Pharm Res (2011) 28:949–961
DOI 10.1007/s11095-011-0414-9

PERSPECTIVE

Coupling of Aggregation and Immunogenicity in Biotherapeutics: T- and B-Cell Immune Epitopes May Contain Aggregation-Prone Regions

Sandeep Kumar • Satish K. Singh • Xiaoling Wang • Bonita Rup • Davinder Gill

Received: 15 November 2010 / Accepted: 1 March 2011 / Published online: 25 March 2011

ABSTRACT Biotherapeutics, including recombinant or plasma-derived human proteins and antibody-based molecules, have emerged as an important class of pharmaceuticals. Aggregation and immunogenicity are among the major bottlenecks during discovery and development of biotherapeutics. Computational tools that can predict aggregation prone regions as well as T- and B-cell immune epitopes from protein sequence and structure have become available recently. Here, we describe a potential coupling between aggregation and immunogenicity: T-cell and B-cell immune epitopes in therapeutic proteins may contain aggregation-prone regions. The details of biological mechanisms behind this observation remain to be understood. However, our observation opens up an exciting potential for rational design of de-immunized novel, as well as follow on biotherapeutics with reduced aggregation propensity.

KEY WORDS aggregation · biotherapeutics · cross β motif · drug development · immunogenicity

ABBREVIATIONS

3D	three-dimensional
ADA	anti-drug antibody
APC	antigen-presenting cell
APR	aggregation-prone region
CDR	complementarity-determining region
CMC	chemistry manufacturing and control
Fab	fragment antigen binding
Fc	fragment crystallizable
IgG	immunoglobulin G
Igm	immunoglobulin M
mAb	monoclonal antibody
MHC	major histocompatibility complex
T_H-cell	T-helper cell
Treg-cell	T-regulatory cell

INTRODUCTION

Biotherapeutics, particularly antibody-based therapeutics, bind their targets with high specificity and affinity. Non-mechanism-based toxicity due to off-target binding is also nearly absent for these drugs (1). However, despite more than 200 biologic drug products on the market, there remains significant and possibly increasing concern about aggregation and immunogenicity in the development of biotherapeutics (2–4).

Here, our focus is on aggregation and immunogenicity risk factors for biotherapeutic candidates during the early

S. Kumar (✉) • X. Wang
Biotherapeutics Pharmaceutical Sciences, Pfizer Inc.
MC6S, 575 Maryville Centre Drive
St. Louis, Missouri 63141, USA
e-mail: Sandeep.Kumar@Pfizer.com

S. K. Singh
Biotherapeutics Pharmaceutical Sciences, Pfizer Inc.
BB3D, 700 Chesterfield Parkway West
Chesterfield, Missouri 63017, USA

B. Rup
Protein Bioanalytics, Pfizer, Inc.
G2001A, One Burtt Road
Andover, Massachussets 08100, USA

D. Gill
Global Biotherapeutic Technologies, Pfizer Inc.
W1009, 87 Cambridge Park Drive
Cambridge, Massachussets 02139, USA

discovery stage. At this stage, computational approaches are more readily applicable and can provide insights not available from the experiments alone. We report our initial observations of overlap between aggregation-prone regions (APRs) and determinants of the immune response, i.e., major histocompatibility complex (MHC) class II T-cell epitopes as well as B-cell epitopes. The biological mechanism leading from dosing of the aggregated protein to immunogenicity against the biotherapeutic is not yet fully understood (5). Nevertheless, our observation indicates that a rational structure-based design may enable biotherapeutics that are de-immunized, less aggregation prone, easier to formulate and, therefore, have a greater probability of success in development.

AGGREGATION AND IMMUNOGENICITY ISSUES IN BIOTHERAPEUTICS

Degradation of biotherapeutics is an important chemistry, manufacturing and control (CMC)-related challenge faced by biotechnology companies during all stages of product development, storage, shipping and administration. Drug product degradation may also have immunological consequences. Physico-chemical degradation may result in sequence as well as structural changes. These changes may generate novel epitopes and cause a breakdown of the immune tolerance to the biotherapeutics. In a worst case scenario, the resultant anti-drug antibody (ADA) may cross-react against a non-redundant endogenous protein, leading to serious consequences for the patient. Modifications to the biotherapeutic molecule may either impact T-/B-cell recognition or alter the antigen-processing events that generate the MHC restricted peptides that are presented on the surface of the antigen-presenting cells (APCs) (6). These changes could also occur simultaneously and may mimic the natural auto-immune response to protein aging.

Protein degradation can occur *via* several physico-chemical routes such as deamidation, oxidation, hydrolysis, fragmentation, isomerization and aggregation (7). Among these, aggregation is one of the most common but also the least understood stability-limiting degradation routes for biotherapeutics. Aggregation refers to the self-association of a number of molecules to form dimers, oligomers, and even submicron or micron-sized particles and display a wide range of morphologies ranging from amorphous units to amyloid like fibrils (8). A theoretical rationale behind the various experimental observations on protein aggregates was recently elucidated by Dill and coworkers (9).

The formation of aggregates is also a concern from a safety and efficacy perspective due to their potential to trigger an immunologic reaction in the patient (2,4). Clinical manifestations of the immunologic response, defined as generation of anti-drug antibodies (ADAs), could range from no apparent impact to partial or complete loss of efficacy (4). Loss of efficacy could potentially also manifest itself over time. The pharmacokinetic and pharmacodynamic profiles of the biotherapeutic could be altered. In the case of therapeutic monoclonal antibodies (mAbs), their design has evolved from being of fully murine origin (Muromomab, 1986) to chimeric (Abciximab, 1994), to humanized (Daclizumab, 1997) to fully human (Panitumumab, 2006). Limitation of efficacy due to manifestation of clinical immunogenicity has been a major driver for technology improvements in this important class of biotherapeutics (10–13).

Clinical immunogenicity of a biotherapeutic is determined by a number of factors related to characteristics of both the product and the patient. From the perspective of product development, it is the molecular characteristics of the product that are important and can be controlled (4,7,10,12,14). Even humanized and fully human mAbs may show immunogenicity in the clinic. It is proposed that the residual immunogenicity may arise from their complementarity-determining regions (CDRs) (12). Experimental evidence from studies in animal models (transgenic and otherwise) suggests that aggregation and immunogenicity may be connected (15,16). The clinical evidence is not as unequivocal due to the multi-factorial nature of the human immune response (4,10,17).

AGGREGATION AND PREDICTION OF AGGREGATION-PRONE REGIONS

Protein molecules can self-associate to form aggregates in several ways with or without the need for significant perturbation to their native states, leading to a wide range of aggregate morphologies. Most aggregation prediction algorithms identify protein sequence regions susceptible to the formation of cross β steric zipper motif, and the discussion below is, therefore, restricted to this motif. However, this may not be a significant limitation if aggregates containing repetitive and arrayed structures of the protein are considered a greater risk for triggering immunogenicity (2).

