JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

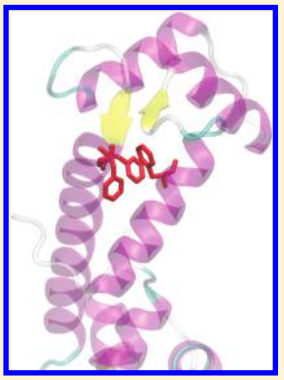
Aromatic−Aromatic Interactions in Proteins: Beyond the Dimer

Esteban Lanzarotti,[†] Rolf R. Biekofsky,[†] Darío A. Estrin,[‡] Marcelo A. Marti,[*,†,‡] and Adrián G. Turjanski[*,†,‡]

[†]Departamento de Química Biológica and [‡]Departamento de Química Inorgánica, Analítica y Química Física/INQUIMAE-CONICET, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 2, Buenos Aires, C1428EHA, Argentina

Supporting Information

ABSTRACT: Aromatic residues are key widespread elements of protein structures and have been shown to be important for structure stability, folding, protein−protein recognition, and ligand binding. The interactions of pairs of aromatic residues (aromatic dimers) have been extensively studied in protein structures. Isolated aromatic molecules tend to form higher order clusters, like trimers, tetramers, and pentamers, that adopt particular well-defined structures. Taking this into account, we have surveyed protein structures deposited in the Protein Data Bank in order to find clusters of aromatic residues in proteins larger than dimers and characterized them. Our results show that larger clusters are found in one of every two unique proteins crystallized so far, that the clusters are built adopting the same trimer motifs found for benzene clusters in vacuum, and that they are clearly nonlocal brining primary structure distant sites together. We extensively analyze the trimers and tetramers conformations and found two main cluster types: a symmetric cluster and an extended ladder. Finally, using calmodulin as a test case, we show aromatic clsuters possible role in folding and protein−protein interactions. All together, our study highlights the relevance of aromatic clusters beyond the dimer in protein function, stability, and ligand recognition.



INTRODUCTION

Within the collection of determined protein structures, there is a wealth of principles governing the complex sequence conformation−function relationships. The rapid increase of protein structural information accumulating in databases offers the opportunity to look at the biomolecular structural space (structurome) as a whole, in order to understand these principles. Studies of several protein families have shown significant sequence and structural conservation in the hydrophobic core and in positions that are involved in substrate recognition, in catalytic activity, or that are otherwise functionally and/or structurally important.[1−5] The interaction among aromatic residues is known to be important in the structural stabilization of proteins, in protein−protein recognition, and in ligand binding and protein folding.[6−17] So far, aromatic dimers have been characterized in great detail both structurally and energetically. Aromatic−aromatic interactions, as those observed in dimers, can be structurally classified in three types depending on the orientation between the planar ring systems: edge to edge, edge to face, or face to face, with of course also the presence of intermediate orientations. Interestingly, surveys of protein structures have shown that dimers prefer the edge to face (or T-shape) type of interaction.[6,18]

Aromatic molecules tend to form higher order clusters, beyond the dimer, that adopt specific conformations, as has been shown for small benzene clusters in vacuum.[17,19−26] The corresponding trimer, tetramer, and even pentamer structures (or conformations) have been determined, and their energetic and structural features have been evaluated both experimentally and with computational methods.[17,19−26] Taking this into account, we thought that similar higher order clusters of aromatic amino acids should also be present in proteins and that they might be important for protein structure and function. To answer this fundamental question we identified and characterized possible clusters formed by aromatic amino acids based on the structural information available in the Protein Data Bank (PDB).[27]

Our results show that aromatic trimers, tetramers, and even larger clusters are found in almost half of the proteins crystallized so far and that they are nonlocal and adopt specific conformations which resemble the most favorable conformation previously observed in isolated benzene clusters.

COMPUTATIONAL METHODS

Protein Structure Selection. For the current analysis, all unique proteins (as defined by Uniprot Id[28] with a structure deposited in the PDB)[27] were considered. The use of Uniprot Id significantly reduces the redundancy of the PDB but does not eliminate the bias due to differential representation of protein families or multiple structures of highly similar proteins. For this reason we applied a sequence similarity filter, i.e., we considered only one structure for all sequences with over 95% identity. The total number of structures used in our study turns out to be 18 547.

Aromatic Cluster Detection and Definition. The aromatic residues considered were phenylalanine (Phe), tyrosine (Tyr),

Received: February 8, 2011
Published: June 11, 2011

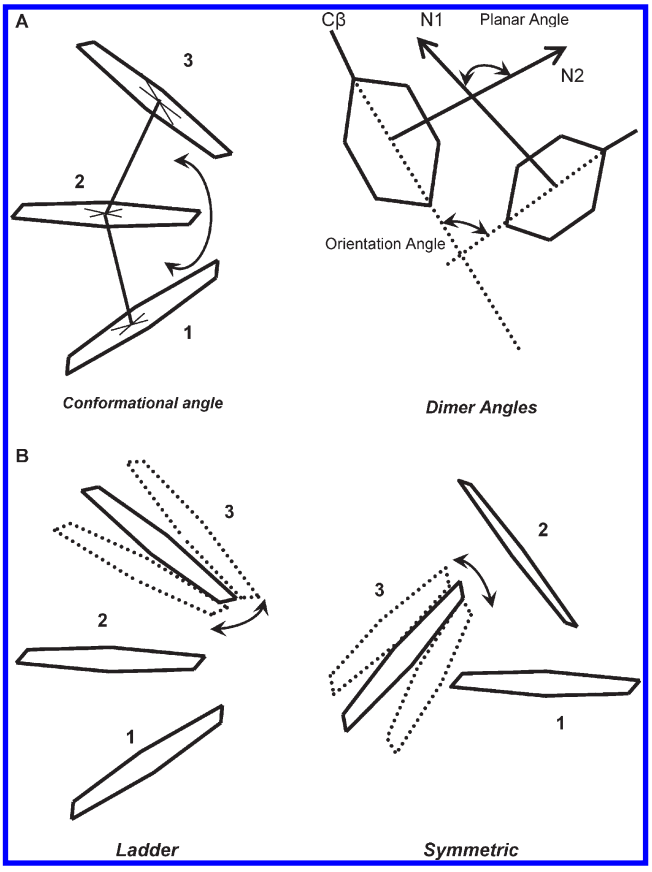


Figure 1. (A) Conformational trimer angle, (upper left panel) and planar and orientational dimer angles (upper right panel). (B) Two main conformations Lad and Sym observed for the aromatic trimers. (lower left and right panels, respectively).

and tryptophan (Trp), and the aromatic center was defined as the center of mass (COM) of the atoms of the corresponding benzene ring. Histidine was not considered since, although it is aromatic, it is also highly polar and hydrophilic and therefore is mainly involved in other types of interactions. To identify the clusters, we looked first for all pair of aromatic residues with COM−COM distance of less than 8 Å. We plotted the distance distribution between aromatic residues (Figure S1, Supporting Information) and calculated the optimal value, 8 Å is about 2 standard deviations from this value. Then, the presence of a third residue at less than 8 Å distance to any of the two residues was added to identify a trimer. The same procedure was used for higher order clusters. This coarse definition allowed us to have a subset of residues that are near in space but do not necessarily form characteristic clusters. To find real clusters we calculated several geometrical properties (see below) and compared them with those computed for the optimal structures of isolated benzene clusters, as determined from ab initio calculations.[21,23,24]

The corresponding analysis is based on properties that were defined for aromatic trimers. As will be shown by our results, larger clusters are built upon trimer structure and therefore can be identified by adding new aromatic residues to the existing trimer. Each trimer is characterized by the three distances between the COMs of the aromatic rings, the angles between the planes containing the aromatic rings (called the planar angle), the orientational angle defined as the angle between the vectors pointing from the β-carbon to the COM of the aromatic ring, and the angles formed by the three COMs of the aromatic residues in the trimer (called the conformational angle). These angles are schematically shown in Figure 1A. Based on the analysis of these defined distances and angles, we divided the clusters in two categories: those with three similar conformational angles and three similar distances where classified as symmetric (Sym) and those with one distance and one angle larger than the other two were classified as ladder (Lad) (Figure 1B).

