2.5.6 Research Area "Molecular interactions in organic and biological systems. Applications and methodological implementations" (E. Sánchez-García)

Involved: K. Bravo-Rodriguez, S. Mittal, P. Sokkar, N. Tötsch, J. Iglesias, M. Fernandez-Oliva, S. Carmignani, G. Gerogiokas, V. Muñoz Robles, J. Mieres, L.Wollny

Objectives: In the Sánchez-García group, molecular interactions are used as a tool to tune the properties of chemical and biological systems. In this context three very much interconnected, general lines are developed (Figure 15). One key topic is the study of ***molecular interactions in biological systems,*** namely protein–ligand interactions, protein–protein interactions, enzymatic activity, and computational mutagenesis, as well as the effect of solvent on these processes. Another research line is the study of ***molecular interactions in chemical reactivity***, where we focus on reactive intermediates and unusual molecules and the effect of molecular interactions on their stability, spectroscopic properties, and reactivity. At the core of these applications lies the use and methodological implementation of ***multi-scale computational approaches***.

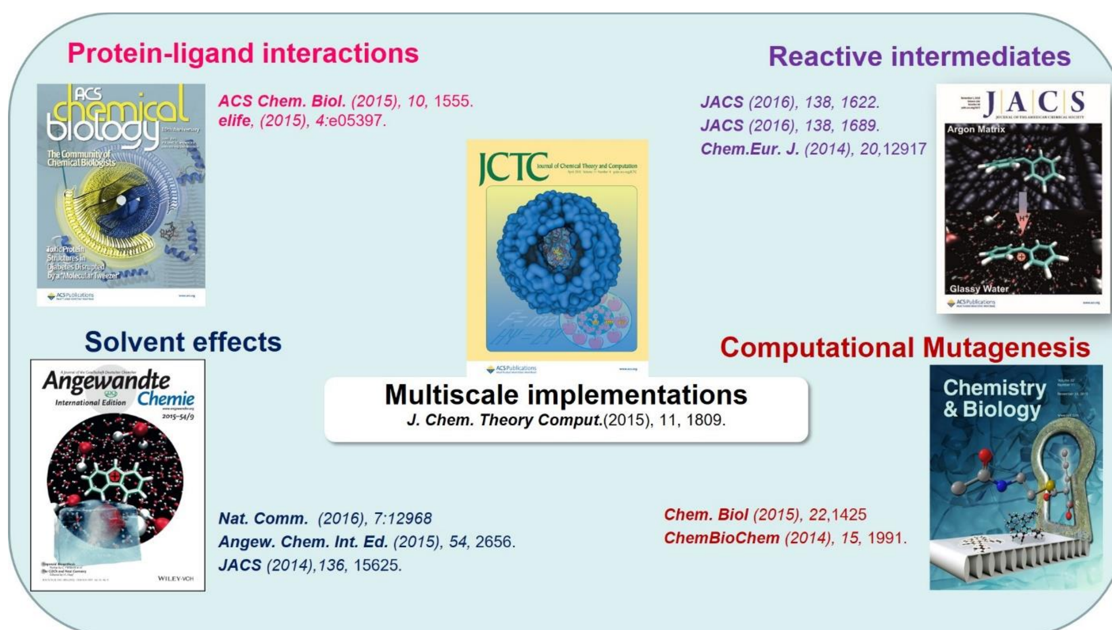

Fig. 15. Main research lines and selected representative publications of the Sánchez-García group in 2014-2016.

Results:

Molecular interactions in biological systems

a) Protein-ligand interactions

Interactions with small molecules can significantly influence the functionality of systems of diverse structural complexity – from amyloidogenic peptides to large proteins and enzymes. In our group we develop computational models of protein-ligand complexes to study their association and how such molecules can modulate protein–protein interactions. The combination of molecular dynamics simulations with free energy calculations and QM/MM methods allows us to predict ligand binding sites in a protein and to identify interactions patterns for an *in silico* design of improved ligands able to reach specific protein regions of biological relevance.

Specifically, we investigate the effect of selective ligands such as molecular tweezers (MT) that bind specifically to lysine and arginine residues, and molecules with a guanidiniocarbonylpyrrole (GCP) moiety targeting negatively charged amino acids. We are also interested in less specific compounds like aromatic heterocyclic derivatives and peptide ligands.

