Basis of the Intrinsic Flexibility of the Cε3 Domain of IgE

ARTICLE *in* BIOCHEMISTRY · MAY 2011

Impact Factor: 3.02 · DOI: 10.1021/bi200019y · Source: PubMed

3 AUTHORS:

Susmita Borthakur
Case Western Reserve University
6 PUBLICATIONS **37** CITATIONS
SEE PROFILE

Gabriela Andrejeva
Vanderbilt University
5 PUBLICATIONS **18** CITATIONS
SEE PROFILE

James McDonnell
King's College London
72 PUBLICATIONS **4,775** CITATIONS
SEE PROFILE

BIOCHEMISTRY
including biophysical chemistry & molecular biology

ARTICLE

pubs.acs.org/biochemistry

Basis of the Intrinsic Flexibility of the Cε3 Domain of IgE

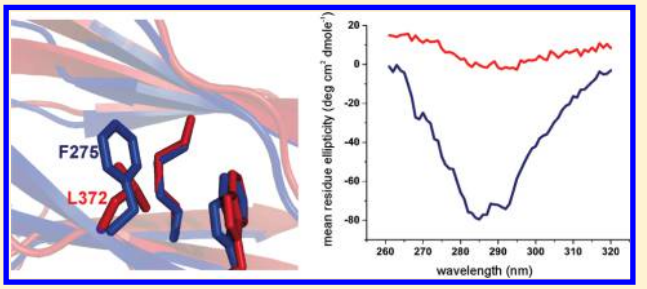
Susmita Borthakur,[†,‡] Gabriela Andrejeva,[†] and James M. McDonnell*[,†,‡]

[†]Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

[‡]Randall Division of Cell and Molecular Biophysics, King's College London, London SE1 1UL, U.K.

Supporting Information

ABSTRACT: Allergic reactions are triggered by the interaction between IgE and its high-affinity receptor, FcεRI. Various studies have mapped the interaction surface between IgE and its cellular receptors to the third constant domain of IgE (Cε3). The isolated Cε3 domain has been shown to exist as a molten globule, and the domain retains significant flexibility within the context of the IgE protein. Here we have analyzed the structural basis of the intrinsic flexibility of this domain. We have compared the sequence of the Cε3 domain to the sequences of other members of the C1 subset of the immunoglobulin superfamily and observed that Cε3 has an unusually high electrostatic charge and an unusually low content of hydrophobic residues. Mutations restoring Cε3 to a more canonical sequence were introduced in an attempt to derive a more structured domain, and several mutants display decreased levels of disorder. Engineered domains of Cε3 with a range of structural rigidities could serve as important tools for the elucidation of the role of flexibility of the Cε3 domain in IgE's biological functions.



Antibodies are adaptor molecules that provide a link between humoral and cellular immune responses; they bind to foreign antigens with their highly variable Fab (antigen-binding fragment) regions and link to specific effector mechanisms through interactions between their conserved Fc regions and cellular or soluble receptors. Different isotypes of antibodies have specialized roles in host defense. The natural function of immunoglobulin E (IgE) is probably in response to parasitic infections,[1−3] but IgE is best known for mediating allergic immune responses. When IgE is bound through its Fc region to its receptor on mast cells (FcεRI), a multivalent "allergen" (allergy-inducing antigen) can bind to its Fab regions and cross-link FcεRI, activating the mast cell to release inflammatory mediators. In addition to its interaction with FcεRI, IgE-Fc also binds to B cell receptor CD23 and surfactant protein D (SP-D), which play regulatory roles controlling IgE expression levels and activities.[4]

The IgE−FcεRI interaction is of uniquely high-affinity for antibody−receptor interactions, with subnanomolar K_D values compared with micromolar binding affinities for a number of analogous IgG−FcγR interactions.[5] Particularly striking is the difference in off rates, with IgE showing a half-life of >10 h compared with seconds for most IgG−receptor interactions.[6] At least some of these differences are related to the physical differences between IgE and IgG, the relevant features of which have been revealed by structural studies of immunoglobulin (Ig) Fc regions, Fc receptors (FcR), and Fc−FcR complexes (reviewed by Woof and Burton[7]). Of particular relevance for this study are the unusual structural characteristics of the third constant domain of IgE (Cε3), the domain responsible for all IgE−receptor binding events. It has been known for many years that IgE-Fc contains a region of low thermal stability, which is stabilized by FcεRI binding.[8] More recently, the isolated Cε3 immunoglobulin domain was shown to be a molten globule using a variety of biophysical methods, including nuclear magnetic resonance (NMR), circular dichroism, and fluorescence spectroscopies;[9] other evidence suggests that the Cε3 domain remains highly labile even in the context of IgE.[10]

Such structural instability is unexpected in an immunoglobulin (Ig) domain; these domains are among the most common motifs found in the human proteome, because of their ability to act as a robust scaffold for the display of variable loop regions. This provides Ig domains with a great capacity for presenting structural variability at the protein surface, exploited for binding to correspondingly diverse ligands, while maintaining a stable core structure.[11] From both theoretical and practical standpoints, Ig domains and the Ig superfamily (IgSF) are among the best characterized systems for defining protein sequence−structure relationships. The practical application of engineered antibodies as experimental reagents, diagnostics, and therapeutics has meant that Ig domains have been the subject of intense and often spectacularly successful efforts in protein engineering.[12]