Characterization of the Aromatic Trimers. Once all trimers (or larger clusters) were found, we characterized them according to the following properties:

(i) The geometry determined by the aromatic center distances and characteristic angles, as defined above.

(ii) The secondary structure, as defined by the secondary structure elements to which each residue belongs. For the definition of secondary structure we used the definition of secondary structure of proteins (DSSP) algorithm[29] and classified all residues as belonging to helix (H), Sheet (E), or loop (L) structure.

(iii) The sequence distance relations, as defined by the distances (or order) in the primary sequence between the residues. The sequence distance is defined as the relative position in primary sequence of two aromatic residues. For example, if the first aromatic residue is at position 10 and the second at position 13, their sequence distance is 3. Sequence distance probability functions were constructed by computing the presence of an aromatic residue in position "*i*" given that an aromatic residue is found at position 0 for a given set of clusters.

(iv) For each cluster we also computed an effective pair interaction energy and a total cluster interaction effective energy E_t (see below).

Effective Aromatic Residue Interaction Energies. To have an estimation of the interaction energy between any pair of aromatic residues and therefore to compute the total cluster effective energy, we used an effective energy function based on the distance probability function between any pair of aromatic residues. Based on this probability distribution, the interaction effective energy between any pair of aromatic at distance "*r*" residues was defined by the following equation (eq 1):

$$E(r) = -k_B T \ln[p(r)/r^2] \quad (1)$$

Where $E(r)$ is the interaction energy of the two aromatic residues whose COM distance is r, k_B is the Boltzmann constant, T is the temperature (set at 300 K), r^2 is the volume normalization factor, and $p(r)$ is the probability function (see Supporting Information). This methodology, which is also called Boltzmann inversion, is widely used to determine contact energies of residues in coarse grain potentials and also to define effective atom−atom interaction energies.[30−33]

Reference Aromatic Cluster Interaction Energies. To compare the effective interaction energy with a more rigorously computed value, the total interaction energy of selected trimers (ΔE) was calculated using quantum mechanics-based methods as follows:

$$\Delta E = E_{clus} - (E_1 + E_2 + E_3) \quad (2)$$

The energies, for the cluster E_{clus} and each individual aromatic residue energy (E_i), were calculated using two methods: (i) the Gaussian-98 package at the MP2 level with a 6-31++G** basis set.

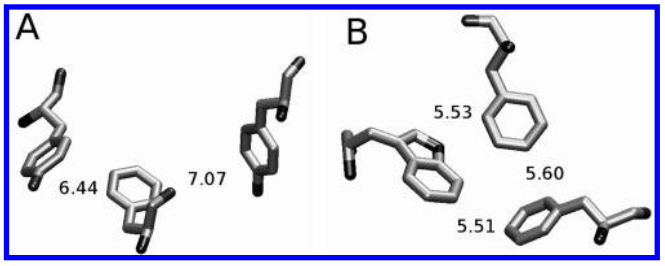


Figure 2. Structure of the two geometrical aromatic trimer types: (A) Lad-type PDB ID: 1VFG, residues 205, 212, and 260. (B) Sym-type PDB ID: 1JDI, residues 176, 153, and 87. The distances (Å) among the geometrical centers of the aromatic rings are shown.

The results were corrected by taking into account the basis sets superposition error; and (ii) the classical Amber99 force field.[34] It is important to remark that different ab initio methods have been tested to calculate isolated aromatic interactions,[35,36] and we found that MP2 with this basis set performs relatively well for the purpose of our study.

Environment Hydropathy index. In order analyze the characteristics of each cluster close environment, we computed the mean hydropathy index between all residues interacting with the corresponding cluster. We defined the interacting residues as those whose side chain COM is at less than 8 Å of any aromatic side chain COM from the cluster residues. The index scale ranges from −4.5 for highly hydrophilic residues, such as arginine, to 4.5 for highly hydrophobic residues, such as leucine.

Aromatic Cofactor Analysis. As cofactors are integral parts of protein structures, we also looked if their aromatic rings are part of aromatic clusters with the aromatic residues of proteins surveyed before. We selected a group of cofactors taken from the CoFactor database.[37,38] We looked in all PDB structures, as defined above, for the following cofactors: Flavin adenine dinucleotide (FAD), flavin mononucleotide (FMN), menaquinone (MQ7), pyridoxal 5′-phosphate (PLP), nicotinamide-adenine dinucleotide (NAD), *S*-adenosylmethionine (SAM), and thiamine diphosphate (TDP), which all show the presence of an aromatic ring. For the FAD and MQ7 the more hydrophobic ring was selected, while for SAM and TDP the outermost six-member ring was considered. A total of 1241 PDB structures contained the selected cofactors.

RESULTS AND DISCUSSION

Aromatic Trimers. *Cluster Identification and Geometrical Characteristics.* Beyond the dimer the first possible cluster is the aromatic trimer. Previous in vacuum ab initio optimization of benzene trimers showed that the preferred conformations can be assigned to two main geometrical types: a triangular, cyclic symmetric conformation with three similar distances between the COM of the aromatic residues and also three similar conformational angles between them, and a stacked (ladder-like) conformation with one COM distance and one conformational angle larger than the other two.[21,23,24] We looked for all trimers in all available unique protein structures by computing the mentioned geometrical properties for all aromatic clusters, as described in the Computational Methods Section. Interestingly, but not totally unexpected, we found that trimers of aromatic residues in proteins strongly resemble and therefore can be clearly assigned to each of these previously observed main structural types: (i) those displaying three aromatic rings in which the central residue interacts with the two other aromatic residues that do not interact with each other because they are too far away (i.e., trimers with two interactions) and therefore have one conformational angle and one COM distance larger than the other two, which we will call lad trimers; and (ii) those trimers where the three residues are forming a ring with circular symmetry where each aromatic residue is in contact with the other two (i.e., trimers with three interactions) and therefore have all three similar conformational angles and all three similar COM distances, which we will call Sym trimers, as depicted in Figure 1B. A typical trimer structure, found in proteins, for each geometrical type is depicted in Figure 2.

Having found the characteristic trimers, we decided to analyze their impact in protein structures and to classify their properties. To determine the impact of the trimers in the available structurome, we initially computed the percentage of protein structures harboring a trimer of each type. Interestingly, almost half of the protein structures display at least one aromatic trimer (44% of all analyzed unique structures). More important, 80% of all dimers are part of a trimer, and about half of the trimers part of a tetramer, showing a clear tendency of the clusters to increase size. Concerning each type, the Lad type is clearly the most abundant trimer with 78%.

To structurally characterize the aromatic trimers, we initially computed the COM−COM distances, r_{12}, r_{13}, and r_{23} for all the identified trimers of each type. We selected residue 2 as the one that has two interactions, and the order of 1 and 3 is defined according to sequence numbering; in the case of Sym trimers only the sequence order was used, since all three residues have two interactions. Sym distance distribution histogram (Figure S2, Supporting Information) yields normal distribution with a mean of 5.8 Å and a standard deviation (SD) of 1.05 Å. However, when we separate the COM−COM distances in groups, minimal, medium, and maximum distances of each aromatic trimer and build the distribution histograms, three overlapping but slightly different curves appear. The peak of the minimal, medium, and maximal distances are at about 5.0, 6.0, and 6.5 Å, showing the tendency to close packing of the aromatic residues but with a noncomplete symmetrical structure. It should be noted that the lack of complete symmetry observed for the Sym trimers in proteins (as reflected by slightly different COM−COM distance probability histograms) possibly reflects constraints imposed by the overall protein structure. However, it is striking that Sym trimers adopt almost the same conformation as the lowest-energy minima described for isolated benzene clusters.[21,23,24] For the Lad type, the COM−COM distance (r_{12}, r_{13}, and r_{23}) distribution histogram yields, as expected, two Gaussian-like curves (Figure S2, Supporting Information). One similar to that of the Sym type, with a mean value of 5.7 Å, corresponding to the distance distribution of the central residue with its neighbors, and the other one representing the distance of the noninteracting aromatic residues with a mean distance of about 10 Å.

