Cation–π–cation interactions in structural biology

Silvana Pinheiro, [1] Ignacio Soteras, [1] José Luis Gelpí, [2] François Dehez, [3] Christophe Chipot, [3,4] F. Javier Luque, [1] and Carles Curutchet[1]

[1]*Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Spain*

[2]*Departament de Bioquimica y Biologia Molecular, Facultat de Biologia, Universitat de Barcelona, Spain*

[3]*Laboratoire International Associé Centre National de la Recherche Scientifique et University of Illinois at Urbana −Champaign, Unité Mixte de Recherche No. 7565, Université de Lorraine, Vandoeuvre-lès-Nancy cedex, France*

[4]*Beckman Institute for Advanced Science and Engineering, University of Illinois at Urbana — Champaign, Urbana, Illinois (USA).*

Corresponding author: * Carles Curutchet; carles.curutchet@ub.edu;

Phone: (+34) 934.029.00 Fax: (+34) 934.035.987

Abstract *- Biological structures are stabilized by a variety of noncovalent interactions, such as hydrogen bonds, π –stacking, salt bridges or hydrophobic interactions. Besides hydrogen bonds and π–stacking, cation–π interactions between aromatic rings and positively charged groups have emerged as one of the most important interactions in structural biology. Although the role and energetic characteristics of these interactions is well established, a special case involving three molecular species, termed cation–π–cation interaction, is still poorly understood. In this contribution, we aim at advancing in the understanding of cation–π–cation interactions and their role in the structure and stability of biosystems via two complementary approaches. The first one consists of a statistical study of the occurrence, composition and geometry of cation–π–cation interactions identified on a non-redundant set of protein structures from the PDB (Protein Data Bank), which demonstrates that cation–π–cation interactions are indeed common in proteins. We also analyze the degree of conservation of the interactions by inspection of similar sequences obtained through the sequence alignment tool BLAST. The second part of the study consists of an energetic analysis of the most relevant interactions at the SCS-MP2/CBS level of theory, as well as an energy decomposition analysis for representative cases performed at the SAPT2 level. Besides the well-known deficiency of standard additive force fields to describe the relevant polarization contribution to cation–π interactions, our results indicate that non-additive three-body contributions are significant in this case, implying that cation–π–cation interactions constitute an even more challenging case expected to be dramatically misrepresented by standard additive force fields. In vacuum, the interactions identified are strongly repulsive. Further work is being carried out in order to understand the strength of cation–π–cation interactions in a protein environment, where the cation–cation repulsion will be strongly screened.*

Keywords: Cation–π interaction; interaction energy; cation– π –cation interactions; three-body; non-polarizable FFs.

I. INTRODUCTION

The understanding of intermolecular forces is extremely important to understand the behavior of chemical systems at the molecular level. It should be noted that such interactions and understanding gain its maximum expression in biological systems. Molecules of life (DNA, RNA, proteins, etc.) are held in their three-dimensional structures by intra and intermolecular interactions. Thus, the three-dimensional molecular structure is responsible for the specific biological activity of these molecules; note the importance of understanding of such interactions. Moreover, the correct description in the classical force field for molecular simulations used is a key to enable a realistic description of such systems.

Hydrogen bonds, π–stacking, salt bridges, and the hydrophobic effect all play roles in folding a protein and establishing its final structure. In addition, the cation–π interactions is a non-covalent interaction between positively charged atoms or groups and a π-electron system have recognized as an important non-covalent binding interaction relevant to structural biology (1-5). The presence of a positive charge near a highly polarizable π-system leads to significant polarization effects, not properly described by standard non-polarizable FFs (6-10). Like this hydrogen bonding and π-π stacking interactions, the cation-π interactions have emerged as the fundamental forces that control the structure and function of macromolecules (2).

A number of studies have established a role for cation-π interactions in biological recognition and the contribution of polarization in such interactions are very important, and may even surpass electrostatic contribution, from a monopole (cation) interacting with a quadruple (aromatic ring) (3). This is explained by the presence of a positive charge near one system as the highly polarizable π system. This type of interaction is common in protein structures.

Although the role and energetic characteristics of these interactions is well established, a special case involving three molecular species, termed cation-π-cation interaction, for example, Arg-Phe-Lys interactions, is still poorly understood, are simultaneous interaction between a π-electron system, the aromatic sidechains of phenylalanine (Phe, F), tyrosine (Tyr, Y) or tryptophan (Trp, W) and two cations than occur between the cationic side chains of either lysine (Lys, K) or arginine (Arg, R)). In which the polarization effect are even more complex due to its nature of the three-body interactions.

In this contribution, we aim at advancing in the understanding of cation-π-cation interactions and their role in the structure and stability of biosystems via two complementary approaches. A structural analysis, statistical study of frequency, composition and structure of cation-π-cation interactions identified on a non-redundant set of proteins from the PDB and energetic analysis, SCS-MP2/CBS high-level calculations of cation-π-cation interaction energies for the most relevant interactions identified.

