Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins

Sebastian Pechmann[a], Emmanuel D. Levy[b], Gian Gaetano Tartaglia[a], and Michele Vendruscolo[a,1]

[a]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom; and [b]Laboratory of Molecular Biology, Medical Research Council, Hills Road, Cambridge CB2 0QH, United Kingdom

Edited by Michael Levitt, Stanford University School of Medicine, Stanford, CA, and approved April 16, 2009 (received for review December 9, 2008)

To maintain protein homeostasis, a variety of quality control mechanisms, such as the unfolded protein response and the heat shock response, enable proteins to fold and to assemble into functional complexes while avoiding the formation of aberrant and potentially harmful aggregates. We show here that a complementary contribution to the regulation of the interactions between proteins is provided by the physicochemical properties of their amino acid sequences. The results of a systematic analysis of the protein–protein complexes in the Protein Data Bank (PDB) show that interface regions are more prone to aggregate than other surface regions, indicating that many of the interactions that promote the formation of functional complexes, including hydrophobic and electrostatic forces, can potentially also cause abnormal intermolecular association. We also show, however, that aggregation-prone interfaces are prevented from triggering uncontrolled assembly by being stabilized into their functional conformations by disulfide bonds and salt bridges. These results indicate that functional and dysfunctional association of proteins are promoted by similar forces but also that they are closely regulated by the presence of specific interactions that stabilize native states.

protein aggregation | protein complexes | protein interfaces | physicochemical properties

The controlled association of proteins into functional complexes is essential to perform the myriad biochemical processes required to maintain homeostasis and promote development in living cells (1–3). By contrast, aberrant assembly can lead to misfolding and aggregation, which are phenomena associated with a variety of severe human neurodegenerative and systemic conditions, including Alzheimer's and Parkinson's diseases, type II diabetes, and dialysis-related amyloidosis (4). It is therefore important to elucidate the principles that enable proteins to form highly specific functional complexes while avoiding misassembly.

Through systematic studies of protein–protein interfaces, it has been established that size, shape, and physicochemical complementarities are key determinants of complex formation (5–9). Remarkably, hydrophobic and electrostatic interactions, which are major factors that stabilize protein–protein interfaces (5–7, 10), have also been identified as the main driving forces for protein aggregation (11–15). Indeed, it has been suggested that it is possible to define an intrinsic propensity for aggregation of amino acid sequences based on their physicochemical properties and that this propensity makes it possible to characterize in detail the aggregation behavior of proteins (11–15). Because protein aggregation puts the quality-control system under severe strain and may lead to cell death, there is a strong evolutionary pressure to avoid it (16). Specific interactions that prevent aberrant assembly, which are often described as negative design principles (17, 18), have been suggested to help β-sheet proteins avoid edge-to-edge aggregation (17) and promote solubility (19). Similarly, negative selection has been found to be a powerful evolutionary mechanism for optimizing specificity in protein interactions (18, 20).

We address here the problem of understanding the close relationship between complex formation and protein aggregation: Although these 2 processes have dramatically different effects on cell viability, they are promoted by similar interactions. This observation has prompted us to identify the specific principles that ought to be present to avoid dysfunctional aggregation and to promote normal protein–protein association. Our results indicate that aggregation propensities are higher at interfaces of protein complexes than at other solvent-exposed surfaces. Indeed, we found that the aggregation propensity is more effective than hydrophobicity at identifying protein–protein interfaces. To explain why these aggregation-prone surfaces do not trigger uncontrolled assembly we characterize some of the specific interactions that are used to regulate the behavior of proteins by favoring protein complex formation over protein aggregation.

Results

Aggregation Propensity as a Driving Force for Macromolecular Assembly. The aggregation propensity profile of a protein (see *Methods*) is defined according to the physicochemical properties of its amino acid sequence, including hydrophobicity, electrostatic charge, secondary structure propensities and the presence of "hydrophobic patterns" formed by regions of the sequence with alternating polar and nonpolar residues (11, 12, 15, 21). Aggregation propensities are normalized to have a zero average and a unitary standard deviation (14, 15, 21); thus, positive peaks in the aggregation propensity profiles indicate regions that promote aggregation, whereas negative peaks indicate regions that tend to prevent aggregation.

Here, we first considered whether a significant difference exists between the aggregation propensities at interfaces and surfaces. Intrinsic aggregation propensity profiles were calculated from the protein sequences for a nonredundant set of 475 homodimers, 237 heterodimers, and 85 homotrimers from the 3DComplex database (www.3DComplex.org) (22, 23). Our analysis revealed a clear difference in the distribution of aggregation-prone regions between interfaces and surfaces of homomeric dimers (Fig. 1*A*, red line). We found that such regions are more likely to appear at interfaces, as indicated by the fact that the probability distribution $P(D)$ of the difference

Author contributions: S.P., E.D.L., G.G.T., and M.V. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: mv245@cam.ac.uk

This article contains supporting information online at www.pnas.org/cgi/content/full/0812414106/DCSupplemental.

