Beta-Strand Interfaces of Non-Dimeric Protein Oligomers Are Characterized by Scattered Charged Residue Patterns

Giovanni Feverati1, Mounia Achoch1,2, Jihad Zrimi1,2, Laurent Vuillon3, Claire Lesieur1,4*

1 Université de Savoie, Annecy le Vieux Cedex, France, 2 Laboratoire de Chimie Bioorganique et Macromoléculaire (LCBM), Faculté des Sciences et Techniques-Guéliz, Université Cadi Ayyad, Marrakech, Maroc, 3 LAMA, Université de Savoie, Le Bourget du Lac, France, 4 AGIM, Université Joseph Fourier, Archamps, France

Abstract

Protein oligomers are formed either permanently, transiently or even by default. The protein chains are associated through intermolecular interactions constituting the protein interface. The protein interfaces of 40 soluble protein oligomers of stoichiometries above two are investigated using a quantitative and qualitative methodology, which analyzes the x-ray structures of the protein oligomers and considers their interfaces as interaction networks. The protein oligomers of the dataset share the same geometry of interface, made by the association of two individual β-strands (β-interfaces), but are otherwise unrelated. The results show that the β-interfaces are made of two interdigitated interaction networks. One of them involves interactions between main chain atoms (backbone network) while the other involves interactions between side chain and backbone atoms or between only side chain atoms (side chain network). Each one has its own characteristics which can be associated to a distinct role. The secondary structure of the β-interfaces is implemented through the backbone networks which are enriched with the hydrophobic amino acids favored in intramolecular β-sheets (MCWIV). The intermolecular specificity is provided by the side chain networks via positioning different types of charged residues at the extremities (arginine) and in the middle (glutamic acid and histidine) of the interface. Such charge distribution helps discriminating between sequences of intermolecular β-strands, of intramolecular β-strands and of β-strands forming β-amyloid fibers. This might open new venues for drug designs and predictive tool developments. Moreover, the β-interfaces of the cholera toxin B subunit interface, when produced individually as synthetic peptides, are capable of inhibiting the assembly of the toxin into pentamers. Thus, their sequences contain the features necessary for a β-interface formation. Such β-strands could be considered as ‘assemblons’, independent associating units, by homology to the foldons (independent folding unit). Such property would be extremely valuable in term of assembly inhibitory drug development.

Citation: Feverati G, Achoch M, Zrimi J, Vuillon L, Lesieur C (2012) Beta-Strand Interfaces of Non-Dimeric Protein Oligomers Are Characterized by Scattered Charged Residue Patterns. PLoS ONE 7(4): e32558. doi:10.1371/journal.pone.0032558

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: lesieur@lapp.in2p3.fr

Introduction

Most proteins are made of more than one polypeptide chain to carry out their biological function [1,2]. They are referred to as protein oligomers and have what is called a quaternary structure. In addition, numerous monomeric proteins associate transiently in binary or in higher stoichiometries (number of chains associated in a protein oligomer) during their life span. The formation of protein oligomer, known as protein assembly, is also a common reaction involving protein oligomers and considers their interfaces as interaction networks. The protein oligomers of the dataset share the same geometry of interface, made by the association of two individual β-strands (β-interfaces), but are otherwise unrelated. The results show that the β-interfaces are made of two interdigitated interaction networks. One of them involves interactions between main chain atoms (backbone network) while the other involves interactions between side chain and backbone atoms or between only side chain atoms (side chain network). Each one has its own characteristics which can be associated to a distinct role. The secondary structure of the β-interfaces is implemented through the backbone networks which are enriched with the hydrophobic amino acids favored in intramolecular β-sheets (MCWIV). The intermolecular specificity is provided by the side chain networks via positioning different types of charged residues at the extremities (arginine) and in the middle (glutamic acid and histidine) of the interface. Such charge distribution helps discriminating between sequences of intermolecular β-strands, of intramolecular β-strands and of β-strands forming β-amyloid fibers. This might open new venues for drug designs and predictive tool developments. Moreover, the β-interfaces of the cholera toxin B subunit interface, when produced individually as synthetic peptides, are capable of inhibiting the assembly of the toxin into pentamers. Thus, their sequences contain the features necessary for a β-interface formation. Such β-strands could be considered as ‘assemblons’, independent associating units, by homology to the foldons (independent folding unit). Such property would be extremely valuable in term of assembly inhibitory drug development.

This is due to the broad diversity of the contact solutions [10,11]. The rationalization of known patterns of protein interfaces is also far from accomplished.

The patterns result from geometrical and chemical complementarities between the two partners. Numerous reports on protein interfaces, based on theoretical and experimental approaches, allow understanding some of the general rules underlying intermolecular contacts (for reviews see [2,10,12]).

First, one needs to distinguish within the interface, the amino acids involved in intermolecular contacts, the so called “hot spots”, from those who are not. Several programs can identify theoretical hot spot residues at interfaces based on: (i) distance cuts-off combined or not with some chemical selection, (ii) solvent accessible surfaces, (iii) geometrical selection (e.g. Voronoi cells) or (iv) evolutionary conserved residues [2,13,14,15]. All require the atomic structure of the protein oligomer. Experimental evidences have also confirmed the presence of hot spot residues in interfaces (for review see [2]). One beautiful example is the selective effect of the mutation of only some of the residues of the interface on the protein assembly of the heptameric co-chaperone cpn10 [16].
Second, the interaction patterns of protein interfaces are related to their secondary and tertiary structures as it was initially described by Sir Francis Crick for α-coiled interfaces with the discovery of the heptad sequences [17,18,19,20,21,22,23,24]. The importance of the structure of the interface in the implementation of a particular motif has been now generalized with high-throughput interaction discovery [25,26].

Third, at the amino acid level, a versatile solution has to be sought rather than a specific one. In fact, even for identical secondary structures, the geometry (triple helix, α-coiled, β-sandwich…) and/or the symmetry of the protein interfaces also affect the patterns at the amino acid levels [11,17,18,20,27,28,29,30].

For a geometry of interface made of interacting β-strands (β-interfaces), dimers are the main stoechiometry studied, particularly when considering dataset analysis [21,31,32,33,34]. Here, we report the analysis of the β-interfaces of 40 soluble protein oligomers whose stoechiometries are from trimers to octamers. We used our tailor made program Gemini to select hot spots and to produce an interaction network -or a graph- of the subset of interactions that composes an interface [15]. Gemini quantitative and qualitative analyses reveal relatively long β-interfaces enriched with charged residues scattered within the interface. More precisely, arginine residues are preferred at N- and C-terminal extremities whereas histidine and glutamic acid residues are more frequent in the middle of the interfaces. Such a broad charge distribution has never been observed previously in dimeric β-interfaces or in intramolecular β-interactions.