We also characterized the angular probability distribution for the conformational angles defined as the angles formed by the three COMs of residues 1, 2, and 3 (there are three conformational angles for each trimer). Analysis of the corresponding distribution, presented in Figure 3A, shows that Sym trimers have a narrow probability distribution around 60°, corroborating the close to symmetric nature of the trimer. We then computed the planar angle for each pair of aromatic residues in the trimer because its value can be used to distinguish between stacked or T-shape types of interaction. Figure 3B shows the corresponding probability distribution of the planar angles. The figure shows a broad distribution for Sym trimers with preferential orientations

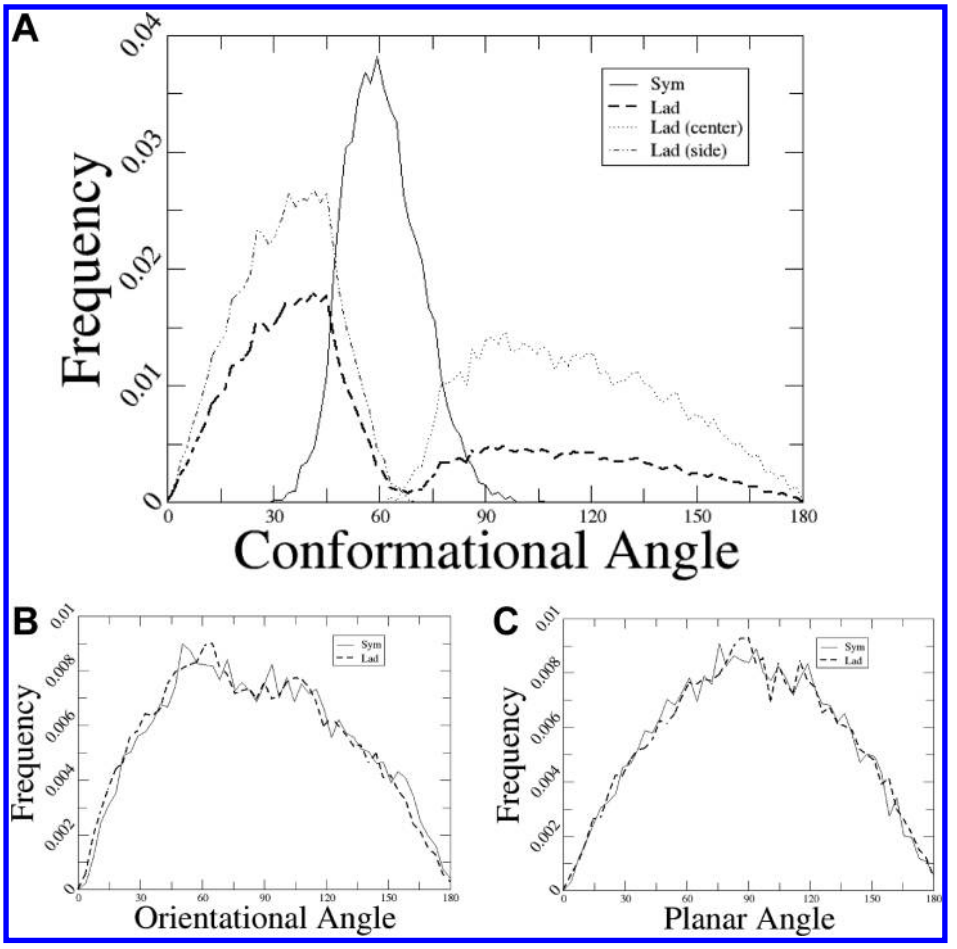


Figure 3. (A) Conformational angle probability distribution for Sym and Lad trimers (solid and dashed lines, respectively), the Lad trimers were separated the angle of the central residue (dotted line), and the side residues (dashed dotted line). (B) Probability distribution for the planar angle in the Sym and Lad trimers (solid and dashed lines, respectively). (C) Probability distribution for the orientational angles in the Sym and Lad trimers (solid and dashed lines, respectively).

closer to the T-shaped (or edge to face type of interactions) with angles between 60° to 120° and a maximum ca. 90°. However, almost no face to face interactions are present (i.e., angles close to 0° or 180°). Finally, analysis of the orientational angle, which measures the relative orientation of the side chain, shows again a broad distribution with a preferred orientation for the aromatic residues in the trimers close to 40°. This is probably related to the fact that many trimer residues (two and rarely even the three residues) belong to the same helix, are in relative sequence position i and $i + 3$, and therefore have a relative orientation near 40°. The role of secondary structure will be further analyzed below.

Analysis of the conformational angular probability distribution for Lad trimers, computed only for the case where the central residue corresponds to the angle vertex, (Figure 3A dotted line) shows a wider distribution from ca. 60° to 180° and displaced two larger values with a maximum around 100°, when compared to the Sym trimers. Interestingly, almost no perfect ladder, i.e., with 180° angles, is present. On the other hand, the conformational angle computed with the side residues as the vertex shows lower angle values compared to the Sym trimers and a very narrow distribution. The other computed characteristic angles, planar and orientational angles, show for the Lad trimer the same results as for the Sym trimers described above.

Our results show that aromatic amino acids tend to form trimers and that their structures resemble the two types of optimized structures reported for isolated benzene clusters but a tendency to establish edge to face like interactions. We will now discuss the interaction energy of the most abundant trimer conformations.

Effective Trimer Interaction Energy. In order to evaluate the difference in trimer interaction energies, we will use an effective residue-based potential, as explained in the Methods Section. To check if the corresponding residue level potential was a valid approximation, we initially computed the interaction energy for six selected trimers using ab initio and classical atomistic-based methods, as shown in Table S1, Supporting Information. The results show that aromatic interactions can be considered as additive, as little electronic cooperativity (due to polarization effects) is observed. This conclusion is based on two observations: First because the classical nonpolarizable AMBER force field reproduces reasonably well the total interaction energy and observed trends. Second, we tested the cooperativity by comparing the MP2 computed binding energies (shown in Table S1, Supporting Information) with the sum of each dimer binding energies calculated separately. The results yield values around 0.2 kcal/mol, which are quite low and suggest that at this level of theory no significant cooperativity is present. Even though higher

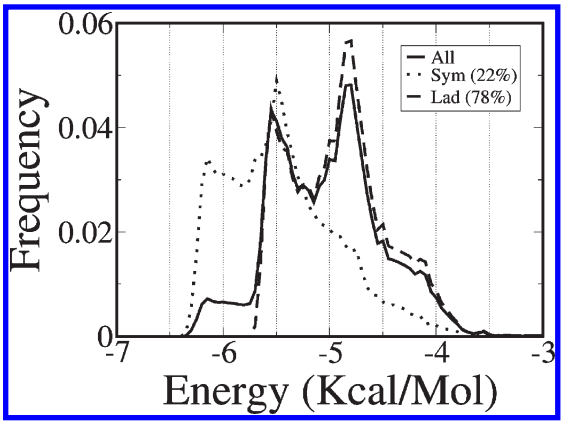


Figure 4. Frequency histogram of trimer interaction energies computed with eq 1. Solid line is all aromatic trimers. Dashed line is Lad-type trimers (78% of the total trimers). Dotted line is Sym-type trimers (22% of the total trimers).

order ab initio methods should be used to study cooperativity in aromatic interactions, to our purpose, the results indicate that we can consider the interactions as additive. It is also important that the effective residue-based energy yields similar values (and trends) as those obtained with the more accurate methods. Although the values are less disperse, the approximate magnitude of the interaction is clearly well described by the effective potential. Taking this into account and the low computational cost of this methodology, it will be used to compute the effective interaction energy of all trimers.

