

Edoardo Sarti

Curriculum Vitae

Identité

Date de naissance 30/04/1987
Lieu de naissance Milano, Italie
Adresse 3, Rue Blacas
06000 Nice

Contact

Adresse professionnelle Inria Sophia Antipolis - Méditerranée, ABS
2004, Route des Lucioles,
06902 Valbonne
Téléphone 07 78 32 80 48
Adresse électronique edoardo.sarti@inria.fr

Mots clés

Apprentissage automatique • Modèles statistiques • Réduction dimensionnelle • Interaction protéine-protéine • Protéines virales • Prédiction évolutive • Analyse et prédiction structurale

Formation et expérience académique

- de oct. 2021 au présent **Chargé de recherche de classe normale,**
ABS, Inria Sophia Antipolis - Méditerranée, Valbonne, France.
Sujet principal : Etudes structurales et évolutives sur les protéines
Recherche : Création d'algorithmes, analyse de données, création de bases de données et approches intégrés pour mieux comprendre les aspects divers de la structure, dynamique et évolution des protéines, en se basant sur approches de mécanique statistique, géométrie et topologie.
- de juin 2020 à juillet 2021 **Chercheur post-doctorant,**
LCQB, Sorbonne Université, Paris, France.
Sujet principal : Interaction entre les protéines de SARS-CoV-2 et leur évolution
Recherche : Analyse de données et création d'algorithmes et approches intégrés pour comprendre les interactions entre les quatre protéines structurales de SARS-CoV-2, en utilisant des modèles géométriques, coévolutifs et mutationnels.
Enseignement : TDs et TPs d'Informatique (18,5 heures eq. TD) et Bioinformatique (40 heures eq. TD) pour les étudiants en Licence d'Informatique et Master de Bioinformatique (UFR 919).
Superviseur : Alessandra Carbone

- de sept. 2019 **ATER - Attaché temporaire d'enseignement et de recherche,**
à août 2020 *Section CNU 65, UFR 927, Profil Bio - Maths/Stats,*
LCQB, Sorbonne Université, Paris, **France.**
Enseignement : CMs, TDs et TPs d'Informatique (85 heures eq. TD), Mathématiques (20 heures eq. TD) et Statistique (52 heures eq. TD) pour les étudiants en Licence de Biologie (UFR 927).
Recherche : Création d'algorithmes et bases de données pour faciliter l'analyse structurale des protéines et pour la prédiction des complexes de protéines à partir de la structure des monomères et de ses informations évolutives contenues dans les alignements de séquences. Les projets sont poursuivis au sein du laboratoire LCQB (groupe de Statistical Genomics and Biological Physics, dirigé par Martin Weigt).
- de juin 2018 **Chercheur post-doctorant,**
à août 2019 *LCQB, Sorbonne Université, Paris, France.*
Sujet principal : Modèles statistiques pour l'interaction protéine-protéine
Recherche : Développement du cadre statistique de la *Direct Coupling Analysis* (DCA) et d'autres procédés adaptatifs pour prédire l'interaction par paire de protéines sans en connaître a priori la structure; création d'une nouvelle base de données de référence pour l'apprentissage et l'évaluation des algorithmes susmentionnés; création d'un environnement Python pour la validation des prédictions des structure des protéines et des complexes protéiques et pour le développement de nouvelles méthodes de prédiction de contact intra- et inter-protéine.
Enseignement : Encadrement de quatre étudiants en L3 dans un projet de modélisation pluridisciplinaire au Polytech Sorbonne; TD du cours de Statistique en Biologie et Algorithmes sur les Séquences (SBAS) pour les étudiants en M1 du Master en Bioinformatique à Sorbonne Université.
Superviseur : Martin Weigt
- de nov. 2015 **Chercheur post-doctorant,**
à mai 2018 *CSB, NINDS, NIH - National Institutes of Health, Bethesda, MD, USA.*
Sujet principal : Alignement de séquences et de structures des protéines transmembranaires
Recherche : Optimisation et développement de techniques d'apprentissage automatiques pour AlignMe, un algorithme d'alignement de séquences de protéines transmembranaires; création et entretien de EncoMPASS, une base de données pour l'analyse des ressemblances structurales et symétriques des protéines transmembranaires.
Superviseur : Lucy R. Forrest

