Published in final edited form as:

J Struct Biol. 2009 November ; 168(2): 223–233. doi:10.1016/j.jsb.2009.06.003.

Characterization of the kringle fold and identification of a ubiquitous new class of disulfide rotamers

Olga A. Ozhogina and Emile L. Bominaar

Department of Chemistry, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA

Abstract

The disulfide-bridged chains in the kringle (K) and fibronectin type II (FN2) domains are characterized using a taxonomy that considers the regularities in both β -secondary structure and cystine cluster. The structural core of the kringle fold comprises an assembly of two β -hairpins (a “ β meander”) accommodating two overlapping disulfides; one cystine is incorporated in adjacent β -strands, whereas the other is located just beyond the ends of non-adjacent β -strands. The dispositions of the (N, C) termini of the two overlapping disulfides in the kringle fold are given as $(m, j+1)$ and $(i-1, k+1)$, in which m, i, j , and k ($m < i < j < k$) are residues fulfilling the relations $m \sim_w j+3$ and $i \sim_n j \sim_w k$, where the relationship $\sim_{n/w}$ associates residues belonging to a narrow/wide hydrogen-bonded pair of an antiparallel β -sheet. This pattern is the structural signature of the kringle fold and is referred to as the “disulfide kringle-cross”. The metrics of this motif are quantified, revealing structural differences between the two families of the kringle fold. The conformations of disulfides in the kringle fold are poorly accommodated by existing classification schemes. To elucidate the nature of these rotamers we have performed density functional theory (DFT) calculations for diethyl disulfide. A new classification for the disulfide conformations in proteins is proposed, consisting of six rotamer types: spiral, trans-spiral, corner, trans, hook, and staple. Its relation with previous classification schemes is specified. A survey of high-resolution X-ray structures reveals that the disulfide conformations are clustered around the averaged conformations for the six classes. Average conformation dihedral and distance values are in excellent agreement with the DFT values. The two overlapping disulfides in kringle domains adopt the trans-spiral conformation that appears to be ubiquitous ($\sim 17\%$) in proteins. One of the disulfides stretches across the β -meander, invoking “strain” in the disulfide conformational state. The relevance of the new classification and the concept of strain are briefly discussed in the context of disulfide bond cleavage in proteins.

Keywords

Cystine; Density functional theory (DFT); Disulfide; Disulfide rotamer classification; Disulfide-rich fold; Fibronectin type 2 domain (FN2); Kringle domain; Neurotrypsin kringle domain; Protein structure comparison

1. Introduction

Since structure is better conserved than sequence, structural similarity between proteins may indicate an evolutionary relationship, even if their sequence homology is low [1-3]. Kringle

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

(K) and fibronectin type 2 (FN2) domains share a sequence identity of 15% or less but are considered to be divergent members of the kringle fold. The K domains are ~ 80 amino acid residue-long proteins with three disulfides. The FN2 domains comprise ~ 40 amino acids residues with two disulfides and have been proposed to have evolved from the K domains through partial deletion of residues from the terminal strands [4,5]. Since the N- and C-termini of the neurotrypsin kringle (NT/K) are shorter than in other K domains, it seemed attractive to hypothesize that NT/K represents a structural intermediate in the evolutionary transition from K to FN2 domains [5]. This example illustrates the dilemma of what constitutes a meaningful definition of “fold” in the evolutionary context: either an exact recapitulation of the chain path or just a common structural core of the chain?

The kringle fold has been characterized as a three-disulfide, triple-loop folding domain with overlapping inner disulfides in conventional protein structure taxonomy [6,7]. However, this definition is too restrictive because it excludes the two-disulfide containing proteins of the FN2 family and too permissive because it includes proteins from other folds, for example the periplasmic binding protein-like II fold (e.g. serum transferrin, milk lactoferrin, and avian ovotransferrin). Alternatively, Cheek et al. [8] have used the concept of a common structural core for assigning both families to the kringle fold, which they described as a pattern of two overlapping disulfides and five anti-parallel β -strands, arranged in “two perpendicular layers”. Although this qualitative description provides a family-specific sketch of the β -structure of the FN2 domain, it incorrectly depicts the conserved structural unit of the kringle superfamily. Classification schemes, such as SCOP [9] and CATH [10], describe the kringle fold in different ways. CATH assigns the kringle fold to the “mainly β/β barrel” of the conventional fold classification. SCOP recognizes the class of small disulfide-rich folds (SDF) and groups the K and FN2 families together into the kringle fold under the denominator “disulfide-rich fold; nearly all β ”, but does not quantify their structural relationship. Harrison and Sternberg [11] have proposed a rigorous SDF classification based on the regularities in cystine clusters and their relationship with β -secondary structure, that they connect. Using this classification, we present here an in-depth analysis of the structural features of the kringle fold, including the geometries, mutual arrangements, and dispositions of the disulfides relative to the β -structure.