Until very recently, it was generally accepted that a well-defined three-dimensional structure is essential for the functional activity of a protein, and the "lock and key" metaphor was often used to describe the molecular recognition process in protein−ligand interactions. In recent years, the fact that even highly

Received: January 6, 2011
Revised: April 4, 2011
Published: April 26, 2011

flexible regions of proteins can be functionally active has become appreciated; in fact, a large number of functional advantages have been proposed for poorly ordered proteins, including greater ligand promiscuity and, perhaps counterintuitively, greater specificity, extended binding surface areas (compared to those of well-folded proteins) allowing high affinities and high stoichiometries for interactions, increased association rates, a particular sensitivity to post-translational modification offering regulatory mechanisms, and many others.[13] One challenge for those groups that study intrinsically unstructured proteins is the generally untestable nature of the postulated advantages of disorder, because of the inability to modify the degree of disorder in the system.

The observed intrinsic flexibility of the isolated Cε3 domain, although reduced in the context of the whole IgE, has been postulated to be essential for the interaction of IgE with its cellular receptors.[9,14] Understanding the basis of the intrinsic flexibility would be helpful in improving our understanding of the mechanisms of IgE-mediated interactions and devising strategies for modulating these interactions. On the basis of comparative structure-based sequence analyses, rational engineering of the Cε3 domain has been performed to construct more rigid versions of the domain. These studies offer insights into the unusual biophysical properties of the Cε3 domain.

EXPERIMENTAL PROCEDURES

Construction of the Database of Human C1-set Ig Domains. Protein sequences of Ig C1-set domains were retrieved from the Pfam,[15] SCOP,[16] and International Immunogenetics Information System (IMGT)[17] databases and sorted to create a database of human, translated, nonredundant sequences. Each sequence was cross-referenced against the corresponding entry in the UniProt database[18,19] to confirm the sequences. T-Cell receptor α chain c-region (TRAC) was excluded from the database because it lacks the typical topology of an Ig C1-set domain,[20] although most databases group TRAC with other Ig C1-set domains. Only C1-set domains with ≥74 amino acid residues were included to create a database consisting of domains from antibodies, MHC class I and II proteins, TCR proteins, and Ig-related proteins.

Calculation of the Structure-Based Multiple-Sequence Alignment. The multiple-sequence alignment was determined using the algorithms included in the T-Coffee software suite.[21] In the first step, T-Coffee Expresso was used to calculate structure-based sequence alignment. The output from T-Coffee Expresso was used as a seed alignment, keeping the key aligned positions fixed to compute the second round of alignment using M-Coffee. The output from M-Coffee was manually adjusted to generate the final multiple-sequence alignment (Figure S1 of the Supporting Information).

Comparison of the Propensity for Folding of C1-set Domains. The propensity for folding of a protein with a given sequence was assessed by using the method proposed by Uversky.[22] FoldIndex[23] is an algorithm that uses this method to predict whether a given sequence will fold. In the first step, the domain boundaries of the sequences were defined by using the structure-based multiple-sequence analysis computed in this study. Each sequence was clipped to give a sequence starting with two residues before the beginning of strand A and ending with two residues after strand G. The Web Service of the FoldIndex tool (http://bioportal.weizmann.ac.il/fldbin/findex) was used to calculate the FoldIndex value for each of the domains of the database. The data computed were fit using a Gaussian model to plot the distribution of the FoldIndex values within Ig C1-set domains. Disorder prediction was also conducted using the algorithms PONDR,[24−26] DisEMBL,[27] and DISOPRED[28] for the Cε3 domain.

Calculation of the van der Waals Contacts Made by Buried Residues. The numbers of van der Waals contacts made by buried hydrophobic residues were computed using the CCP4 program NCONT.[29] The Protein Data Bank (PDB) coordinates were taken from 1O0V and 1E4K for the Cε3 and Cγ2 domains, respectively. The contacts were computed using a cutoff value of 4.0 Å.

Site-Directed Mutagenesis To Engineer More Structured Versions of the Cε3 Domain. On the basis of the comparative analyses, 15 different mutant constructs of the Cε3 domain were produced. Two complementary oligonucleotides were designed for each mutation investigated, with 12−14 bases on either side of the substitution site. The primers for the mutagenesis were obtained from Sigma-Genosys (Sigma Lifescience). Site-directed mutagenesis was conducted using a QuikChange II site-directed mutagenesis kit (Stratagene) according to the manufacturer's instructions.

Expression and Purification of the Wild-Type and Mutant Cε3 Domains. The wild-type and mutant Cε3 domains (comprising residues Ser328−Ser437) in a pET28a vector were expressed in BL21(DE3) cells as His tag fusion proteins. Expression and purification were performed using protocols described by Price et al.[9] The pooled sample that eluted from a nickel column was concentrated to ≤13 mL in a 20 mL centrifugal concentrator with a 5000 Da molecular mass cutoff (Sartorius) and applied to a HiLoad 26/60 Superdex 75 gel filtration column (GE Healthcare) to separate correctly folded Cε3 from misfolded aggregates. Fractions containing monomeric Cε3 were pooled and concentrated for further analysis.