- de oct. 2011 **Doctorant,**
à oct. 2015 *SBP, SISSA - International School for Advanced Studies, Trieste, Italie.*
Sujet principal : Optimisation de modèles physiques et statistiques pour le repliement et l'interaction des protéines
Recherche : Optimisation et développement de BACH, une fonction d'évaluation de qualité structurale des repliements et interactions des protéines; étude et optimisation de champs de forces tout-atomes et à gros grains pour la dynamique moléculaire des protéines; évaluation de méthodes d'échantillonnage amélioré pour la reconstruction de paysages énergétiques de systèmes complexes.
Intitulé de la thèse : Evaluation de la structure des protéines et des complexes protéiques par des approches physiques et statistiques
Directeur : Alessandro Laio
Rapporteurs : Guido Tiana, Bruno Correia
Membres du jury : Giovanni Bussi (Président du jury), Alessandra Magistrato, Angelo Rosa, Antonio Celani, Giorgio Colombo, Francesco Gervasio, Cristian Micheletti (Secrétaire du jury)
- nov. 2014 **Sejour doctorat,**
LBT, IBPC - Institut de Biologie Physico-Chimique, Paris, France.
Sujet principal : Collaboration avec le groupe de Philippe Derreumaux pour le devis d'un nouveau terme pour champs de forces à gros grains modélisant le interaction de structure secondaire dans le squelette des protéines.
Superviseur d'accueil : Philippe Derreumaux
- de mars 2014 **Sejour doctorat,**
à mai 2014 *Clementi Lab, Rice University, Houston, TX, USA.*
Sujet principal : Collaboration avec le groupe de Cecilia Clementi pour la mise au point d'une coordonnée de réaction pour les changes conformationnels dans les protéines
Superviseur d'accueil : Cecilia Clementi
- 2013 **Tutor,**
Università Cattolica del Sacro Cuore, Milano, Italy.
Main topic : Teaching a Physics-101 course for students of Physical Education and Sport Sciences. The course was complementary to a Human Anatomy course and was worth 3 credits.
Main lecturer : Alberto Granato
- de oct. 2009 **Master en Physique,**
à nov. 2011 *Università degli Studi Milano-Bicocca, Milano, Italie.*
grade : 110/110 cum laude
Spécialisation en Physique de la Matière Condensée et Physique Statistique Théorique.
Le projet final a été en collaboration avec l'Université di Padova et SISSA (Trieste), ayant comme sujet la généralisation d'un potentiel statistique qui ensuite est devenu le sujet de mon travail de thèse en SISSA.
Sujet : Optimisation d'un potentiel empirique pour l'interaction protéine-protéine
Superviseurs externes : Alessandro Laio, Flavio Seno, Antonio Trovato
Superviseur interne : Giuseppe Chirico
- de oct. 2006 **Licence en Physique,**
à juil. 2009 *Università degli Studi Milano-Bicocca, Milano, Italie.*
grade : 108/110
Sujet : Mesures de photoluminescence et de microscope à force atomique d'atomes artificiels obtenus par épitaxie de faisceau moléculaire (MBE).
Superviseur : Stefano Sanguinetti

Enseignement

Sorbonne Université - Vacataire

- 2020-2021 **Science des Données et Intelligence Artificielle**, 18,5h TD,
L3 Sorbonne Université, Paris, France.
18,5 HETD **Responsable de cours** : Christophe Marsala
- 2020-2021 **Algorithmes sur les Arbres et les Graphes en Bioinformatique**, 20h TD,
M1 Sorbonne Université, Paris, France.
20 HETD **Responsable de cours** : Alessandra Carbone
- 2020-2021 **Statistiques en Bioinformatique et Algorithmes sur le Séquences**, 24h TD,
M1 Sorbonne Université, Paris, France.
24 HETD **Responsables de cours** : Martin Weigt, Juliana Bernardes

Sorbonne Université - ATER

- 2019-2020 **Éléments de programmation I**, 38,5h TD,
L1 Sorbonne Université, Paris, France.
38,5 HETD **Responsable de cours** : Romain Demangeon
- 2019-2020 **Mathématiques et Statistiques pour la Biologie**, 40h TD,
L2 Sorbonne Université, Paris, France.
40 HETD **Responsables de cours** : Stéphane Genet, Céline Ellien
- 2019-2020 **Bioinformatique : les outils indispensables**, 2h CM, 17h TP,
L2 Sorbonne Université, Paris, France.
20 HETD **Responsable de cours** : Stéphane Le Crom
- 2019-2020 **Statistiques et mathématiques pour la Biologie**, 36h TD,
L3 Sorbonne Université, Paris, France.
36 HETD **Responsables de cours** : Dominique Lamy, Lorette Noiret
- 2019-2020 **Modélisation, algorithmique et programmation pour la Biologie**, 8h CM, 16h TP,
L3 Sorbonne Université, Paris, France.
28 HETD **Responsable de cours** : Mathilde Carpentier