We find that the conformations of the overlapping disulfides are conserved in the kringle fold but that they are poorly described by the disulfide rotamer classes proposed in the protein literature. Previous analyses of the disulfide conformations in proteins employed visual inspection for classification with an emphasis on the side-chain dihedral angles, in particular χ_2 , χ_3 and χ_2' [11-13]. Richardson sorted the disulfides into two major classes: left-handed spirals and right-handed hook [12]. Hutchinson and Thornton loosely grouped disulfides based on the signs of these angles into four categories: left-handed and right-handed spirals, right-handed hook and short right-handed hook [13]. Harrison and Sternberg termed the short right-handed hook as the right-handed staple and identified the right-handed corner conformation [11]. The cystine rotamer classification in proteins presented here differs from previous ones in that there are six categories on the basis of density functional theory calculations for diethyl disulfide. The relevance of the new classification is discussed in the context of disulfide bond cleavage in proteins, in particular the potential role of disulfide rotamer strain in the release of angiostatin from plasmin.

2. Methods

2.1. Conservation scores for residues in multiple sequence alignments

A conservation score was calculated for each column of the multiple alignment within families using Scorecons [14]. The scores range from 0 to 1, in which columns with all identities are assigned the highest value and the others get a score according to the degree of similarity between their residues.

2.2. Density functional theory calculations

Density functional theory calculations were performed using Becke's three parameter hybrid functional (B3LYP) and basis set 6-311G* provided by the Gaussian 03 (release B.05) software package [15]. The geometry optimizations were terminated upon reaching the default convergence criteria.

2.3. Data set construction

A non-redundant data set of disulfide-bridged chains was constructed from the PDB database as of January 14, 2009. The high resolution structures with disulfide bond count ≤ 20 were selected using 30% sequence identity. Only single chain (monomeric) proteins were included in our analysis. This list was further refined using resolution ≤ 2.0 Å and R-factor ≤ 0.30 (198 structure hits).

3. Results and discussion

3.1. Characterization of the kringle fold

3.1.1. Amino acid sequence alignments—Multiple sequence alignments (Figure 1) show that (i) the members of the K family share a sequence identity of $\sim 30\%$; (ii) The members of the FN2 family share a sequence identity of $\sim 50\%$; (iii) The sequence identity between the two families is below 15%. The two inner disulfide bridges cystine *b* and cystine *c* occur at identical positions in all sequences, and rank among the five strictly conserved residues in the kringle superfamily (Figure 1). The outer disulfide bridge, cystine *a*, connecting the residues near the N- and C-termini, is a unique feature within the K family, but is not present in the FN2 family. In NT/K the N- and C-termini are five and three residues shorter, respectively, than in other proteins within the K family. The shortening does not affect the specifically conserved positions in the K family (blue in Figure 1), however. NT/K fits the profile of a true member of an ancient K lineage, having $\sim 30\%$ identity in sequence alignment.

3.1.2. Kringle topology—In Figure 2 X-ray structures of the plasminogen kringle 4 [6] (PGN/K4) and the second type II module from matrix metalloproteinase 2 [16] (MMP2/2) domains, distinct examples representing the K and FN2 families, have been chosen for pictorial comparison with the NT/K structure recently determined by NMR [17]. To facilitate our understanding of hydrogen bonding pattern and disposition of disulfides with respect to β -secondary structure in kringles, we have projected the primary sequence of NT/K on a plane and aligned hydrogen bonds horizontally such that the residues appear in arrays (Figure 3A^a). Residues that belong to the same array are referred to as being “in register” (for example, residues N10–C18–Y60–D68). The secondary-structure elements have been assigned using the Kabsch & Sander notations [18]. An inspection of structures reveals that the core, consisting of two antiparallel β -bridges (B and C) and β -ladder (D), is shared by members of the K and FN2 families.

Cystines *b* and *c* have a conserved disposition relative to the β -secondary structure that they connect. The N- and C-termini of both cystines *b* and *c* are separated by one array of residues (Figure 3A). The N-terminal half-cystine residue Cys18_b is in register with conserved residue Tyr60, which is two residues up along the backbone with respect to the C-terminal half-cystine residue Cys58_b. The N-terminal half-cystine residue Cys47_c is in register with Gly70, which is two residues down along the backbone with respect to the C-terminal half-cystine residue Cys72_c (Figure 3A). The two N-terminal half-cystine residues (Cys18_b and Cys47_c) are

^aAs a note of caution, we emphasize that Figure 3A illustrates the *topology* of the network of hydrogen bonds and disulphide bridges but not protein *geometry*. For example, one should not be misguided by the “antiparallel” and a “parallel” orientations for the flanking β -strands of cystines *b* and *c* in Figure 3A, respectively.

separated by the loop 1. The N-terminal half-cystine residue Cys47_c is in register with the C-terminal half-cystine residue Cys58_b. Residues Cys18_b and Cys58_b of cystine *b* are incorporated in adjacent β -strands, whereas residues Cys47_c and Cys72_c of cystine *c* are both located just beyond the ends of non-adjacent β -strands (Figure 3A). Cystine *b* links β -bridge B to the core β -ladder D whereas cystine *c* crosses all over the β -meander. These relationships can be formalized as follows. If we use the symbol $\sim_{n/w}$ to indicate a pair of residues forming a narrow/wide hydrogen-bonded pair of an antiparallel β -sheet, then the dispositions of the (N, C) termini of cystines *b* and *c* are given as (m, j+1) and (i-1, k+1), where m, i, j, and k ($m < i < j < k$) are residues fulfilling the relations $m \sim_w j+3$ and $i \sim_n j \sim_w k$ (in FN2, only one of the two hydrogen bonds between residues m and j+3 is present, see below). These relationships have been summarized in Figure 3B where arrays of residues that are in-register have been given by dashed lines.