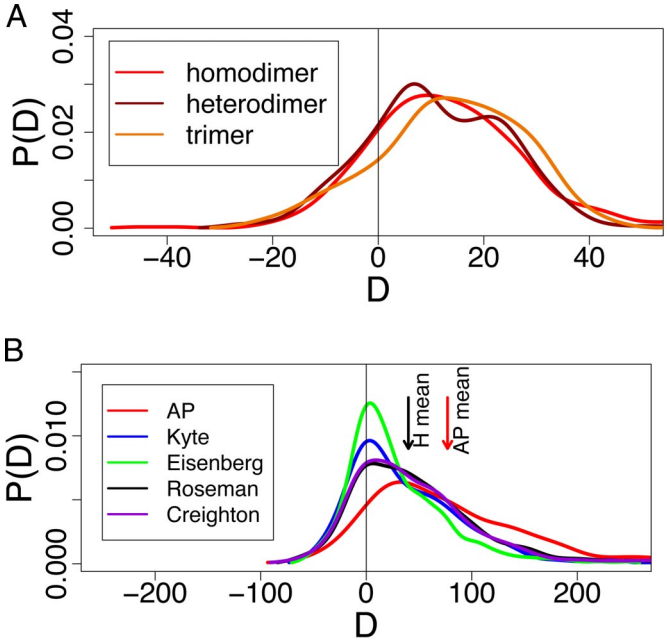

Fig. 1. Aggregation propensities of interfaces. (*A*) Interfaces are more aggregation prone than surfaces. We demonstrate this result by analyzing the distribution $P(D)$ of the difference D between the aggregation propensity Z_i^{agg} (see *Methods*) at interfaces and at surfaces; positive values of D indicate a higher aggregation propensity at interfaces. Results are shown for 475 homodimers (red line, mean: 13.1; 3rd quantile: 21.6); 237 heterodimers (purple line, mean:11.1; 3rd quantile: 21.2), and 85 cyclic trimeric protein complexes (orange line, mean: 14.3; 3rd quantile: 23.5). (*B*) Interfaces are identified more effectively through their surface aggregation propensity S_i^{agg} (see *Methods*) than through their surface hydrophobicity S_i^{hyd} (see *Methods*). This result is obtained by comparing the distribution $P(D)$ of the difference D between the surface aggregation propensity at interfaces and surfaces with the analogous distributions of the difference D between the surface hydrophobicity at interfaces and surfaces; positive values of D indicate a higher aggregation propensity (or a higher hydrophobicity, respectively) at interfaces. The D values for the aggregation propensity are more shifted toward positive values (AP, mean: 77.1; 3rd quantile: 119.4) than the D values for the hydrophobicity [Kyte and Doolittle (ref. 24) (39.1, 68.3), Eisenberg et al. (ref. 25) (29.7, 52.3), Roseman (ref. 26) (46.8, 74.4), and Creighton (ref. 27) (44.7, 73.1)]; arrows indicate the average values of hydrophobicity (black) and aggregation propensity (red).

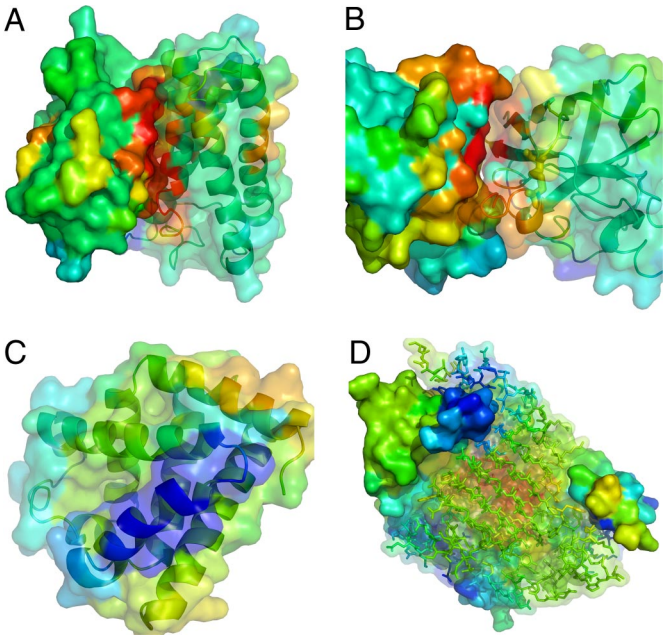

Fig. 2. Representative examples of surface aggregation propensities. The aggregation-prone portions of the surface are shown in red and aggregation-resistant portions in blue in a gradient coloring method from high to low surface aggregation propensity. (*A*) Aggregation-prone interface of a mainly-α homodimer complex (PDB ID code 1BBH). (*B*) Aggregation-prone interface of a mainly-β homodimer complex (PDB ID code 1XSO). (*C*) Aggregation-prone interface of a cyclic trimeric protein complex (1PDB ID code 1KRR). (*D*) Aggregation propensity surface of the aggregation-resistant and monomeric human myoglobin protein (2PDB ID code 2MM1).