One of the key properties related with the energetic interaction of the trimers when compared with the dimer lies in the fact that trimer formation from an existing dimer can be synergic due to the additional increase in the total number of interactions. For example, a nonsynergic trimer would be formed by the addition of an aromatic residue to an existing dimer interacting with only one of the dimer members, so addition of 1 residue results in only 1 new aromatic−aromatic interaction (as in the Lad trimers). On the other hand, if the incoming residue interacts with two residues, the number of interaction increases from 1 to 3, producing a synergistic effect in the total interaction energy, as in the Sym trimer. To analyze the energetic properties of the aromatic trimers, we computed the interaction energy probability distribution using the effective energy function. The resulting histogram for the effective interaction energy of all analyzed trimers is shown in Figure 4. The resulting energies range from −4.0 to −6.5 kcal/mol. As expected, Sym-type trimers have lower energies with peaks at about −5.5 kcal/mol but a significant population around −6.0 kcal/mol. On the other hand, Lad-type trimers have slightly higher energies with a main peak at around −4.8 kcal/mol, which corresponds to the expected value for only two interactions.

The use of a simplified energy function is also supported by the comparison of our results with literature data obtained with higher quality methods. The aromatic pair−pair interaction energies were analyzed with more rigorous approaches by several other groups.[14,18,21,23,24,36] For example, Kollman and colleagues obtained an interaction energy of −2.72 kcal/mol for the T-shape toluene pair in its optimum configuration,[39] consistently with the obtained value for the Lad type. The structures of isolated benzene trimers, tetramers, and even pentamers have also been the subject of theoretical studies.[14,18,21,23,24,36] These works show

Table 1. Distributions for Secondary Structure and Sequence Distance Classifications

classification	% Lad-type	% Sym-type
	Secondary Structure	
HHX	41.66	43.51
EEX	28.71	28.19
other	29.16	28.19
	Sequence Distance	
close	2.02	1.72
S1	10.2	12.07
S2	10.85	10.07
S3	11.51	10.26
S4	11.88	9.57
far	53.52	56.28

several stable structures for the trimer cluster. Interestingly, the minimum energy trimer corresponding to a perfect Sym trimer shows a binding energy of around −5.0 kcal/mol, in good agreement with our results.

Our results indicate that Sym trimers are mainly stabilized because they contain more inter-residue contacts than the Lad trimers. We now examine the trimer composition, the secondary structure preference for residues forming trimers and the sequence order are what gives a detailed analysis of this new motif.

Aromatic Trimer Composition. The analysis of trimer composition shows that Tyr percentage is 31−39%, Phe percentage is 45−50%, and Trp percentage 15−17%, values which are very similar to the natural abundance of these amino acids, showing no particular residue preference. In order to analyze aromatic interaction types in the different clusters, we computed how many dimers are formed with each possible pairs. The corresponding results are: Phe-Tyr, 31%; Phe-Phe, 26.%; Phe-Trp, 15%; Tyr-Tyr, 11%; Tyr-Trp, 9%; and Trp-Trp, 3%. The most abundant interaction is between Phe and Tyr, closely followed by Phe residues. Interestingly, Phe-Trp interactions are more abundant than Tyr-Tyr or Tyr-Trp, probably reflecting the fact that Tyr residues are hydrophilic. Trp-Trp interaction rarely occurs, probably as a consequence of low Trp abundance in proteins.

Secondary Structure Characteristics. Residues involved in structural motifs may belong to specific secondary structure elements. We used DSSP[29] to determine three types of secondary structure elements: H, E, or L. We classified all possible trimers in groups based in these three elements. The complete results are shown in Table 1. Interestingly, secondary structure distribution is similar for both geometrical types. The data show that most trimers have the three residues in helices (HHH type about 17%), and adding together all groups having at least two residues in helix, (HHX, with X meaning any type) represents ≈42% of all trimers. L-type trimers (EEX) on the other hand altogether represent ≈28%. These results clearly show a preference for the presence of aromatic trimers in helical structures. This result is more significant considering the known preference of aromatic residues to be part of β sheets over helices.[40] In this sense, we found that the three aromatic residues studied distribute more evenly among the secondary structure elements when they are part of an aromatic clusters beyond the dimer, in contrast to the β sheet preference when isolated (Table S4, Supporting Information).

Sequence Distance Classification. An interesting and very important question about the sequence distribution of the residues

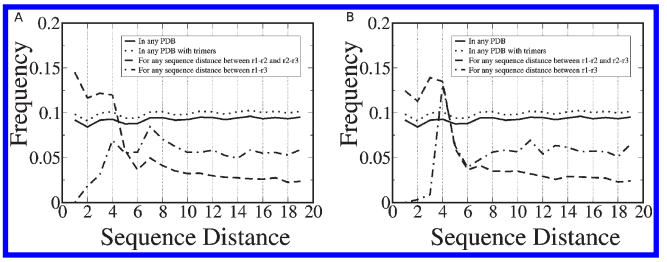


Figure 5. Sequence distance frequency for any pair of aromatic residues found in any protein structure in the PDB (solid line) and any structure containing an aromatic trimer (dotted line). Contiguous aromatic residues in sequence in any aromatic trimer (dashed line) and non-contiguous residues in sequence in any aromatic trimer (dashed dotted line) for Lad-type left panel (A) and Sym-type right panel (B). Contiguous aromatic residues in sequence are defined as two residues belonging to the same cluster that lack any aromatic residue between them belonging to the same cluster.

forming aromatic trimers is whether or not their distribution is random. To analyze this issue, we computed sequence distance probability functions as described in the Computational Methods Section. Figure 5A shows the corresponding sequence distance probability distribution for any pair of aromatic residues found in Sym and Lad trimers, together with that for any randomly chosen pair of aromatic residues found in all of the PDB structures that we analyzed (i.e., completely random pairs, which includes both interacting and non interacting pairs of aromatic residues) (black line). The results for any randomly selected pairs of aromatic residues show that the probability is the same at all distances, meaning that given an aromatic residue is in position 0, the position of any other aromatic residue is completely random. Analysis of the results for the dimers (Figure S3, Supporting Information) shows a clear preference for aromatic residues that are going to form a dimer to be close in sequence; with significant probability of finding aromatic residues in dimers at sequence distances of 1−4 residues, with a maximum at position 4. The same trend is observed for the sequence probability distributions of residues in the Lad and Sym trimers, considering the pairs s12 and s23. These results probably reflect that aromatic trimers arise from the presence of existing dimers. On the other hand, the s13 probability in the Sym trimer shows a clear and unique preference for position 4. Interestingly, in Lad trimers (Figure 5B), the s13 probability is below the random distribution for close residues, suggesting that many Lad trimers are nonlocal.

Based on the previous analysis we classified the trimers according to the sequence distribution pattern in the following categories: close as those having the three residues less than 5 residues apart, s1−s4, and as those having at least two of the aromatic residues at sequence distance of 1−4 and not belonging to close. Finally, the ones that have all three residues further than 4 residues apart were consider as far. The distribution of the trimers in these categories is shown in Table 1. Again, both Sym- and Lad-types present very similar results. Interestingly more than half of the trimers of both geometrical types fall in the far category, while s1−s4 show and average of 10% each group. The group close represents only about 2% of the total trimers, highlighting the nonlocal characteristic of the trimers as opposed to the dimers, for which the close group is 21%. We have found that trimers are nonlocal in sequence, meaning that they tend to integrate different structural elements probably important for protein stability and folding. In the next section we calculate the cross probabilities among the previous determined characteristics that allows a deeper understanding of this motif.