At the molecular level, the physico-chemical forces involved in protein aggregation are the same as those involved in native state protein folding and association. These forces lead to solvent exclusion (hydrophobicity) and shape as well as charge complementarity among the persistently associating partners (Fig. 1). Consistently, aggregation-prone regions (APRs) often overlap with the interfacial regions in functionally active protein-protein complexes (18). Protein aggregation can be seeded by the surface exposed edge β-strands (19) in the native-like conformations or may involve combinations of domain swapping and α-helix/random coil to β-strand transitions

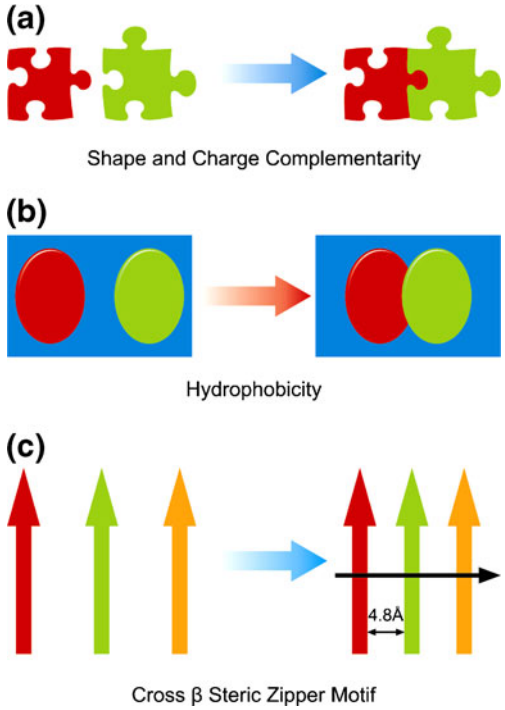

Fig. 1 Schematic illustration of molecular level considerations in protein self-association. (**a**) Proteins may self-associate due to charge and shape complementarity among the associating molecular surface regions. (**b**) Large surface-exposed hydrophobic patches could also drive protein self-association. These considerations are not exclusive of each other and, instead, can work in tandem. A good example is the cross β steric zipper motif. (**c**) Formation of the cross β steric zipper motif by β-strands from different molecules. Aggregates containing this motif can be immunogenic (26,38). In this regard, the amyloid $\beta_{1\text{-}42}$ peptide is a good example of coupling between aggregation and immunogenicity (Fig. 2).

(20). The cross β steric zipper motif, whose molecular structure was determined by Eisenberg lab (20–22), is often implicated in protein aggregation. Although it was first implicated in neurodegenerative diseases, the cross β motif and amyloid fibril-like aggregate formation (Figs. 1c and 2) are almost universal among proteins (23). For example, certain proteins in bacterial inclusion bodies also aggregate *via* this route (24). The cross β motif formation typically involves only a few residues (5–10 amino acids), and overall protein sequence and structures of the proteins need not be related. Recent literature indicates that amyloid-like structures may also have functional roles such as storage of hormones and skin pigmentation (23). A recent data mining exercise uncovered experimental evidence for short sequence regions found in more than eighty different proteins which aggregate *via* formation of cross β steric zipper motif (25). The cross β motif may seed molecular self-association which then may propagate to form fibrils and plaques. In several instances, self-association of proteins does not propagate enough to form fibrils within a reasonable time scale, but the cross β motif can still be detectable by marker dyes, namely, Thioflavin T and Congo Red. Thioflavin T has been shown to bind to expired biotherapeutics (26) and to mAbs under stress conditions (27).

In recent years, a number of APR prediction algorithms have been developed to predict potential APRs in proteins (25). These tools identify short (5–10 residues) regions in protein sequences based on their ability to form the cross β steric zipper motif. We have utilized these tools to predict APRs in amino acid sequences of biotherapeutics. We also mapped the APRs on to the three-dimensional (3D) crystal structures or homology models of the biotherapeutics.

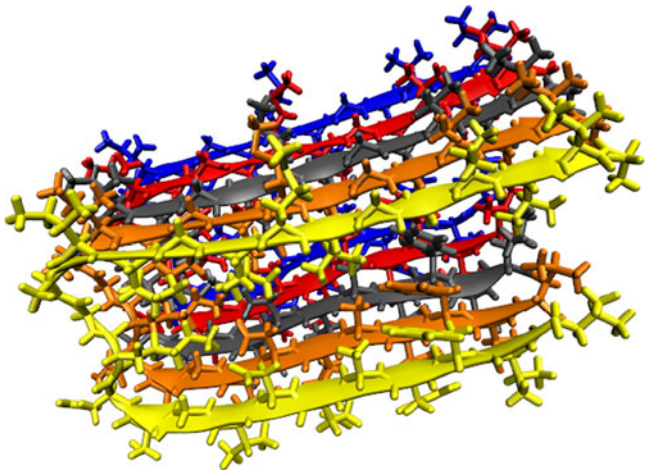

Fig. 2 Structure of the fibril formed by residues 18–42 in amyloid $\beta_{1\text{-}42}$ peptide obtained from the Protein Data Bank (PDB) entry 2BEG (50). The sequence of the amyloid $\beta_{1\text{-}42}$ peptide is 1-DAEFRHDSGYEVHHQK**LVFFA**EDVGSNKG**AIIGLMVGGVVI**A- 42. Soluble aggregates of this peptide are immunogenic and the region 15–42 of this peptide contain three immune epitopes (underlined in the sequence), namely, 15-QKLVFFAEDV-24, 25-GSNKGAIIGL-34 and 35-MVGGVVIA-42 as per the information available in IEDB database (Table II). The Tango/PAGE-predicted APRs for this peptide are 17-LVFFA-21, 30-AIIGLMV-37 and 38-GGVVI-41 (shown in bold letters in the sequence). The portion of the amyloid $\beta_{1\text{-}42}$ peptide shown in this diagram is both aggregation-prone and immunogenic.

Elimination of such an APR from light chain framework 2-CDR2 region of a mAb candidate led to its reduced aggregation tendency and improved solubility (unpublished data). A complementary approach, called spatial aggregation propensity (SAP), identifies surface-exposed hydrophobic patches in therapeutic antibodies (28,29). Variants of mAbs designed to disrupt these surface hydrophobic patches showed reduced aggregation behavior (28). However, the APRs predicted by SAP are not related to the cross β motif.

Therapeutic monoclonal antibody (mAb) sequences often contain predicted APRs (30). In the variable regions of mAbs, such APRs are found in CDR loops and adjoining framework region β-strands. These APRs also contribute significantly towards antigen recognition, thereby linking aggregation with potential loss of antibody function (31). Point mutations to address aggregation can therefore be made with a systematic analysis of their potential impact on antibody function.

An APR represents a region of the protein that may be susceptible and drive self-association when exposed to solvent due to the changes in the native protein sequence or structure. These changes may be caused by the physico-chemical stresses such as denaturation, freeze/thaw, agitation, pH, temperature variations and aging. The same APR may respond to multiple stresses. Hence, mutations aimed at disruption of a single (or a few) APR could potentially mitigate aggregation resulting from several stresses. The number of APRs in a biotherapeutic drug candidate will not necessarily indicate its overall aggregation propensity relative to another molecule. Additional factors such as location of the APRs in the protein structure and the relative tendency of the APRs to drive self-association would also have to be considered. A given molecule with certain APRs or certain number of surface-exposed hydrophobic patches may or may not actually show aggregation under pharmaceutically relevant solution conditions. However, the knowledge of potential APRs enables a rational aggregation mitigation strategy, if experimentally observed solution aggregation is a problem, and the stage of development allows sequence mutations to be considered. It also enables comparison of aggregation behavior of a biotherapeutic candidate with its close variants and rational selection among them.