Materials and Methods

Interfaces by Gemini

The computer programs (Gemini) relevant to the present paper have been described previously [15]. In summary, Gemini characterizes an interface as a subset of amino acids in interaction, or “hot spots”. They emerge after a purely geometrical analysis of the 3D atomic structure of the protein, well described in the indicated publication. Gemini is equipped with an effective tool (GeminiGraph) that represents interfaces by (bipartite) graphs (Fig. 1). Throughout the paper, the graphs-and so the interfaces-are also referred to as ‘interaction networks’ or simply as ‘networks’. Briefly, the two segments S1 and S2, of an interface are represented by two parallel rows. The interacting amino acids selected by Gemini are indicated by ‘X’ and the non interacting ones by dots ‘.’ (Fig. 1C). The ‘X’ amino acids are the hot spots of the interface. The interactions (I) are illustrated by lines connecting two ‘X’. The version used here includes the name of the amino acids at positions ‘X’, following the one-letter code. In few cases, the β-interface is so intimately close to a different interface geometry that Gemini keeps them together in the same interface region (see Table S2 and Dataset S1). In the present work only the β-interface part has been used; the corresponding graphs have therefore been manually annotated (supplementary material).

A supplementary feature has been added to Gemini, which describes the interfaces as two interaction sub-networks. One of them only includes interactions between backbone atoms (BB sub-network), the other interactions with at least one side chain atom (SC sub-network). The interactions of the BB sub-network (Ibb) are represented with dashed lines whereas those of the SC sub-network (Isc) are represented with solid lines. Xbb and Xsc are the side chain and backbone hot spots, respectively.

Circular proteins

This is also a new addition to Gemini especially relevant to the present work. The goal of this part of the code is to recognize circular homo-oligomers (oligomers made of the same protein chain). The program classifies proteins into two classes: circular homo-oligomers and the rest that can contain hetero-oligomers and non circular homo-oligomers. For short, we call it non-circular (NC). The input information is the three-dimensional structure of PDB. No other database or author’s annotation is used. The first step in the classification recognizes as NC those proteins whose chains are composed of different numbers of residues. Actually, given that in PDB files there can be additional or missing residues, an error of 25% is tolerated on the differences in the number of residues. The remaining proteins are therefore good candidates to be homo-oligomeric. In a second step, the program tries to find the first amino acid common to all the subunits. From it, five other common amino acids must be found, located at 15%, 30% and so on, of the sequence. If this step fails, the protein is NC. If it succeeds, the protein is very likely to be a homo-oligomer so a third step is needed to evaluate the spatial organization of the subunits. This is simply done by comparing the circular homo-oligomers (oligomers made of the same protein chain). The program classifies proteins into two classes: circular homo-oligomers and the rest that can contain hetero-oligomers and non circular homo-oligomers. For short, we call it non-circular (NC). The input information is the three-dimensional structure of PDB. No other database or author’s annotation is used. The first step in the classification recognizes as NC those proteins whose chains are composed of different numbers of residues. Actually, given that in PDB files there can be additional or missing residues, an error of 25% is tolerated on the differences in the number of residues. The remaining proteins are therefore good candidates to be homo-oligomeric. In a second step, the program tries to find the first amino acid common to all the subunits. From it, five other common amino acids must be found, located at 15%, 30% and so on, of the sequence. If this step fails, the protein is NC. If it succeeds, the protein is very likely to be a homo-oligomer so a third step is needed to evaluate the spatial organization of the subunits. This is simply done by comparing the
distances of the Cα of the six common amino acids already found. If
the protein is a circular n-oligomer, there must be n identical
distances (a tolerance of 5 Ångstrom is used) otherwise the protein
is NC. This algorithm is effective in finding circular homo-
oligomers but is not enough to fully discriminate within the NC
class. There are some false negatives, namely proteins that are
circular homo-oligomers but are recognized as NC. This has the
only effect of slightly reducing the size of our dataset. We did not
observe false positives.

Cytoscape (http://www.cytoscape.org/)

It is an open source bioinformatics software platform for
visualizing molecular interaction networks and biological pathways
and integrating these networks with annotations, gene expression
profiles and other data. Although Cytoscape was originally
designed for biological research, now it is a general platform for
complex network analysis and visualization. Among the several
types of interaction data supported, the format SIF (simple
interaction format) was used for the present paper.

RING (Residue Interaction Network Generator)

It is a web server with software for transforming a protein
structure (in PDB format) into a network of interactions. Nodes
represent single amino acids in the protein structure, while the
edges represent the non-covalent bonding interactions that exist
between them [35,36,37]. The interaction network and the edge
attributes are stored in files with the SIF format. These files can
then be easily loaded into CYTOSCAPE to visualize and
manipulate the network [35,36,37]. In the present study, RING
and CYTOSCAPE were used to produce and visualize the
network of hydrogen bonds for the proteins of the dataset.

Statistics

Median, quartile- The median is the value that splits the dataset
into two equally populated subsets (above and below the median).
For example, for 40 cases and a median of 180 amino acids in size,
there are 50% of the cases with a length above 180 and 50% with a
length below 180 amino acids. The quartile is the value at which the
dataset it divided into four parts, equally populated with the 25% of
the samples. The lower separation point is the first quartile, the
middle one is the median and the higher is the third quartile.

Global and Local propensity

The ratio between the amino acid frequency in a domain and the
amino acid frequency in a database is called “global propensity”. If the
global propensity is above 1, the amino acid is “preferred” in the
domain and if the propensity is below 1, the amino acid is
“disfavored” in the domain. The “local propensity” is defined by the
ratio between the amino acid frequency in a particular position (e.g.
corner) of a sub-domain (e.g. β-interface) and its frequency in all the
other positions in the sub-domain. A local propensity above 1 means
the amino acid is preferred in that position than anywhere else in the
sub-domain [38]. On the contrary, a local propensity below 1 means
the amino acid is disfavored in that position compared to elsewhere
in the sub-domain. The corner positions are the amino
acids located at the four outer positions on a segment: two outer
positions on each side of the segment. So each segment has four
amino acids positioned on corners and two outer interactions. The
central positions are anywhere else on the segment.

Secondary-structure prediction

GOR IV software was used to perform the secondary structure
prediction of the segments of the proteins of the dataset. The
secondary structure of each segment of the dataset was predicted
(40×2 cases) considering all the wild-type amino acids of the
segments and not only the ‘X’. Then, a residue was mutated and the
secondary structure prediction was performed again. When a
mutation affected the wild-type original secondary structure
prediction, the mutated residue was considered important for the
secondary structure of the segment. Hydrophobic residues of the BB
or of the SC sub-networks, centrally located or at corners were
mutated to charged residues (e.g. K, D, R, E, H). If one of the
mutations affected the secondary structure prediction, mutation to
other charged amino acids was not essayed. Polar and charged
residues of the BB sub-networks centrally located in the full network,
were also mutated to either polar or hydrophobic residues.