D between the aggregation propensity Z_i^{agg} (see *Methods*) at surfaces and at interfaces is not centered around zero. We repeated this analysis for a range of choices of the value of the minimum solvent-accessible surface area (SASA) above which a residue is considered to be surface-exposed [supporting information (SI) Fig. S1]. The results are shown in Fig. 1*A* for the 3 major interface types: isologous interfaces from homodimers (Fig. 1*A*, red line), heterologous interfaces from heterodimers (Fig. 1*A*, purple line), and heterologous interfaces from homotrimers (Fig. 1*A*, orange line). We observed a similar behavior for all 3 interfaces types across our dataset that spans a wide range of interfaces sizes.

These results thus suggest that the presence of surfaces with large aggregation propensities may promote protein–protein interactions resulting in the formation of both functional complexes and aberrant aggregates.

Surface Aggregation Propensities. Aggregation propensity profiles provide an estimate for the different regions of a polypeptide chain to form intermolecular interactions (12, 14, 15, 21). In globular proteins under native conditions, most of these aggregation-prone regions are buried in the structural core because their exposure on the surface can reduce solubility and cause aggregation (15). As we have seen, however, aggregation-prone regions also play a role in the functional association of proteins. To facilitate the analysis of the solvent-exposed aggregation-prone regions, we project the aggregation propensity profiles on the protein surfaces.

In defining the surface aggregation propensity scores, S_i^{agg} (see *Methods*), we take into account the fact that the sizes of protein–protein interfaces and of binding pockets vary quite substantially (22). To have an approach applicable to both transient complexes with small interfaces and stable complexes with large interfaces, we select a fairly large surface patch size of A = 1,000 Å^2 but introduce a distance weighting function so that neighboring residues are contributing more to the local aggregation propensity than more distant ones. This score for a given residue on the surface is defined as the sum of the aggregation propensities of its solvent-exposed neighbors weighted by their distance (see *Methods*).

The analysis of surface aggregation propensities illustrates that interfaces tend to be aggregation prone, whereas surfaces tend to be aggregation resistant (Fig. 2). The analysis discussed above of the differences between the surface aggregation propensity scores of surfaces and interfaces confirms that such scores are highly effective in identifying aggregation-prone interfaces (Fig. 3).

Comparison of Aggregation Propensity and Hydrophobicity Profiles. Hydrophobicity is one of the best indicators for identifying protein–protein interfaces (5–7, 10), and hydrophobic forces are also among the major determinants of protein aggregation (11–15).

Because hydrophobicity profiles and aggregation propensity profiles are significantly different form each other (see Table

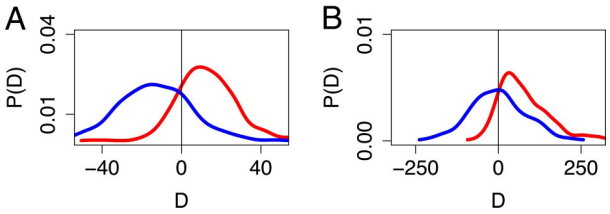

Fig. 3. Aggregation-prone and aggregation-resistant regions at surfaces and interfaces. (*A*) Analysis of aggregation propensity profiles at protein interfaces. We first considered the distribution *P*(*D*) (red line) of the difference *D* of the aggregation propensity at interfaces and at surfaces for positive Z_i^{agg} scores (see *Methods*); the result indicates that interfaces contain more aggregation-prone regions than surfaces. To complement these findings, we then considered the distribution *P*(*D*) (blue line) calculated for negative Z_i^{agg} scores; in this case the results indicate that interfaces contain also more aggregation-resistant regions than surfaces. (*B*) Analysis of surface aggregation propensity scores, S_i^{agg} (see *Methods*), at protein interfaces. We repeated the analysis presented in *A* for surface aggregation propensities; in this case we found that interfaces contain more aggregation-prone portions than surfaces (red line) but also that surfaces contain a similar number of aggregation-resistant portions as interfaces.