IMMUNOLOGICAL MECHANISM AND PREDICTION OF IMMUNE EPITOPES

The induction of classical immune response against foreign antigen is broadly understood to be driven by the interaction between B- and T-cells and APCs (Fig. 3). In the context of therapeutic proteins, ADAs are produced by B-cells which are activated when epitopes present on the protein bind specifically to B-cell antigen receptors, and additional signals are obtained from other components of the immune system, typically, T-helper cells and cytokines. In contrast, T-regulatory cells and other cytokines may play a role in diminishing B-cell activation and ADA responses. Activation of B-cells can also occur in a T-cell-independent manner. However, such activation usually requires recognition of repeating epitopes on the antigen and results in a weaker degree of activation than the one which occurs with T-cell help. This is because the T-cell-independent antibody responses tend to result in transient low affinity IgM responses, while T-dependent responses often result in IgM switch to other immunoglobulin isotypes, increased affinity of secreted antibody, and production of memory cells. Analysis of ADA response from clinical studies shows that serious side effects are mainly driven by IgG antibodies, i.e. *via* the T-cell-dependent pathway (4). Both the mechanisms require T-cell and/or B-cell epitopes to be present for recognition. T-cell receptors bind to linear peptide epitopes in a therapeutic protein that have been generated from processing by APCs and presented in a binding site of MHC antigens (MHC). B-cells bind to epitopes on the protein surface with specific 3D conformations.

Generation of ADAs against human therapeutic proteins presents a conundrum in this context. How do such proteins break tolerance? One hypothesis suggests that aggregates may be the key (5). Analogous to the presence of repetitive arrays of proteins, polysaccharides, or lipids on the surface of microbial or viral pathogens, it has been proposed that protein aggregates also present an array of native-like protein epitopes. These repetitive structures on protein aggregates may also lead to B-cell receptor cross-linking and generation of (low affinity IgM) immune response by the T-cell-independent pathway (2,5). A second signal would still be required to induce isotype switching and maturation of the B-cell to generate IgG antibodies. Presence of aggregates in therapeutic proteins might also impact the T-cell-dependent mechanism through enhanced uptake by APCs. Once taken up by the APC, aggregated forms could affect proteolytic digestion or binding to the MHC. Numerous factors, such as the anatomical location in which the T-cell receptor encounters the MHC-peptide complex, affinity of the T-cell receptor/MHC-peptide interaction, the presence of other co-stimulatory and cytokine signals, determine the T-cell response that will ensue. For example, in the thymus, immature T-cell clones that either do not bind or bind with high affinity to the presented peptides die. The T-cell clones that recognize the MHC-peptide complex with weak or moderate affinity survive, undergo further maturation and have potential to become T-helper or regulatory cells. In other anatomical

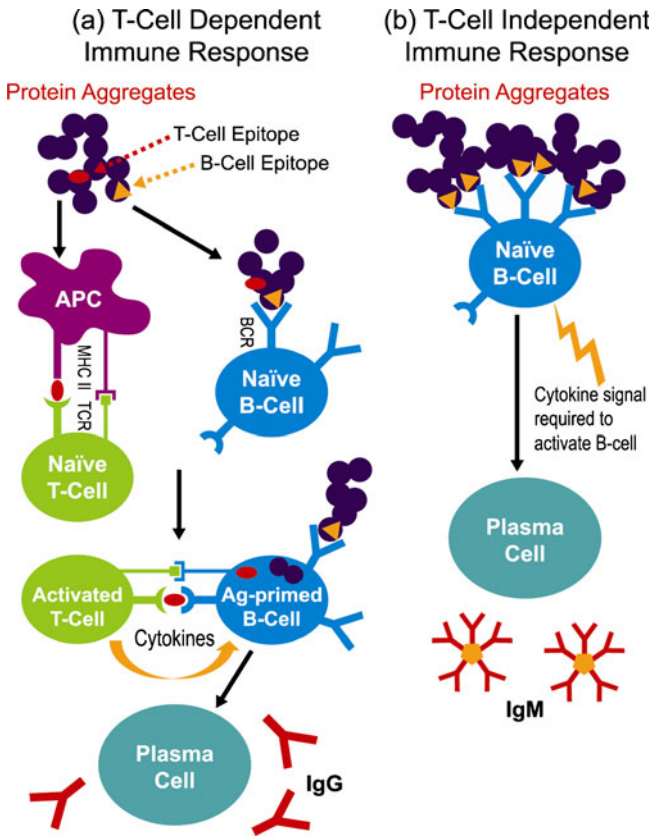

Fig. 3 Simplified hypothetical pathways for immune cell activation by aggregated proteins are shown. (**a**) The T-cell-dependent pathway. (**b**) The T-cell-independent pathway. Generation of immune response by the T-cell-dependent pathway would require presence of both B-cell and T-cell epitopes. Recognition of the B-cell epitope by BCR would drive uptake and processing by the B-cell and presentation of the T-cell epitope in the context of MHC class II molecule on its surface. In parallel, non-specific uptake and processing by professional APCs would lead to the presentation of the T-cell epitope on MHC class II to naïve T-cells. These activated T-cells, on encountering the antigen-primed B-cell, deliver the cytokine signal required to cause the B-cells to convert to IgG-secreting plasma cells. The T-cell-independent response occurs as a result of cross-linking of BCRs by repetitive epitopes on the antigen/aggregate. A cytokine signal is required to enable the B-cells to mature into plasma cells.

locations, T-cells that encounter APCs displaying the specific MHC-peptide will be activated if the appropriate co-stimulatory signals (provided by the APCs) are also present, or they may become non-responsive (anergized) in the absence of such co-stimulation. Drug-specific T-cells with helper cell phenotype could promote ADA responses, while drug-specific T-cells with regulatory cell phenotype could diminish such responses. Regardless of the mechanism, it is clear that binding to the T-cell and/or B-cell receptors is a key step.

Computational tools and databases have been developed to identify likely T-cell (both MHC class I and II) and B-cell epitopes in proteins (32). T-cell epitopes are generally linear and contiguous in amino acid sequence. On the other hand, B-cell epitopes are conformational and consist of residues that are close in the 3D structure of the antigen but may be distant in amino acid sequence (sequentially discontinuous). Hence, it has been more straightforward to predict T-cell epitopes *via* a number of available prediction tools (33). Steady progress is being made towards prediction of the B-cell epitopes as well (34). In some cases, co-localization between B-cell and T-cell epitopes has been observed (32,33). Moreover, manually curated epitope databases are also being developed due to their potential applications towards several biochemical problems, including identification of epitopes in proteins from infectious agents and pathogens and development of prophylactic and therapeutic vaccines. For example, Immune Epitope Database (IEDB) is the largest publically available NIH-supported database for peptide and non-peptide T- and B-cell epitopes (35).

Bryson *et al.* have analyzed the available immune epitope computational prediction tools and databases (36). They report that computational methods tend to predict a greater number of immune epitopes than validated by experiments. Hence, the computational predictions should be combined with *in vitro* assays (36). The T-cell-dependent immunogenicity prediction algorithms identify potential immune epitopes based on the binding of peptides to MHC (class I and class II) molecules. The recognition of the

MHC-peptide complexes by the T-cell receptors (TCRs) is not accounted for. Hence, these methods do not distinguish between T-helper and T-regulatory cell receptors. The computational methods also do not account for the additional factors present in the cellular milieu that may significantly modulate the T-cell response *in vivo*. These gaps may explain the over-prediction of potential immune epitopes by *in silico* methods. Availability of the structural models for MHC-peptide-TCR ternary complexes is required to better understand the interactions between the peptides and TCRs.