Probability

Let’s call \( p_c \) the probability to find in an interface, a charged
amino acid. We now evaluate \( p_{cc} \), the probability to have at least
one charged amino acid in (at least) one of the corners. This is evaluated as follows:

\[
p_{cc} = 4 \cdot p_c \cdot (1 - p_c)^3 + 6 \cdot p_c^2 \cdot (1 - p_c)^2
\]

\[
+ 4 \cdot p_c^3 \cdot (1 - p_c) + p_c^4 = 1 - (1 - p_c)^4
\]

where each addendum is respectively the probability to find: a
charged amino acid in one corner only, a charged amino acid in
two corners, a charger amino acid in three corners, a charged
amino acid in all corners. Everything holds true for the corner probability
within one of the sub-networks, provided \( p_c \) is the
corresponding probability.

Reagents and buffers

Cholera toxin B pentamer (CtxB5) and all other chemicals were
obtained from Sigma. McIlvaine buffer (0.2 M disodium hydrogen
phosphate, 0.1 M citric acid, pH 7.0), PBS and 0.1 M KCl/HCl
at pH 1.0 were used. All buffers were filtered through sterile
0.22 μm filter before use. Synthetic peptides were ordered from
proteogenix (www.proteogenix.fr).

SDS-PAGE analysis

SDS-PAGE (15% or 12%) were performed with a Bio-Rad
mini-Protean 3 system using the Laemmli method [39]. The gels
were stained with Coomassie blue. 1 μg of sample was loaded
on each lane of the gel.

Reassembly of CtxB into native pentamer

The conditions used for reassembly were adapted from elsewhere
[40]. Briefly, native CxtB5 was acidified in 0.1 M HCl/KCl at
pH 1.0 for 15 min at a final toxin concentration of 86 μM, to induce
the toxin dissociation into monomers (MW ~ 11 600 kDa). The
toxin was subsequently diluted to a final concentration of 8.6 μM, in
McIlvaine buffers at pH 7.0 to promote reassembly. The samples
were incubated for 15 min at 25°C before analysis by SDS-PAGE.
The reassembly into native CtxB pentamer was inferred from SDS-
PAGE analyses since CtxB5 is stable in SDS-containing buffers and
migrates in a gel, run on ice,with an apparent molecular weight
characteristic of the B-subunit pentamer (MW ~ 55 000 kDa). Only
the native pentamer is SDS-resistant. The CtxB concentration for all
experiments refers to the monomeric concentration.

Reassembly of CtxB in presence of peptides

The toxin reassembly was measured in presence of synthetic
peptides whose sequences correspond to the toxin β-interfaces
sequences (segments 1 and 2). The peptides were added in the neutralizing buffer at a molar ratio peptide to protein of 20. The reassembly conditions were identical to the one used for the toxin alone.

**Results**

The primary goal of the analysis is to seek protein interface features within a dataset of protein oligomers sharing only a common geometry of interfaces. This is inspired by the success obtained for α-coiled interfaces [17,18,19]. The second objective is to see if the features can be rationalized in terms of assembly mechanisms. The interfaces are analyzed using our tailor made program Gemini, which considers interfaces as interaction networks and allows both quantitative and qualitative studies [15].

**The dataset**

The dataset was built by screening the Protein DataBank (PDB) [41]. First, cyclic protein oligomers were selected so all the cases had identical symmetry (circular, Cn). To this purpose a program called “Circular” (materials and methods) was made.

In total 502 protein oligomers were identified with stoichiometries from 3 (trimer) to 8 (octamer) (Table 1). Stoichiometries above 8 contained too few cases to be considered. Second, the secondary structure of the protein interface was chosen as two interacting β-strands at least 4 amino acids apart on the β-coiled interfaces [17,18,19]. The second objective is to see if the features can be rationalized in terms of assembly mechanisms. The interfaces are analyzed using our tailor made program Gemini, which considers interfaces as interaction networks and allows both quantitative and qualitative studies [15].

The dataset was built by screening the Protein DataBank (PDB) [41]. First, cyclic protein oligomers were selected so all the cases had identical symmetry (circular, Cn). To this purpose a program called “Circular” (materials and methods) was made. In total 502 protein oligomers were identified with stoichiometries from 3 (trimer) to 8 (octamer) (Table 1). Stoichiometries above 8 contained too few cases to be considered. Second, the secondary structure of the protein interface was chosen as two interacting β-strands at least 4 amino acids apart on the individual chain. The two interacting β-strands had to be different in their amino acid sequences (Fig. 1). Each strand is called a segment. Segment 1 (S1) appears first (N-terminal side) followed by segment 2 (S2) (C-terminal side) on the primary sequence. This geometry is referred to as a β-interface throughout the paper. Third, dimers, hetero-oligomers, transient oligomers, viral and membrane proteins were discarded from the dataset as their interfaces are likely to be differently programmed. After selection, the dataset was made of 40 protein interfaces but the list is non-exhaustive.

**Properties of the whole chain proteins of the dataset**

The protein oligomers are produced by organisms from the three super-kingsdoms of life with 2% of archea, 75% of bacteria and 23% of eukaryotes (Table S1). For comparison, there are 8%, 54% and 38% of archea, bacteria and eukaryotic protein oligomers for the stœchiometries from 3 to 8 in the PDB. The circular trimers are the most represented (67%) against an average of 7±4% for the other stœchiometries (Table 1). The abundance of trimers might be related to the fact that the PDB over-represents low stœchiometries, dimer and trimer in particular, owing to the difficulties in crystallization. The β-interface geometry represents on average 8% of the circular protein oligomers (40/502) in good agreement with a previous measurement in dimers [21].

In summary, the protein oligomers of the dataset are produced by diverse organisms and cover a variety of functions, folds, amino acid lengths and stoichiometries (Table S1). Not surprisingly, the alignment of their amino acid sequences has no worthy of notice homology (not shown). Hence the dataset is characterized by a large heterogeneity.

**Global beta interface characteristics**

Gemini’s interaction networks (or graphs) of the β-interfaces are in Dataset S1. The length and the number of hot spots (N-X) of each β-interface, are determined using the Gemini graphs (materials and methods). Both are counted considering the two segments, S1 and S2, of the interface (Table S2). The statistics on hot spots, interface length and number of interactions are summarized in Table 2. The average length and number of hot spots for the segment S1 or for the segment S2, are similar, indicative of indistinguishable characteristics of the two β-strands of the β-interfaces. The number of interactions between two hot spots (X) involved in the β-interfaces (IXY) is also provided by Gemini (Table S2 and Table 2).