II. COMPUTATIONAL METHODS

Identification and selection of cation-π-cation interactions

In this section, we detail our strategy for identifying and ranking cation-π-cation interactions in proteins. To conduct the analysis of cation-π-cation interactions was performed over a representative list of proteins of Cluster90

database. Given the big amount of cation-π-cation interactions complexes found and characterized in this database which consists of 20.759 non-redundant protein structures taken from the PDB, criterions accurate was used. A distance cutoff of d_1+d_2 < 7,5 Å and angle cutoff: $\alpha_1+\alpha_2$< 30°. The letter is defined between the normal of the ring and a vector between the centre of the ring and the positively charged site was considered. In case of Lys and Arg, the distance is taken to N and C atoms respectively. In case of Trp, both rings are considered, and in the case of His, the two nitrogens of the five membered ring were considered. The results were filtered in order to remove duplicities in each interaction due to symmetry considerations within the pdb.

In order to do that, an analysis of geometries was carried out looking up the distribution of distances and angles found in the interactions. Based on this criterion we selected 51 interactions for energic analysis. Were classified the trimers systems in accordance with the type of aromatic (Phe, Tyr and Trp) and cationic (Lys, Arg and His) aminocids, as well as ions (Na^+, Ca^{2+}, Mg^{2+}, K^+, Li^+, Cu^{2+}). Among all the structures analyzed, the representative structure of cation-π-cation interactions multiple is 1A22 (Human growth hormone), that is a special case with three consecutive cation-π-cation interactions K379-W386-R411, R411-F425-R413, R413-Y422-K415.

Calculation of conservation and accessibility

We have evaluated the conservation of cation-π-cation interactions in each protein with the aid of the BLAST (Basic Local Alignment Search Tool). That compares the query sequence against a large number of protein sequences, which search database non-redundant protein sequences (nr) using algorithm BLASTp (protein-protein BLAST) [11]. To investigate the conservation of interaction in proteins, was considered sequence of proteins above >80% of identity. For example, in an alignment of 20 sequences that all share >90% pairwise identify conservation may be interesting. In contrast, if the pairwise identify is below 30% then lower conservation scores will be informative.

The computations of solvent accessibility and protonation of residues involved in cation-pi-cation interaction were calculated by NACCESS and ProPka programs, that calculates the accessible area of a molecule from PDB. The pattern was categorized into buried, partially buried and exposed based on ranges of accessible surface area.

Construction of the model Systems

From the representative pdb coordinates was built simplified model systems of each amino acid aromatic and cationic on which we have carried out theoretical calculations to study cation-π-cation interactions. Each trimer then was reduced to a system that could be studied computationally. Lys and Arg were represented as ammonium and guanidinium ions, and Phe, Tyr, and Trp were represented as benzene, phenol, and indole, respectively. The geometries of the structures of such monomers were optimized at *MP2/aug-cc-pVDZ* level using the Gaussian 09 suite of programs. Thereafter was performed the alignment the structures, overlapping the optimized structures of the monomers with the structures of pdb reference.

Calculation of interaction energy

To evaluate the non-additive contribution to three bodies interactions in cation-π-cation by comparing the contributions, we conducted the calculations the interaction energies of both systems as models and pdb's, apply different basis sets incorporating polarization and diffuse functions, a total of 6 basis sets were also tested to examine the convergence of the interaction energy with basis set size, we have used the basis sets 6-31G, 6-31+G(d), 6-31++G(d,p), 6-311++G(d,p) and also the correlation consistent basis sets aug-cc-pVDZ and aug-cc- pVTZ, the last two were considered as the reference to calculate the errors on the interaction energies obtained with the other basis sets. The calculations were developed using the ab initio level SCS/MP2, the new method improves upon MP2, thereby rectifying many of its problems, indicates significant robustness and suggests it as a valuable quantum chemical method of general use.[12]

$$E_{SCS/MP2} = E_{HF} + 1/3(E_{corr(\alpha-\alpha)} + E_{corr(\beta-\beta)}) + 6/5E_{corr(\alpha-\beta)}$$

In principle, it could be possible to extrapolate the energies of the systems (and hence the interaction energy) to the complete basis set limit (CBS) using the Dunning DZ-TZ basis set series, will combine the calculations MP2/cc-pVDZ and MP2/cc-pVTZ level by formula Helgaker.[13-15] Such a formula allows obtain CBS E_{XY}^{∞} energy, combining the calculated energies cc-pVXZ level according to the expression.

$$E_{XY}^{\infty} = \frac{E_x^{corr} X^3 - E_x^{corr} Y^3}{X^3 - Y^3}$$

To obtain more accurate energetics, were applied and basis set superposition errors (BSSE) were subtracted from the computed BDEs using the correction counterpoise (CP), that is calculations of monomers using the base set trimer. Counterpoise corrections for basis set superposition error (BSSE) were included in the interaction energy calculations, using the Gaussian 09 software.

$$E_{\infty}^{tot} = \frac{3\alpha}{3\alpha-2\alpha} E_3^{HF} - \frac{2\alpha}{3\alpha-2\alpha} E_2^{HF} + \frac{3\beta}{3\beta-2\beta} E_3^{corr} - \frac{2\beta}{3\beta-2\beta} E_2^{corr}$$