Table 2. Groups of Aromatic Trimer with Significant Cross Probabilities between Different Classifications

cross prob.	value	cross prob.	value
	Sym Trimers		
HHX-close	2.0	EEX-close	0.0
HHX-*i* + (1,3,4)	1.5−1.7	EEX-*i* + (1,3,4)	0.2
HHX-*i* + 2	0.1	EEX-*i* + 2	2.2
	Lad		
HHX-*i* + (1,3,4)	1.4−1.7	EEX-*i* + (1,3,4)	0.2−0.6
HHX-*i* + 2	0.1	EEX-*i* + 2	2.1

Cross Probabilities. Given the above-described classification, we analyzed if the probability of a trimer being in one group concerning its secondary structure is independent or not on the sequence distance classification. To analyze this, we computed the observed probability of belonging to the two groups simultaneously, as defined in Computational Methods Section, and divided the result by the expected probability assuming that both classifications are independent, for all possible pairs for Sym and Lad trimers. Therefore, a value greater than unity implies that belonging to the two classifications simultaneously is preferred, while a value lower than one suggest that both characteristics tends to be avoided. Table 2 shows only those cases where a significant differences are observed.

Even though not totally unexpected, the results shown in Table 2 provide a cross quality check of the results, as evidenced by the relations between primary and secondary structure classification, which shows clear tendencies for the three secondary structure-type cluster groups. In both Lad- and Sym-type of trimers, the HHX group shows a clear preference for belonging to the s1, s3, or s4 groups, while very rarely the trimers of the HHX group are found in s2. The observed preference probably reflects the fact that if two aromatic residues are in the same helix they have to be either contiguous (s1) or at least at three residues apart (s3) to interact, given that a helix turns in 3.6 residues. On the contrary, trimers belonging to the EEX group show a clear preference for the s2 group over all other groups. This probably reflects the fact that in β sheet side chains alternating residues are oriented in the same direction. The Sym trimers belonging to the close group are mostly found in HHX and never in EEX secondary structures. These results show, as expected, that primary and secondary structure groups of trimers are biased due to intrinsic structural properties of the secondary structure elements. The analysis shown before is useful to classify families of proteins that have aromatic clusters as structural motifs with a relevant role in their function.

We have shown that aromatic trimers are formed in proteins, that their conformations resemble the ones observed in isolated benzene clusters with clear synergistic effects, that there are no preferences in aminoacid composition, and that the secondary structure determines how they interact when they are close in sequence but that mostly in timers are nonlocal in sequence. We now perform an analysis of the location and environment of the trimers.

Analysis of the Clusters Environment. In order to study the relative location inside the protein of each aromatic trimer, we

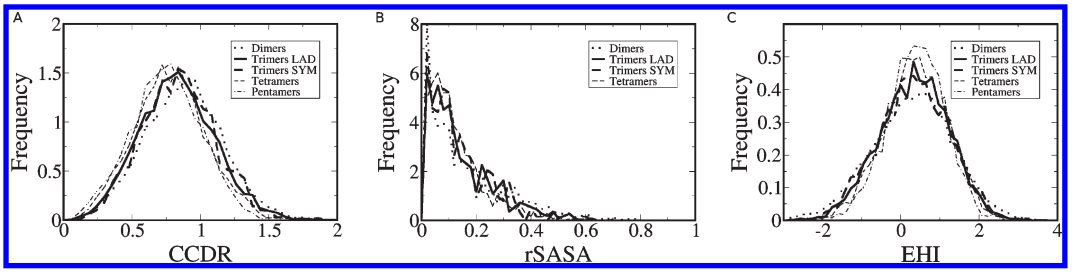


Figure 6. Analysis of the environment of the aromatic clusters. (A) Relation between radius of gyration of each protein and the distance between the center of mass (COM) of each aromatic cluster and the COM of the protein that it belongs; referred to as CCRD. (B) Relation of the SASA for each cluster with respect to the SASA it should have in isolation; referred as relative SASA (rSASA). (C) EHI as defined in Methods Section.

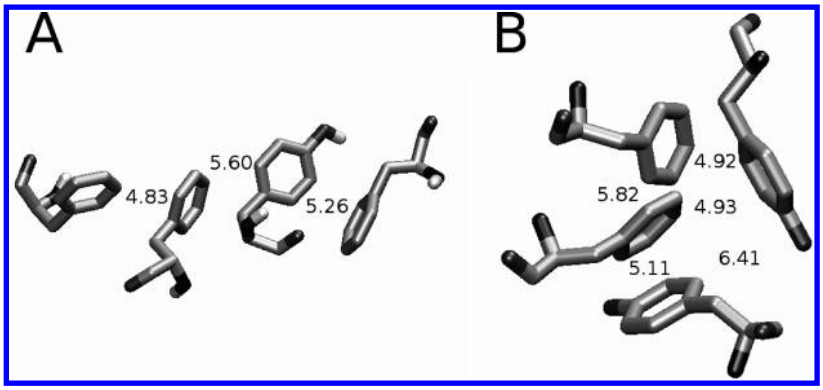


Figure 7. Aromatic tetramer structures. (A) Chain-like (nonsynergic) tetramer with 3 interactions in PDB ID: 1AQ8, residues 64, 99, 124, and 134. (B) High-synergic tetramer structure with 6 interactions PDB ID: 3BPM, residues 64, 132, 72, and 74.

calculated first the radius of gyration of each protein; and second, the distance between the COM of each aromatic trimer to the corresponding protein COM. The relation between these two quantities, referred as trimer protein center relative value (CCRD), gives the idea of the location of each cluster relative to the global protein structure. Accordingly, values close to zero mean that the corresponding clusters are near the protein COM, values close to unity would mean the cluster is located in the interior of the protein, while values greater than one would refer to clusters located close or in the protein surface. The results of Figure 6A shows that most trimers display CCRD values below one, clearly meaning that they tend to be buried forming part of the protein hydrophobic core. Interestingly, the same results are obtained for Lad or Sym trimers. Second, we computed the relation of the solvent accessible surface area (SASA) for each cluster, with respect to the SASA it would have if the trimer were isolated. The values, named relative SASA (rSASA), are shown in Figure 6B for both trimer types. These results show that trimers are deeply buried inside the proteins as almost all of them have values below 0.4, and only a small percentage has values greater than 0.2. In summary, both results show that aromatic trimers, as expected, are buried inside the protein interior possibly forming part of the hydrophobic core of proteins.

Now, in order to analyze the nature of each cluster environment, we computed for each trimer their environment hydropathy index (EHI) as defined in Methods Section and shown in Figure 6C. The EHI measures the average hydropathy of each trimer neighboring residues. Accordingly, an EHI negative value corresponds to highly hydrophilic environments (Arg index is −4.5), while positive values correspond to highly hydrophobic environments (Leu index is 4.5), and the random expected EHI value is, −0.26, computed as the hydropathy index multiplied by the natural abundance of each amino acid in the PDB. The results show that on average the trimer environments tend to be more hydrophobic (the mean value is about 0.3, and the histogram decays faster on the negative side), although the shift is small.

Taken together the analysis of cluster relative position in the global protein structure and the chemical nature of their environment points to a clear preference for aromatic trimers as being part of the hydrophobic cores of the proteins, although no negligible presence of clusters is found closer to the protein surface (i.e., those with CCDR > 1).

In the next section we examine the presence of higher order clusters with a more detailed analysis of the tetramer.