The length, the hot spot number and the interaction number (Ixy) have medians and interquartile ranges fairly similar to their respective average and standard deviation values indicative of a relative homogeneity of these features throughout the dataset (Table 2). Yet there is no visible common topological feature within the graphs of the β-interfaces or any specific chemical composition compared to the whole chains (Table 3). A slightly different chemical composition appears when the hot spots are considered instead of all the amino acids of the two segments S1 and S2 (Table 3). No particular sequence homology was observed upon alignments of the S1 and S2 segments (not shown).

It was then assumed that common features might be somehow diluted in a ‘background’ noise.

As the backbone atoms are identical for the twenty amino acids, it was possible that counting them in the chemical properties of the β-interfaces ‘hid’ some chemical specificity only distinguishable on the side chain atoms. Likewise, only the backbone atoms might

**Table 1. Circular protein oligomers containing a β-interface.**

| Category          | Trimer | Tetramer | Pentamer | Hexamer | Heptamer | Octamer | Total |
|-------------------|--------|----------|----------|---------|----------|---------|-------|
| Circular oligomers| 339    | 39       | 54       | 43      | 22       | 5       | 502   |
| β-interface       | 13     | 6        | 11       | 4       | 4        | 2       | 40    |
| Circular oligomers (%)| 67 (339/502) | 8 | 11 | 9 | 4 | 1 | 100 |

doi:10.1371/journal.pone.0032558.t001
carry topological information. Moreover, previous studies on protein interfaces had indicated the importance of distinguishing main chain (backbone atoms) contacts from side chain contacts [2,43,44].

Accordingly, the graphs of the \( \beta \)-interfaces were partitioned in two sub-graphs, one made of the backbone interactions (one atom of the backbone per segment, BB sub-networks) and one made of the side chain interactions (one atom of the side chain per segment or one atom of the side-chain on a segment and one atom of the backbone on the other segment, SC sub-network). They are shown in supplementary material 1 (Dataset S1). The interactions within the BB sub-networks are illustrated with dashed lines whereas the interactions within the SC sub-networks are illustrated with solid lines (see also materials and methods). It is important to note that the BB and SC sub-graphs can be considered individually (not considering the whole graphs) or within the whole graph. This nuance is important and when the two sub-networks are considered together, we will refer to as the “full” graph or the full network.

Characteristics of the BB sub-networks

The discrimination of the BB and SC sub-networks revealed significant features shared by the \( \beta \)-interfaces. The BB sub-networks appeared characterized by common topological features but not by chemical specificities. First, different patterns of interactions show up in the BB sub-graphs. The first one, which appears in 19 graphs, is referred to as the “ladder” pattern because the BB interactions are running parallel to one another (Fig. 5). The second pattern which appears in 8 graphs is referred to as the “V-shape” pattern because it’s a triplet interaction in the shape of a -V- (Fig. 6). The patterns are defined by elementary interaction blocks. One block “X.X” on one segment interacts with one block “X.X” on the other segment in the ladder pattern. One “X” on one segment interacts with one block “X.X” on the other segment in the V-shape pattern. The elementary blocks appear singly or in multiple copies. Single versions of the ladder pattern appear in 1PVN, 2OJW, 1U1S and 1HX5 and in multiple copies in 1PM4, 1SNR, 1HI9, 1WUR, 2BCM, 2RCF, 2GJV, 2GVH, 2P90, 1J8D, 1WNR, 2RAQ, 1EEI and 1EFI. There are slightly altered versions of the ladder pattern. One graph (1FB1) is made of one block “X.X” on one segment interacting with one block “X X X” on the other segment. Two graphs (2I9D and 2RCF) have one block of “XX” on one segment interacting with one block “XX” on the other segment. Single version of the V-shape pattern can be observed in 2A7R and 2V9U and in multiple copies in 1SJN, 2BAZ, 1L3A, 1NQU, 1OEL and 1Q3S.

There are also 5 graphs made of a mix of ladder and V-shape patterns (1Y13, 2I9D, 2H5X, 3BFO, 2GJV).

The second topological information of the BB sub-networks is the fact that the ladder and the V-shape patterns appear related to the arrangement of the secondary structures of the \( \beta \)-interfaces. Indeed, they are observed mostly in anti-parallel and in parallel intermolecular \( \beta \)-strand interactions, respectively, and the pattern shapes are reminiscent of the anti-parallel and parallel intramolecular main chain hydrogen bond networks found in \( \beta \)-sheets (Figs. 5B & 5C and 6B & 6C). To determine whether Gemini’s BB networks were related to intermolecular hydrogen bonds, the program RING (materials and methods) was used, showing that out of the 100 atoms detected by RING as participating in hydrogen bonds, 98 are Gemini’s backbone atoms. This is likely due to the selection process of Gemini which retains the closest atoms [15]. Gemini detects slightly more backbone atoms and bonds than RING (139 against 100) due to the fact that Gemini is
able to detect the double interactions per amino acids observed in the hydrogen bond network of intramolecular β-sheets (Fig. 5B & 5C and 6B & 6C). Thus, the BB sub-networks describe intermolecular β-sheets. This is confirmed by the observation that the graphs which have no BB interaction (1JN1, 1T0A, 2JCA, 1B09, 2XSC, 1SAC) or only one BB interaction (2BT9 and 2BVC) are not intermolecular β-sheets but are two rather perpendicular interacting β-strands, as can be seen on their respective PDB.

The BB sub-networks (X_{BB}) cannot be distinguished from the whole chains by a specific chemical composition (charged, polar and hydrophobic amino acids). Yet, they are dominated by hydrophobic properties: half of the amino acids of the BB sub-networks are hydrophobic and a third of the interactions are purely hydrophobic (Table 3 and table 4).

The global propensity (materials and methods) of the hydrophobic amino acids of the BB sub-networks was measured to evaluate which hydrophobic amino acids were over-represented in the β-interfaces compared to the whole chains (Table 5). A global propensity above 1 indicates a hydrophobic amino acid “preferred” in the BB sub-networks and on the contrary, a global propensity below 1, indicates a hydrophobic amino acid depleted in the BB sub-networks. Methionine (M), cysteine (C), tryptophane (W), isoleucine (I) and valine (V) are preferred in the BB sub-networks whereas proline (P), alanine (A), glycine (G) and leucine (L) residues are not favored in the BB sub-networks. The phenylalanine is equally present in the BB sub-networks and in the whole chains of the dataset (Global propensity around 1).