CLR01 is a lysine- and arginine-specific hydrogen phosphate tweezer able to inhibit the self-assembly and toxicity of several amyloid proteins *in vitro* and *in vivo*. In this context, we studied the interactions of the islet amyloid polypeptide (IAPP) with CLR01. Here, the selective binding to critical lysine residues is the molecular basis for the effect of the tweezer that causes IAPP to adopt an off-pathway conformation not found in the absence of CLR01 [117].

$PAP_{248\text{-}286}$ is the prototype amyloidogenic peptide in semen, closely related to the transmission of HIV-1. We found that CLR01 is able to bind all positively charged residues of $PAP_{248\text{-}286}$ in a conserved manner. Conversely, a control molecule consisting of the charged core of CLR01 only features labile interaction patterns with $PAP_{248\text{-}286}$. Thus, we were able to explain the lack of experimental effect of the spacer vs. the inhibition of toxicity by the tweezer [118]. Notably, the experimental studies indicated that CLR01 has a dual activity, namely destroying diverse enveloped viruses (including HIV) and remodeling amyloid fibrils in semen. To clarify how CLR01 can exhibit these two distinct activities, we also studied the molecular tweezer CLR05 that acts as potent anti-viral activity agent with no anti-amyloid activity. Unlike CLR01, the substituents in CLR05 are methylene carboxylate groups. Our previous studies with single amino acids and small peptides indicated that the hydrogen phosphate tweezer CLR01 threads lysine

or arginine side chains very efficiently through its tweezer cavity, while the carboxylate derivative CLR05 is only able to weakly bind the free amino acid outside its cavity by simple ion pairing. Hence, by investigating the CLR05 interaction with $PAP_{248-286}$ we could show that CLR05 is less able to form inclusion complexes with lysine or arginine compared to CLR01. The global minima on the peptide-tweezer free energy surfaces obtained from adaptive biasing force calculations indicated that binding of CLR05 to residues at the N- and C-terminal regions of $PAP_{248-286}$ is not favored. In addition, free energy perturbation calculations predicted that, in $PAP_{248-286}$, CLR01 forms better inclusion complexes than CLR05 for almost all Lys/Arg residues. Thus, we proposed that CLR05 may lack the anti-amyloid activity displayed by CLR01, in agreement with the experimental results (manuscript submitted).

We also studied the interactions of CLR01 with the Huntingtin protein exon-1. This protein is a key target in therapeutic strategies against Huntington's disease (HD), a neurodegenerative disorder without cure. We showed that the lysine residues found at low concentration in the N-terminal fragment of the exon-1 sequence (N17) are crucial for htt aggregation since binding of CLR01 induces structural rearrangements within the htt exon-1 monomer. In a joint experimental and computational study, we also demonstrated that CLR01 potently inhibits htt exon-1 aggregation, underpinning the key role of N17 in modulating htt exon-1 toxicity (manuscript submitted).

We previously reported studies on the selectivity of CLR01 towards Lys residues in a 14-3-3 protein. Now, we wanted to investigate also Arg complexation by molecular tweezers on proteins. In a combined experimental (P. Bayer, Essen) and computational study, we revealed the affinity profile of the tweezers to preferred lysine and arginine residues on the surface of the N-terminus region of the p97 protein (p97-N). Our QM/MM calculations confirmed the preferred complexation sites but also allowed us to discriminate between ambiguous host residues derived from NMR data. The binding of the tweezer to p97-N resulted in the inhibition of the protein-protein interaction between p97 and its cofactor UBXD1 [123].

In another multidisciplinary study using protein crystallography, biophysical affinity determination, and biomolecular simulations, we revealed the structural details of how the molecular tweezer CLR01 influences the 14-3-3/Cdc25CpS216 protein-protein interaction (PPI). CLR01 acts as a supramolecular "Janus" ligand that can bind simultaneously to a flexible peptidic PPI recognition motif and to a well-structured adapter protein (Figure 16). This binding "freezes" one of the conformational states of

the intrinsically disordered Cdc25C protein partner and enhances the apparent affinity of the interaction (manuscript submitted).