The β -secondary structure of the core is conserved in the kringle superfamily, apart from a minor variation between the K and FN2 families. In the K family structures both hydrogen bonds 19 $\rightarrow$ 59 and 61 $\rightarrow$ 17 within β -bridge B are conserved whereas in the FN2 module family only one of the two hydrogen bonds (19 $\rightarrow$ 59) is present (Figure 3A). The conservation of at least one hydrogen bond ensures the alignment of the linked residues into the arrays of Figure 3B throughout the kringle superfamily. Such close structural correspondence is striking given that the residues involved in hydrogen bonds 19 $\rightarrow$ 59 and 61 $\rightarrow$ 17 are not conserved between the two families (Figure 1). This observation highlights the importance of the interplay between disulfides and adjacent hydrogen bonding of the backbone in the stabilization of the kringle fold.

The divergence in β -secondary structure between the K and FN2 families is mainly due to a drastic reduction in the length of the loop 1 in the FN2 family, i.e. 12 residues compared to 26–29 residues in the K family (Table S1). Unlike NT/K and PGN/K4, MMP2/2 comprises an antiparallel β -sheet within loop 1 (Figure 2), which is a conserved feature within the FN2 family. An overlapping pattern of a left-handed polyproline II (PPII) helix (residues 17–20, containing a conserved half-cystine at position 18), with β -bridges A and B (11–17 and 18–60, containing conserved residues at positions 18, 59 and 60) is a conserved feature within the K family. In the FN2 family there is a shorter PPII-helix containing residue Cys18, but there is no equivalent of β -bridge A and only one hydrogen bond of B, viz. 19 $\rightarrow$ 59 (Figure 3A).

3.1.3. Core geometry—Three-dimensional alignments based on the superposition of the C $^{\alpha}$ atoms of the residues that participate in the conserved hydrogen bonding of disulfide-bridged β -sheet (Figure 3A) show differences in the core structure (Table S2). This motif contains the residues at positions 18, 19, 47, 48, 56–60, 68–72 (Figure 1); only five of them are strictly conserved (residues 18, 47, 57, 58 and 72) while the other residues of the motif show a wide range of conservancy scores between and within the families. The average RMSDs between the K and FN2 families (0.76 Å) are consistently larger than the averages within the families (0.41 Å). The difference of the averages (0.35 Å) for the inter- and intra-family comparisons represents a structural discrepancy between the K and FN2 families. The same pattern is observed when NT/K is compared with members of the K and FN2 families: NT/K shares consistently lower RMSDs with the K family. [N.B. The RMSDs for comparisons of NT/K (an NMR structure) are $\sim$ 0.1 Å larger than the RMSDs for the comparisons between the other structures (all X-ray structures). However, the difference of 0.34 Å between the average RMSDs for the NT/K–K and the NT/K–FN2 comparisons is about equal to the difference between the average RMSDs within and between the K and FN2 families.]

Comparison of the backbone orientation angles, κ , in the kringle superfamily shows that κ is $\sim$ 55° and $\sim$ 107° for cystines *b* and *c*, respectively (Table 1). Hence, the kringle fold conserves the parallel and orthogonal orientations for the flanking backbone segments of cystines *b* and

c, respectively. The polypeptide chain follows a right-screw path from Cys47_c to Cys72_c via Cys58_b.

The dihedral angles and distances for the cystine-pair cluster (Table 2), while confined to narrow ranges, exhibit significant family-specific differences in the averages. In the K family cystines *b* and *c* meet the two-cystine distance clustering criterion [11] (three D_{CαCα} values ≤ 7.5 Å) for closely clustered cystine pairs with average minimum distance of ~ 5.2 Å between Cys47_c and Cys58_b and negative cluster dihedrals $v_{bc}^{C^{\alpha}} \sim 90^{\circ}$ and $v_{bc}^{S^{\gamma}} \sim -146^{\circ}$, and $\theta_{bc}^{S^{\gamma}} \sim 31^{\circ}$. In the FN2 family the cystine pairs tend to be loosely clustered (two D_{CαCα} values ≤ 7.5 Å) with $v_{bc}^{C^{\alpha}} \sim -74^{\circ}$, $v_{bc}^{S^{\gamma}} \sim -111^{\circ}$, and $\theta_{bc}^{S^{\gamma}} \sim 56^{\circ}$. The K and FN2 families differ consistently in the cluster dihedrals, namely by ~ 16°, 35°, and 25° in $v_{bc}^{C^{\alpha}}$, $v_{bc}^{S^{\gamma}}$, and $\theta_{bc}^{S^{\gamma}}$, respectively (Table 2). In NT/K the dihedral angles and distances for the cystine-pair cluster are within the standard deviation from the average for the K family.