Characteristics of the SC sub-networks

In contrast to the BB sub-networks, the SC sub-networks have no topological information but some chemical specificity. In fact the SC sub-networks present an average chemical composition significantly different from the whole chains with a decrease of the percentage of hydrophobic amino acids in favor of an increase of the percentage of charged amino acids (Table 3). The percentage of polar residues remains similar for the SC sub-networks and the whole chains. This observation is even more obvious when the interactions (I_{SC}) are considered instead of the individual amino acids (X_{SC}), as the SC sub-networks have 5 times more purely charged interactions (Ch-Ch) than the BB sub-networks (Table 4). The SC sub-networks also have twice less purely hydrophobic interactions (F-F) than the BB sub-networks (Table 4).

| Interfaces | whole | S1+S2 | S1+S2 ‘X’ | X_{SC} | X_{BB} |
|-----------|-------|-------|-----------|-------|-------|
| Charged   | 24±17 | 24±10 | 28±14     | 30±17 | 23±16 |
| Polar     | 23±15 | 26±14 | 29±16     | 29±17 | 27±24 |
| Hydrophobic | 53±34 | 50±12 | 45±15     | 41±15 | 50±14 |

Table 3. Average chemical composition, in percentage, of the amino acids of the whole chain of the protein dataset, of the two segments of the interface S1+S2 and of the hot spots of S1 and S2. SC and BB stand for side chain and backbone amino acids, respectively.
The global propensity (materials and methods) of the charged residues of the SC sub-networks compared to the whole chains is reported in Table 6. A charged amino acid with a global propensity above 1 is "preferred" in the SC sub-networks whereas a charged amino acid with a propensity below 1 is depleted. Apart from the histidine, which has a global propensity slightly above 1.0, all the charged residues of the SC sub-networks have a global propensity around 1.

The local propensity of the charged amino acids in the SC sub-networks was analyzed considering corner (the four outer SC amino acids) and central (non corner) positions (Table 7 and table 8, respectively). The local propensity (material and methods) is the ratio of the frequency of an amino acid in a particular position (e.g. corner) within a local structure (e.g. the β-interfaces) and of the frequency of the same amino acid in any other position within that local structure [38]. There are almost as much charged amino acids at corners than at central positions (44% in corner positions). But the two positions are made of different types of charged residues. Arginine (R) residues are more frequent at corners (local propensity above 1 in table 7) whereas it is glutamic acid and histidine residues which are favored centrally (local propensity above 1 in table 8). The lysine and aspartic acid residues have no local preferences (local propensity around 1 in both table 7 and table 8).

Comparison of BB and SC sub-networks
There exist several differences between the BB and the SC sub-networks (Table S3). There are $6 \pm 3$ I$_{SC}$ interactions for only $4 \pm 2$ I$_{BB}$ interactions. Additionally, there are $9 \pm 4$ X$_{SC}$ amino acids for only $5 \pm 3$ X$_{BB}$ amino acids. An amino acid with one atom involved in a BB interaction and one atom involved in a SC interaction is counted twice, one per network. But an amino acid having several atoms participating to the same network is counted only once. Thus, on average, the SC sub-network is bigger than the BB sub-network with roughly 60% of the interface amino acids and interactions devoted to it.

When considering the full graphs, it appears that the BB sub-networks are depleted of interactions and of hot spots at corners having only two graphs with two I$_{BB}$ in the outer positions (1NQU and 2Z9H) and only 11 with one I$_{BB}$ in the outer position (1Y13, 2BCM, 1PVN, 2A7R, 2H5X, 3BFO, 1EFL, 2QJW, 1U1S, 1WNR AND 1Q3S). In contrast, 28 graphs have two SC interactions in the outer positions and 39 (out of 40) have at least one. Likewise, the SC sub-networks are depleted of interactions and of hot spots.

### Table 4. Chemical composition of the interactions (amino acid i of segment 1 with amino acid j of segment 2 or vice-versa, data are added together) in the SC and in the BB (bracket) networks of the β-interaces.

| Chemical properties | Charged | Polar | Hydrophobic |
|---------------------|---------|-------|-------------|
| Charged             | 17% (35%) | 10% (9%) | 16% (30%) |
| Polar               | 13% (13%) | 25% (21%) |            |
| Hydrophobic         | 18% (23%) |       | 16% (30%) |

doi:10.1371/journal.pone.0032558.t004
at central positions. There are 86 ISC centrally located for a total of 240 ISC (36%) and 143 XSC centrally located for a total of 374 XSC (38%). In the BB sub-networks, there are 86 IBB centrally located for a total of 156 IBB (55%) and 131 XBB centrally located for a total of 219 XBB (60%). This means that in a typical arrangement, the SC sub-network spatially contains and surrounds the BB one.

Consequently, the corners of the SC sub-networks are enriched with charged residues (32 graphs out of 40, 80%) while those of the BB sub-networks are depleted (10 graphs out 34: 29%). Similarly, the BB sub-networks are enriched centrally with hydrophobic residues (72 central hydrophobic residues for 110 in total: 65%) while the SC sub-networks are depleted (41 central hydrophobic residues for 101 in total: 41%).

Hence, the relative position of the sub-networks provides enrichment (or depletion) of a chemical property without having to vary the absolute number of amino acids of that property in the sub-networks. For example, there are 110 and 101 hydrophobic residues in the BB and SC sub-networks, respectively. Also, the probabilities of finding a charged residue in the corner of the SC or of the BB sub-networks, based on their respective chemical properties (Table 3), are indeed very similar 76% and 65%, respectively (materials and methods). Yet by positioning the XBB centrally, the charged XSC appear more frequently at corners.

Rationalization of the BB and SC features

Once common features are identified within the β-interfaces of the dataset, the next question is: can those features be rationalized in term of protein assembly or interface formation?

The first argument in that direction, is the weight of the β-interactions (Table S2). Iβ are the interactions involved in the β-interface region of the protein oligomers of the dataset. Now, the total number of intermolecular interactions (Iiα) in a whole chain is the number of interactions in all the interface regions. Iiα is provided by Gemini. The average number of intermolecular interactions (Iiα) per chain is the total number of interactions (Iiα) divided by the number of interface regions. The weight of the β-interactions is measured by the ratio -Iβ/Iiα- which gives the amount of interactions in a β-interface compared to the average number of interactions in the whole chains. On average, there are twice more interactions in the β-interfaces than in the whole interface (1.8±0.6). The high number of interactions due the beta geometry is consistent with a role of the β-interfaces in the assembly mechanism.

The data indicate that the BB sub-networks are related to the secondary structures of the interfaces and that they are enriched in hydrophobic residues and hydrophobic interactions. In order to test the involvement of the hydrophobic residues in the secondary structure of the interface, the effect of their mutation on secondary structure prediction was investigated.