S1), we investigated whether aggregation propensity profiles are more effective than hydrophobicity profiles in identifying the protein–protein interfaces; we calculated surface hydrophobicity scores, S_i^{hyd} (see *Methods*), using 4 of the most commonly used hydrophobicity scales [Kyte and Doolittle (24), Eisenberg (25), Roseman (26), and Creighton (27)]. The difference *D* between the surface hydrophobicity at interfaces and surfaces (Fig. 1*B*) was calculated for different hydrophobicity scales in the same way as for the surface aggregation propensity S_i^{agg} (see *Methods*). Our results suggest that whereas surface hydrophobicity scores can be used to distinguish between interfaces and surfaces (Fig. 1*B*), surface aggregation propensity scores are better at discriminating between the 2 (Fig. 1*B*). These results are supported by the fact that the distribution of the values of *D* for the surface aggregation propensity scores is more shifted toward positive values than the corresponding distribution for surface hydrophobicity scores (Fig. 1*B*).

Gatekeepers Residues on Surfaces. It has been suggested that charged residues positioned along the amino acid sequences of proteins in proximity of aggregation-prone regions can prevent aggregation; for this role, these charged residues have been referred to as "gatekeepers" (15, 28, 29).

Protein surfaces, including interfaces, tend to contain more charged amino acids than structural cores, and it is well known that electrostatic forces provide specificity in protein–protein interactions (10, 30). Here, instead of searching for gatekeepers along the amino acid chain, we carried out an analysis of the positions of charged residues in proximity of hydrophobic patterns across the interfaces; we considered surface hydrophobic patterns to be present if hydrophobic residues were found within a 4-Å radius from hydrophilic ones, either along the sequence or through space, and vice versa for hydrophilic residues. We found that such patterns at interfaces were, in all cases, in proximity of charged residues.

As discussed above, interfaces tend to exhibit both the most aggregation-resistant and the most aggregation-prone residues in the entire surfaces of proteins. Charges tend to reduce the aggregation propensities, and negative aggregation propensity scores are mostly found at the rim of interfaces, whereas the core tends to be more aggregation prone; the interface rim is defined here as formed by the residues that bury <25 Å^2 upon binding.

We illustrate this effect on the structure of a T cell receptor Vα homodimer (PDB ID code 1AC6, Fig. 4). The side chains of residues with negative aggregation propensity scores are contributing only weakly to the interactions stabilizing the interface. Instead, they are shielding the aggregation-prone interface to the outside from competing or unwanted binding partners. For comparison, we note that for the different hydrophobicity scales, this region has weakly positive scores but no peaks (Fig. 4, blue line).

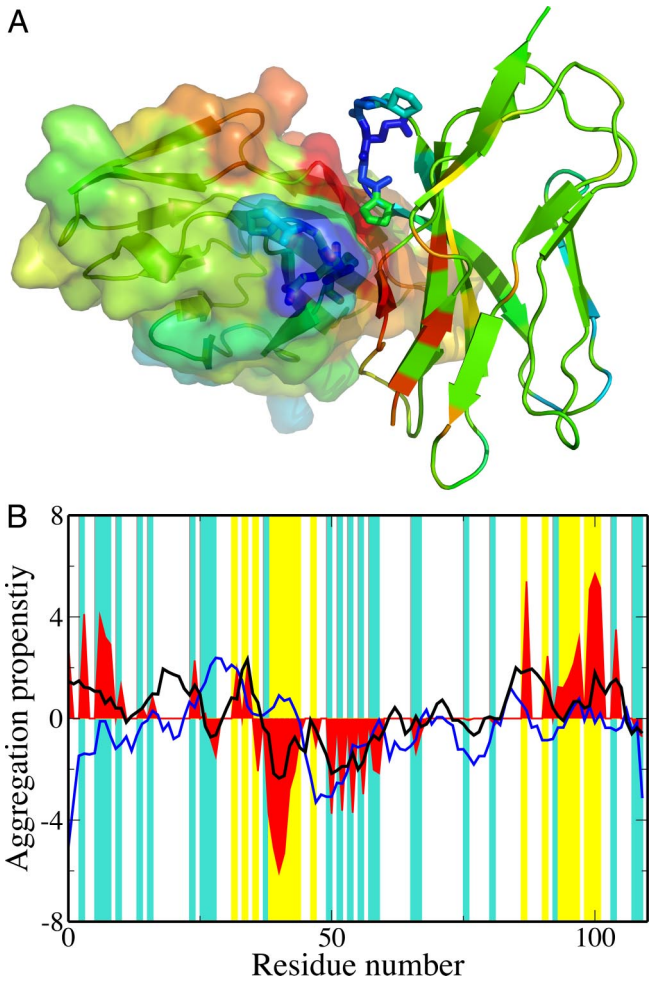

Fig. 4. Example of a complex (the T cell receptor Vα homodimer, PDB ID code 1AC6) that exhibits an aggregation-prone interface flanked by an aggregation-resistant region. (*A*) The aggregation-resistant loop at the interface is shown in blue; the loop is formed by 2 proline residues, 2 glycine residues, and a glutamine residue with the charged side chain pointing away from the aggregation-prone interface, which is shown in red. (*B*) Comparison of the intrinsic aggregation propensity profile (black), the surface aggregation propensity profile (red), and the Creighton hydrophobicity profile (blue). Interface regions are denoted as yellow bands, and surface regions as blue bands.