3.1.4. The disulfide kringle-cross—The analysis of the kringle fold given in the preceding sections can be summarized as follows: (1) there are two overlapping disulfides; the first half-cystine in the primary sequence (N1) is bridged to the third (C1) and the second (N2) to the fourth (C2). (2) The C-termini of the two disulfide bridges belong to the same β-hairpin and are separated by an array of residues. (3) The N-terminal half-cystine residue N2 is in register with the C-terminal half-cystine residue C1. (4) The backbone exhibits a right-screw path from N2 to C2 via C1. (5) The N1–C1 and N2–C2 cystines connect marginally parallel and nearly orthogonal flanking chain segments, respectively. (6) The two cystines are clustered and have negative values for the cluster dihedrals $v_{bc}^{C^{\alpha}}$ and $v_{bc}^{S^{\gamma}}$.

The arrangement of the disulfides in the kringle fold is reminiscent of the disulfide β-cross, which occurs in non-homologous domains with a variety of folds [11]: the two motifs share properties (1), partially (2), (5), and partially (6). However, there are also a number of distinctive differences: the C-terminal half-cystine residues Cys58_b and Cys72_c are out of register in terms of wide-pair hydrogen-bonding position^b; cystines *b* and *c* form a cluster with negative $v_{bc}^{C^{\alpha}}$ and $v_{bc}^{S^{\gamma}}$ values, unlike the majority of β-crosses, which have positive cluster dihedrals; the handedness of the backbone screw path is opposite to that in β-crosses. The two overlapping disulfides appear as a cross symbol when viewed along the line connecting the midpoints of the bonds (Figure 3C). To distinguish the disulfide pair in the kringle fold from the disulfide β-cross we have coined the term “disulfide kringle-cross”. The comparative analysis shows that the disulfide kringle-cross motif is present in both the K and FN2 families and is the defining feature of the kringle fold.

3.2. Disulfide conformations

3.2.1. Dihedrals of disulfides in kringles and other disulfide-rich proteins—Table 3 lists dihedrals (χ_1 , χ_2 , χ_3 , χ_2' , and χ_1') and distances (C^α–C^α and C^β–C^β) of disulfides in high-resolution kringle structures. To place these values in the context of the protein data at large, we have compared them with the phenomenological distributions given by Petersen et al. [19]. The following observations can be made: (a) $|\chi_3| \approx 90^{\circ}$ with $\chi_3 > 0$ and $\chi_3 < 0$ appearing with equal probabilities. Thus, from a general perspective, both the left- and right-handed cystines conformers are admissible in kringle structures. (b) The maxima in the distributions for χ_1 and χ_1' , which are quite similar, are centered at ~ -60°, ~ 60°, and 180°, with -60° being, by far, the most likely value. The χ_1 and χ_1' values for cystine *b* and the χ_1' for cystine

^bAccording to the definition of a disulfide β-cross the C-terminal half-cystine residues are in register with respect to the main-chain hydrogen-bonding of the β-hairpin at wide pair positions [5].

c ($\sim -60^\circ$) are in the preferred domain. The value $\chi_1' > 0$ for cystine c is less frequent and appears to be a special feature conserved throughout the kringle superfamily (Table 3). (c) For $\chi_3 < 0$ there is a high probability that both χ_2 and χ_2' are $\sim -60^\circ$ and a minor likelihood for finding $\sim 180^\circ$. The value $\chi_3 < 0$ yields in combination with the preferred values $\chi_2, \chi_2' \approx -60^\circ$ the most frequent disulfide rotamer, i.e. the left-handed spiral (class ggg, see below).

The conformations of the two overlapping disulfide bridges, cystine b and cystine c , do not belong to any of the rotamer types described in the protein literature [11-13]. To obtain a better understanding of the disulfide conformations, we have performed density functional theory (DFT) calculations (sections 3.2.2 and 3.2.3) [15] for the model system diethyl disulfide (EtSSEt). These calculations give also insight into the origin of the systematic differences (see Table 3) in the values of the dihedrals and $C^\alpha-C^\alpha$ distances for the two inner disulfides (section 3.2.6).