The secondary structure of the segments (S1 and S2) with the wild-type (WT) sequence was predicted using GOR IV and compared to the prediction of the same segment after a point mutation of one hydrophobic residue. The mutation of centrally located hydrophobic residues to a charged residue (e.g. K, D, R, E, H) altered the secondary-structure prediction in 83% of the cases. The mutation of hydrophobic residues located at corners to

| Hydrophobic | Number in the BB sub-networks | Percentage in the BB sub-networks | Number in the Whole chains | Percentage in the whole chains | Global propensity |
|-------------|-------------------------------|----------------------------------|-----------------------------|-------------------------------|------------------|
| I           | 26                            | 0.20                             | 577                         | 0.13                          | 1.6              |
| L           | 17                            | 0.13                             | 700                         | 0.16                          | 0.8              |
| V           | 27                            | 0.21                             | 714                         | 0.16                          | 1.3              |
| A           | 13                            | 0.10                             | 756                         | 0.17                          | 0.6              |
| C           | 4                             | 0.03                             | 95                          | 0.02                          | 1.5              |
| M           | 11                            | 0.09                             | 180                         | 0.04                          | 2.1              |
| F           | 9                             | 0.07                             | 295                         | 0.07                          | 1.1              |
| G           | 15                            | 0.12                             | 714                         | 0.16                          | 0.7              |
| P           | 4                             | 0.03                             | 351                         | 0.08                          | 0.4              |
| W           | 3                             | 0.02                             | 82                          | 0.02                          | 1.3              |
| Total       | 129                           |                                   | 4464                        |                               |                  |

| Charged | Number in the SC sub-networks | Percentage in the SC sub-networks | Number in the whole chains | Percentage in the whole chains | Global propensity |
|---------|-------------------------------|----------------------------------|-----------------------------|-------------------------------|------------------|
| R       | 19                            | 0.16                             | 403                         | 0.18                          | 0.9              |
| E       | 31                            | 0.27                             | 584                         | 0.27                          | 1.0              |
| K       | 26                            | 0.22                             | 497                         | 0.23                          | 1.0              |
| D       | 24                            | 0.21                             | 507                         | 0.24                          | 0.9              |
| H       | 13                            | 0.11                             | 188                         | 0.09                          | 1.3              |
| Total   | 113                           |                                   | 2179                        |                               |                  |

doi:10.1371/journal.pone.0032558.t005
doi:10.1371/journal.pone.0032558.t006
charged residue, also disturbed the secondary-structure prediction but to a much lesser extent (44% of the cases). In the same way, the mutation of polar or of charged residues of the BB sub-networks centrally located, to hydrophobic, charge or polar amino acids affected the secondary-structure prediction in only 44% of the cases.

We then measured the local propensity of the hydrophobic residues located centrally in the BB sub-networks and affecting the 2D structure prediction (Table 9). It appears that among the secondary-influencing hydrophobic residues centrally located, the valine (V) and the phenylalanine (F) are preferred (local propensity above 1). The leucine (L), the isoleucine and the methionine (M) appear neutral in the central position (local propensity around 1). Tryptophan (W), proline (P), glycine (G), alanine (A) and cysteine (C) are not favored (local propensity below 1).

The local propensity results were tested using secondary-structure prediction again. Mutations of central hydrophobic amino acids of the BB sub-networks to hydrophobic amino acids which have a local propensity above 1 were expected to have a secondary-structure prediction identical to the wild-type one. This is referred to as the amino acid having a positive versatility (act as wild-type amino acid). On the contrary, mutations to amino acid with a local propensity below 1 were expected to alter the wild-type secondary-structure prediction. These amino acids are referred to as having a negative versatility. In total 331 mutations-predictions have been performed and on average 69% behave as expected (229/331). Both the versatilities are giving similar results with 67% (116/172) of the mutations to amino acids of positive versatility not affecting the secondary structure and 71% (113/159) of the mutations to amino acids of negative versatility affecting it.

This is consistent with the involvement of the features of the BB sub-networks in the secondary structure formation of the β-interfaces.

The SC sub-networks have no topological information and therefore cannot be related to geometrical features. But they have enrichment in charged residues and more precisely a specific distribution of the type of charges along the interface. This suggests a chemical role of the SC sub-networks in the formation of the β-interfaces, via electrostatic interactions.

We have seen that the local positions of the hydrophobic and of the charged residues of the BB and SC sub-networks were connected to the relative position of the two sub-networks. Now, remarkably for the 11 graphs which have one outer BB interaction, 7 have one charged BB residue at a corner. Following the same drift, the graphs with a low content of SC interactions but made of a majority of BB interactions have a charged BB residue in a corner in 44% of the case (7 out 16 graphs) whereas this occurs only in 12% of the graphs made of a minority of BB interactions (3/24).

So even if having a charged residue in a corner appears a trademark of the SC sub-networks, a corner charged residue is maintained via the BB sub-networks if necessary. This looks like a compensatory or a substitutive mechanism.

A similar phenomenon can be observed for the hydrophobic property of the graphs. On average twice more SC hydrophobic residues are located centrally (1,1 central SC hydrophobic) in graphs made of a minority of BB interactions than in graphs made of a majority of BB interactions (0,45 central SC hydrophobic). More precisely, the number of centrally located hydrophobic residues is maintained at a value of 2,8 ± 0,6 across the dataset with 2,2 ± 0,5 of them affecting the secondary structure predictions (Fig. 7). This value is kept constant using either BB or SC residues, or a balance of both. The mutation of the centrally located hydrophobic residues of the SC sub-networks to charged residue affects the secondary prediction in 83% of the case, as for the BB sub-networks. Thus the regulation of the secondary structure

| Charged | Number in the corner position | Percentage in the corner position | Number in the SC sub-networks | Percentage in the SC sub-networks | Local propensity |
|---------|------------------------------|-----------------------------------|-------------------------------|-----------------------------------|-----------------|
| R       | 12                           | 0.24                              | 19                            | 0.17                              | 1.4             |
| E       | 11                           | 0.22                              | 26                            | 0.23                              | 1.0             |
| K       | 12                           | 0.24                              | 26                            | 0.23                              | 1.0             |
| D       | 11                           | 0.22                              | 24                            | 0.21                              | 1.0             |
| H       | 4                            | 0.08                              | 13                            | 0.12                              | 0.7             |
| Total   | 51                           |                                   | 113                           |                                   |                 |

doi:10.1371/journal.pone.0032558.t007

| Charged | Number in a NOT corner position | Percentage in a NOT corner position | Number in the SC sub-networks | Percentage in the SC sub-networks | Local propensity |
|---------|---------------------------------|-------------------------------------|-------------------------------|-----------------------------------|-----------------|
| R       | 7                               | 0.11                                | 19                            | 0.17                              | 0.7             |
| E       | 20                              | 0.32                                | 31                            | 0.27                              | 1.2             |
| K       | 14                              | 0.22                                | 26                            | 0.23                              | 1.0             |
| D       | 13                              | 0.21                                | 24                            | 0.21                              | 1.0             |
| H       | 9                               | 0.14                                | 13                            | 0.12                              | 1.2             |
| Total   | 64                              |                                     | 113                           |                                   |                 |

doi:10.1371/journal.pone.0032558.t008
through hydrophobic amino acids located centrally is organized by the BB sub-networks in most cases. But the BB sub-networks can be substituted by the SC sub-networks as an alternative. Such compensatory or substitutive phenomenon is also in favor of the features being involved in the formation of the interface.