Sorbonne Université - Vacataire

- 2018-2019 **Modélisation pluridisciplinaire**, 15h TD,
L3 Sorbonne Université, Paris, France.
15 HETD **Responsable de cours** : Frédérique Charles
- 2018-2019 **Statistiques en Bioinformatique et Algorithmes sur le Séquences**, 24h TD,
M1 Sorbonne Université, Paris, France.
24 HETD **Responsables de cours** : Martin Weigt, Juliana Bernardes

Italie - Vacataire

- 2012-2015 **Summer School on Atomistic Simulation Techniques**, 3x15h TP,
M1 SISSA, Trieste, Italie.
Responsables de cours : Giovanni Bussi, Cristian Micheletti, Alessandro Laio
- 2012-2013 **Introduction à la Physique**, 28h CM,
L1 Università Cattolica del Sacro Cuore, Milano, Italie.
Responsables de cours : Alberto Granato

Publications (RICL) (8)

- J*₁ (2020) M. Muscat, G. Croce, **E. Sarti**, M. Weigt, *FilterDCA : Interpretable supervised contact prediction using inter-domain coevolution* PLoS computational biology 2020 10.1371/journal.pcbi.1007621
- J*₂ (2020) A. A. Aleksandrova, **E. Sarti**, L. R. Forrest *MemSTATS : a benchmark set of membrane protein symmetries and pseudo-symmetries* Journal of Molecular Biology 10.1016/j.jmb.2019.09.020
- J*₃ (2019) **E. Sarti***, A. A. Aleksandrova*, S. K. Ganta, A. S. Yavatkar, L. R. Forrest *EncoMPASS : an online database for analyzing structure and symmetry in membrane proteins* Nucleic acids research, 47(D1), D315-D321 10.1093/nar/gky952
- J*₄ (2018) A. Battisti, S. Zamuner, **E. Sarti**, A. Laio *Toward a unified scoring function for native state discrimination and drug-binding pocket recognition* Physical Chemistry Chemical Physics, 20(25), 17148-17155 10.1039/C7CP08170G
- J*₅ (2016) **E. Sarti**, I. Gladich, S. Zamuner, B. E. Correia, A. Laio *Protein-protein structure prediction by scoring molecular dynamics trajectories of putative poses* Proteins Structure Function and Bioinformatics, 84(9), 1312-1320 10.1002/prot.25079
- J*₆ (2015) **E. Sarti**, D. Granata, F. Seno, A. Trovato, A. Laio *Native fold and docking pose discrimination by the same residue-based scoring function* Proteins Structure Function and Bioinformatics, 83(4), 621-630 10.1002/prot.24764
- J*₇ (2013) **E. Sarti**, S. Zamuner, P. Cossio, A. Laio, F. Seno, A. Trovato *BACHSCORE. A tool for evaluating efficiently and reliably the quality of large sets of protein structures* Computer Physics Communications, 184(12), 2860-2865 10.1016/j.cpc.2013.07.019
- J*₈ (2011) S. Bietti, C. Somaschini, **E. Sarti**, N. Koguchi, S. Sanguinetti, G. Isella, D. Chrastina, A. Fedorov *Photoluminescence study of low thermal budget III-V nanostructures on silicon by droplet epitaxy* Nanoscale research letters, 5(10), 1650 10.1007/s11671-010-9689-8
- * auteurs à contributions équivalentes

Séminaires (CICL) (3)

- C*₁ (2018) *Analyzing the Structure and Symmetry of Membrane Proteins through the Systematic Online Database EncoMPASS*, **Biophysical Society 2018**, San Francisco (CA)
- C*₂ (2017) *EncoMPASS : an Encyclopedia of Membrane Proteins Analyzed by Structure and Symmetry*, **ISMB/ECCB 2017**, Prague (CZ)
- C*₃ (2017) *EncoMPASS : an Encyclopedia of Membrane Proteins Analyzed by Structure and Symmetry*, **GRC Membrane protein folding 2017**, Easton (MA)

Posters (CICL) (6)

- P*₁ (2019) *A supervised but interpretable coevolutionary predictor of protein-protein interactions*, **ISMB/ECCB 2019**, Basel (CH)
- P*₂ (2019) *A supervised but interpretable coevolutionary predictor of protein-protein interactions*, **Statistical Physics Approaches to Systems Biology 2019**, La Habana (CU)
- P*₃ (2018) *Transmembrane topology alignment for identifying related protein structures*, **ISMB 2018**, Chicago (IL) USA
- P*₄ (2018) *A supervised but interpretable coevolutionary predictor of protein-protein interactions*, **Approches Interdisciplinaires de l'Évolution Moléculaire**, Villeneuve d'Ascq (FR)
- P*₅ (2018) *Analyzing the structure and symmetry of membrane proteins through the systematic online database EncoMPASS*, **Biophysical Society 2018**, San Francisco (CA) USA
- P*₆ (2017) *EncoMPASS : an Encyclopedia of Membrane Proteins Analyzed by Structure and Symmetry*, **ISMB/ECCB 2017**, Prague (CZ)