3.2.2. Rotamers of diethyl disulfide—Here we propose a six-class taxonomy for the disulfides in proteins on the basis of the rotamers for diethyl disulfide. The geometry of diethyl disulfide is characterized by the dihedrals directing the ethylene moieties, χ_2 and χ_2' , with values of $\sim \pm 60^\circ$ (*gauche*, $g^\pm$), $\sim 180^\circ$ (*trans*, t) and the central dihedral $\chi_3 \sim \pm 90^\circ$ (*gauche*, $g^\pm$), cf. Figure 4A. There are 18 combinations for $(\chi_2, \chi_3, \chi_2')$; the dihedrals χ_1 and χ_1' are not considered here because the backbone atoms adjacent to the C^α atoms have been replaced by hydrogens in the diethyl disulfide model. As shown in an elegant study by Görbitz [20], these combinations can be divided into six classes (Figure 4B). Three of these classes contain each 4 combinations representing rotamers of C_1 symmetry and the remaining three classes contain each 2 combinations representing rotamers of C_2 symmetry (Table 4); for example, class $tgg' = \{tg^+g^-, tg^-g^+, g^+g^-t, g^-g^+t\}$, and class $ggg = \{g^+g^+g^+, g^-g^-g^-\}$ (N.B. g and g' indicate dihedrals with opposite signs). Any two rotamers in which the dihedrals appear in reversed orders, e.g., as in the pair tg^+g^+ and g^+g^+t , represent identical diethyl disulfide molecules but are distinguishable in the context of the protein environment (with the left-hand symbol referring to N-terminus and the right-hand symbol to C-terminus of the disulfide). Thus, the classes for the diethyl disulfide model contain each 2 rotamers, namely, the left- and right-handed forms (giving a total of 12 rotamers), whereas the classes for the disulfides in proteins contain 4 and 2 rotamers for classes of C_1 and C_2 symmetry, respectively (giving a total of 18 rotamers). The 18 protein rotamers will be denoted: R-ggg, L-ggg (Class 1 or ggg); R-tgg, L-tgg, R-ggt, L-ggt (Class 2 or tgg); R-ggg', L-ggg', R-g'gg, L-g'gg (Class 3 or ggg'); R-tgt, L-tgt (Class 4 or tgt); R-tgg', L-tgg', R-g'gt, L-g'gt (Class 5 or tgg'); R-g'gg', L-g'gg' (Class 6 or g'gg').

The proposed six-class taxonomy for the disulfides in proteins is based on the following definitions for g , g' , and t : in the R (right-handed) form, $\chi_3 \in (0, 180^\circ)$ and $\chi_2, \chi_2' \in (0, 120^\circ)$ are denoted g , and $\chi_2, \chi_2' \in (-120^\circ, 0)$ are denoted g' . In the L form, $\chi_3 \in (-180^\circ, 0)$ and $\chi_2, \chi_2' \in (-120^\circ, 0)$ are denoted g , and $\chi_2, \chi_2' \in (0, 120^\circ)$ are denoted g' . Irrespective of the L- or R-form, $\chi_2, \chi_2' \in (120^\circ, 240^\circ)$ [or, equivalently, $\chi_2, \chi_2' \in (-240^\circ, -120^\circ)$] are denoted t . These definitions imply that any disulfide belongs to one, and only one, of the 18 rotamers and, consequently, to one, and only one, of the 6 rotamer classes.

The geometry of diethyl disulfide has been optimized for one representative from each of the six classes, using density functional theory (see Methods). The relative energies of the classes have been listed in Table 4 in the order of increasing energy. Figure S1 shows the minima for the classes ggg, tgg, and tgt in the potential energy surface of diethyl disulfide as a function of χ_2 and χ_2' in the angular range relevant for the kringle fold. Table 4 gives also the values of dihedrals and selected distances in the optimized structures. The listed dihedrals, although deviating from the ideal values given above, confirm the classification scheme. Figure 4B shows the optimized geometries for the six classes. Interestingly, ggg is the conformation of

diethyl disulfide with the lowest energy, suggesting that the ubiquity of this rotamer in proteins (see below) derives from an intrinsic property of the disulfide bridge. Class g'gg' is highest in energy, due to the steric repulsion arising from the short C^α–C^α distance. Class g'gg' is the only class for which the C^α–C^α distance is shorter than C^β–C^β (Table 4). The disulfide conformations span a remarkably broad range of C^α–C^α distances (~ 3 Å) at the expense of only minor energy penalties (Table 4). Given the smallness of the energy differences, these quantities inevitably depend on the choice of the basis set and density functional, affecting the results in details but not in essence. Thus, while it is tempting to consider the disulfides as the principal determinants of the kringle fold, it should not be overlooked that the great flexibility of these entities leaves considerable room for other interactions, e.g. hydrogen bonding, to shape the structure. The rotameric energies show a high degree of additivity with respect to changes in χ_2 and χ_2' : $E(\text{tgt}) \approx 2 E(\text{tgg})$ and $E(\text{tgg}') \approx E(\text{tgg}) + E(\text{ggg}')$. However, $E(\text{g'gg}') > 2 E(\text{ggg}')$, due to the short C^α–C^α distance in g'gg' (Table 4).

3.2.3. Potential energies for diethyl disulfide—The rotameric energies as a function of C^α–C^α distance for the classes with minima in the 5 – 7 Å range is presented in Figure 5. The lowest system energy stays, throughout the distance domain shown, within a narrow, sub kcal-per-mole energy range, attesting the flexibility of the disulfide bond. Compression of the tgt conformer gives, after passing a narrow transition range around ~ 5.85 Å, the tgg' conformer (Figure S2). The potential energy curve for ggg crosses the curves for tgg and tgt without changing conformation. The definition ranges for the potential wells in Figure 5 were limited by a lack of convergence in the geometry optimizations outside these ranges. Except for C^α–C^α distances beyond 6.7 Å (Figure 5), the disulfide bridge can exist in several, generally C^α–C^α-strained, conformations. The effect of C^α–C^α strain on the dihedrals for the ggg and tgg conformations is illustrated in Figure 6 (*cf.* section 3.2.6). Disulfides possess the conformational and structural flexibility needed for insertion into a wide range of dispositions with respect to the hydrogen bonding patterns of protein backbones.

