Patrick Aloy

Structural bioinformatics and network biology group

Proteins are the main perpetrators of most cellular tasks. However, they seldom act alone and most biological processes are carried out by macromolecular assemblies and regulated through a complex network of protein-protein interactions. Thus, modern molecular and cell biology no longer focus on single macromolecules but now look into complexes, pathways or even entire organisms. The many genome-sequencing initiatives have provided a near complete list of the components present in an organism, and post-genomic projects have aimed to catalogue the relationships between them. The emerging field of systems biology is now centred mainly on unraveling these relationships. However, all these interaction maps lack molecular details: they tell us who interacts with whom, but not how. A full understanding of how molecules interact can be attained only from high resolution three-dimensional (3D) structures, since these provide crucial atomic details about binding. These details allow a more rational design of experiments to disrupt an interaction and therefore to perturb any system in which the interaction is involved. Our main scientific interests are in the field of structural bioinformatics and network biology, in particular, the use of protein sequences and high-resolution 3D structures to reveal the molecular bases of how macromolecular complexes and cell networks operate.

Novel peptide-mediated interactions derived from high-resolution 3D structures

Many biological responses to intra- and extra-cellular stimuli are regulated through complex networks of transient protein interactions where a globular domain in one protein recognises a linear peptide from another, creating a relatively small contact interface. These peptide stretches are often found in unstructured regions of proteins and they contain a consensus motif complementary to the interaction surface displayed by their binding partners. While most current methods for the *de novo* discovery of such motifs exploit their tendency to occur in disordered regions, our work focuses on another observa-

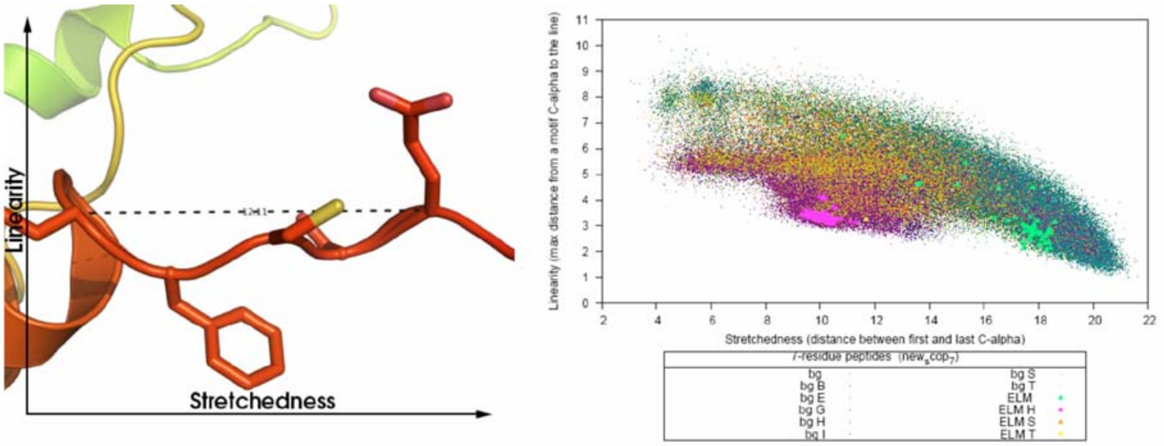
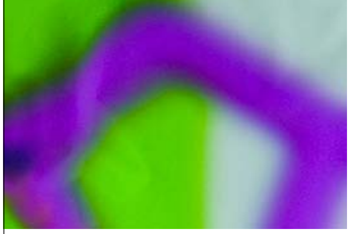


Figure 1. ***Linearity and stretchedness of linear motifs****. Linearity is defined as the maximum deviation of any C_α in the motif from the line through the first and last C_α. Comparison of the linearity of known peptides [11] with that of random peptides shows that the linear motifs tend to be more linear, but there is no clear distinction between the two distributions. Stretchedness values for known Eukaryotic Linear Motifs (ELM) peptides tend to be higher than those for random peptides, but again there is no clear distinction. Linearity and stretchedness divide into groups on the basis of the most frequently assigned secondary structure class (DSSP [42]) shown for 7 residue peptides. Combining linearity, stretchedness and secondary structure class with data from the SCOP background (bg), shown as dots, we can observe that known linear motifs fall into distinct regions of the parameter space.*