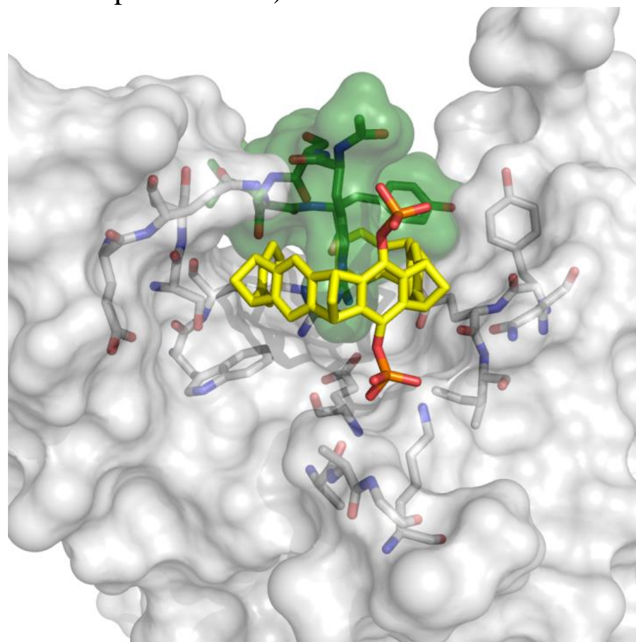

Fig. 16. CLR01 traps Arg 208 of Cdc25C (green surface) inside its cavity and simultaneously establishes contacts with 14-3-3ζ (white surface) via its hydrophobic sidewalls (yellow).

In another study on the 14-3-3ζ protein, we presented the first example of a small molecule binding to the 14-3-3ζ dimerization interface. This compound, featuring a GCP motif, was designed by rational *in silico* optimization of a peptidic ligand identified from a biochemical screening of a peptidic library. The binding was characterized by UV/Vis, MST, multiscale simulations, and X-ray crystallography. QM/MM MD simulations allowed us to investigate the binding of the ligand in solution and confirmed the dimer interface as preferred binding site (manuscript submitted).

The ribosome is a highly relevant but also complex system for ligand design. Macrolides, which are commonly used as antibiotics, are ligands targeting the ribosome. They can selectively bind at the prokaryote ribosome, inhibiting its function. Due to the increased antibiotic resistance of pathogenic strains, there is high interest in designing new synthetic macrolide molecules with enhanced binding to the ribosome. We used molecular dynamics simulations, free energy calculations, and QM/MM methods to study the binding of antibiotic derivatives to the ribosome in explicit solvent. This allowed us to establish which modifications of the macrolide core result in binders with better affinity and to clarify the role of the hydration free energy and conformational entropy on the binding events (unpublished work).

b) Protein-protein interactions and computational mutagenesis

Polyketides are natural products frequently used for the treatment of various diseases, but their structural complexity hinders efficient derivatization. In this context, we introduced enzyme-directed mutasynthesis to incorporate non-native extender units into the biosynthesis of erythromycin. More recently, we extended the molecular rationalization of enzyme-substrate interactions through modeling, to investigate the incorporation of substrates with different degrees of saturation of the malonic acid side chain. This allowed the biosynthesis of new erythromycin derivatives and the introduction of additional mutations into the AT domain for a further shift of the enzyme's substrate scope [113].

We are also interested in the study of disease-causing mutations in complex systems like *high temperature requirement A* (HTRA) serine proteases. Our free energy perturbation calculations predicted which of such mutations strongly destabilize the HTRA1 trimer. Molecular dynamics simulations of the wild type HTRA1 and the mutated systems allowed us to identify key interactions for the integrity of the enzyme. Our data suggested the presence of an intricate network of interactions composed of a hydrophobic cluster and two salt bridges that mediate trimer formation in this enzyme (unpublished work).

In addition to the human HTRA1, we also studied the bacterial serine protease DegP, this time as protein guest in a DNA origami host. Two models were considered for the binding of the 24-mer of DegP ($DegP_{24}$) inside the origami cage. In one model, $DegP_{24}$ interacts with opposite sides of the hexagonal cage while in the second model $DegP_{24}$ interacts with consecutive sides of the hexagonal cage. For each setup two binding motifs were considered. Our atomistic geometric models along with MD simulations suggested that the presence of three ligands per origami face should provide the maximal probability for binding to occur and that all DegP forms, although with distinct space-filling capabilities, can be hosted inside the DNA prisma (manuscript accepted, Nature Communications).

CXCR4 is a receptor protein of the chemokine receptor family. The CXCR4/CXCL12 signaling pair is associated with a variety of diseases like cancer cell metastasis or chronic inflammation. EPI-X4 is a peptide that specifically interacts with the receptor, thereby blocking CXCR4 (X4)-tropic HIV-1 infection and CXCL12 signaling. Our computational studies allowed us to propose binding sites of the peptide on CXCR4. The molecular environment was explicitly considered by embedding the protein in a full

atomistic membrane model and explicit water molecules. Our work revealed which residues of EPI-X4 are essential for receptor binding. On this basis, we made specific predictions for a next generation of EPI-X4 derivatives with improved binding efficiencies. These predictions were experimentally proven by the group of J. Münch (Ulm) and resulted in the generation of even more potent leads (unpublished work).