Most immune epitope prediction tools have been trained on the experimental binding assay data obtained from the overlapping peptides derived from antigens. The typical peptide length is 15 residues. Different MHC class II alleles may have different binding affinities for a given peptide. Hence, it is important to perform the binding assays for a large number of alleles and feed such data into the training sets (36). As compared to MHC class I molecules, the binding site groove of MHC class II molecules is more open. Hence, flanking residues besides the 9-residue core regions are also important. These flanking residues may be part of the cleavage sites for antigen-processing proteases (37). The nature of flanking residues may also alter the binding of the peptides with TCR (37). Hence, the rules for computational detection of the MHC class II binding peptides are more complicated than those for MHC class I. This also results in lower accuracy for prediction of peptides binding to MHC class II molecules (36).

In summary, it is possible to determine MHC binding propensity of peptide sequences from a protein therapeutic using *in silico* (and *in vitro)* methods. However, the actual impact of such binding on ADA development would require much further research in the context of treatment with the specific biotherapeutic. The current computational methods for immunogenicity do not predict experimentally measurable quantities, such as percent or magnitude of ADA that a biotherapeutic candidate could produce *in vivo*. Instead, they predict the regions of protein sequence or structure which are likely to be immunogenic, should a concern for immunogenicity arise. In this sense, they are analogous to the tools available to predict aggregation-prone regions, discussed above.

LINKAGE BETWEEN AGGREGATION AND IMMUNOGENICITY

A number of studies using transgenic animals have been published showing the role of protein aggregates in generating enhanced immunologic response (5,15,16). With aggregates implicated as a key factor in breaking of tolerance to human proteins, it is apparent that molecular determinants of aggregation, such as APRs and the consequent structural motifs could be important to understanding the relationship between aggregation and immunogenicity. In this regard, the cross β motif (38), formed due to inter-molecular association of β-strands coded by the APRs in protein sequence, is of particular interest. It represents a type of repetitive array of sequence-structural motif required by the T-cell-independent pathway. Alternatively, the cross β motif may simply be a construct that allows enhanced uptake by APCs for the T-cell-dependent pathway, as discussed earlier (2,39). Here, we asked the following question: Could potential APRs and immune epitopes lie in the same sequence/structural region of a molecule? Below, we provide our *initial* observations. These bioinformatics-derived observations are consistent with the experiments of Maas *et al.* (26), who probably first proposed that the cross β motif containing biotherapeutic aggregates can break tolerance and cause immunogenicity. Consistently, the proteins in this above-mentioned experimental study were predicted to contain APRs (30).

COUPLING BETWEEN APRs AND IMMUNE EPITOPES

APRs and Immune Epitopes in Therapeutic Proteins

We chose to interrogate human interferon β (IFN-β) for coupling between APRs and immune epitopes because sequence, structure and biophysical properties of this helical protein have been relatively well studied (40). Furthermore, evidence linking clinical immunogenicity of an MHC class II allele, DRB1*0701, in multiple sclerosis patients treated with IFN-β is available (41). For APR predictions, we used the method described in our earlier work (31). This method combines the predictions made by two programs: Tango (42) and PAGE (43). These predictions were also cross-validated with those from AmylPred server (http://biophysics.biol.uoa.gr/AMYLPRED/). For T-cell and antibody epitopes, the predictions were made by following the methods and tools available from IEDB organization (www.iedb.org) (35). Table I shows the potential APRs in IFN-β and their overlap with the sequence regions that are strongly predicted to contain MHC class II T-cell epitopes in IFN-β. Figure 4 further illustrates these observations by mapping the MHC class II T-cell epitopes and APRs on the protein sequence. In this case, the predicted linear and 3D structural (discontinuous) B-cell epitopes also overlap with two APRs (Table I). The peptides from the IFN-β C-terminus have been consistently implicated in the clinical immunogenicity of IFN-β, and one of these was found

Table I Overlap Between Predicted APRs and Immune Epitopes in IFN-β

APRs		MHC class II T-cell immune epitopes		B-cell immune epitopes	
Tango/PAGE[a]	AmylPred[b]	Immune epitope region[c]	HLA Alleles[d]	Linear[e]	Discontinuous[f]
4-NLLGFLQ-10	3-YNLLGF LQ-10	1-MSYNLLGFLQRSSNFQCQKLLW-22	DRB1*0301 DRB1*0401 DRB1*0701 DRB1*1101 DRB1*1301 DRB3*0101 DRB5*0101 DRB5*0105	4-NLLGFLQRS-12	1, 2, 4, 5, 166
63-QNIFAIFR-71	55-AALTIY-60	38-FDIPEEIKQLQQFQKEDAALTIYEMLQ-64	DRB1*0101 DRB1*0301 DRB1*0401 DRB1*0701	56-ALTIYEML-63 65-NIFAIFRQ-72	None
	62-MLQNFAIFR-71	58-TIYEMLQNIFAIFRQDSSSTGWNETIV-84	DRB1*1101 DRB1*1301 DRB1*1501 DRB3*0101 DRB4*0101 DRB5*0101 DRB5*0105		
141-AWTIVRVEIL-151	139-SHCAWTIVRVEILRNFYFINR-159	137-EYSHCAWTIVRVEILRNFYFINRLTGYLRN-166	DRB1*0301 DRB1*0701 DRB1*1101 DRB1*1301 DRB1*1501 DRB3*0101 DRB4*0101	117-MSSLHLKRYYGRILHYLKAKEYSHCAWTIVRVEILRNF-154	15,16, 19, 23–35, 72–86, 131, 134–141

Evidence for aggregation–immunogenicity coupling in IFN-β is presented here.

[a] Aggregation-prone regions (APRs) predicted using a combination of Tango and PAGE (31).

[b] Corresponding predictions from AmylPred server (http://biophysics.biol.uoa.gr/AMYLPRED/) are shown. Note that AmylPred server predicts additional APRs, and these are also contained in MHC class II T-cell receptor immune epitope regions.

[c] Regions containing the strong predictions for MHC class II T-cell epitopes. The predictions were made using the 'peptide binding to MHC class II molecule' tool available at IEDB web site (www.iedb.org). The method outlined at the website was followed. Besides this, we chose to apply at cut-off 20 percentile for the consensus score to identify potential immune regions. The minimum consensus percentile score was used to pick up the best immune epitopes shown in Fig. 4b.

[d] The predictions were made for commonly found HLA alleles in human populations. The following alleles were considered: HLA-DRB1*0101, HLA-DRB1*0301, HLA-DRB1*0401, HLA-DRB1*0701, HLA-DRB1*0801, HLA-DRB1*1101, HLA-DRB1*1301, HLA-DRB1*1501, HLA-DRB3*0101, HLA-DRB4*0101, HLA-DRB5*0101 and HLA-DRB5*0105.

[e] Linear B-cell epitope predicted using the method of Kolaskar and Tongaonkar available at IEDB website (49).

[f] 3D structural (discontinuous) B-cell epitopes were predicted by Ellipro3D method, available at IEDB web site. The predicted residues are indicated by their residue numbers in the protein sequence.