tion: upon binding to their partner domain, motifs adopt a well-defined structure. Indeed, through the analysis of all peptide-mediated interactions of known high-resolution 3D structure, we found that the structure of the peptide may be as characteristic as the consensus motif and may help identify target peptides even though they do not match the established patterns. Our analyses of the structural features of known motifs reveal that they tend to have a particular stretched and elongated structure, unlike most other peptides of the same length. Accordingly, we have implemented a strategy based on a support vector machine that uses these features, along with other structure-encoded information about interaction interfaces, to propose novel peptide-mediated interactions. Whenever enough information has been available, we have also derived consensus patterns for these interactions -and compared our results with established linear motif sequences and their binding domains. Finally, we have cross-validated our newly derived patterns on interactome network data from several model organisms, and presented a list of 64 peptide-mediated interactions, 47 of which have not been described before, involving 46 distinct domains, along with their respective high-resolution 3D structures and consensus motifs.

Pushing structural information into the yeast interactome by high-throughput protein docking experiments

Recent years have seen the consolidation of high-throughput proteomics initiatives to identify and characterise protein interactions and macromolecular complexes in model organisms. In particular, more that 10,000 high-confidence protein-protein interactions have been described in the roughly 6,000 proteins encoded in the budding yeast genome (Saccharomyces cerevisiae). However, unfortunately, high-resolution 3D structures are available for fewer than one hundred of these interacting pairs. In this project, we expand this structural information on yeast protein interactions by running the first-ever high-throughput docking experiment with some of the best state-of-the-art methodologies. To increase the coverage of the interaction space, we also explore the possibility of using homology models of varying quality in the docking experiments, instead of experimental structures, and assess how they affect the

Research Group Members

Group Leader:
Patrick Aloy

Research Associate:
Roberto Mosca

Postdoctoral Fellows:
Arnaud Ceol, Albert Pujol, Guillermo Suñé, Andreas Zanzoni

PhD Students:
Manuel Alonso, Rodrigo Arroyo, Clara Berenguer, Marc Duocastella, Roland Pache, Amelie Stein

Research Assistant:
Ricart Lluís

Visiting Students:
Rafael Pedret (Spain), Joan Marc Seoane (Spain), Francesc Tresserres (Spain), Josep Lluís Villanueva (Spain)

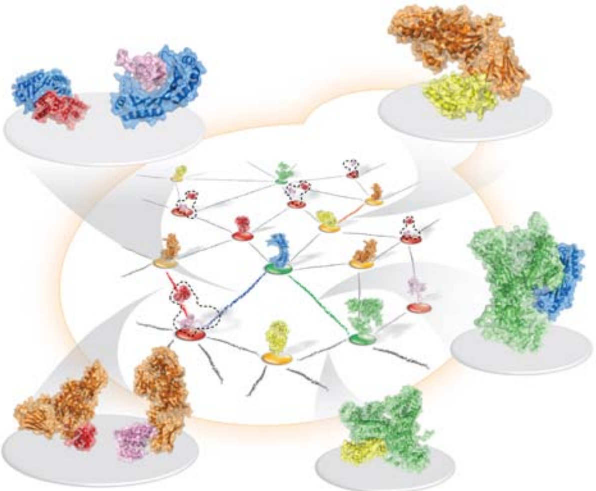


Figure 2. Artistic representation of the structured yeast interactome.

global performance of the methods. In total, we have applied the docking procedure to 217 experimental structures and 1,023 homology models, providing putative structural models for over 3,000 protein-protein interactions in the yeast interactome. Finally, we analyse in detail the structural models obtained for the interaction between SAM1-anthranilate synthase complex and the MET30-RNA polymerase III, to illustrate how our predictions can be used straightforwardly by the scientific community. The results of our experiment will be integrated into the general 3D-Repertoire pipeline, a European initiative to solve the structures of protein complexes in yeast at the best possible resolution. All docking results are available at http://gatealoy.pcb.ub.es/HT_docking/.