3.2.4. Statistics for the six-class taxonomy of disulfide conformations in proteins—Table 5 presents the results of a statistical analysis of protein disulfides on the basis of the six-class taxonomy proposed in section 3.2.2. We have analyzed a test set of 453 disulfides in 198 high-resolution X-ray structures, selected with the criteria specified in Methods. The PDB codes of the crystal structures are given in Table S3 of the Supporting Information, together with the dihedrals, C^α–C^α distances, and rotamer type for each disulfide of the test set. The number of incidences, the corresponding percentages of the total number, and the averages of the inner dihedrals and C^α–C^α distances are listed in Table S4 for each of the 18 rotamers. The data for the 18 rotamers have subsequently been combined to obtain the statistics for the six rotamer classes presented in Table 5. The spiral (ggg) is the most frequent class of rotamers (42%), followed by the corner (ggg', 20%) and within a narrow margin the trans–spiral (tgg, 17%). Thus, although the trans–spiral is ignored in existing classification schemes (see section 3.2.5), it is actually an ubiquitous rotamer, which comes as no surprise given that tgg appears as the class with the second lowest energy in Table 4. The remaining three classes give a total of 21%. The deviations of the dihedral averages in Table 5 from the idealized values ($\pm 60^\circ$, 180° for χ_2 and χ_2' , and $\pm 90^\circ$ for χ_3) are closely matched by those of the DFT values listed in Table 4. For example, the large value found for χ_3 in the staple is in excellent agreement with the DFT result for χ_3 in this rotamer. Also the dependence of C^α–C^α distance on rotamer class (Table 4) is remarkably well reproduced by the DFT calculations (Table 5). Thus, while the conformations of the individual disulfides are distorted by interactions with the protein environment (Table S3), the class averages (Table 5) appear to have conserved the characteristics of the diethyl disulfide rotamers, which are the basis of the proposed six-class taxonomy (Table 4). To investigate to what extent the members of the classes are clustered around the averaged conformations we have refined the six-class

taxonomy by requiring the dihedral angles to belong to a set of narrower ranges centered at the averages, χ^i , for rotamers $i = 1 - 18$: $\chi_2 \in (\chi_2^i - \Delta\chi_2, \chi_2^i + \Delta\chi_2)$, $\chi_3 \in (\chi_3^i - \Delta\chi_3, \chi_3^i + \Delta\chi_3)$, and $\chi_2' \in (\chi_2'^i - \Delta\chi_2, \chi_2'^i + \Delta\chi_2)$. Using $\Delta\chi_2 = 40^\circ$ and $\Delta\chi_3 = 30^\circ$, instead of the original ranges $\Delta\chi_2 = 60^\circ$ and $\Delta\chi_3 = 90^\circ$, only 10% of the disulfides is found to elude classification (the classes are affected comparably, cf. Supporting Information, Tables S5 and S6), compared to 85% for the hypothetical case that the disulfides were dispersed randomly. Hence, the proposed six-class taxonomy is strongly supported by the protein structure data.

3.2.5. Comparison of six-class taxonomy with earlier classifications—Several classification schemes for the disulfide conformations in proteins have been proposed in the literature [11,13,21]. These schemes have in common that the side-chain dihedrals, χ_2 , χ_3 , and χ_2' , are classified according to their signs (+ or −), leading to a total of 3 classes (Table 6), while in our taxonomy χ_2 and χ_2' are classified as g, g', or t, resulting in a total of 6 classes (Table 4). The correlation between the three-class and six-class taxonomies is shown in Table 6. For example, the hook of the three-class taxonomy includes disulfides of the trans–spiral, corner, trans, and hook of the six-class taxonomy.

In consulting the literature one should be aware that the trivial names used for the rotamer classes in proteins depend on the source. Table 7 lists recurrent trivial names and how they are defined by various authors. For example, the terms “corner” as defined by Harrison and Sternberg [11] and “hook” as defined by Schmidt, Ho, and Hogg [21] refer to the same rotamer of the three-class taxonomy but describe distinct species in the six-class taxonomy. The motif unanimously called “staple” by the latter authors is featured as a “short hook” by Hutchinson and Thornton [13]. By definition, the use of the descriptors “trans” and “trans–spiral” is reserved for the six-class taxonomy (Table 7).

3.2.6. Disulfide rotamer strain in the kringle fold—The values for χ_2 , χ_3 , and χ_2' listed in Table 3 indicate that cystines *b* and *c* belong to the class 2, tgg = {R-tgg, L-tgg, R-ggt, L-ggt}, of trans–spiral rotamers in both the K and FN2 families. The cystine *b* is invariably L-ggt, while cystine *c* is L-ggt with two exceptions (R-ggt in *lpk4* and *2fd6*).