Circular Dichroism (CD) Spectroscopy of Wild-Type and Mutant Cε3 Domains. The CD spectra of wild-type Cε3 and its mutants were recorded on a Jasco J-720 spectropolarimeter at 293 K. Spectra in the far-UV region (260−195 nm) were recorded in cylindrical cells with a path length of 0.2 mm using a protein concentration of 0.2 mg/mL in PBS (pH 5.0). Spectra in the near-UV region (320−260 nm) were recorded in a cell with a path length of 5 mm, using a protein concentration of 3 mg/mL. In each case, six scans (recorded at a scan rate of 20 nm/min with a time constant of 2 s) were averaged and corrected by subtraction of the spectrum of buffer alone. Mean residue ellipticities were calculated using a value of 112 for the mean residue weight of the protein (calculated from the amino acid sequence). The far-UV spectra were analyzed by the K2D2 algorithm to give an estimate of the contribution of elements of regular secondary structure.[30]

ANS Binding Study of Wild-Type and Mutant Cε3 Domains. 8-Anilinonaphthalene-1-sulfonate (ANS) binding to wild-type and mutant Cε3 was studied using an ANS concentration of 17.3 μM. ANS fluorescence emission was measured between 430 and 550 nm, following excitation at 370 nm. ANS fluorescence emission was measured for samples of wild-type Cε3, mutant Cε3 domains, denatured Cε3, and bovine serum albumin (BSA). Emission spectra were corrected for the effect of dilution by performing control experiments in which buffer alone was added. All fluorescence measurements were taken on a Perkin-Elmer luminescence spectrometer (LS50B) at 293 K.

Nuclear Magnetic Resonance (NMR) Spectroscopy of Wild-Type and Mutant Cε3 Domains. NMR samples (350 μL),

with 5% (v/v) D_2O and 0.02% (w/v) sodium azide, were transferred into Shigemi tubes (Sigma-Aldrich, St. Louis, MO). All NMR experiments were conducted in PBS (pH 5.0) at 293 K. The spectra were recorded on a Bruker Avance 500 MHz spectrometer with a cryoplatform equipped with a TCI cryoprobe. The one-dimensional ^{1}H NMR spectra were recorded with a sweep width of 12048.2 Hz, collected over 4K points with 32 scans, and referenced to DSS at 0 ppm. Spectra were processed using NMRPipe[31] and plotted using OriginLab (OriginLab, Northampton, MA).

RESULTS

Sequence Analyses Show That the Cε3 Domain Has a Propensity for Disorder. The amino acid sequences of intrinsically unstructured proteins share some common characteristics: a higher than average number of charged residues, a smaller than average number of hydrophobic residues, and use of low-complexity amino acid sequences.[32] We have performed structure-based sequence analyses of the Ig C1-set family. The Pfam database has 7404 entries for Ig C1-set sequences from 328 species. We decided to restrict our studies to human, nonredundant, translated sequences to maximize the efficiency of computing the sequence alignments. The database created contained 66 full-length Ig C1-set sequences. The sequences were aligned on the basis of structural alignment as described in Experimental Procedures, and the full alignment is shown in Figure S1 of the Supporting Information.

The immunoglobulin domain fold is characterized by a sequence of 70−110 amino acids folded as a β-sandwich. Previous comparative studies of the structure and sequence of domains of the IgSF have established that the domains carry highly variable sequences, but the individual families do share a common structural core centered around a nearly universally conserved "central pin",[33] comprised of two cysteine residues that form a disulfide bond and a tryptophan residue. Highly conserved residues are also found in β-strands B, C, and F, and hydrophobic residues in alternating positions of the strands is a signature of the fold[34,35] (Figure S1 of the Supporting Information). Unusually, the first constant domain of immunoglobulin heavy chains (CH1) was observed to have a proline within strand A, in a position predominantly occupied by aliphatic residues in other C1-set domains. In addition, the Cε3 domain (as well as the SIRBL, SHPS1, and HLA-DMA domains) has an alanine in position 3 of strand A; most other domains carry a hydrophobic residue with greater side chain volume in that position. A statistical analysis of the sequence composition, performed using the algorithm SAPS,[36] revealed some unusual features in the sequence of the Cε3 domain. The domain carries a net charge of +9, which accounts for 8.7% of the total sequence. In contrast, the Cγ2 domain, the equivalent domain in IgG that is known to fold well as an isolated domain,[37,38] carries no net charge. In addition, the hydrophobic residues form only 22.0% of the sequence of the Cε3 domain, whereas hydrophobic residues form 28.8% of the sequence of the Cγ2 domain. Overall, the mean hydrophobic content for the domains in the Ig C1-set database in the SAPS analysis was 25 ± 4%, and the average net charge was |2.8| ± 2.4. In the C1-set database, the Cε3 domain had one of the lowest percentages of hydrophobic residues and one of the highest values for net electrostatic charge.

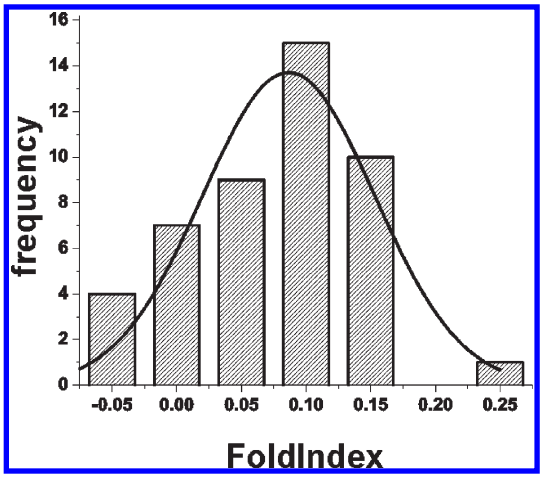


Figure 1. Distribution of FoldIndex values of the Ig C1-set domains excluding the MHC proteins. The FoldIndex values were calculated for all the domains in the database of the Ig C1-set, and the distribution was plotted and fit using a Gauss model. All the domains in the database excluding the MHC class I proteins were included in the plot. The Cε3 domain is one of the outliers, with an unfavorable FoldIndex value of −0.017.