Our results thus indicate that the presence of aggregation-prone surface regions promote the formation of interfaces, but also that aggregation-resistant surface regions are present at interfaces to create a tradeoff between stability and specificity (Fig. 3) (31).

Specific Interactions That Prevent Aberrant Protein Assembly. Globular proteins under native conditions are usually protected against aggregation because the most aggregation-prone regions are buried within the structural cores (15). Solvent-exposed aggregation-prone regions, when present, can lead to the formation of either complexes or aggregates. As a result, complex formation and aberrant assembly appear to be determined by similar physicochemical properties.

To keep these competing processes under strict regulation, one expects specific interactions to be present to stabilize the native states and disfavor misfolding and aggregation (17). A recent study of the aggregation process of β-lactoglobulin, which is normally found as a dimer, has identified one of these interactions, which promotes dimerization while avoiding the formation of potentially harmful aggregates (32). The interface of the β-lactoglobulin dimer is formed by the most amyloidogenic β-strand of the protein, which is stabilized in the native state by a disulfide bond preventing its unfolding. Therefore, this highly aggregation-prone region promotes dimerization but is unable to trigger further intermolecular association. This result is supported by the finding that by reducing the disulfide bond, β-lactoglobulin readily forms aggregates (32). Taken together, these results indicate that the dimerization of β-lactoglobulin is promoted by a combination of positive and negative design principles, namely by the presence of an aggregation-prone surface that is stabilized in its native state by the formation of a disulfide bond.

It has been suggested that protein aggregates, especially those in form of amyloid fibrils, are stabilized by a network of hydrogen bonds (33). Specific interactions against aggregation should thus be of a strength comparable with or greater than that of hydrogen bonds. Interactions that can play this role are salt bridges, which are ionic interactions between acidic and basic side chains, and disulfide bonds, which are covalent bonds between cysteine residues. We carried out a systematic analysis of these potentially protective interactions. In the dataset of homodimers that we used (see *Methods*), 85% of complexes containing disulfide bonds and 74% of complexes containing salt bridges have one or more stabilizing interactions located at, or close to, the interface.

We assessed whether disulfide bonds are preferentially found in proximity of interfaces by comparing the ratio of the number M_o of observed disulfide bonds at the interface and the number M_p of all potential disulfide bonds at the interface to the ratio of the number N_o of all observed disulfide bonds in the whole protein and the total number N_p of possible sites for disulfide bonds

$$\frac{M_o}{M_p} > \frac{N_o}{N_p}. \quad [1]$$

We applied a similar definition to assess whether salt bridges are found preferentially near interfaces (see below).

We found a clear preference for both disulfide bonds and salt bridges to be near interfaces (Fig. 5 and Table 1). For the disulfide bonds, the large majority of the points lie above the diagonal in Fig. 5*A*. For salt bridges we also found (Fig. 5*B*) that for the majority of protein–protein complexes there is a clear preference for them to be at interfaces, although the bias is weaker than in the case of disulfide bonds (Fig. 5*A*). Because salt bridges are weaker than disulfide bonds, their use as specific interactions against aggregation can indeed be expected to be less effective. A disulfide bond (or a salt bridge) is considered to be in proximity of an interface if at least 1 of the 2 residues forming the bond itself is at <6 residues away along the sequence from residues at this interface; we assessed in this way whether a disulfide bond (or salt bridge) can be capable of stabilizing an interface because the persistence length along a polypeptide chain is ≈6 residues (34). The distributions shown in Fig. 5 are robust against changes in the cut-off distance; in this case the number of proteins with disulfide bonds (or salt bridges) at the interface changes, but the overall trend remains the same.

In this section, we have shown that 440 of 475 (92%) homodimer complexes have either disulfide bonds or salt bridges in proximity of their interfaces, with the great majority of these directly flanking aggregation-prone regions. We thus suggest that the presence of disulfide bonds and salt bridges can protect aggregation-prone interfaces from aggregation.