The magnitude of the different components of the interaction energy was examined by means of the Symmetry-adapted perturbation theory (SAPT) as implemented in the SAPT2006 program. In particular, we have used the SAPT2 computational scheme, which should provide an interaction energy U_{tot}^{SAPT2} equivalent to the supermolecular MP2 one. Within the SAPT2 framework the interaction energy is decomposed as shown in eqn, where the first four terms stand for the electrostatic U_{ele} , induction U_{ind}, exchange U_{exch} and dispersion U_{disp} components. In addition, the SAPT expansion yields two terms corresponding to the coupling between exchange and induction $U_{exch-ind}$ and between exchange and dispersion $U_{exch-disp}$. Finally, the last term δHF accounts for a collection of higher order induction and exchange-induction terms.[16]

$$U_{tot}^{SAPT2} = U_{ele} + U_{ind} + U_{exch} + U_{disp} + U_{exch-ind} + U_{exch-disp} + \delta HF$$

III. RESULTS AND DISCUSSION

Using the above criteria, we scanned a larger dataset of representative protein crystal structures, taken from the Protein Data Bank (PDB). Were identified in the *Cluster90* database 20.759 non-redundant protein structures. The results point out to 2.328 structures which have one or more cation-π-cation interaction corresponding 11.3%. In total, 2.898 interactions were identified.

Structural analysis

- Among the cation-π-cation interactions identified on the PDB, the frequency of aromatic amino acids found (Tyr > Phe > Trp) is similar to the frequency observed for 2-body cation-p complexes.

- Because Trp amino acids are less frequent in proteins than Tyr or Phe, the comparatively large number of cation-π-cation interactions identified suggest a preference for Trp over Tyr or Phe.

Energetic analysis

The non-additive contribution (3-body) is significant (up to 7 kcal/mol). Thus, cation-π-cation interactions are expected to be badly described by standard additive FFs. In a gas phase environment, cation-π-cation interactions are strongly repulsive, in the range 20-35 kcal/mol. Further work is needed to understand the strength of cation-π-cation interactions in a protein environment, where the cation-cation repulsion will be strongly screened.

IV. CONCLUSION

Our results demonstrates that cation–π–cation interactions are indeed common in proteins and one every ten proteins contain at least one cation-π-cation interaction. Indicate that non-additive three-body contributions are significant in this case, implying that cation-π–cation interactions constitute an even more challenging case expected to be dramatically misrepresented by standard additive force fields. In vacuum, the interactions identified are strongly repulsive. Further work is being carried out in order to understand the strength of cation-π–cation interactions in a protein environment, where the cation–cation repulsion will be strongly screened.

REFERENCES

1. Ma, J. C. & Dougherty, D. A. *Chem. Rev.* **97** (1997) 1303–1324
2. A. Subha Mahadevi and G. Narahari Sastry. *Chem. Rev.* **113** (2013) 2100.
3. E. Cubero, F. J. Luque, M. Orozco. *Proc. Natl. Acad. Sci. U.S.A.* **95** (1998) 5976.
4. H. Minoux, C. Chipot. *J. Am. Chem. Soc.* **121** (1999) 10366.
5. I. Soteras, C. Curutchet, A. Bidon-Chanal, F. Dehez, J. G. Ángyán, M. Orozco, C. Chipot, F. J. Luque. *J. Chem. Theory Comput.* **3** (2007) 1901.
6. Cubero, Luque & Orozco PNAS 95, 5976 (1998)
7. Minoux & Chipot, JACS 121, 10366 (1999)
8. Soteras et al. JCTC 3, 1901 (2007)
9. François Dehez, János G. Ángyán, Ignacio Soteras, F. Javier Luque, Klaus Schulten, and Christophe Chipot; Modeling Induction Phenomena in Intermolecular Interactions with an Ab Initio Force Field *J. Chem. Theory Comput.* 2007, 3, 1914-1926.
10. Ignacio Soteras, Modesto Orozco and F. Javier Luque; Induction effects in metal cation–benzene complexes. *Phys. Chem. Chem. Phys.*, 2008, 10, 2616–2624.
11. Stephen F. Altschul, Warren Gish, Webb Miller, Eugene W. Myers and David J. Lipman; Basic Local Alignment Search Tool. J.*Mol. Biol*. 1990, 215, 403-410.
12. Grimme, 2003. *The Journal of Chemical Physics* 118 (9095).
13. ASGER HALKIER, TRYGVE HELGAKER, POUL JORGENSEN, WIM KLOPPER, HENRIK KOCH, JEPPE OLSEN, ANGELA K. WILSON. Basis-set convergence in correlated calculations on Ne, N2 and H2O. *Chemical Physics Letters*, 286 (1998) 243-252.
14. D. G. Truhlar .(1998), *Chemical Physics Letters* . **294** (1998) 45-48.
15. T. Helgaker, Trygve; et al (1997). *The Journal of Chemical Physics*. 106 (9639).
16. Jerziorski, Moszynski, Szalewicz Chem. Rev. (1994), 94, 1887.