FoldIndex is an algorithm that predicts the propensity of a particular sequence to adopt a folded structure.[23] The FoldIndex values were calculated for all the domains in the database. The unfavorable net charge and hydrophobicity of the Cε3 domain were also reflected by the FoldIndex value computed for the domain; the Cε3 domain has a FoldIndex value of −0.017. In addition to the Cε3 domain, the Cγ3 domains of IgG2 and IgG4, the Cδ1 domain of IgD, and some major histocompatibility complex (MHC) class I proteins also exhibited negative FoldIndex values. The mean of the FoldIndex values computed for Ig C1-set domains is 0.099 ± 0.074, and the variance is 0.005. Except for the Cε3 domain, all the other domains with unfavorable charge and hydrophobicity are dimers. The MHC proteins are expressed as obligate dimers, and it has been shown that the dimerization status is essential for the stability of these proteins.[39,40] The unfavorable charges observed in the monomeric proteins are compensated by the formation of homo- or heterodimeric interfaces. For example, the Cγ3−Cγ3 dimer interface is stabilized by two salt bridges.[41] Similarly, in MHC class I proteins, the interface between β2-microglobulin and α3 has four salt bridges.[42] In contrast, the Cε3 domain is not a dimer; it does pack against Cε2 at its amino terminus and the Cε4 domain at its C-terminus, but no salt bridges are formed in these interactions. A distribution of the FoldIndex values of the Ig C1-set domains excluding the MHC proteins placed the Cε3 domain among the outliers with a negative FoldIndex value (Figure 1).

A number of other protein disorder predicting algorithms (PONDR, DISOPRED, and DisEMBL) also predict regions of disorder within Cε3. PONDR[24−26] is a neural network-based predictor, which uses sequence attributes like hydropathy and fractional composition of particular amino acids to calculate disorder in a window of 9−21 amino acids. DisEMBL[27] is an artificial neural network-based algorithm that predicts disorder according to many different definitions. DISOPRED[28] is also a neural network-based algorithm, which gives the probability of individual residues of a domain being disordered. All of the

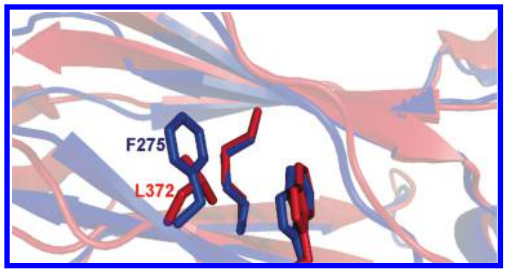


Figure 2. Superposition of the three-dimensional structures of Cε3 and Cγ2. Cε3 is shown as a red cartoon representation and Cγ2 as a blue representation. The nearly universally conserved pair of cysteines and the tryptophan residue that comprise the Ig domain central pin are shown as sticks. Additionally, the conserved hydrophobic residue L372 of Cε3 and the residue at the equivalent position in Cγ2, F275, are highlighted as an example of a difference in the hydrophobic cores of Cε3 and Cγ2.

algorithms used have predicted disorder for the same sequence of Cε3, in a region comprised of strands C and D, and the C−D and D−E loops, by analogy to other Ig C1-set domains (Figure S2a,b of the Supporting Information).

Engineering of More Rigid Versions of the Cε3 Domain. The sequence analyses described above showed that Cε3 differs from the rest of the Ig C1-set domains mainly in its unfavorable net charge and hydrophobicity. We postulated that replacing noncanonical residues found in Cε3 with more canonical Ig C1-set residues would result in a more structured Cε3 domain. Two different classes of mutations were employed. The first class included mutations to reduce the net charge on the domain, and the second class of the mutations involved replacing residues in the Cε3 domain with canonical residues found in other Ig C1-set domains. The individual mutations were designed using the information gathered from the structure-based sequence analysis. We observed that Cε3 substituted several conserved hydrophobic residues with residues smaller than those used in closely related Ig domains. It is not uncommon to observe mutation of conserved hydrophobic residues in IgSF, but there is almost always a second mutation that compensates for the effect of the first. Perhaps the best known example of this is β2-microglobulin (β2m). β2m is the only C1-set protein that lacks the conserved "central pin" tryptophan residue in strand C; it compensates with a mutation in strand E that replaces a residue that is typically valine with tyrosine (Figure S1 of the Supporting Information). The structural alignment of Cε3 with Cγ2 shows that Cε3 has a leucine residue in position 2 of strand C (C2) whereas Cγ2 has a phenylalanine in the equivalent position (Figure 2); Cε3 does not have a compensating mutation elsewhere to offset the loss of side chain volume at this position. A comparison of the interstrand contacts made by Leu372 within the Cε3 domain and by Phe275 within the Cγ2 domain confirms that Leu372 makes far fewer van der Waals contacts. Leu372 makes six contacts compared to 15 contacts made by Phe275 (Table S1 of the Supporting Information). A Cε3 domain carrying an L372F mutation was constructed to evaluate the effect of a larger side chain in that position on the packing of the hydrophobic core of Cε3. The side chain of Arg387 is buried inside the hydrophobic core of Cε3; uncompensated charges within the core of a protein can be highly destabilizing. An R387A mutant construct was made to replace the charged side chain with a hydrophobic side chain. Arg376 and Lys380 are clustered close together; the proximity of residues with similar charges might result in electrostatic repulsion and hence influence the structural rigidity of a domain. To weaken the effect of electrostatic repulsions on the domain, the positively charged side chains of both Arg376 and Lys380 were mutated to alanine in a double mutant (R376A/K380A) of Cε3. Similarly, a number of mutant domains were created with both single and multiple mutations of positively charged residues to decrease the net positive charge on the domain. A total of 15 mutants were constructed and characterized. Most of the mutations had no effect or a very modest effect on the flexibility of the Cε3 domain. A number of the mutations that demonstrated significantly different biophysical profiles compared to that of the wild-type Cε3 domain are described below.