Aromatic Tetramers, Pentamers, and Beyond. *Cluster Identification and Relevant Characteristics.* So far we extensively analyzed the aromatic trimers found in proteins. As mentioned in the Introduction Section, analysis of isolated benzene clusters showed that also tetramers and even pentamers may adopt specific optimal cluster-like structures. Based on these results and our identified trimers, we looked first whether more aromatic residues could be found interacting with at least one residue from an existing trimer and therefore forming a tetramer. Strikingly tetramers are found in 28% of all proteins. Tetramers structure is diverse, but an easy classification, based on the trimer structure, can be performed. Similarly to what is observed for the trimers, most abundant tetramers (57%) correspond to a chain (or Lad-like) cluster, where all residues are piled on each other with a conformation similar to face to edge interactions, with planar and orientational angles close to 90°. These tetramers are formed when an additional aromatic residue is brought top (or bottom) of an existing lad trimer, an example is shown in Figure 7A; 32% of

tetramers shows a structure that is a combination of Sym and a Lad trimers. In these tetramers three residues could be regarded as a Sym trimer, while two of them and the fourth for a Lad-like trimer. More interesting is the Sym-like tetramers, only 10% of tetramers structures shows this compact structure as depicted in Figure 7B. The appearance of conformations similar to Sym and Lad trimers can also be seen if the configurational angle is plotted (Figure S4, Supporting Information). The distribution is similar to that observed for trimers. The analysis of the binding energy clearly follows the structural pattern presenting several peaks which correspond to a Lad tetramer with the lowest overall interaction energy and almost no synergy and the compact tetramer with a total interaction energy showing the highest synergy of all possible tetramers, i.e., 5 interactions with an energy below 10 kcal/mol (see Figure S5, Supporting Information).

To analyze the secondary structure of tetramers, we defined four types: predominantly helical (HHXX), predominantly β (EEXX), predominantly unstructured (X4), and half-helix half-β (EEHH). The results, shown in Table S2 Supporting Information, are very similar to those observed for the trimers; a predominance of helical structures, with over 40%, followed by sheet structures and then only a minor fraction of mixed type tetramers. Finally, we also classified the tetramers according to the sequence distribution of the residues. We grouped them as close (whenever all four sequence distances are smaller than 4 residues apart) and as s2, s3, and s4, when the residues belong to 2, 3, or 4 different sequence regions, defining that 2 residues belong to the same region when they are less than 4 residues apart. Interestingly, most clusters either bring three or even four sequence regions together, showing the nonlocal characteristic of tetramers implying an important role of these clusters in stabilizing tertiary structure. The fact that all classifications yield similar results for the tetramers when compared to the trimers is a strong indicator that tetramers are built upon trimers, and therefore trimers could be considered as the basic aromatic cluster unit.

Pentamers and Higher Order Clusters. In a similar way, as described above when tetramers are formed by the addition of another aromatic residue to an existing trimer, pentamers are formed by the addition of an aromatic residue to an existing tetramer, hexamer by the addition of a residue to a pentamer, and so on. The energy distribution for all aromatic clusters is depicted in Figure S6, Supporting Information. Pentamers are found in about 18% of proteins, hexamers in about 11%, and even heptamers and octamers can be found in about 5% of all proteins. Two examples of big clusters, size 10 and 20, are shown in Figures S7A and S7B, Supporting Information. All these larger size clusters can be subdivided and therefore structurally characterized as trimers of both Lad- and Sym-type. Two important questions concerned us about these larger clusters: (i) The cluster overall structure, if they can display a chain-like structure (like the Lad trimers) or a more compact structure, like the Sym trimers; and (ii) The synergy of the additional residues that is added to the smaller clusters. A nonsynergic addition corresponds to a Lad-like addition, where the new residue forms only one additional interaction. On the other hand, if the new residue is added in a Sym-like fashion, then two additional interactions are formed. To analyze both aspects of larger size clusters, we computed the histogram of each aromatic residue COM distance to the whole cluster COM (Figure S7C, Supporting Information). The results show that for trimers, as expected, two peaks appear corresponding to Sym- and Lad-type of structures. Interestingly, for tetramers and pentamers there is still significant probability of forming tight clusters. The same conclusion is obtained when analyzing the number of interactions. Significant synergistic effect is observed for all clusters, since the number of established interactions is in all cases larger than the minimum number for the corresponding size, implying a clear trend to form tight groups (Table S3, Supporting Information). These results are not unexpected as proteins have finite size, and the probability of having a Lad kind of clusters for bigger clusters sizes is small. In summary, analysis of larger size clusters shows that they are built using the trimer (Sym and Lad) as basic structures and that tight clusters with significant synergy in the number of interactions can be found with up to eight aromatic residues.

We have shown that trimers and tetramers clusters are important motifs in protein structures and that higher order cluster can be identified, even though their classification is difficult due to the small size of proteins relative to the cluster size. We now show if aromatic clusters are found within the aromatic part of cofactors and aromatic residues in proteins.

Cofactor Aromatic Rings As Integral Part of Aromatic Clusters. As many protein structures have been crystallized with cofactors, we also looked at whether aromatic rings of these cofactors, which are integral part of proteins structures, are found establishing aromatic clusters as those described above. We analyzed the aromatic rings of the following cofactors as described in the methodology section: FAD, FMN, MQ7, PLP, NAD, SAM, and TDP. The results (presented in the Supporting Information, Tables S5 and S6) show that in ca. 33% of the protein structures that present a cofactor, its aromatic ring is an integral part of a dimer, being 19% part of a trimer and 11% of a tetramer. As for the trimer clusters, when one of the rings derives from a cofactor, there is a clear preference for Lad-type, probably because cofactors are big and because it is difficult to establish a Sym-type of trimer. Two examples of Sym-type clusters, one trimer for FAD and one tetramer for NAD, are depicted in Figure S8, Supporting Information. The cofactors that are well represented in the PDB are FAD, NAD, FMN, PLP, and SAM, with at least 90 structures. Interestingly, NAD has a high propensity to form dimers, 54% of the total structures that have a NAD, well above the average of 33%. On the other hand FAD and PLP had a lower tendency to form dimers, 19% and 26%, respectively. However, they all have a similar tendency to form trimers and tetramers.

Our results show that cofactor's aromatic rings tend to interact with aromatic residues of proteins and that a relevant subset of them, around 30%, tend to form higher order clusters, probably stabilizing cofactor protein interactions.

Having analyzed the clusters of aromatic residues beyond the dimer, we now depict an interesting example of a protein family that has a conserved trimer, which can be also extended to a tetramer. This example shows how identification of this motive points out possible relevant structure to function relationships.

Calmodulin Case. We focus on calmodulin aromatic trimer interactions.[41] Calmodulin presents conserved aromatic amino acids which have been shown previously to be important for protein stability and ligand recognition.[41] Interestingly, these residues are usually forming trimers according to our definition. The calmodulin protein is shaped like a dumbbell with two lobes that are connected by a seven-turn α helix. Each of the four Ca^{2+}-binding domains in calmodulin consists of the typical EF-hand conformation displaying a helix−loop−helix motif.[42,43] This characteristic fold defines the whole calmodulin-like family as defined by SCOP.[44] Our analysis reveals that Sym-type of trimers of the HHH subgroup appeared in many of these proteins. The

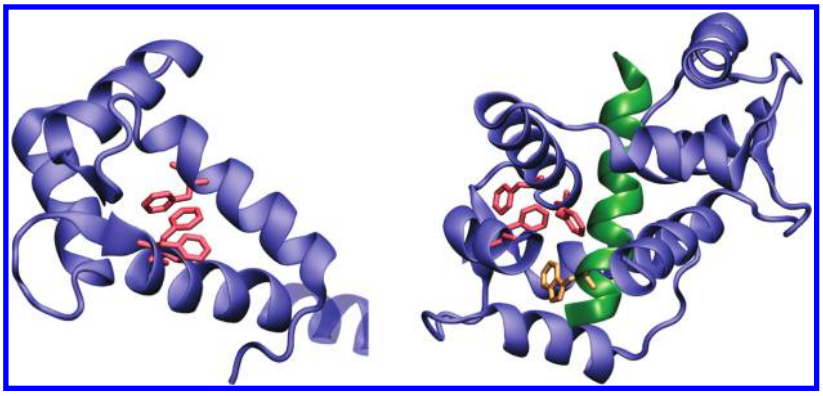


Figure 8. Left panel: One lobe of the calmodulin structure highlighting the presence of an aromatic trimer (red residues), taken from PDB ID: 1ahr. Right panel: Structure of calmodulin with a peptide from MLCK, highlighting the formation of the tetramer (red residues) due to the addition of an aromatic residue from MLCK, shown as green ribbons. Taken from PDB ID: 1cdl.