Séminaires (CNCL) (1)

*C*₄ (2020) *Statistical analysis of coevolutionary signals on the SARS-CoV-2 Spike protein*, **Journée ASIM SARS-CoV-2 et Bioinformatique Structurale 2020**, en ligne

Publications en révision (2)

- J*₉ R. Staritzbichler*, **E. Sarti***, E. Yaklich, M. Stamm, A. A. Aleksandrova, K. Khafizov, L. R. Forrest *Refining pairwise sequence alignments of membrane proteins by the incorporation of anchors* bioRxiv 10.1101/2020.09.16.299453
- J*₁₀ A. A. Aleksandrova*, **E. Sarti***, L. R. Forrest *EncoMPASS : an Encyclopedia of Membrane Proteins Analyzed by Structure and Symmetry* bioRxiv 10.1101/391961
- * auteurs à contributions équivalentes

Fiches logiciel

- L*₁ **Locusts** : Family=utility; Audience=community; Evolution=lts; Duration=<1; Contribution=leader; Url=https://github.com/EdoardoSarti/locusts
Locusts est un paquet en Python3 / bash permettant de distribuer de très grandes quantités de travaux courts sur les systèmes HPC. Chaque travail s'exécute dans un environnement protégé et les données sont transférées en toute sécurité d'une machine locale vers le HPC et vice versa. Locusts peut également être appliqué à n'importe quelle machine locale ou cluster local sans qu'il soit nécessaire de s'appuyer sur un système de planification. Le paquet est né dans le cadre du projet EncoMPASS et a été largement testé sur différents HPC. Je suis le créateur et le seul développeur et testeur de cet algorithme.
- L*₂ **Pfam Interactions** : Family=vehicule; Audience=partners; Evolution=lts; Duration=1; Contribution=leader; Url=https://github.com/infernet-h2020/pfam_interactions
Pfam Interactions est une plateforme en Python3 pour l'entraînement et le test d'algorithmes qui utilisent des alignements de séquences multiples (MSA) pour la prédiction de la structure des protéines et de ses complexes. Il exploite les MSA contenus dans la base de données Pfam, qui classifie plus de 18000 domaines protéiques par des modèles de Markov cachés, et de la PDB (Protein Data Bank), qui fournit les fichiers de coordonnées 3D des protéines. Pfam Interactions effectue un mapping très précis entre les ressources de séquence et de structure, et est capable d'évaluer la précision de tout algorithme de prédiction de structure des protéines, en calculant différentes métriques (positive predictive value ou PPV, area under the curve ou AUC) et en combinant des analyses et des outils visuels et interactifs. J'ai développé cet ensemble d'outils dans le cadre de FilterDCA, et il a déjà été utilisé dans 3 autres travaux, en faisant désormais partie de la bibliothèque d'outils d'analyse du groupe de Martin Weigt.
- L*₃ **EncoMPASS** : Family=research; Audience=community; Evolution=lts; Duration=4; Contribution=leader; Url=https://github.com/Lucy-Forrest-Lab
EncoMPASS est une base de données en ligne (200 utilisateurs par mois) pour analyser et relier les structures et les symétries des protéines membranaires. Pour ce faire, ce logiciel en Python3 combine les informations de différentes bases de données, standardise leur description, les insère dans un modèle bicouche lipidique, les classe en termes de structure et de topologie, et trouve un ensemble exhaustif de symétries internes. EncoMPASS utilise plusieurs ressources externes, qui sont fournies dans un conteneur Singularity associé. De manière complètement automatisée, il renvoie un système de fichiers où les données, les chiffres et le contenu interactif sont stockés et prêts à être téléchargés sur le serveur web. EncoMPASS comprend également une procédure de mise à jour pour l'intégration de nouvelles données, qui est exécutée tous les 2 mois. J'ai développé, avec Antoniya Aleksandrova, le cadre général, les bibliothèques, les algorithmes originaux et le contenu web d'EncoMPASS. À ce jour, la version actuelle basée sur le gestionnaire de tâches HPC Locusts est dans sa phase de test tardive et, conformément au protocole de sécurité fédéral américain, elle ne peut pas encore être rendue publique. Il apparaîtra bientôt sur l'adresse GitHub fournie. N'hésitez pas à me contacter en cas de besoin.