CD spectroscopy and ANS binding studies of the isolated Cε3 domain show that Cε3 has near-native-like secondary structure but lacks the rigid packing of the side chains characteristic of well-folded proteins.[9] The biophysical profiles of the mutant Cε3 domains were compared with the wild-type profile to assess effects on Cε3 structure. ANS (1-anilino-8-naphthalenesulfonate) is a hydrophobic dye that binds specifically to molten globules, and binding increases the fluorescence intensity of the dye.[43] The high capacity to bind ANS as measured by the increased relative fluorescence intensity of the dye is used as an experimental criterion to characterize a protein as a molten globule.[44] The wild-type Cε3 domain binds to and strongly enhances ANS fluorescence,[9] whereas in the presence of mutants L372F, R351A, R387A, and R431S, ANS exhibited a 5−10-fold reduction in fluorescence intensity compared that of wild-type Cε3 (Figure 3A). A decrease in the level of ANS binding indicates a loss of molten globule structures, but this could occur by either stabilizing or destabilizing the structure; therefore, it was important to confirm the stabilizing effect of mutations using other methods.

The secondary structure and tertiary structure for the Cε3 domain and Cε3 mutants were assessed using CD spectroscopy. The comparison of the secondary structure content computed using the K2D2 algorithm[30] from far-UV CD spectra of the mutant domains with the wild-type domain indicated comparable values of ∼33% β-sheet structure. The only exception was the mutant L372F, which showed a small increase in estimated β-sheet content [to ∼40% (data not shown)]. The significance of this change in secondary structure must be considered questionable because of the well-known difficulties in making precise estimates of β-sheet content from far-UV CD measurements.[45,46] In near-UV CD spectra, wild-type Cε3 shows very little ellipticity, indicating a lack of tertiary structure[9] (Figure 3B). In contrast, the L372F and R431S mutants exhibited distinctive signals for the aromatic region, indicating the packing of aromatic residues in an asymmetric environment. Figure 3B shows a comparison of the near-UV CD spectra of wild-type Cε3 with the mutants, indicating the presence of aromatic ellipticity for the mutants.

The one-dimensional ^{1}H NMR spectrum of wild-type Cε3 shows the broad lines and poor signal dispersion typically observed for molten globule-like structures.[9] Three mutants (R387A, R431S, and R376A/K380A) exhibited upfield-shifted methyl peaks in one-dimensional ^{1}H NMR spectra, which is a characteristic of folded proteins. In addition, two distinct signals for the two tryptophan residues of Cε3 were observed in the spectra of these mutants. A single averaged chemical shift position is observed for the tryptophan residues in the spectrum of wild-type Cε3. The one-dimensional ^{1}H NMR spectra of the