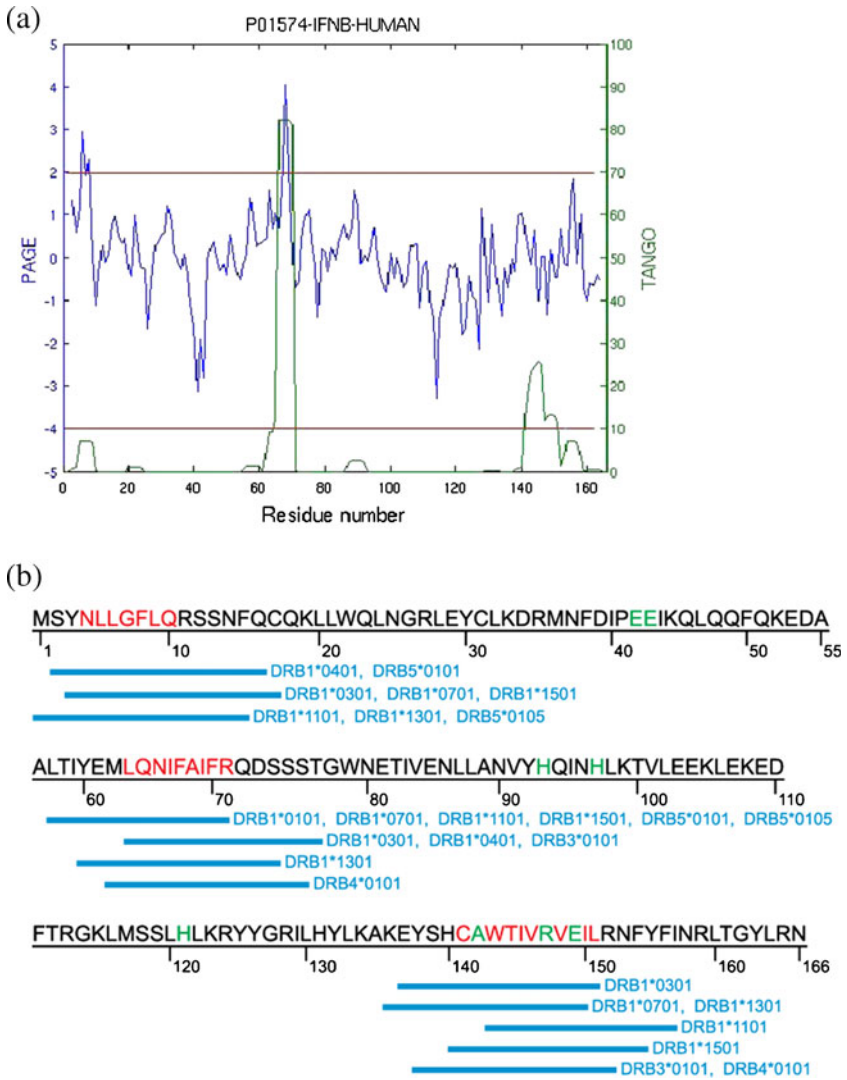

Fig. 4 (**a**) Prediction of APRs for human Interferon-β. The plot shows Tango and PAGE profiles for the mature IFN-β sequence (P10574) from UniProt knowledgebase (www.uniprot.org). The X-axis shows the residue number. The left Y-axis and blue colored curve indicate the Z-score computed from PAGE InP values (31). The right Y-axis and green curve indicate the Tango percent aggregation. The upper horizontal red line is for PAGE Z-score = 1.96. The lower horizontal red line is for Tango aggregation percentage = 10%. (**b**) Overlap between APRs and MHC class II T-cell receptor immune epitopes in IFN-β amino acid sequence. The APRs predicted by our method are shown in red color, and the functional residues are shown in the green color in the sequence. The predicted immune epitopes for different MHC class II alleles which overlap with the APRs are also shown by sky blue lines below the sequence. These predictions were made using the tools and methods described at IEDB website (www.iedb.org). We selected the potential MHC class II immune regions with consensus percentile score of 20 or better. The blue lines indicate the immune epitopes with the best percentile score within these regions.

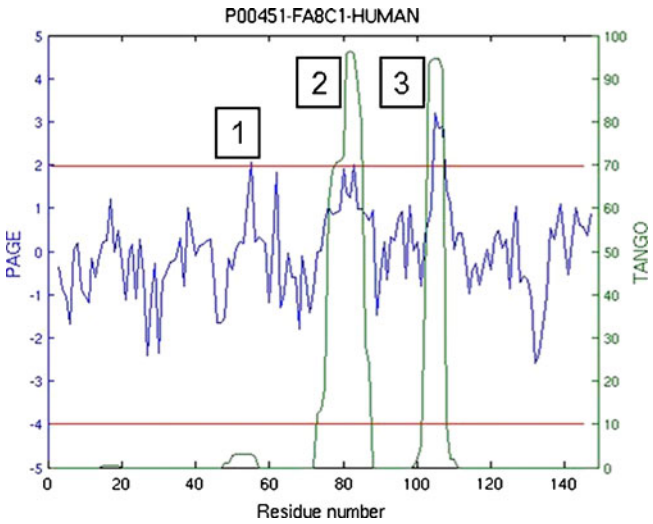

Fig. 5 Prediction of APRs for C1 domain of human Factor VIII is presented. The X-axis shows the residue number in the C1 domain. Note that this is different from the sequence number in the whole FVIII. The two Y-axes and the horizontal red lines denote the same quantities as in Fig. 4a. Three potential APRs are predicted: **1.** VDLLA, **2.** FSSLYISQFIIMYSL and **3.** TLMVFFGN. The second APR, strongly predicted by both Tango and PAGE, overlaps with an experimentally validated immune epitope (46).

to aggregate at higher concentrations (41). The APR and the immune epitope at the C-terminal of IFN-β also contain the functional residues Ala142, Arg147 and Glu149 (44).

We have also evaluated APRs in Factor VIII (FVIII), another human protein with a well-characterized immunogenicity profile (45). There is an experimentally proven promiscuous CD4+ T-cell epitope in the C1 domain of FVIII: the peptide 402 (2098-ISQFII MYSLDGKKW-2112) (46). It overlaps with a strongly predicted APR in our evaluations (2092-FSSLYISQFII MYSL-2107). Figure 5 shows predictions for APRs in human FVIII C1 domain.

APRs and Immune Epitopes in Therapeutic mAbs

Prediction of potential MHC class II T-cell receptor epitopes *via* computational and experimental means may be performed for biotherapeutics to aid in the selection of candidates in the discovery stages; in some cases, variable portions (Fv) of therapeutic mAb candidates are checked for presence of immune epitopes. Humanization of the candidates reduces the risk of immunogenicity. However, MHC class II T-cell epitopes can be found in the CDRs and adjoining framework regions; thus, the risk is not completely eliminated (12). When combined with our observations about incidence of APRs in the CDRs and adjoining regions of the mAbs and their contribution toward antigen binding (30,31), a three-way coupling among MHC class II T-cell epitopes, APRs and antigen recognition is possible for antibody-based therapeutics. Consistently, 13 of 21 experimentally validated MHC class II T-cell epitopes located in CDRs and adjacent framework regions of mAbs were predicted to contain at least one APR, when analyzed by Tango and PAGE (unpublished data).

Tregitopes and APRs in mAbs

We found one example in literature where the experimental evidence for both T-regulatory cell epitopes and APRs is available in the *same regions* of monoclonal antibodies. Two highly conserved human regulatory T-cell epitopes ("Tregitopes") (47), namely, hTregitope 167 (167-PAVLQSS GLYSLSSVVTVPSSSLGTQ-192) and hTregitope 289 (289-EEQYNSTYRVVSVLTVLHQDW-309), found in C_H1 and C_H2 domains of human immunoglobulin G (IgG), respectively, were identified by De Groot *et al.* (47). Figure 6 shows the location of these Tregitopes in the 3D structure of a full-length human IgG1 mAb. Both hTregitope167 and hTregitope289 have been experimentally shown to activate human natural regulatory T-cells and suppress immune response to the immunogenic peptides (47).