trimer is formed by two aromatic residues located in one helix and (in postions i and $i + 4$, i.e., the s3 group) and by a third residue located in another helix which is not contiguous to the first, as shown in Figure 8 left panel, clearly stabilizing the fold, consistent with previously mutagenesis studies.[45] Even more interesting, when calmodulin binds peptide targets by joining the two lobes, one aromatic residue from the target adds to the trimer therefore forming a tetramer, as shown in Figure 8 right panel.[46] This example shows that the found motif, a trimer, seems to be key for calmodulin stability and function, possibly explaining the high conservation of these residues in the whole EF-hand family[47] and highlighting the relevance of aromatic clusters beyond the dimer.

DISCUSSION

Aromatic interactions in proteins have been the subject of many studies since the beginning of structure determinations. Back in the 1980s, Thornton and colleagues[48] surveyed 62 structures for aromatic−aromatic interactions and found that the geometry of aromatic pair interactions is essentially random. Our work extends the idea of describing and classifying the aromatic−aromatic interactions beyond the dimer in an extensive and a statistically significant way. Our results reveal that the impact of aromatic trimers in proteins is extremely high with presence in almost half of them and tetramers in one-third of all structures. Careful analysis of the trimer clearly shows that it can be thought as the basic unit of aromatic clusters. Sym trimers are the smallest clusters that have interaction synergy, an important issue that determines the total cluster interaction energy. We have shown that all possible structures of larger clusters and/or chains of aromatic residues can be understood based on the trimer definition. Analysis of secondary structure shows a slight preference for the aromatic residues in clusters as being part of helices and especially as bringing together secondary structure elements that are far apart in sequence. Indeed, sequence distance analysis shows that clusters tend to be clearly nonlocal and bring together 2−4 different protein regions close in space possibly stabilizing tertiary structure, a fact that is further confirmed by the localization of the aromatic clusters as part of protein cores. Our detailed structural analysis of the two basic trimer types, Sym and Lad, also allows their easy identification in protein structures, which with the present data allows estimating their energetic contribution to protein stability and points to possible important roles for the residues forming the cluster. In addition to the identification of clusters of aromatic residues, we performed an analysis of cofactors as they are usually cocrystallized with proteins and are relevant for enzymatic activity. Interestingly, our results show that cofactors also have a high tendency to interact with aromatic residues as part of aromatic clusters beyond the dimer. We believe that that this may help stabilize the interactions of these relevant molecules with proteins.

The fact that aromatic trimers and tetramer structures observed in proteins closely match the optimized structures observed for benzene and toluene clusters in vacuum are very interesting and suggest that protein structure evolved for optimizing the relative position and conformation of the aromatic residues belonging to small aromatic clusters. Last but not least, their widespread occurrence and conserved structure suggests these clusters could be important for protein function either through folding, stability, or protein−protein or protein−ligand interactions, a fact that warrants a closer inspection of these motifs in each particular family, as highlighted for calmodulin-like proteins.

CONCLUSIONS

Our results, based on a detailed analysis of all aromatic clusters found in protein structures shows that: (i) beyond the dimer clusters (trimers and beyond) are found in about half of all unique protein structures; (ii) clusters are built adopting the same trimer motifs (Sym and Lad) found for benzene clusters in vacuum; and (iii) clusters are clearly nonlocal brining primary structure distant sites together.

ASSOCIATED CONTENT

Supporting Information. Additional figures and tables can be found. This information is available free of charge via the Internet at http://pubs.acs.org/.

AUTHOR INFORMATION

Corresponding Author
*E-mail: marcelo@qi.fcen.uba.ar; adrian@qi.fcen.uba.ar.

ACKNOWLEDGMENT

This work was partially supported by grants from Universidad de Buenos Aires 08-X625 to M.A.M., 08-X499 to A.G.T. and

08-X074 to D.A.E., ANPCYT 07-1650 to M.A.M., 06-5203 to A. G.T. and Raices 157 to D.A.E., Conicet PIP 01207 and Guggenheim Foundation grant awarded to D.A.E. D.A.E., A.G.T. and M.A.M. are members of CONICET. Data analysis was performed at the Centro de Bioinformática de Buenos Aires, FCEN, UBA with a homemade database.

REFERENCES

(1) Bashford, D.; Chothia, C.; Lesk, A. M. Determinants of a protein fold. Unique features of the globin amino acid sequences. *J. Mol. Biol.* **1987**, *196* (1), 199–216.

(2) Lesk, A. M.; Fordham, W. D. Conservation and variability in the structures of serine proteinases of the chymotrypsin family. *J. Mol. Biol.* **1996**, *258* (3), 501–537.

(3) Chothia, C.; Gelfand, I.; Kister, A. Structural determinants in the sequences of immunoglobulin variable domain. *J. Mol. Biol.* **1998**, *278* (2), 457–479.

(4) Michnick, S. W.; Shakhnovich, E. A strategy for detecting the conservation of folding-nucleus residues in protein superfamilies. *Folding Des.* **1998**, *3* (4), 239–251.

(5) Larson, S. M.; Davidson, A. R. The identification of conserved interactions within the SH3 domain by alignment of sequences and structures. *Protein Sci.* **2000**, *9* (11), 2170–2180.

(6) Burley, S. K.; Petsko, G. A. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science* **1985**, *229* (4708), 23–28.

(7) Serrano, L.; Bycroft, M.; Fersht, A. R. Aromatic-aromatic interactions and protein stability. Investigation by double-mutant cycles. *J. Mol. Biol.* **1991**, *218* (2), 465–475.

(8) Bhattacharyya, R.; Samanta, U.; Chakrabarti, P. Aromatic-aromatic interactions in and around alpha-helices. *Protein Eng.* **2002**, *15* (2), 91–100.

(9) Aravinda, S.; Shamala, N.; Das, C.; Sriranjini, A.; Karle, I. L.; Balaram, P. Aromatic-aromatic interactions in crystal structures of helical peptide scaffolds containing projecting phenylalanine residues. *J. Am. Chem. Soc.* **2003**, *125* (18), 5308–5315.

(10) Johnson, R. M.; Hecht, K.; Deber, C. M. Aromatic and cation-pi interactions enhance helix-helix association in a membrane environment. *Biochemistry* **2007**, *46* (32), 9208–9214.

(11) Eidenschink, L. A.; Kier, B. L.; Andersen, N. H. Determinants of fold stabilizing aromatic-aromatic interactions in short peptides. *Adv. Exp. Med. Biol.* **2009**, *611*, 73–74.

(12) Espinoza-Fonseca, L. M.; Garcia-Machorro, J. Aromatic-aromatic interactions in the formation of the MDM2-p53 complex. *Biochem. Biophys. Res. Commun.* **2008**, *370* (4), 547–551.

(13) Pereira de Araujo, A. F.; Pochapsky, T. C.; Joughin, B. Thermodynamics of interactions between amino acid side chains: experimental differentiation of aromatic-aromatic, aromatic-aliphatic, and aliphatic-aliphatic side-chain interactions in water. *Biophys. J.* **1999**, *76* (5), 2319–2328.

(14) Chelli, R.; Gervasio, F. L.; Procacci, P.; Schettino, V. Stacking and T-shape competition in aromatic-aromatic amino acid interactions. *J. Am. Chem. Soc.* **2002**, *124* (21), 6133–6143.