No distinction between the stoichiometries was found for any of the properties of the β-interfaces (not shown).

**Autonomous β-interface segments**

As mentioned earlier, the features describing the β-interfaces are rather homogeneous compared to the heterogeneity observed for their whole chains. In addition, it seems possible to associate the β-interface features to geometrical and chemical properties. This hinted the possibility that the β-interfaces had some autonomous capacity to associate in absence of the whole chain. This was further supported by the narrow distribution of the β-interface lengths and by the absence of proportion between the lengths of the β-interface and the length of their respective whole chain (Fig. 8). To test that possibility, a simple experiment was carried out using the pentamer of the cholera toxin B (CtxB) as a prototype of the β-interfaces (Fig. 1). Conditions to follow the assembly of the CtxB in vitro had been established previously and are indicated in material and methods [40]. Briefly, the native toxin (Fig. 9, lane 2) is acidified for 15 min at room temperature (RT) to lead to its dissociation into monomers (Fig. 9, lane 3). Subsequently, it is neutralized for 15 min at RT, time during which the reassembly into pentamer takes place (Fig. 9, lane 4). In subsequent experiments, 9mer (P1) or/and 8mer (P2) synthetic peptides with sequences corresponding to S1 (23KIFSYTESL31) and S2 (96IAAISMAN103), respectively, of the wild-type CtxB β-interface were added to the neutralizing buffer. The amounts of CtxB reassembled into pentamer under the different conditions, Table 9.

**Table 9. Local propensity of the central hydrophobic residue of the BB sub-networks affecting the 2D-structure prediction.**

| Hydrophobic | Number in central position | Percentage in the central position | Number in BB sub-networks | Percentage in BB sub-networks | Local propensity |
|-------------|---------------------------|-----------------------------------|---------------------------|--------------------------------|-----------------|
| I           | 11                        | 0.20                              | 26                        | 0.20                           | 1.0             |
| L           | 8                         | 0.15                              | 17                        | 0.13                           | 1.1             |
| V           | 18                        | 0.33                              | 27                        | 0.21                           | 1.6             |
| A           | 3                         | 0.06                              | 13                        | 0.10                           | 0.6             |
| C           | 1                         | 0.02                              | 4                         | 0.03                           | 0.6             |
| M           | 4                         | 0.07                              | 11                        | 0.09                           | 0.9             |
| F           | 5                         | 0.09                              | 9                         | 0.07                           | 1.3             |
| G           | 3                         | 0.06                              | 15                        | 0.12                           | 0.5             |
| P           | 1                         | 0.02                              | 4                         | 0.03                           | 0.6             |
| W           | 0                         | 0.00                              | 3                         | 0.02                           | 0.0             |
| Total       | 54                        | 129                               |                           |                                |                 |

doi:10.1371/journal.pone.0032558.t009

**Figure 7. Central hydrophobic residues and percentage of BB interactions.** The number of hydrophobic amino acids of the BB (I) or of the SC sub-networks (●) located centrally in the full networks are plotted against the percentage of BB interactions.

doi:10.1371/journal.pone.0032558.g007

**Figure 8. Absence of correlation between the lengths of the whole chains and of the β-interfaces.** The length of the β-interface (sum of the amino acids of the two segments) of each protein of the dataset is plotted against the length of its respective whole chain (●, ‘all amino acids’ and ○, ‘X’, respectively). If there was a correlation between the size of the whole chain and the size of its interface or the size of its hot spot numbers, the points would appear on the dashed line.

doi:10.1371/journal.pone.0032558.g008
However, the study entirely focuses on the search of a sequence of a biological function (e.g. active site). Thus the search of a sequence of an interface cannot be done as features are made through strategic positioning of chemical observation of interdigitated networks in which the interface (the secondary structure predictions of the whole length sequences illustration, we have seen that the mutations of the central involved in the assembly of the whole chains of the dataset. As an formation of the CtxB residues on the formation of the secondary structure and on the result of the hydrophobic residues disfavored in the β-interfaces (AGP) are disfavored in intramolecular β-sheets (AGP) [34,45,46,47]. There are some discrepancies for the leucine and phenylalanine residues which are favored in intramolecular β-sheets but disfavored or neutral in the β-interfaces, respectively. Intriguingly, these two amino acids are enriched in amyloid β-fiber (LIF) [33]. The role of hydrophobic forces in interfaces (dimers) was previously reported but not in connection with the geometry of the interface [21,48,49] and for review see [2,12,33].

The hydrophobic amino acids of the BB sub-networks are thus devoid of ‘intermolecular’ specificity since they are shared with intramolecular interactions.

In contrast, the charged amino acids favored in the SC sub-networks present some specificity. First, intra-molecular β-interactions as well as dimeric β-interfaces are rather depleted in charged residues, apart from arginine for the dimeric interfaces ([21,32,33,45,46,50] and for review [2]). On the contrary, in the β-interface side chains, charged residues represent a third of the interfacial amino acids and have only a slight preference for histidine residues. It is interesting that the histidine residue stands out as it is the only amino acid charged under physiological conditions. It is also an amino acid already shown to take part in the assemblies of several protein oligomers [51,52,53]. Second, the β-interfaces of our dataset have an average net charge of ~0.5 which differs from the one required for the formation of amyloid β-fiber (net charge of ±1), another type of β-interface [54,55,56].

The third and most practical information about the charge specificity, resides in the distribution of the charged residues. The arginine residues are frequent at both the corners (N- and C-terminal caps) of the β-interfaces whereas histidine and glutamic acid are favored centrally. Lysine and aspartic acid residues have no preferred position in the β-interfaces.

This is in contrast to parallel intramolecular β-sheet in which positively charged residues (KR) are located at the N-terminal extremities only and negatively charged residues (DE) are present at the C-terminal extremities only [47]. The presence of charges at the N- or C-terminal extremities is believed to act as β-breakers [45,47]. Additionally, the formation of amyloid β-fiber is promoted with positively charged residues (KR) located at the N-terminal extremities of the amyloid β-strands and negatively charged residues (DE) at both the N- or C-terminal extremities [54,55]. Finally, charged residues centrally located are observed in intra-molecular edge β-strands and are thought to prevent their aggregation [34]. Hence, the scattered distribution observed on the β-interfaces differentiates them from other types of intramolecular and intermolecular dimeric β-interactions (Fig. 10).