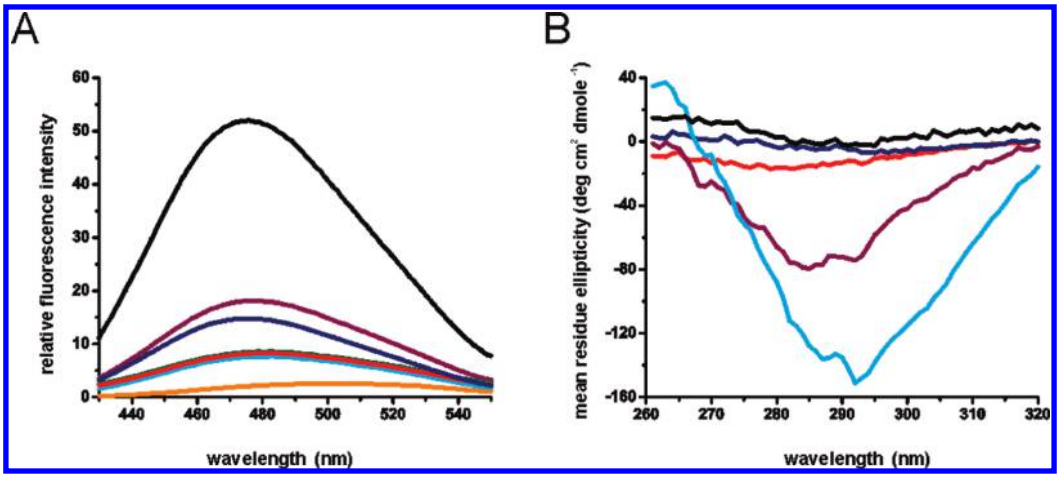


Figure 3. Effects of mutations on the biophysical characteristics of the Cε3 domain. (A) Comparison of the relative fluorescence intensity of ANS in the presence of wild-type and mutant Cε3 domains. Emission spectra of 25 μM free ANS (orange) and ANS in the presence of 20 μM wild-type Cε3 (black), Cε3 L372F (purple), Cε3 R376A/K380A (blue), Cε3 R351A (green), Cε3 R387A (red), and Cε3 R431S (cyan). (B) Overlay of near-UV CD spectra of the wild-type Cε3 domain and some mutant Cε3 domains: wild-type Cε3 (black), Cε3 R376A/K380A (blue), Cε3 R387A (red), Cε3 L372F (purple), and Cε3 R431S (cyan). Spectra were recorded at a protein concentration of 3 mg/mL in 0.01 M phosphate (pH 5.0) with a cuvette path length of 5 mm. The mutant Cε3 domains give distinctive near-UV signals, consistent with the formation of a hydrophobic core and packing of aromatic residues into a unique asymmetric environment.

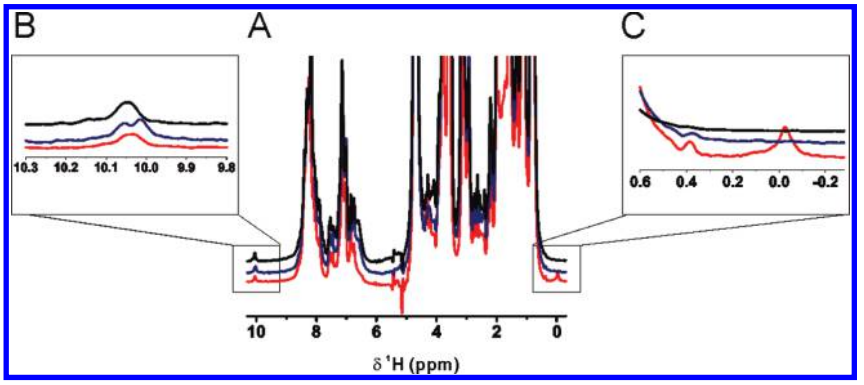


Figure 4. Overlay of one-dimensional ^{1}H NMR spectra for wild-type and mutant Cε3 domains. (A) Overlay of the one-dimensional ^{1}H NMR spectra of Cε3 R387A (red), Cε3 R376A/K380A (blue), and wild-type Cε3 (black), recorded with a 200 μM sample in PBS (pH 5.0) at 293 K in a magnet at 500 MHz. (B) Magnified region of the overlaid spectra showing chemical shifts for the side chains of the two tryptophan residues. (C) Region of the spectra showing chemical shifts for the upfield-shifted methyl peaks.

mutants R387A and R376A/K380A are compared with the wild-type spectrum in Figure 4.

DISCUSSION

The IgSF is one of the largest and most diverse of protein superfamilies. It has evolved to mediate protein−protein interactions in diverse interaction networks. The members of the family share a common structural motif, the Ig fold. The Cε3 domain retains the highly conserved Ig fold disulfide bond and residues forming the hydrophobic core. However, residues at some positions within the hydrophobic core are exchanged for residues with smaller volumes. Residues with smaller side chains would form less optimal van der Waals contacts, and this appears to affect the packing of the Cε3 core.

A striking feature of the sequence of the Cε3 domain is the presence of a large number of positively charged residues. The presence of high net charge on a molecule interferes with the formation of a compact hydrophobic core due to electrostatic repulsion. High net charge, low sequence complexity, and low hydrophobicity make up the hallmark signature of intrinsically unstructured proteins.[22,47,48]

Protein structures can be remarkably tolerant to mutations.[49] Evolution optimizes functional factors such as the folding rate and stability of a particular fold. There are specific positions in every fold that are highly conserved.[50] The sequence of a native protein is usually optimized to form favorable folding contacts. Mutations that replace the residues involved in forming the essential contacts can be highly destabilizing.[49] An understanding of the universally conserved residues of a particular fold can therefore be used to stabilize or destabilize a particular domain. Rational engineering conducted to stabilize the Cε3 domain demonstrated the influence of net charge and the canonical hydrophobic residues of Ig C1-set domains on the stability of the Cε3 domain. A number of point mutations that reduced the net charge on the domain increased the degree of structural stability

of the domain. Similarly, replacing the residues within Cε3 with canonical residues found in the Cγ2 domain (e.g., L372F) also resulted in a decrease in the flexibility of the domain. However, it was often observed that combining two or three mutations, which were effective as individual mutations, did not have an additive effect on the reduction of the flexibility of the domain, suggesting that the effects of the mutations are not linear.

This study has shown that the intrinsic flexibility of the Cε3 domain is encoded by its sequence. Cε3 within the context of IgE is less flexible than the isolated domain, but there is evidence that suggests that Cε3 retains an unusual level of flexibility within the whole protein. We propose that the intrinsic flexibility of the Cε3 domain has evolved under positive pressure, because it provides functional advantages to IgE. The Cε3 domain contains the binding sites for both the high-affinity receptor, FcεRI, and the low-affinity receptor, CD23. The two interactions have very different binding modes and thermodynamic signatures. Cε3's structural plasticity may impart versatility to the interactions mediated by this domain. Further, the binding of FcεRI has been shown to induce folding of the isolated Cε3 domain.[9] Coupled folding and binding has been suggested to provide a number of advantages in protein−protein interactions: it allows proteins to bind with high specificity and to multiple partners. Both of these abilities are deemed highly advantageous for proteins involved in complex interaction networks. The engineered domains of Cε3 with various levels of rigidity could serve as important tools in determining the role of flexibility of the Cε3 domain in IgE's biological functions.