(15) Samanta, U.; Pal, D.; Chakrabarti, P. Packing of aromatic rings against tryptophan residues in proteins. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1999**, *55* (Pt 8), 1421–1427.

(16) Mitchell, J. B.; Nandi, C. L.; McDonald, I. K.; Thornton, J. M.; Price, S. L. Amino/aromatic interactions in proteins: is the evidence stacked against hydrogen bonding? *J. Mol. Biol.* **1994**, *239* (2), 315–331.

(17) Kannan, N.; Vishveshwara, S. Aromatic clusters: a determinant of thermal stability of thermophilic proteins. *Protein Eng.* **2000**, *13* (11), 753–761.

(18) Hunter, C. A.; Singh, J.; Thornton, J. M. Pi-pi interactions: the geometry and energetics of phenylalanine-phenylalanine interactions in proteins. *J. Mol. Biol.* **1991**, *218* (4), 837–846.

(19) Easter, D. C.; Terrell, D. A.; Roof, J. A. Monte Carlo studies of isomers, structures, and properties in benzene-cyclohexane clusters: computation strategy and application to the dimer and trimer, $(C6H6)(C6H12)n$, $n = 1-2$. *J. Phys. Chem. A* **2005**, *109* (4), 673–689.

(20) Morimoto, T.; Uno, H.; Furuta, H. Benzene ring trimer interactions modulate supramolecular structures. *Angew. Chem., Int. Ed. Engl.* **2007**, *46* (20), 3672–3675.

(21) Gonzalez, C.; Lim, E. C. Ab initio study of the intermolecular interactions in small benzene clusters: The equilibrium structures of trimer, tetramer, and pentamer. *J. Phys. Chem. A* **2001**, *105* (10), 1904–1908.

(22) Dang, L. X. Molecular dynamics study of benzene-benzene and benzene-potassium ion interactions using polarizable potential models. *J. Chem. Phys.* **2000**, *113* (1), 266–273.

(23) Engkvist, O.; Hobza, P.; Selzle, H. L.; Schlag, E. W. Benzene trimer and benzene tetramer: Structures and properties determined by the nonempirical model (NEMO) potential calibrated from the CCSD-(T) benzene dimer energies. *J. Chem. Phys.* **1999**, *110* (12), 5758–5762.

(24) Krause, H.; Ernstberger, B.; Neusser, H. J. Binding energies of small benzene clusters. *Chem. Phys. Lett.* **1991**, *184* (5−6), 411–417.

(25) De Meijere, A.; Huisken, F. CO2-laser induced photodissociation studies of size-selected small benzene clusters. *J. Chem. Phys.* **1990**, *92* (10), 5826–5834.

(26) Tauer, T. P.; Sherrill, C. D. Beyond the benzene dimer: an investigation of the additivity of $\pi-\pi$ interactions. *J. Phys. Chem. A* **2005**, *109* (46), 10475–10478.

(27) Kouranov, A.; Xie, L.; de la Cruz, J.; Chen, L.; Westbrook, J.; Bourne, P. E.; Berman, H. M. The RCSB PDB information portal for structural genomics. *Nucleic Acids Res.* **2006**, *34* (Database issue), D302–305.

(28) Wu, C. H.; Apweiler, R.; Bairoch, A.; Natale, D. A.; Barker, W. C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; Martin, M. J.; Mazumder, R.; O'Donovan, C.; Redaschi, N.; Suzek, B. The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res.* **2006**, *34* (Database issue), D187–191.

(29) Hooft, R. W.; Sander, C.; Scharf, M.; Vriend, G. The PDBFINDER database: a summary of PDB, DSSP and HSSP information with added value. *Comput Appl Biosci.* **1996**, *12* (6), 525–529.

(30) Clementi, C. Coarse-grained models of protein folding: toy models or predictive tools? *Curr. Opin. Struct. Biol.* **2008**, *18* (1), 10–15.

(31) Camacho, C. J.; Zhang, C. FastContact: Rapid estimate of contact and binding free energies. *Bioinformatics* **2005**, *21* (10), 2534–2536.

(32) Zhang, C.; Vasmatzis, G.; Cornette, J. L.; DeLisi, C. Determination of atomic desolvation energies from the structures of crystallized proteins. *J. Mol. Biol.* **1997**, *267* (3), 707–726.

(33) Miyazawa, S.; Jernigan, R. L. An empirical energy potential with a reference state for protein fold and sequence recognition. *Proteins: Struct., Funct., Genet.* **1999**, *36* (3), 357–369.

(34) Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26* (16), 1668–1688.

(35) Sherrill, C. D.; Takatani, T.; Hohenstein, E. G. An Assessment of Theoretical Methods for Nonbonded Interactions: Comparison to Complete Basis Set Limit Coupled-Cluster Potential Energy Curves for the Benzene Dimer, the Methane Dimer, Benzene-Methane, and Benzene-H(2)S (dagger). *J. Phys. Chem. A* **2009**, *113* (38), 10146–10159.

(36) Hill, J. G.; Platts, J. A.; Werner, H. J. Calculation of intermolecular interactions in the benzene dimer using coupled-cluster and local electron correlation methods. *Phys. Chem. Chem. Phys.* **2006**, *8* (35), 4072–4078.

(37) Fischer, J. D.; Holliday, G. L.; Rahman, S. A.; Thornton, J. M. The Structures and Physicochemical Properties of Organic Cofactors in Biocatalysis. *J. Mol. Biol.* **2010**, *403* (5), 803–824.

(38) Fischer, J. D.; Holliday, G. L.; Thornton, J. M. The CoFactor database: organic cofactors in enzyme catalysis. *Bioinformatics* **2010**, *26* (19), 2496–2497.

(39) Chipot, C.; Jaffe, R.; Maigret, B.; Pearlman, D. A.; Kollman, P. A. Benzene Dimer: A Good Model for $\pi-\pi$ Interactions in Proteins? A Comparison between the Benzene and the Toluene Dimers in the Gas

Phase and in an Aqueous Solution. *J. Am. Chem. Soc.* **1996**, *118* (45), 11217–11224.

(40) Malkov, S. N.; Zivkovic, M. V.; Beljanski, M. V.; Hall, M. B.; Zaric, S. D. A reexamination of the propensities of amino acids towards a particular secondary structure: classification of amino acids based on their chemical structure. *J. Mol. Model.* **2008**, *14* (8), 769–775.

(41) Chin, D.; Means, A. R. Calmodulin: a prototypical calcium sensor. *Trends Cell Biol.* **2000**, *10* (8), 322–328.

(42) Crivici, A.; Ikura, M. Molecular and structural basis of target recognition by calmodulin. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 85–116.

(43) Yap, K. L.; Ames, J. B.; Swindells, M. B.; Ikura, M. Diversity of conformational states and changes within the EF-hand protein superfamily. *Proteins* **1999**, *37* (3), 499–507.

(44) Lo Conte, L.; Ailey, B.; Hubbard, T. J.; Brenner, S. E.; Murzin, A. G.; Chothia, C. SCOP: a structural classification of proteins database. *Nucleic Acids Res.* **2000**, *28* (1), 257–259.

(45) Ohya, Y.; Botstein, D. Structure-based systematic isolation of conditional-lethal mutations in the single yeast calmodulin gene. *Genetics* **1994**, *138* (4), 1041–1054.

(46) Ikura, M.; Clore, G. M.; Gronenborn, A. M.; Zhu, G.; Klee, C. B.; Bax, A. Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science* **1992**, *256* (5057), 632–638.

(47) Sammut, S. J.; Finn, R. D.; Bateman, A. Pfam 10 years on: 10,000 families and still growing. *Briefings Bioinf.* **2008**, *9* (3), 210–219.

(48) Singh, J.; Thornton, J. M. The interaction between phenylalanine rings in proteins. *FEBS Lett.* **1985**, *191* (1), 1–6.