Unveiling the role of network and systems biology in drug discovery

Network and systems biology offer a novel way to approach drug discovery by developing models that consider the global physiological environment of protein targets and the effects derived from tinkering with them, without losing the key molecular de-

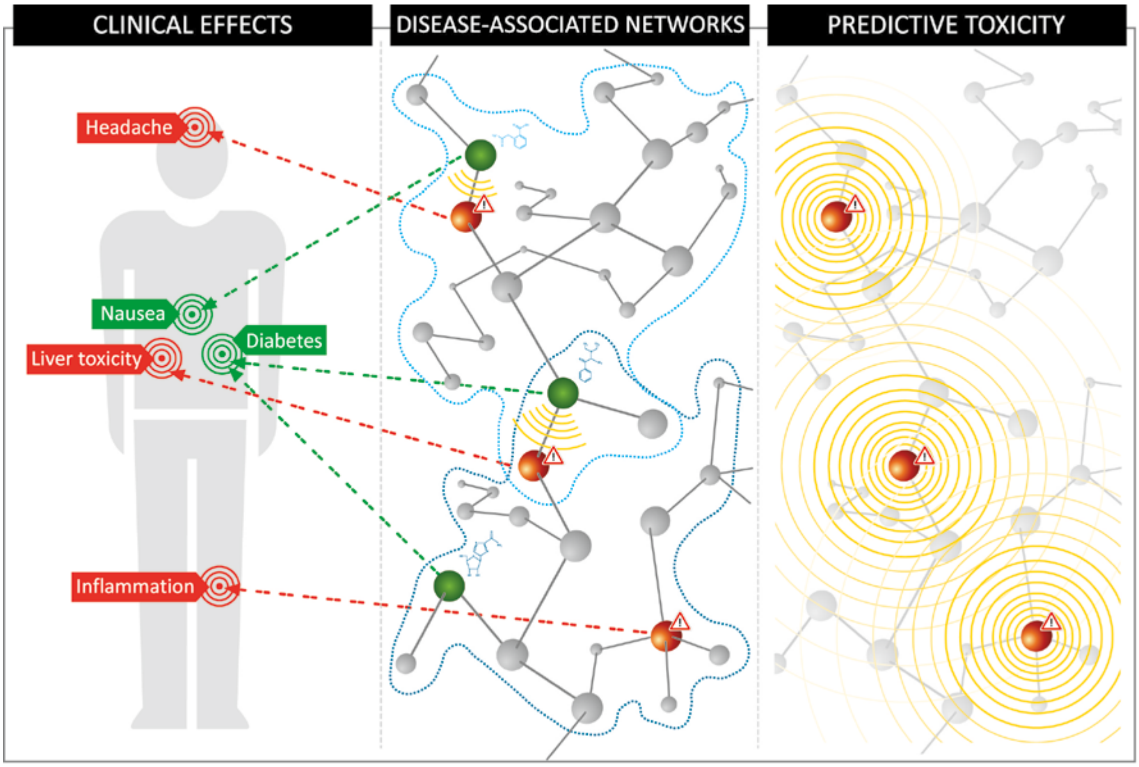


Figure 3. ***Network biology applied to predictive toxicology and drug repurposing****. The disease-associated networks for diabetes (dark blue dashed lines) and nausea (light blue dashed lines) contain several proteins that have been reported to be possible causes of some frequent adverse effects when their normal functioning is affected (red nodes). In addition, the networks contain drug targets annotated to their specific diseases (green nodes). Intense research is carried out to develop models with the capacity to identify the areas of influence of proteins leading to undesired effects and to explore how they are related to network connectivity. If successful, these models could help to discard potential drug targets that are likely to trigger severe adverse reactions at early stages of the discovery process, and to rationally design the toxicity tests required to check the safety of other under the area of influence of a certain red node. In addition, a detailed description of the molecular networks associated with certain diseases can unveil the existence of validated drug targets for a given therapeutic indication in key enclaves of the network that describe a distinct disease, thereby suggesting candidates for drug repurposing (*ie*, finding new indications for a target).*