■ ASSOCIATED CONTENT

Supporting Information. A figure showing structure-based sequence alignment of the Ig C1-set domains, a figure showing a comparison of the results of disorder prediction by different algorithms, a figure showing the tabulated data of DISOPRED prediction for Cε3, and a table comparing the total number of van der Waals contacts made by Leu372 of the Cε3 domain and Phe275 of the Cγ2 domain. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*Telephone: +44 (0) 207 848 6970. Fax: +44 (0) 207 848 6435. E-mail: james.mcdonnell@kcl.ac.uk.

■ ABBREVIATIONS

IgE, immunoglobulin E; IGSF, immunoglobulin superfamily; ANS, 8-anilinonaphthalene-1-sulfonate; CD, circular dichroism; NMR, nuclear magnetic resonance.

■ REFERENCES

(1) Capron, A., Dessaint, J. P., Haque, A., and Capron, M. (1982) Antibody-dependent cell-mediated cytotoxicity against parasites. *Prog. Allergy 31*, 234–267.

(2) Capron, M., and Capron, A. (1994) Immunoglobulin E and effector cells in schistosomiasis. *Science 24*, 1876–1877.

(3) Dombrowicz, D., and Capron, M. (2001) Eosinophils, allergy and parasites. *Curr. Opin. Immunol. 13*, 716–720.

(4) Gould, H. J., and Sutton, B. J. (2008) IgE in allergy and asthma today. *Nat. Rev. Immunol. 8*, 205–217.

(5) Gould, H., Sutton, B. J., Beavil, A. J., Beavil, R. L., McCloskey, N., Coker, H. A., Fear, D. A., and Smurthwaite, L. (2003) The biology of IgE and the basis of allergic disease. *Annu. Rev. Immunol. 21*, 579–628.

(6) McDonnell, J. M., Calvert, R., Beavil, R. L., Beavil, A. J., Henry, A. J., Sutton, B. J., Gould, H. J., and Cowburn, D. (2001) The structure of the IgE Cε2 domain and its role in stabilizing the complex with its high-affinity receptor FcεRIα. *Nat. Struct. Biol. 8*, 437–441.

(7) Woof, J. M., and Burton, D. R. (2004) Human antibody-Fc receptor interactions illuminated by crystal structures. *Nat. Rev. Immunol. 4*, 89–99.

(8) Dorrington, K. J., and Bennich, H. H. (1978) Structure-function relationships in human immunoglobulin E. *Immunol. Rev. 41*, 3–25.

(9) Price, N. E., Price, N. C., Kelly, S. M., and McDonnell, J. M. (2005) The Key Role of Protein Flexibility in Modulating IgE Interactions. *J. Biol. Chem. 280*, 2314–2330.

(10) Hunt, J., Beavil, R. L., Calvert, R. A., Gould, H. J., Sutton, B. J., and Beavil, A. J. (2005) Disulfide linkage controls the affinity and stoichiometry of IgE Fcε3−4 binding to FcεRI. *J. Biol. Chem. 280*, 16808–16814.

(11) Lesk, A. M., and Chothia, C. (1982) Evolution of proteins formed by β-sheets II. The core of the immunoglobulin domains. *J. Mol. Biol. 160*, 325–342.

(12) Maynard, J., and Georgiou, G. (2000) Antibody engineering. *Annu. Rev. Biomed. Eng. 2*, 339–376.

(13) Tompa, P. (2005) The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett. 579*, 3346–3354.

(14) Wurzburg, B. A., Garman, S. C., and Jardetzky, T. S. (2000) Structure of the human IgE-Fc Cε3-Cε4 reveals conformational flexibility in the antibody effector domains. *Immunity 13*, 375–385.

(15) Finn, R. D., Tate, J., Mistry, J., Coggill, P. C., Sammut, S. J., Hotz, H.-R., Ceric, G., Forslund, K., Eddy, S. R., Sonnhammer, E. L. L., and Bateman, A. (2008) The Pfam protein families database. *Nucleic Acids Res. 36*, D281–D288.

(16) Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol. 247*, 536–540.

(17) Giudicelli, V., Chaume, D., Bodmer, J., Muller, W., Busin, C., Marsh, S., Bontrop, R., Marc, L., Malik, A., and Lefranc, M. P. (1997) IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res. 25*, 206–211.

(18) UniProt Consortium (2010) The Universal Protein Resource. *Nucleic Acids Res. 38*, D142–D148.

(19) Jain, E., Bairoch, A., Duvaud, S., Phan, I., Redesachi, N., Suzek, B. E., Martin, M. J., McGarvey, P., and Gasteiger, E. (2009) Infrastructure for the life sciences: Design and implementation of the UniProt website. *BMC Bioinf. 10*, 136.