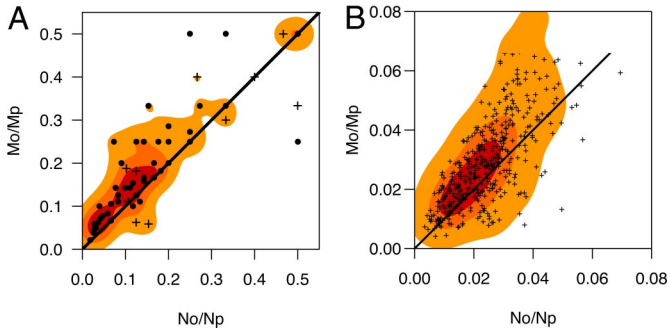

Fig. 5. Disulfide bonds and salt bridges are found preferentially in the proximity of interfaces in protein homodimer complexes. For the homodimer complexes in the 3DComplex database, the probability of finding disulfide bonds (*A*) or salt bridges (*B*) at the interface (M_o/M_p, see text) is, on average, greater than the probability of finding them anywhere in the proteins (N_o/N_p, see text), as demonstrated by the fact that the majority of the points lie above the diagonal; density estimates of the distributions describing 25%, 50%, and 90% of the data are shown as contour maps with increasing brightness. (*A*) Preference of disulfide bonds to be at interfaces for the 62 homodimers that have one more disulfide bond at the interface ($P < 2.5 \cdot 10^{-3}$); the 10 complexes with the highest interface-to-surface ratio are marked as crosses. (*B*) Preference of salt bridges to be at interfaces ($P < 2.3 \cdot 10^{-16}$).

Discussion

Protein–protein interfaces have been studied intensively to shed light on the principles determining macromolecular assembly and recognition (5–9, 35). To understand how proteins can have interfaces that are attractive enough to find their specific binding partners in the crowded cellular environment while at the same time maintaining sufficient specificity to avoid aggregation, we investigated the distribution of aggregation propensities across protein complex interfaces and surfaces. An analysis of the aggregation propensities at surfaces of protein complexes has revealed that interfaces tend to be more aggregation prone than other surface regions.

To explain how solvent-exposed aggregation-prone regions do not actually promote dysfunctional aggregation, we have suggested that specific protective interactions regulate the competition between protein aggregation and protein complex formation. We have found that charged residues act as gatekeepers to disrupt hydrophobic patterns at interfaces, and regions of negative aggregation propensity on the interface rim can protect the aggregation-prone core. We have also shown that 440 of 475 homodimer complexes (92%) have at least a disulfide bond or a salt bridges in proximity of their interface, with the great majority of these flanking aggregation-prone

Table 1. Distribution of disulfide bonds and salt bridges at protein complex interfaces

Type	*N*	N_{ss}	I_{salt}, %	I_{ss}, %
Homodimers	475	78	74	85
Heterodimers	237	122	83	91
Cyclic trimers	85	8	86	100

N is the total number of complexes in the database for each type considered (homodimers, heterodimers. and cyclic trimers). N_{ss} is the number of complexes with disulphide bonds. "I" denotes the percentage of complexes in which salt bridges (salt) or disulphide bonds (ss), if present, are at interfaces.

regions. The enrichment of these protective interactions close to the interface thus helps prevent aggregation.

Conclusions

By analyzing a large set of protein complexes encompassing a wide variety of different types of interfaces, we have identified a series of interactions that promote the formation of functional complexes and the avoidance of uncontrolled aggregation. By characterizing aggregation propensities in terms of specific physicochemical properties of amino acid sequences, including hydrophobicity, electrostatic charge, and secondary structure propensities (11–15), we have shown that protein interfaces are more aggregation prone than surfaces, thus providing a driving force for the formation of functional complexes. In addition, because these interactions can also potentially give rise to nonspecific aggregation, we have identified disulfide bonds and salt bridges as interactions that can stabilize aggregation-prone interfaces in their native conformations, thus preventing the conformational rearrangements required for misfolding and aggregation. Our results provide insight into the way in which the interactions responsible for the behavior of proteins are finely tuned to promote folding and complex formation rather than misfolding and aggregation.

Methods

Database of Protein Complexes. The set of protein complexes used in this work was compiled by using the 3DComplex database (www.3DComplex.org) (22, 23), from which we omitted designed complexes and complexes with large cofactors at the interfaces, such as S–Fe clusters or heme groups as well as metal ions coordinated across the interface. Homodimers provided a group of isologous interfaces (i.e., where a surface of the protein interacts with itself), homotrimers provided a set of heterologous interfaces (i.e., where a surface of the protein interacts with a different surface on itself), and heterodimers provided another set of heterologous interfaces between different proteins.

Aggregation Propensity Profiles. The aggregation propensity profile of an amino acid sequence is provided by the intrinsic aggregation propensity scores, Z_i^{agg}, for each residue i, which is calculated as a function of its physicochemical properties (including hydrophobicity, electrostatic charge, and secondary structure propensities) by using the Zyggregator method (15, 21).