The hTregitope289 sequence (289-EEQYNSTYRVVS VLTVLHQDW-309) contains a strongly predicted APR (298-VVSVLTVL-305) (30). This APR is highly conserved in human IgG. This region was also found to contain an aggregation-prone conformational motif by the Trout group (28,29). AmylPred sever (http://biophysics.biol.uoa.gr/AMYLPRED/), which predicts amyloidogenic regions based on consensus from a number of different computational methods, also identifies the region 296-YRVVSVLTVLHQ-307 as strongly aggregation prone. In experiments from the Trout lab, disruption of the aggregation motif *via* L305K mutation led to reduced aggregation behavior by the antibody (28). Incidentally, this mutation also eliminates the strong APR mentioned above (case study #3 in the supplementary material for Wang *et al.* (31)). The APR prediction is not as consistent with hTregitope167. Tango and PAGE do not predict APRs in this region. However, two computational methods, namely average packing density and 3D profile, available as part of AmylPred server, identify

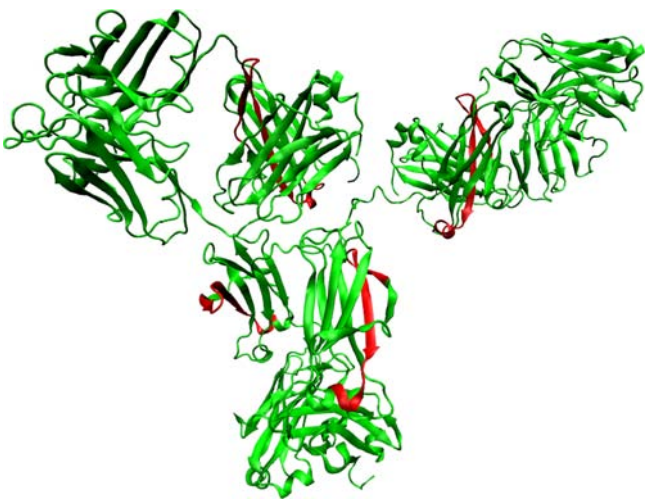

Fig. 6 Ribbon representation of human IgG1 mAb (PDB entry: 1HZH). The hTregitope 167, in C_H1 domain, and hTregitope 289, in C_H2 domain, are shown in the red color. hTregitope 289 (289-EEQYNSTYRVVSVLTVLHQDW-309) also contains a strong experimentally validated aggregation-prone region (298-VVSVLTVL-305).

the region 178-LSSVVTVPS-186 as aggregation prone. The Trout lab's spatial aggregation propensity (SAP) method also identifies an aggregation-prone conformational motif (169-VL-170 and Y-176) in this region (29). Both the Tregitopes fall in the highly conserved constant domains of human IgGs.

APRs and Immune Epitopes in IEDB Database

Our initial survey of peptide epitopes in IEDB (35) indicates that approximately 30 percent (1130 out of 3764) of human protein derived 15 residues or longer peptides contain at least one APR. The peptide epitopes are labeled as 'positive' in IEDB to indicate a successful experimental result in at least one of the following types of assays: T-cell response, B-cell response, MHC binding and MHC ligand elution. Moreover, IEDB contains 1028 MHC class II epitopes which yield positive results in at least one of the MHC class II binding, T-cell response, B-cell response and MHC ligand elution assays. Human is both the source and the host organism for these epitopes; that is, these peptides are potentially auto-immunogenic. However, depending on the context in which exposure to proteins containing these sequences first occurs, deletion of potentially self-reactive clones in the thymus or induction of T-regulatory cell pathways, rather than induction of T-helper cells may be the outcome. 246 (23.9%) of these 1028 human epitopes contain at least one APR. Table II contains the examples of experimentally validated MHC class II T-cell immune epitopes which also contain strongly predicted aggregation-prone regions. These surveys were carried out in August 2010, and the IEDB database is growing rapidly. The parent proteins for these peptides are involved in cellular processes including signaling, apoptosis, autoimmunity, etc. Many of these proteins have been known to form amyloids and have been implicated in neurodegenerative as well as auto-immune diseases.

DISCUSSION

Aggregation and immunogenicity are among the most important factors assessed in the discovery and development of biotherapeutic drugs. Both are the outcomes of multiple complex underlying phenomena. The characteristics of therapeutic molecules and their formulation, dosage, storage, shipping and delivery devices, procedures and routes of administration as well as the genetic characteristics and immune status of the patients need to be considered carefully.

Table II Predicted APRs in Experimentally Validated Immune Epitopes in IEDB

Immune Epitope Number/Sequence		Source Protein	Summary of evidence for immunogenicity	APRs	Strength of prediction
5516	AVPVYIYFNTW TTCQSIAFP	Human proteolipid protein (myelin) 1 involved in Multiple Sclerosis	MHC class II binding assay: IC50 = 70 nM for HLA-DRB5*0101. T-cell proliferation assay: Positive for HLA-DRB1*1501 & DRB5*0101	VYIYFNTWTT	>90% aggregation predicted by Tango
18488	FYYTTGAVRQI FGDYKTTICG	Human proteolipid protein (myelin) 1 involved in Multiple Sclerosis	MHC class II binding assay: IC50 = 22 nM for HLA-DRB1*1501 & 41 nM for HLA-DR53. T-cell proliferation assays: Positive for HLA-DRB1*1501 & DRB5*0101	KTTIC	Highly significant prediction by PAGE
41007	MAATYNFAVL KLMGRFTKF	Human proteolipid protein (myelin) 1 involved in Multiple Sclerosis	MHC class II binding assay: IC50 = 12 nM for HLA-DRB1*1501 T-cell proliferation assays: Positive for HLA-DRB1*1501 & DRB5*0101	NFAVL	Highly significant prediction by PAGE
65528	TPDFIVPLTDL RIPS	Human Apolipoprotein B-100 precursor	MHC class II binding assay: IC50 = 8 nM for HLA-DRB1*0401 & 70 nM for HLA-DRB1*0402.	FIVPL	Highly significant prediction by PAGE
116813	QKLVFFAEDV	Human Amyloid beta A4 protein precursor involved in Alzheimer's disease	T-cell proliferation assays: Positive for HLA-DRB1*1502 & DRB5*0301	LVFFA	>90% aggregation predicted by Tango and significant prediction by PAGE
	GSNKGAIIGL MVGGVVIA			AIIGLMV GGVVI	>90% aggregation predicted by Tango

A few representative examples of experimentally validated immune epitopes containing potential APRs are listed. These immune epitopes show strong binding affinity towards MHC class II molecules and are positive for T-cell proliferation assays. Data were taken from IEDB database (35) in August 2010. These immune epitopes also contain at least one strongly predicted APR. IEDB database contains 1028 such potentially auto-immunogenic peptide epitopes which yield positive results in at least one of the following assays: MHC class II binding, T-cell response, B-cell response and MHC ligand elution. 246 of these peptides are predicted to contain at least one APR by our method.