Molecular interactions on chemical reactivity

Carbenes and carbenium ions are challenging molecules and among the most important reactive intermediates in chemistry. They play key roles in a large number of reactions. In nucleophilic solvents such as alcohols, they can be extremely short-lived (lifetimes in the order of picoseconds), and it was believed that the corresponding cations could be stabilized only in super-acidic, non-nucleophilic solvents. We recently used QM MD and QM/MM MD approaches to investigate the reaction of diphenylcarbene (DPC), an archetypical triplet state carbene, with water in argon matrices and in water ice at 3 K. The combined matrix isolation (W. Sander, Bochum) and computational study allowed us to establish that, in the complex with a single water molecule, the triplet ground state of DPC is switched to its singlet state, stabilized by a strong hydrogen bond with water [109]. A similar effect was found for fluorenylidene (FY), where we also demonstrated that hydrogen bonds with protic solvents like water strongly influence the reactivity of the carbene by selectively stabilizing the singlet state and thus inverting the singlet-triplet gap [114].

The interactions between DPC and the halogen bond donor CF_3I were studied using QM and QM/MM calculations. CF_3I forms very strong complexes with the singlet state of DPC, but interacts only weakly with triplet DPC. This results in a switching of the spin state of DPC, with the singlet complex becoming more stable than the triplet complex. CF_3I forms a second complex (type II) with DPC that is thermodynamically slightly more stable. Our calculations predicted that in this second complex the DPC···I distance is shorter than the F_3C···I distance, whereas in the first complex (type I) the DPC···I distance is, as expected, larger. The type II complex could be only found as a minimum in the matrix environment (QM/MM calculations) and the interconversion was temperature-dependent. We also performed a 2-dimensional potential energy scan with the halogen bond distance and angle as reaction coordinates to explore the relative stability of these structures. The type II complex is characterized by a C-I distance of 2.3 Å. It is stable over a range of C-I-C angles while the type I structure is characterized by a nearly linear C-I-C angle and is stable over a range of C-I distances. Our study of intersystem crossing in the reaction of DPC and CF_3I indicated that it may occur when

the C-I distance is between 3.25 and 3.90 Å. The large calculated spin-orbit coupling may facilitate the intersystem crossing [67].

Unlike DPC and FY, bis(*p*-methoxyphenyl)carbene is the first carbene to be isolated in both its lowest-energy singlet and triplet states. We studied the influence of the C-C-C bond angle at the carbene center and of the conformational flexibility of the methoxy groups on the singlet-triplet gap. Unlike the carbene angle, the orientation and rotation of the methoxy groups have basically no influence on the relative stability of the conformers in the singlet or triplet state. In addition, to assess the impact of water on the singlet-triplet gap, several water complexes were computed considering not only the carbene center as a potential H-bond acceptor, but also both oxygen atoms of the methoxy groups. We found that hydrogen bonding with the methoxy groups shows a small tendency to stabilize triplet states over singlets, which is however not pronounced enough to overcome the larger effect of the interaction of water with the carbene center that strongly stabilizes the singlet [122].

In addition to the interactions with water, we were also interested in the effect of organic solvents and their mixtures on singlet-triplet gaps and carbene reactivity. In a combined broadband femtosecond transient absorption (P. Nürnberger, Bochum) and QM/MM study, we showed that for DPC the decision-maker is not the nearest solvent molecule but its neighbor. Therefore, variation of the solvent mixing ratio allows control over the reactivity of DPC. Using QM/MM molecular dynamics simulations, we also proposed two mechanisms for OH insertion into DPC by methanol [70] (Figure 17) and predicted possible side reactions.