Altogether the data lead us to propose some hypothesis on the construction mechanism of the β-interfaces following two principles: (i) interfaces are built via geometrical and chemical recognition of the interacting domains and (ii) there are a recognition phase (‘binding’) and a stabilization phase. The BB sub-networks, via the hydrophobic residues, could provide the geometrical recognition whereas the side chain charged residues could provide the chemical one. It is tempting to speculate that the long arginine residue located at the extremities is employed as a
hook to promote encounter. The central smaller histidine and glutamic acid residues could act as clips to stabilize the interface. Alternatively, they might, as proposed for the β-edge strands, maintain the two domains soluble prior the recognition.

Some experimental data are consistent with a relation between Gemini’s hotspot residues and their involvement in the process of a β-interface formation. For example, the heat labile enterotoxin B (LTB₅) and the cholera toxin B (CtxB₅) pentamers, which shares 84% sequence identity and almost superimposable x-ray structures, have nevertheless different assembly mechanisms and different β-interface graphs (1EFI and 1EEI, respectively). The two toxin pentamers have only 14 different amino acids and one of them is in the β-interface (Leu 25 and Phe 25 in 1EFI and 1EEI, respectively). Residue 25 is involved in a IBB in both graphs but leucine and phenylalanine have been measured with different global propensities (Table 5). There are 6 IBB for 4 ISC in LTB₅ leucine and phenylalanine have been measured with different tually with partially folded CtxB chains capable of associating more ‘chemically’-regulated assembly also observed experimentally since only folded LTB chains associate [57]. On the other hand, there are 5 ISC for 5 ISC in CtxB₅ consistent with a more ‘chemically’-regulated assembly also observed experimentally with partially folded CtxB chains capable of associating [40,52]. The presence of a ISC, involving a lysine residue only in CtxB₅ (K23-N103) also supports a more ‘chemically’-regulated assembly. Similarly, shiga-like toxin I and II have different stabilities and different graphs (2XSC and not shown) [58]. In the bacterial hexameric (1U1S) from Pseudomonas aeruginosa, the mutation of His 57, to alanine (Ala) or to threonine (Thr) destabilizes the hexamer by disturbing the side chain hydrogen bond network of the His 57 with the side chains of Lys 56 and Ile 59 of the adjacent chain [59]. The His 57 side chain hydrogen bond network is properly seen on the Gemini graph of the β-interface of Hfq (Dataset 1, 1U1S). Disappearance of that network (or changes of that network) for mutant Ala 57 (or for mutant Thr 57) is also seen properly on the Gemini graphs of the mutated Hfq (not shown). Moreover, the conserved main chain hydrogen bond network made of the residues Met 53 and Tyr 55 of chain M with the residues Val 62 and Ser 60 of the adjacent chain is also identified by Gemini (not shown) [60]. However, cautious is necessary with interpreting the graph features. At this stage, they should be used as a tool to formulate hypotheses for experimental tests.

There are several arguments, mentioned in the result section, supporting the idea that the β-interfaces are independent assembly unit. The most indicative one is the experimental observation that the CtxB β-interface peptides recognize the CtxB individual chains. Such peptides could be called “assemblons” by homology to the foldons [61,62]. Some peptides have been found to lead to the trimerization of proteins when genetically added to their sequence, supporting the ‘assemblons’ concept [63,64,65].

**Supporting Information**

**Dataset S1** Gemini Graphs of the 40 β-interfaces. Each graph appears on a separate page. The stereochemistry and the PDB code of the concerned protein oligomer is indicated on the box in the left hand side of the image. The amino acid number is indicated with the type of amino acid at position X. Segments 1 and 2 appear on two parallel rows. X indicates amino acids involved in atomic interactions according to Gemini. SC and BB interactions are illustrated by solid and dashed lines, respectively [15]. The graphs which interfaces have been annotated manually are indicated with a straight line above the segments. A top (left) and a side view (right) of the x-ray structure of the protein oligomer is shown above its respective graph. (PDF)

**Table S1** Features of the protein oligomers of the dataset. (DOC)

**Table S2** Features of the β-interfaces. (DOC)

**Table S3** Properties of the two sub-graphs. (DOC)

**Author Contributions**

Conceived and designed the experiments: CL GF LV. Performed the experiments: MA JZ. Analyzed the data: CL. Contributed reagents/materials/analysis tools: CL GF JZ LV. Wrote the paper: CL.

**References**

1. Goodsell DS, Olson AJ (2000) Structural symmetry and protein function. Annu Rev Biophys Biomol Struct 29: 103–153.
2. Janin J, Bahadur RP, Chakrabarti P (2008) Protein-protein interaction and quaternary structure. Q Rev Biophys 41: 133–180.
3. Iacovache I, van der Goot FG, Pernot L (2008) Pore formation: An ancient yet complex form of attack. Biochim Biophys Acta.
4. LeGay C, Vescov-Semjen B, Abrahim L, Fivaz M, Gioux van der Goot F (1997) Membrane insertion: The strategies of toxins (review). Mol Membr Biol 14: 45–64.
5. Kirkland J, Dibat DB (2002) Paradigm shifts in Alzheimer’s disease and other neurodegenerative disorders: the emerging role of oligomeric assemblies. J Neurosci Res 69: 567–577.
6. Soto C (2003) Unfolding the role of protein misfolding in neurodegenerative diseases. Nature Reviews Neuroscience 4: 49–60.
7. Klein W, Stine W (2004) Small assemblies of unmodified amyloid [beta]-protein are the proximate neurotoxin in Alzheimer’s disease. Neurobiology of Aging 25: 569–580.
8. Harrison RS, Sharpe PC, Singh Y, Fairlie DP (2007) Amyloid peptides and proteins in review. Rev Physiol Biochem Pharmacol 159: 1–77.
9. Miller Y, Ma B, Nussinov R (2010) Polymorphism in Alzheimer A amyloid organization reflects conformational selection in a rugged energy landscape. Chemical reviews.
10. Larsen TA, Olson AJ, Goodsell DS (1998) Morphology of protein-protein interfaces. Structure 6: 421–427.
11. Gruminger D, Treiber N, Ziegler MOP, Koetter JWA, Schulze MS, et al. (2008) Designed protein-protein association. Science 319: 206.
12. Tuncbag N, Kar G, Keskin O, Gursoy A, Nussinov R (2009) A survey of available tools and web servers for analysis of protein-protein interactions and interfaces. Briefings in Bioinformatics 10: 217.