Although cystines *b* and *c* belong to the same class (tgg), their C^α–C^α distances (5.9 Å for *b* and 6.5 Å for *c*) are remarkably different (Table 3). The distance in cystine *b* nearly coincides with the energy minimum for conformation tgg (Figure 5) located at 5.84 Å (Table 4) and is thus “relaxed”. The C^α–C^α distance in cystine *c* is 0.6 Å longer than at the tgg minimum because the disulfide stretches across the β-meander (Figure 3A), invoking “strain” in the disulfide conformational state. A similar situation is found in all kringle domains (Table 3). Cystine *b* in NT/K has a conformation with $\chi_2 = -69^\circ$, $\chi_3 = -80^\circ$, $\chi_2' = -172^\circ$ versus -71° , -90° , -176° obtained by the DFT calculations at 5.9 Å for class tgg (Figure 6), yielding an average deviation of only 5° for the three dihedrals. Cystine *c* in NT/K has a conformation with $\chi_2 = -97^\circ$, $\chi_3 = -96^\circ$, $\chi_2' = -157^\circ$ versus -94° , -96° , -165° calculated by DFT at 6.5 Å (Figure 6), giving an average deviation of 4°. A comparison of cystines *c* and *b* reveals significant changes in the dihedrals: $\Delta\chi_2 = -28^\circ / -23^\circ$, $\Delta\chi_3 = -16^\circ / -6^\circ$, and $\Delta\chi_2' = +15^\circ / +11^\circ$ (observed / calculated). The observed trends in both the magnitude and sign of the strain-induced changes in the dihedrals are consistent with the DFT calculations.

Figure 5 shows that the observed tgg conformations are not the lowest-energy conformations for diethyl disulfide at the C^α–C^α distances of cystines *b* and *c*: ggg and tgt are the lower in energy than tgg at the C^α–C^α values for cystine *b* (5.9 Å) and cystine *c* (6.5 Å), respectively, albeit by small margins of 100–300 cm^{−1}. Given the smallness of these energies it is not surprising that the inclusion of interactions ignored in the diethyl disulfide model, such as hydrogen bonding with the backbone, may alter the order of the conformational energies.

3.2.7. Disulfide class and function—Disulfides in proteins have been functionally classified as structural, catalytic, or allosteric bonds [21]. Structural bonds function as stabilizing units of protein structure; catalytic bonds are involved in electron storage by cycling between oxidized (disulfide) and reduced (dithiol) conformations, and allosteric bonds act as potential redox switches for protein function. An attempt to establish a relationship between functional and structural classes has been reported by Hogg and coworkers [21]. Pursuing the idea that a high conformational energy presents a favorable condition for allosteric function, they have evaluated the energies of a large number of disulfides, using the AMBER force field [22], and proposed that the allosteric and catalytic bonds belong, respectively, to the structural classes of the “staple” and “hook” rotamers. This proposal is tempting but also raises a number of issues. (1) AMBER is not a reliable tool for calculating conformational energies of disulfide bonds. The AMBER force field accounts only for torsion energies but ignores distance-dependent contributions due to steric interactions. As a result, the AMBER force field underestimates the steric repulsion arising from the close contact between the C $^{\alpha}$ atoms in the staple conformation of the six-class taxonomy, which sets this rotamer apart in terms of potential energy among the six classes (Table 4, *cf.* section 3.2.2). (2) The meaning of the term “strain energy” is ambiguous. In section 3.2.6 strain energy is defined as the distortion energy of a disulfide relative to the equilibrium conformation for the class to which the disulfide belongs. In our definition, a disulfide with a relaxed staple conformation has zero strain energy. Confusingly, the same disulfide appears as severely “strained” if the equilibrium energy of the spiral is taken as a reference [21]. We note that among the fold-defining overlapping disulfides in the plasminogen kringle 5 (PGN/K5), *viz.* the trans-spirals *b* and *c* in the six-class taxonomy, it is the strained disulfide (cystine *c*, see section 3.2.6) that is subject to cleavage. This observation lends support to the relevance of the strain concept as defined in this paper in the context of disulfide cleavage. (3) The structural classes used in ref. [21] refer to the inadequate three-class taxonomy (Table 6). For example, the “staple” of the three-class taxonomy comprises disulfides from the trans, hook, and staple classes in the six-class taxonomy (Table 6). As disulfide conformations are distinct in geometry and relative energy (Figure 4 and Table 4), the six-class taxonomy is more accurate than the three-class taxonomy for exploring the structural origin of the functional differentiation of disulfides. (4) There are notable exceptions to the proposed rule. For example, the “allosteric” disulfide cystine *c* in PGN/K5, which is cleaved in the formation of angiostatin and microplasmin from plasmin [23], is a spiral and not a staple in the context of the three-class taxonomy (N.B. Cystine *c* is trans-spiral in the six-class taxonomy, *cf.* point (3)).