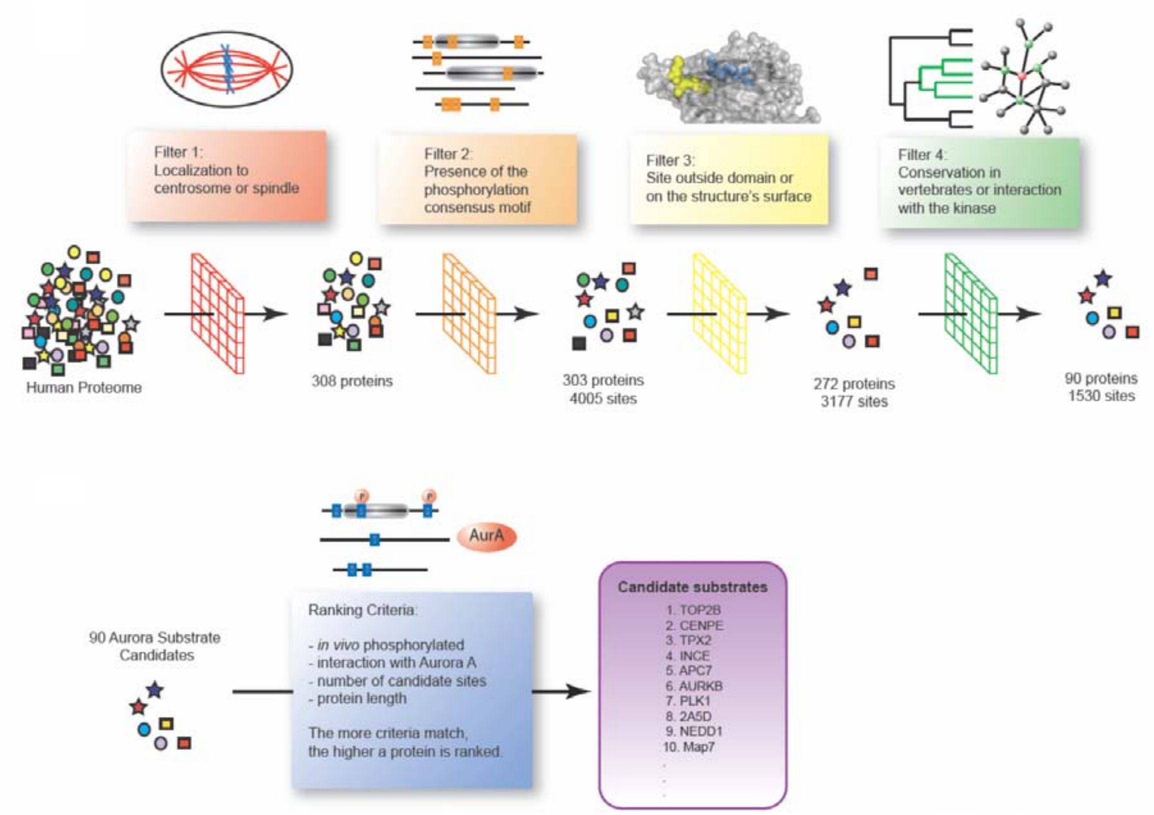

Filter 1: Localization to centrosome or spindle
Filter 2: Presence of the phosphorylation consensus motif
Filter 3: Site outside domain or on the structure's surface
Filter 4: Conservation in vertebrates or interaction with the kinase
Human Proteome
308 proteins
303 proteins 4005 sites
272 proteins 3177 sites
90 proteins 1530 sites
P
P
AurA
90 Aurora Substrate Candidates
Ranking Criteria:
- in vivo phosphorylated
- interaction with Aurora A
- number of candidate sites
- protein length
The more criteria match, the higher a protein is ranked.
Candidate substrates
1. TOP2B
2. CENPE
3. TPX2
4. INCE
5. APC7
6. AURKB
7. PLK1
8. 2A5D
9. NEDD1
10. Map7

Figure 4. ***Schematic representation of the bioinformatics approach developed to uncover new Aurora kinase substrates****. Candidate substrate selection. Aurora substrate candidates were selected based on a series of filters applied to the whole human proteome: presence of an Aurora phosphorylation motif in the sequence, localisation to the centrosome or the spindle, accessibility of the consensus motif and conservation of the potential phosphorylation site among vertebrates. The 90 proposed Aurora substrates were ranked following several criteria.*

tails. In this paper, we reviewed some recent advances in the fields of network and systems biology applied to human health, and discussed their impact on some of the hottest areas of drug discovery. In particular, we claim that network biology will play a central role in the development of novel polypharmacological strategies to fight complex multi-factorial diseases, where efficacious therapies will need to centre on bringing down entire pathways rather than single proteins. In this area of research, we focus mainly on developing novel strategies in the two fields in which we consider network and system biology strategies are most likely to make an immediate contribution: predictive toxicology and drug repurposing.