Aggregation Propensity Scores at Surfaces and Interfaces. For each complex, we constructed a histogram H^{int} for describing the distribution of the Z_i^{agg} scores of residues at interfaces; in a similar manner we constructed a histogram H^{sur} for residues at surfaces. To identify surface and interface residues, we calculated the SASA for each residue with NACCESS (36). Residues with different SASA values in the free and bound states were considered as interface residues, and residues with SASA values $>50\text{Å}^2$ were considered as surface residues.

We then calculated the difference D between interface and surface aggregation propensity in terms of the differences of the 2 histograms H^{int} and H^{sur}

$$D = \sum_k k \cdot (H_k^{\mathrm{int}} - H_k^{\mathrm{sur}}), \qquad [2]$$

where the index k runs over the bins of the histograms. With these definitions, positive values of D indicate that interfaces are more aggregation prone than surfaces, and negative values that interfaces are more aggregation resistant than surfaces. The sum in Eq. **2** is restricted to positive values of Z_i^{agg} because only positive aggregation propensity scores are used for the prediction of absolute aggregation rates (15). We also considered D values with the sum restricted to negative values of Z_i^{agg} to characterize the presence of aggregation-resistant regions at interfaces; in this case, the sign was adjusted so that negative values of D identify interfaces that are more aggregation resistant than surfaces.

Surface Aggregation Propensity Scores. The surface aggregation propensity score for residue i is defined as

$$S_i^{\mathrm{agg}} = \sum_j Z_j^{\mathrm{agg}} d_{ij}, \qquad [3]$$

where d_{ij} is the distance between the Cα atoms of residues i and j, Z_j^{agg} is the aggregation propensity score of residue j (15, 21), and the sum is extended to all of the solvent-exposed residues j within a radius r_A from residue i. The difference D between interface and surface for aggregation propensity surface scores was calculated as in Eq. **2** by using the residue aggregation propensity surface scores S_i^{agg} instead of the intrinsic aggregation propensity score Z_i^{agg}.

Hydrophobicity. Hydrophobicity profiles, Z_i^{hyd}, were obtained by averaging amino acid hydrophobicities, taken from standard hydrophobicity scales (24–27), over 7-residue sliding windows with shorter window sizes at the termini. The sign was adapted so that hydrophobic residues had positive and polar residues negative values. Hydrophobicity scales were first normalized by rescaling the original values so that the resulting profiles had a zero expectation value and a unitary standard deviation, in analogy to the aggregation propensity profiles (15). This property was verified for a large set of random sequences with the amino acid frequencies of the SWISS-PROT database for each hydrophobicity scale (original and normalized values are listed in Table S2). The difference D between interface and surface hydrophobicity was computed as in Eq. **2** for the aggregation propensity, with Z_i^{hyd} replacing Z_i^{agg} in the definitions of H^{int} and H^{sur}. Surface hydrophobicity scores S_i^{hyd} were computed in the same way as surface aggregation propensity scores S_i^{agg} in Eq. **3** by replacing Z_i^{agg} with Z_i^{hyd}.

Disulfide Bonds and Salt Bridges. In the calculations of M_p and N_p in Eq. **1**, we considered residue pairs as potentially capable of forming disulfide bonds (or salt bridges) if their Cβ atoms were at a distance in the range 3.2–4.0 (or 5–10) Å, as calculated from the atomic coordinates in the PDB files. Because the number of salt bridges in proteins scales with the length of amino acid sequences (with a correlation coefficient in the present dataset of 0.79), we have used the length L_p of the protein to approximate an upper limit N_p for the number of salt bridges in a protein, and the number L_i of interface residues to approximate an upper limit M_p for the number of salt bridges at the interface. P values were obtained from standard t tests of the distribution of $(M_o/M_p - N_o/N_p)$.

ACKNOWLEDGMENTS. This work was supported by the Engineering and Physical Sciences Research Council (S.P.), the European Molecular Biology Organization (S.P. and M.V.), the Medical Research Council (MRC) Laboratory of Molecular Biology (E.D.L.), MRC (G.G.T.), the Leverhulme Trust (G.G.T. and M.V.), and the Royal Society (M.V.).