Realization of the linkage between aggregation and immunogenicity at the molecular level makes it feasible to simultaneously optimize the biotherapeutic candidates for reduced aggregation and immunogenicity via rational design. The molecular sequence and structural optimization could also potentially impact the drug product characteristics, formulation components, process-related impurities, dosage and delivery options in a favorable manner.

The observations reported here are preliminary. A detailed follow-up study is currently under progress to document the incidence of immune epitope-APR coupling and its statistical significance. Besides these, the detailed biological mechanism behind the coupling between aggregation and immunogenicity has to be understood. Hydrophobicity and binding to molecular chaperones may be the common factors among the peptides that are recognized by the MHC molecules and those that aggregate. However, detailed 3D structural models are needed to understand how aggregating peptides are simultaneously recognized by their partners and the MHC molecules. This fundamental understanding may be critical for applications aimed at the rational design of safe, soluble and stable biotherapeutics.

Not all immune epitopes contain APRs, and not all aggregates enhance immunogenicity of biotherapeutics. There may also be a wide variation for the immunogenicity, potentially induced by aggregates, among different patient populations due to the MHC class II allele differences. Different patient allotypes may bind the aggregated peptides with varying affinities which may result in different presentations by the APCs. Some aggregates may even turn out to be tolerogenic because the aggregated peptides bound by MHC class II molecules may be recognized by T-regulatory cell receptors, instead of the T-helper cell receptors. Furthermore, as stated earlier, our analysis does not take into account protein aggregates which could originate from sequence-structural regions other than the cross β steric zipper motif. The coupling between aggregation and immunogenicity described above is not absolute, and further studies are needed to fully understand the relationship between the two. The initial observations reported here should lead to further studies into the role of the molecular sequence and structural properties in aggregation and immunogenicity of the therapeutic proteins (4,48). Novel and follow-on biotherapeutics, that are de-immunized and less susceptible to aggregation, can be rationally designed. This may help improve the efficiency of biotherapeutic drug discovery and development cycles.

CONCLUSIONS AND NEXT STEPS

Molecular computational studies such as those described here can help develop rational structure-based design strategies to address CMC-related issues in biotherapeutics. Here, we have described our initial observations about a potential coupling between aggregation-prone regions and MHC class II T-cell receptor and B-cell receptor epitopes. These observations are significant because both immunogenicity and aggregation of biotherapeutics are complex phenomena. These observations may also help us explain why aggregates containing the cross β motif may be immunogenic. Other questions about immunological mechanisms, by which the protein aggregates may be recognized *in vivo*, remain unanswered. Clearly, further studies aimed at analysis of co-incidence between the potential immune epitopes and APR need to be performed, along with experimental validation. Computational as well as experiment-based mechanistic approaches that explain how aggregated peptides are recognized by different MHC class II alleles, T-helper and T-regulatory cell receptors can considerably improve our understanding of the relationship between biotherapeutic aggregation and immunogenicity. On a practical side, these observations open up the possibility of combining the design strategies intended to reduce aggregation with those intended to de-immunize by focusing on mutations aimed at disrupting both the APRs and immune epitopes. Only a few carefully selected residues may need to be re-engineered. Mutations at such residues may simultaneously disrupt promiscuous high binding affinity immune epitopes and strong aggregation-prone regions at or near the protein surface. Because MHC binding and T-cell engagement could lead to either immunogenicity or tolerance, experimental assays that measure both immunogenicity and aggregation propensity of the variants will be required to validate the designed variants. The above observations may also find applications towards vaccine research where controlled aggregation may be potentially used to stimulate immunogenicity (39).

ACKNOWLEDGMENTS

We thank Drs. Patrick Buck, Mark Mitchell, Ned M. Mozier, Herbert A. Runnels, Bruce Thompson, Phoebe Baldus and Vesselin Mitaksov for several helpful discussions regarding the aggregation-immunogenicity coupling. Drs. Sandeep Nema, Kevin King and Graeme Bainbridge are thanked for their interest in applications of computational modeling towards biotherapeutics discovery and development. A postdoctoral fellowship to Xiaoling Wang from Pfizer Biotherapeutics Pharmaceutical Sciences Research and Development is gratefully acknowledged. Sandeep Kumar acknowledges discussions with participants of the Buzz Session entitled "The relationship between aggregation and immunogenicity" on Tuesday 11 January 2011 during CHI's Pep Talk: The Protein Science Week at Hotel Del Coronado in San Diego, CA. His discussions with Dr.

Bjoern Peters of IEDB on the side lines of the above conference are also deeply appreciated. Dr. Neeti Sinha is acknowledged for critical reading of the manuscript.