(20) Hennecke, J., and Wiley, D. C. (2002) Structure of a complex of the human α/β T cell receptor (TCR) HA1.7, influenza hemagglutinin peptide, and major histocompatibility complex class II molecule, HLA-DR4 (DRA*0101 and DRB1*0401): Insight into TCR cross-restriction and alloreactivity. *J. Exp. Med. 195*, 571–581.

(21) Notredame, C., Higgins, D. G., and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol. 302*, 205–217.

(22) Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000) Why are "natively unfolded" proteins unstructured under physiological conditions? *Proteins 41*, 415–427.

(23) Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., Silman, I., and Sussman, J. L. (2005) FoldIndex: A simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics 21*, 3435–3438.

(24) Romero, O., and Dunker, K. (1997) Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family. *Genome Informatics. Workshop on Genome Informatics 8*, 110–124.

(25) Li, X., Romero, P., Rani, M., Dunker, A. K., and Obradovic, Z. (1999) Predicting Protein Disorder for N-, C-, and Internal Regions. *Genome Informatics. Workshop on Genome Informatics 10*, 30–40.

(26) Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) Sequence complexity of disordered protein. *Proteins 42*, 38–48.

(27) Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J., and Russell, R. B. (2003) Protein disorder prediction: Implications for structural proteomics. *Structure 11*, 1453–1459.

(28) Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F., and Jones, D. T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol. 337*, 635–645.

(29) Collaborative Computational Project, Number 4. (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D50*, 760–763.

(30) Perez-Iratxeta, C., and Andrade-Navarro, M. A. (2008) K2D2: Estimation of protein secondary structure from circular dichroism spectra. *BMC Struct. Biol. 8*, 25.

(31) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR 6*, 277–293.

(32) Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., and Obradovic, Z. (2001) Intrinsically disordered protein. *J. Mol. Graphics Modell. 19*, 26–59.

(33) Bork, P., Holm, L. A., and Sander, C. (1994) The Immunoglobulin Fold Structural Classification, Sequence Patterns and Common Core. *J. Mol. Biol. 242*, 309–320.

(34) Poupon, A., and Mornon, J. P. (1998) Populations of hydrophobic amino acids within protein globular domains: Identification of conserved "topohydrophobic" positions. *Proteins 33*, 329–342.

(35) Halaby, D. M., Poupon, A., and Mornon, J. (1999) The immunoglobulin fold family: Sequence analysis and 3D structure comparisons. *Protein Eng. 12*, 563–571.

(36) Brendel, V., Bucher, P., Nourbakhsh, I., Blaisdell, B. E., and Karlin, S. (1992) Methods and algorithms for statistical analysis of protein sequences. *Proc. Natl. Acad. Sci. U.S.A. 89*, 2002–2006.

(37) Feige, M. J., Walter, S., and Buchner, J. (2004) Folding Mechanism of the CH2 Antibody Domain. *J. Mol. Biol. 344*, 107–118.

(38) Prabakaran, P., Vu, B. K., Gan, J., Feng, Y., Dimitrov, D. S., and Ji, X. (2008) Structure of an isolated unglycosylated antibody CH2 domain. *Acta Crystallogr. D64*, 1062–1067.

(39) Pamer, E., and Cresswell, P. (1998) Mechanisms of MHC class I-restricted antigen processing. *Annu. Rev. Immunol. 16*, 323–358.

(40) Maenaka, K., and Jones, E. Y. (1999) MHC superfamily structure and the immune system. *Curr. Opin. Struct. Biol. 9*, 745–753.

(41) Sondermann, P., Huber, R., Oosthuizen, V., and Jacob, U. (2000) The 3.2 Å Crystal Structure of the Human IgG1 Fc Fragment-Fc-γ-RIII Complex. *Nature 406*, 267.

(42) Borbulevych, O. Y., Do, P., and Baker, B. M. (2010) Structures of native and affinity-enhanced WT1 epitopes bound to HLA-A*0201: Implications for WT1-based cancer therapeutics. *Mol. Immunol. 47*, 2519–2524.

(43) Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas', A. F., and Gilmanshin, R. I. (1991) Study of the 'molten globule' intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers 31*, 119–128.

(44) Bychkova, V. E., Berni, R., Rossi, G. L., Kutyshenko, V. P., and Ptitsyn, O. B. (1992) Retinol-binding protein is in the molten globule state at low pH. *Biochemistry 31*, 7566–7571.

(45) Wallace, B. A., and Teeters, C. L. (1987) Differential absorption flattening optical effects are significant in the circular dichroism spectra of large membrane fragments. *Biochemistry 26*, 65–70.

(46) Whitmore, L., and Wallace, B. A. (2008) Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers 89*, 392–400.

(47) Wootton, J. C. (1994) Non-globular domains in protein sequences: Automated segmentation using complexity measures. *Comput. Chem. 18*, 269–285.

(48) Romero, P., Obradovic, Z., and Dunker, A. K. (1999) Folding minimal sequences: The lower bound for sequence complexity of globular proteins. *FEBS Lett. 462*, 363–367.

(49) Tiana, G., Broglia, R. A., Roman, H. E., Vigezzi, E., and Shakhnovich, E. (1998) Folding and misfolding of designed proteinlike chains with mutations. *J. Chem. Phys. 108*, 757–761.

(50) Mirny, L. A., and Shakhnovich, E. I. (1999) Universally conserved positions in protein folds: Reading evolutionary signals about stability, folding kinetics and function. *J. Mol. Biol. 291*, 177–196.