifier leur fonction, leur structure 3D, etc.

Nous proposons ici une approche interactive tirant parti des nombreux outils disponibles pour prédire les structures secondaires d'ARN et d'interactions ARN-ARN pour prédire des structures de complexes composés de plus de deux ARN.
Nous avons formulé le problème de prédiction comme la détermination des meilleures combinaisons de structures secondaires d'ARN et d'interactions ARN-ARN.
Ce problème nous a permis de développer un premier outil mono-objectif, appelé RCPred (RNA Complex Prediction), trouvant les combinaisons de plus basses énergies.
Notre méthode est basée sur un problème d'optimisation combinatoire, et plus particulièrement sur un problème de graphe visant à trouver la clique pondérée maximum.
Ce problème est connu comme étant NP-difficile à résoudre (Bomze et al., 1999) et difficile à approximer (Hastad, 1997). Nous avons alors basé notre méthode sur une heuristique de recherche locale appelée Breakout Local Search (Benlic et Hao, 2013).

Nous avons ensuite développé une seconde version de cet outil, appelé C-RCPred (Constrained-RNA Complex Prediction), qui est multi-objectif.
Dans cette version, les meilleures combinaisons répondent à plusieurs critères, assimilés à des fonctions objectif : tout d'abord l'énergie libre, mais aussi l'accord avec des données structurales (telles que des données SHAPE) et le respect de contraintes utilisateurs.
Les meilleures combinaisons correspondent aux solutions de l'ensemble de Pareto, or, il n'existe pas de mé-

thode permettant de générer un ensemble de Pareto du problème de la clique maximum en temps polynomial. Nous proposons donc une heuristique permettant de trouver un ensemble de Pareto approché de ce problème.

L'outil RCPred est disponible en tant que webserver sur la plateforme EvryRNA (https://evryrna.ibisc.univ-evry.fr/) et l'outil C-RCPred le sera très prochainement.

Références : Una Benlic and Jin-Kao Hao. Breakout local search for maximum clique problems. Computers and Operations Research, 40(1) : 192Ű206, 2013.

Immanuel M Bomze, Marco Budinich, Panos M Pardalos, and Marcello Pelillo. The maximum clique problem. In Handbook of combinatorial optimization, 1Ű74. Springer, 1999.
Johan Håstad. Clique is hard to approximate within $n^{(1-\epsilon)}$ .1997.
Audrey Legendre, Eric Angel et Fariza Tahi. RCPred : RNA complex prediction as a constrained maximum weight clique problem. BMC bioinformatics, 20(3) : 128, 2019.

Sergei Grudinin

CNRS / Inria

Novel methods for integrative structural bioinformatics

Collaborateurs : Maria Kadukova, Karina Machado, Pablo Chacon

Résumé : Very recently, data-driven algorithms equipped with artificial intelligence (AI) have made a big leap forward in protein structure prediction. This was made possible thanks to the rapid progress in multiple experimental and computational disciplines, mostly algorithms for the analysis of sequence profiles, and AI methods. However, many other computational challenges are still open. These include predictions of multiple functional states of the same protein, prediction of protein flexibility, prediction of how proteins interact with each other, with small molecules, how they form assemblies and many more. All of these problems are very difficult to solve without using additional information. A tractable approach will be to complement ab-initio computations with (i) sparse experimental data (these can be pieces of information from multiple experiments) and (ii) information from genomic and structural databases. This constitutes the essence of modern integrative structural bioinformatics - a combination of physics-based and knowledge-based approaches with all available experimental information, which can be sparse and ambiguous.

I will briefly present computational pipelines for integrative structural bioinformatics developed in our team. I will also describe in more detail 2 on-going team developments. The first one is a novel knowledge-based potential for protein-ligand interactions. The novelty of our approach is the coarse-grained backbone-only representation of the protein. The second one is a novel method to construct a multi-level representation of protein flexibility. This method is based on a binary tree representation of rigid domains, which are computed by a comparison of the dynamics of rigid bodies with the dynamics of individual protein's atoms.

Vaitea Opuu

Ecole Polytechnique

Computational design of proteins and enzymes

Collaborateurs : Thomas Simonson

Résumé : Structure-based computational protein design (CPD) addresses the inverse folding problem, exploring a large space of amino acid sequences and selecting ones predicted to adopt a chosen fold. We recently showed that a PDZ domain can be entirely redesigned using CPD with a "physics-based" energy function that combines molecular mechanics with a continuum electrostatic solvent model. Many thousands of sequences were generated by Monte Carlo simulation, using our Proteus software. Among the lowest-energy sequences, three were tested experimentally and all shown to fold into the correct, PDZ structure, despite having 50 of 83

amino acids mutated. This represents a striking validation of our "physics-based" CPD approach.

Next, we applied CPD to enzyme design. A designed enzyme should satisfy multiple criteria : stability, substrate binding, transition state binding. Such multiobjective design is computationally challenging. We recently proposed a new method based on adaptive importance sampling. By first flattening the energy landscape of the apo protein, we obtained positive design for the bound state and negative design for the unbound. We have now extended the method to design an enzyme for its specific transition state binding, ie, its catalytic power. We consider methionyl-tRNA synthetase (MetRS), which attaches methionine (Met) to its cognate tRNA,establishing codon identity. We redesigned MetRS computationally to bind several ligands : the Met analog azidonorleucine, the natural ligand methionyl-adenylate (MetAMP), and the activated ligands that form the transition state for MetAMP production. Enzyme mutants known to have azidonorleucine activity were recovered, and 17 mutants predicted to bind MetAMP were characterized experimentally and all found to be active. Mutants predicted to have low activation free energies for MetAMP production were found to be active and the predicted reaction rates agreed well with the experimental values. We expect that in the future, the present method will become the paradigm for computational enzyme design.

References

Villa, Panel, Chen, Simonson (2018) Journal of Chemical Physics ; Adaptive landscape flattening in amino acid sequence space for the computational design of protein :peptide binding (invited article ; cover article).
Opuu, di Nigro, Gaillard, Schmitt, Mechulam, Simonson (2019) submitted ; Adaptive landscape flattening allows the design of both enzyme :substrate binding and catalytic power.
Opuu, Sun, Hou, Panel, Ichikawa, Corbi-Verge, Kim, Noyes, Fuentes, Simonson (2019) submitted ; A physics-based energy function allows the computational redesign of a PDZ domain.