REFERENCES

1. Giezen TJ, Mantel-Teeuwisse AK, Straus SMJM, Schellekens H, Leufkens HGM, Egberts ACG. Safety-related regulatory actions for biologicals approved in the United States and the European Union. JAMA. 2008;300(16):1887–96.
2. Rosenberg AS. Effects of protein aggregates: an immunologic perspective. AAPS J. 2006;8(3):E501–7.
3. Carpenter JF, Randolph TW, Jiskoot W, Crommelin DJ, Middaugh CR, Winter G, *et al.* Overlooking subvisible particles in therapeutic protein products: gaps that may compromise product quality. J Pharm Sci. 2009;98(4):1201–5.
4. Singh SK. Impact of product-related factors on immunogenicity of biotherapeutics. J Pharm Sci. 2011;100(2):354-87.
5. Sauerborn M, Brinks V, Jiskoot W, Schellekens H. Immunological mechanism underlying the immune response to recombinant human protein therapeutics. Trends Pharmacol Sci. 2010;31(2):53–9.
6. Moss CX, Matthews SP, Lamont DJ, Watts C. Asparagine deamidation perturbs antigen presentation on class II major histocompatibility complex molecules. J Biol Chem. 2005;280 (18):18498–503.
7. McNally EJ, Hastedt JE, editors. Protein formulation and delivery. New York: Informa Healthcare; 2008.
8. Laurence JS, Middaugh RC. Fundamental structures and behaviors of proteins. In: Wang W, Roberts CJ, editors. Aggregation of therapeutic proteins. Hoboken: Wiley; 2010. p. 1–61.
9. Schmit JD, Ghosh K, Dill K. What drives amyloid molecules to assemble into oligomers and fibrils? Biophys J. 2011;100(2):450–8.
10. Clark M. Antibody humanization: a case of the 'Emperor's new clothes'? Immunol Today. 2000;21(8):397–402.
11. Presta LG. Engineering of therapeutic antibodies to minimize immunogenicity and optimize function. Adv Drug Deliv Rev. 2006;58(5–6):640–56.
12. Harding F, Stickler MM, Razo J, DuBridge R. The immunogenicity of humanized and fully human antibodies: residual immunogenicity resides in the CDR regions. mAbs. 2010;2(3):1–10.
13. Beck A, Reichert JM, Wurch T. 5th European antibody congress 2009: November 30–December 2, 2009, Geneva. Switzerland mAbs. 2010;2(2):108–28.
14. Kumar S, Singh SK, Gromiha MM. Temperature dependent molecular adaptations, microbial proteins. In: Flickinger MC, Drew SW, Spier RE editors. Encyclopedia of Industrial Biotechnology. John Wiley & Sons; 2010. p. 4647–4661.
15. Hermeling S, Schellekens H, Maas C, Gebbink MF, Crommelin DJ, Jiskoot W. Antibody response to aggregated human interferon alpha2b in wild-type and transgenic immune tolerant mice depends on type and level of aggregation. J Pharm Sci. 2006;95 (5):1084–96.
16. Fradkin AH, Carpenter JF, Randolph TW. Immunogenicity of aggregates of recombinant human growth hormone in mouse models. J Pharm Sci. 2009;98(9):3247–64.
17. Foulkes R. Preclinical safety evaluation of monoclonal antibodies. Toxicology. 2002;174(1):21–6.
18. Pechmann S, Levy ED, Tartaglia GG, Vendruscolo M. Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. Proc Natl Acad Sci U S A. 2009;106(25):10159–64.
19. Siepen JA, Radford SE, Westhead DR. β Edge strands in protein structure prediction and aggregation. Protein Sci. 2003;12 (10):2348–59.
20. Nelson R, Eisenberg D. Recent atomic models of amyloid fibril structure. Curr Opin Struct Biol. 2006;16(2):260–5.
21. Nelson R, Sawaya MR, Balbirnie M, Madsen AO, Riekel C, Grothe R, *et al.* Structure of the cross-b spine of amyloid-like fibrils. Nature. 2005;435(7043):773–8.
22. Sawaya MR, Sambashivan S, Nelson R, Ivanova MI, Sievers SA, Apostol MI, *et al.* Atomic structures of amyloid cross-b spines reveal varied steric zippers. Nature. 2007;447(7143):453–7.
23. Maurer-Stroh S, Debulpaep M, Kuemmerer N, de la Paz ML, Martins IC, Reumers J, *et al.* Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. Nat Methods. 2010;7(3):237–42.
24. Wang L, Maji SK, Sawaya MR, Eisenberg D, Riek R. Bacterial inclusion bodies contain amyloid-like structure. PLoS Biol. 2008;6 (8):1791–801.
25. Kumar S, Wang X, Singh SK. Identification and impact of aggregation prone regions in proteins and therapeutic mAbs. In: Wang W, Roberts CJ, editors. Aggregation of therapeutic proteins. Hoboken: Wiley; 2010. p. 103–18.
26. Maas C, Hermeling S, Bouma B, Jiskoot W, Gebbink MFBG. A role for protein misfolding in immunogenicity of biopharmaceuticals. J Biol Chem. 2007;282(4):2229–36.
27. Brummitt RK, Nesta DP, Chang L, Chase SF, Laue TM, Roberts CJ. Nonnative aggregation of an IgG1 antibody in acidic conditions: Part 1. Unfolding, colloidal interactions, and formation of high-molecular-weight aggregates. J Pharm Sci. 2011; In press.
28. Chennamsetty N, Voynov V, Kayser V, Helk B, Trout BL. Design of therapeutic proteins with enhanced stability. Proc Natl Acad Sci U S A. 2009;106(29):11937–42.
29. Chennamsetty N, Helk B, Voynov V, Kayser V, Trout BL. Aggregation-prone motifs in human immunoglobulin G. J Mol Biol. 2009;391(2):404–13.
30. Wang X, Das TK, Singh SK, Kumar S. Potential aggregation prone regions in biotherapeutics: a survey of commercial monoclonal antibodies. mAbs. 2009;1(3):1–14.
31. Wang X, Singh S, Kumar S. Potential aggregation-prone regions in complementarity-determining regions of antibodies and their contribution towards antigen recognition: a computational analysis. Pharm Res. 2010;27(8):1512–29.
32. De Groot AS, McMurry J, Moise L. Prediction of immunogenicity: in silico paradigms, *ex vivo* and *in vivo* correlates. Curr Opin Pharmacol. 2008;8(5):620–6.
33. De Groot AS, Moise L. Prediction of immunogenicity for therapeutic proteins: state of the art. Curr Opin Drug Discov Devel. 2007;10(3):332–40.
34. Greenbaum JA, Andersen PH, Blythe M, Bui H-H, Cachau RE, Crowe J, *et al.* Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools. J Mol Recognit. 2007;20(2):75–82.
35. Vita R, Zarebski L, Greenbaum JA, Emami H, Hoof I, Salimi N, *et al.* The immune epitope database 2.0. Nucleic Acids Res. 2010;38(Database issue):D854–862.
36. Bryson CJ, Jones TD, Baker MP. Prediction of immunogenicity of therapeutic proteins: validity of computational tools. BioDrugs. 2010;24(1):1–8.
37. Godkin AJ, Smith KJ, Willis A, Tejada-Simon MV, Zhang J, Elliott T, *et al.* Naturally processed HLA class II peptides reveal highly conserved immunogenic flanking region sequence preferences that reflect antigen processing rather than peptide-MHC interactions. J Immunol. 2001;166(11):6720–7.
38. Gebbink MF, Bouma B. In: Office USPT, editor. Method for detecting and/or removing protein comprising a cross-beta

structure from a pharmaceutical composition. USA: UMC Utrecht Holding B.V; 2007.
39. Filipe V, Hawe A, Schellekens H, Jiskoot W. Aggregation and Immunogenictiy of Therapeutic Proteins. In: Wang W, Roberts CJ, editors. Aggregation of therapeutic proteins. Hoboken: Wiley; 2010. p. 403–33.
40. Basu A, Yang K, Wang M, Liu S, Chintala R, Palm T, *et al.* Structure-function engineering of interferon-beta-1b for improving stability, solubility, potency, immunogenicity, and pharmacokinetic properties by site-selective mono-PEGylation. Bioconjug Chem. 2006;17(3):618–30.
41. Barbosa MDFS, Vielmetter J, Chu S, Smith DD, Jacinto J. Clinical link between MHC class II haplotype and interferon-beta (IFN-beta) immunogenicity. Clin Immunol. 2006;118(1):42–50.
42. Fernandez-Escamilla A-M, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat Biotechnol. 2004;22 (10):1302–6.
43. Tartaglia GG, Cavalli A, Pellarin R, Caflisch A. Prediction of aggregation rate and aggregation-prone segments in polypeptide sequences. Protein Sci. 2005;14(10):2723–34.
44. Karpusas M, Nolte M, Benton CB, Meier W, Lipscomb WN, Goelz S. The crystal structure of human interferon beta at 2.2-A resolution. Proc Natl Acad Sci U S A. 1997;94 (22):11813–8.
45. Lacroix-Desmazes S, Navarrete A-M, Andre S, Bayry J, Kaveri SV, Dasgupta S. Dynamics of factor VIII interactions determine its immunologic fate in hemophilia A. Blood. 2008;112 (2):240–9.
46. Jones TD, Phillips WJ, Smith BJ, Bamford CA, Nayee PD, Baglin TP, *et al.* Identification and removal of a promiscuous CD4+ T cell epitope from the C1 domain of factor VIII. J Thromb Haemost. 2005;3(5):991–1000.
47. De Groot AS, Moise L, McMurry JA, Wambre E, Van Overtvelt L, Moingeon P, *et al.* Activation of natural regulatory T cells by IgG Fc-derived peptide "Tregitopes". Blood. 2008;112(8):3303–11.
48. Richard J, Prang N. The formulation and immunogenicity of therapeutic proteins: product quality as a key factor. IDrugs. 2010;13(8):550–8.
49. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett. 1990;276(1–2):172–4.
50. Lührs T, Ritter C, Adrian M, Riek-Loher D, Bohrmann B, Döbeli H, *et al.* 3D structure of Alzheimer's amyloid-b(1-42) fibrils. Proc Natl Acad Sci U S A. 2005;102(48):17342–7.