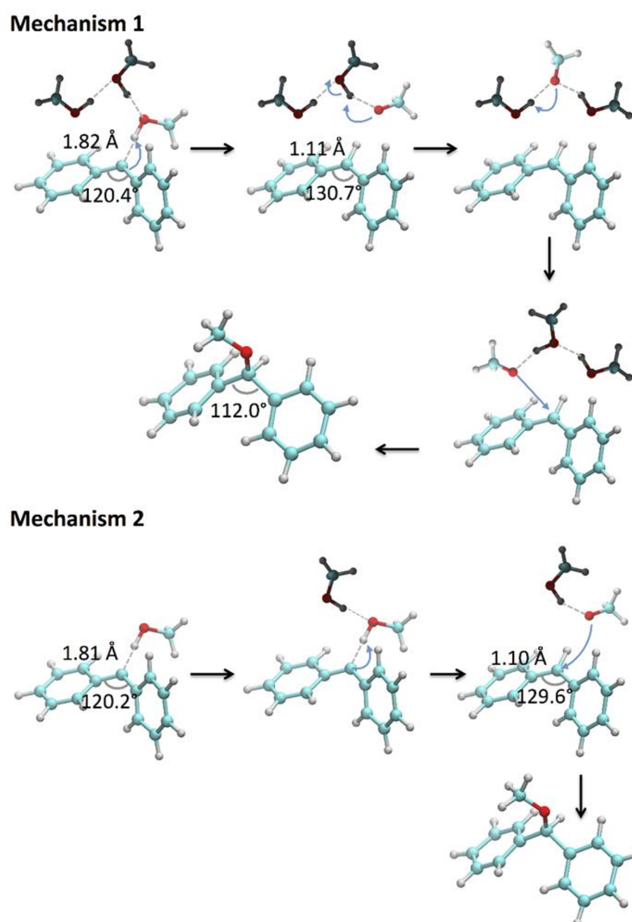

Fig. 17. Mechanisms 1 and 2 of O-H insertion for the reaction of singlet DPC with methanol, as observed in the QM/MM MD simulations. Average distances and angles are given.

Multi-scale computational approaches

Hybrid methods are at the core of our research. ChemShell is a QM/MM modular package that allows the combination of several QM and MM methods. Currently, there is considerable interest in the development of coarse-grained (CG) force fields to improve the performance and sampling in MD simulations and geometry optimizations. Although the CG methodology has been successfully applied to very large molecular systems, it does not allow the study of fine structural details due to the approximate CG representation. In this context, we have implemented a QM/MM/CG protocol in ChemShell. This approach was validated using two enzymes: chorismate mutase (CM) and p-hydroxybenzoate hydroxylase (PHBH). We also evaluated the role of CG modeling on biocatalysis. In CM, the inclusion of an atomistic MM water layer was necessary for a correct description of the energy profile. In the case of PHBH, the use of the polarizable CG model for the outer water did not affect the stabilization of the highly charged FADHOOH-pOHB transition state compared to the fully atomistic QM/MM calculations. A detailed performance analysis in a glycine−water model

system indicated that computation times for QM energy and gradient evaluations at the density functional level are typically reduced by 40−70 % for QM/MM/CG relative to fully atomistic QM/MM calculations [49].

We are currently working on the implementation of an interface to GROMACS in ChemShell and on the implementation of grid cell theory at the multiscale level. We are also implementing an approach to explore potential energy surfaces and to find thermally allowed intersystem crossings (ISC) in reactive intermediates based on QM molecular dynamics simulations.

Future directions: I plan to continue working on protein-ligand interactions. This will be extended to molecular tweezers with specific binding anchors designed to target certain regions of interest in the protein. Multivalent guanidiniocarbonylpyrrole ligands and novel peptide derivatives will also be investigated. The influence of the solvent on protein-protein interactions, enzymatic activity, and catalysis will be in the focus of our research. New reactive intermediates and their interactions with other organic molecules will be explored, and our QM MD approach for ISC will be implemented at the multiscale level to account for solvent and environmental effects.

Publications resulting from this research area: 49, 67, 70, 95, 108-123

External funding: Chemical Industry Funds (Liebig Stipend), German Research Foundation: RESOLV Cluster of Excellence (EXC 1069), Collaborative Research Center "Supramolecular Chemistry on Proteins" (SFB 1093), Boehringer Ingelheim Foundation (Plus-3 grant)

Cooperations:; P. Bayer (Essen, DE); G. Bitan (Los Angeles, US); R. Crespo-Otero (London, U.K); S. Ebbinghaus (Bochum, DE); M. Ehrmann (Essen, DE); J. Münch (Ulm, DE); P. Nürnberger (Bochum, DE); C. Ottmann (Eindhoven, NL); B. Sacca (Essen, DE); W. Sander (Bochum, DE); F. Schulz (Bochum, DE); T. Schrader (Essen, DE); W. Thiel (Mülheim/Ruhr, DE); E. Wanker (Berlin, DE)