1. Hartwell LH, Hopfield JJ, Leibler S, Murray AW (1999) From molecular to modular cell biology. *Nature* 402:C47–C52.
2. Bhattacharyya RP, Remenyi A, Yeh BJ, Lim WA (2006) Domains, motifs, and scaffolds: The role of modular interactions in the evolution and wiring of cell signaling circuits. *Annu Rev Biochem* 75:655–680.
3. Robinson CV, Sali A, Baumeister W (2007) The molecular sociology of the cell. *Nature* 450:973–982.
4. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid and human disease. *Annu Rev Biochem* 75:333–366.
5. Jones DT, Thornton JM (1996) Principles of protein–protein interactions. *Proc Natl Acad Sci USA* 93:13–20.
6. Jones DT, Thornton JM (1997) Analysis of protein–protein interaction sites using surface patches. *J Mol Biol* 272:121–132.
7. Conte LL, Chothia C, Janin J (1999) The atomic structure of protein–protein recognition sites. *J Mol Biol* 285:2177–2198.
8. Tuncbag N, Gursoy A, Guney E, Nussinov R, Keskin O (2008) Architectures and functional coverage of protein–protein interfaces. *J Mol Biol* 381:785–802.
9. Levy ED, Pereira-Leal JB (2008) Evolution and dynamics of protein interactions and networks. *Curr Opin Struct Biol* 18:349–357.
10. Chothia C, Janin J (1975) Principles of protein–protein recognition. *Nature* 256:705–708.
11. Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM (2003) Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* 424:805–808.
12. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotech* 22:1302–1306.
13. Dubay KF, et al. (2004) Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. *J Mol Biol* 341:1317–1326.
14. Pawar AP, et al. (2005) Prediction of ''aggregation-prone'' and ''aggregation-susceptible regions in proteins associated with neurodegenerative diseases. *J Mol Biol* 350:379–392.

15. Tartaglia GG, et al. (2008) Prediction of aggregation-prone regions of structured proteins. *J Mol Biol* 380:425–436.
16. Tartaglia GG, Pechmann S, Dobson CM, Vendruscolo M (2007) Life on the edge: A link between gene expression levels and aggregation rates of human proteins. *Trends Biochem Sci* 32:204–206.
17. Richardson JS, Richardson DC (2001) Natural β-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc Natl Acad Sci USA* 99:2754–2759.
18. Baker D (2006) Prediction and design of macromolecular structures and interactions. *Phil Trans R Soc Ser B* 361:459–463.
19. Doye JPK, Louis AA, Vendruscolo M (2004) Inhibition of protein crystallization by evolutionary negative design. *Phys Biol* 1:9–13.
20. Zarrinpar A, Park SH, Lim WA (2003) Optimization of specificity in a cellular protein interaction network by negative selection. *Nature* 426:676–680.
21. Tartaglia GG, Vendruscolo M (2008) The Zyggregator method for predicting protein aggregation propensities. *Chem Soc Rev* 37:1395–1401.
22. Levy ED, Pereira-Leal JB, Chothia C, Teichmann SA (2006) 3D complex: A structural classification of protein complexes. *PLoS Comp Biol* 2:1395–1406.
23. Levy ED (2007) PiQSi: Protein quaternary structure investigation. *Structure (London)* 15:1364–1367.
24. Kyte J, Doolittle RF (1982) A simple method for displaying the hydrophatic character of a protein. *J Mol Biol* 157:105–132.
25. Eisenberg D, Schwarz E, Komarony M, Wall R (1984) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J Mol Biol* 179:125–142.
26. Roseman MA (1988) Hydrophilicity of polar amino acid side chains is markedly reduced by flanking peptide bonds. *J Mol Biol* 200:513–522.
27. Creighton TE (1993) *Proteins. Structures and Molecular Properties* (Freeman, New York).
28. Rousseau J, Serrano L, Schymkowitz J (2006) How evolutionary pressure against protein aggregation shaped chaperone specificity. *J Mol Biol* 355:1037–1047.
29. Vendruscolo M, Dobson CM (2007) Chemical biology—More charges against aggregation. *Nature* 449:555.
30. Brock K, Talley K, Coley K, Kundrotas P, Alexovy E (2007) Optimization of electrostatic interactions in protein–protein complexes. *Biophys J* 93:3340–3352.
31. Bolon DN, Grant RA, Baker TA, Sauer RT (2005) Specificity versus stability in computational protein design. *Proc Natl Acad Sci USA* 102:12724–12729.
32. Hamada D, et al. (2009) Competition between folding, native state dimerisation and amyloid aggregation in beta-lactoglobulin. *J Mol Biol* 386:878–890.
33. Knowles TPJ, et al. (2007) Kinetics and thermodynamics of amyloid formation. *Proc Natl Acad Sci USA* 104:10016–10021.
34. Narang P, Bhushan K, Bose S, Jayaram B (2005) A computational pathway for bracketing native-like structures for small α helical globular proteins. *Phys Chem Chem Phys* 7:2364–2375.
35. Janin J, Bahadur RP, Chakrabarti P (2008) 2008) Protein–protein interaction and quaternary structure. *Q Rev Biophys* 41:133–180.
36. Hubbard SJ, Thornton JM (1993) NACCESS computer program. Tech rep, (Department of Biochemistry and Molecular Biology, University College, London).