Uncovering novel substrates for Aurora A kinase

Aurora A is a serine/threonine kinase that is essential for cell cycle progression, centrosome maturation and spindle assembly. Although the participation of Aurora A in these events is well established, its mechanism of action is poorly understood in most cases. Moreover, the relatively small number of known substrates for this kinase does not account for its many roles. In this study, we present and validate a novel strategy to identify Aurora A substrates, along with their specific phosphorylation sites. We have developed a computational approach that integrates distinct types of biological information to generate a ranked list of 90 potential Aurora substrates, of which 76 are novel. Experimental validation on a randomly selected group of candidates, using *in vitro* kinase assays and mass spectrometry analyses, indicates a prediction accuracy of about 80%. Our results open the way to a better understanding of Aurora A function during cell division and point to novel unexpected roles for the Aurora kinase family. We estimate that our approach can be readily applied to more that 30 human kinases.

Scientific output

Publications

Aloy P and Oliva B. Splitting statistical potentials into meaningful scoring functions: testing the prediction of near-native structures from decoy conformations. *BMC Struct Biol,* **16**, 9-71 (2009)

Mosca R, Pons C, Fernández-Recio J and Aloy P. Pushing structural information into the yeast interactome by high-throughput protein docking experiments. *PLoS Comput Biol*, **5**(8), e1000490 (2009)

Pache RA, Babu MM and Aloy P. Exploiting gene deletion fitness effects in yeast to understand the modular architecture of protein complexes under different growth conditions. *BMC Syst Biol*, **18**, 3-74 (2009)

Stein A, Pache RA, Bernardó P, Pons M and Aloy P. Dynamic interactions of proteins in complex networks: a more structured view. *FEBS J*, **276**(19), 5390-405 (2009)

Stein A, Panjkovich A and Aloy P. 3did Update: domain-domain and peptide-mediated interactions of known 3D structure. *Nucleic Acids Res*, **37**, D300-4 (2009)

Zanzoni A, Soler-López M and Aloy P. A network medicine approach to human disease. *FEBS Lett*, **583**(11), 1759-65 (2009)

Research networks and grants

A bioinformatics approach to the study of contextual-specificity in protein interaction networks and potential applications to biomedicine and biotechnology
Spanish Ministry of Science and Innovation, BIO2007-62426 (2007-2010)
Principal investigator: Patrick Aloy

A multidisciplinary approach to determine the structures of protein complexes in a model organism
European Commission, LSHG-CT-2005-512028 (2006-2010)
Principal investigator: Patrick Aloy

Grup de recerca emergent
Generalitat de Catalunya, 2009 SGR 1519 (2009-2013)
Principal investigator: Patrick Aloy

Identification of secondary targets and drug design through the structural and functional analyses of biological networks
Spanish Ministry of Science and Innovation, PSE-010000-2009-1 (2009-2010)
Principal investigator: Patrick Aloy

Identification and validation of novel drug targets in Gram-negative bacteria by global search: a trans-system approach
European Commission, 223101 (2009-2011)
Principal investigator: Patrick Aloy

4th CAPRI evaluation meeting
Genoma España (2009)
Principal investigator: Patrick Aloy

Collaborations

Novel strategy for network-based therapeutics
José Manuel Mas, Infociencia & Anaxomics Biotech (Barcelona, Spain)

Novel ways of assessing protein-DNA interactions
Anastassis Perrakis, Nederlands Kanker Instituut (Amsterdam, The Netherlands)

Structural systems biology
Juan Fernández-Recio, Barcelona Supercomputing Center (Barcelona, Spain); M Madan Babu, LMB-MRC (Cambridge, UK); Baldomero Oliva, Pompeu Fabra University (Barcelona, Spain); Miquel Pons, IRB Barcelona (Barcelona, Spain)