It is important to bear in mind that insight into quantitative structure–function relationship is only attainable by tackling the problem from a mechanistic perspective. In this context, our preliminary DFT calculations on disulfides within a protein are revealing, namely by pointing toward the importance of hydrogen bonding interactions between the backbone amide groups and the disulfides in determining its energy and conformation. As hydrogen bonds are known to have a major impact on redox potentials, for example those of iron–sulfur proteins [24], they are expected to be one of the key determinants of allosteric or catalytic function of disulfides. Additionally, hydrogen bonding may play a role in ensuring the reversibility of disulfide cleavage by keeping the resulting thiols in the immediate spatial vicinity.

4. Conclusions

We have characterized the kringle fold on the basis of both the disposition of cystines relative to the conserved hydrogen bonding pattern of the backbone and cystine descriptors. The metrics of the kringle fold are quantified, revealing structural differences in the geometrical descriptors of the core, including dihedral angles and distances for the cystine-pair cluster, between the K and FN2 families of the kringle superfamily. The geometrical descriptors of the core together with family specific secondary structure features provide quantitative criteria for protein

structure comparison and evolutionary relations between families. By reason of these criteria, the neurotrypsin kringle domain (i) belongs to the kringle fold, (ii) has the conserved arrangement of cystine clusters and the secondary structure of the K family, and (iii) lacks the distinct conserved β -structure of the FN2 family. Hence, NT/K is an unambiguous member of the K family and not a structural intermediate between the K and FN2 families as it was previously suggested [5].

A new classification for the disulfide conformations in proteins is proposed on the basis of density functional theory calculations for diethyl disulfide. This taxonomy differs from previous three-class taxonomy in that there are six rotamer types: spiral, trans-spiral, corner, trans, hook, and staple. The six-class taxonomy is confirmed by statistical analysis of high-resolution X-ray protein structures. The trans-spiral conformation, adopted by the disulfides of the kringle fold, is identified as a ubiquitous new class of disulfide rotamers in proteins.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was sponsored by NIH Grants HL-29409 and EB-0001474.

References

1. Doolittle RF. Similar amino acid sequences - chance or common ancestry. *Science* 1981;214:149–159. [PubMed: 7280687]
2. Chothia C, Lesk AM. The Relation between the divergence of sequence and structure in proteins. *EMBO Journal* 1986;5:823–826. [PubMed: 3709526]
3. Holm L, Sander C. New structure - novel fold? *Structure* 1997;5:165–171. [PubMed: 9032077]
4. Patthy L. Evolution of the Proteases of Blood-Coagulation and Fibrinolysis by Assembly from Modules. *Cell* 1985;41:657–663. [PubMed: 3891096]
5. Ozhogina OA, Trexler M, Banyai L, Llinas M, Patthy L. Origin of fibronectin type II (FN2) modules: Structural analyses of distantly-related members of the kringle family identify the kringle domain of neurotrypsin as a potential link between FN2 domains and kringles. *Protein Science* 2001;10:2114–2122. [PubMed: 11567102]
6. Mulichak AM, Tulinsky A, Ravichandran KG. Crystal and molecular structure of human plasminogen kringle 4 refined at 1.9 Å resolution. *Biochemistry* 1991;30:10576–10588. [PubMed: 1657148]
7. Gehrmann M, Briknarova K, Banyai L, Patthy L, Llinas M. The col-1 module of human matrix metalloproteinase-2 (MMP-2): Structural/functional relatedness between gelatin-binding fibronectin type II modules and lysine-binding kringle domains. *Biological Chemistry* 2002;383:137–148. [PubMed: 11928808]
8. Cheek S, Krishna SS, Grishin NV. Structural classification of small, disulfide-rich protein domains. *Journal of Molecular Biology* 2006;359:215–237. [PubMed: 16618491]
9. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP - a structural classification of proteins database for the investigation of sequences and structures. *Journal of Molecular Biology* 1995;247:536–540. [PubMed: 7723011]
10. Cuff AL, Sillitoe I, Lewis T, Redfern OC, Garratt R, Thornton J, Orengo CA. The CATH classification revisited-architectures reviewed and new ways to characterize structural divergence in superfamilies. *Nucleic Acids Research* 2009;37:D310–D314. [PubMed: 18996897]
11. Harrison PM, Sternberg MJE. The disulphide beta-cross: From cystine geometry and clustering to classification of small disulphide-rich protein folds. *Journal of Molecular Biology* 1996;264:603–623. [PubMed: 8969308]
12. Richardson JS. The anatomy and taxonomy of protein structure. *Advances in Protein Chemistry* 1981;34:167–339. [PubMed: 7020376]

13. Hutchinson EG, Thornton JM. PROMOTIF - A program to identify and analyze structural motifs in proteins. *Protein Science* 1996;5:212–220. [PubMed: 8745398]
14. Valdar WSJ, Thornton JM. Protein-protein interfaces: Analysis of amino acid conservation in homodimers. *Proteins-Structure Function and Genetics* 2001;42:108–124.
15. Frisch, MJ.; Trucks, GW.; Schlegel, HB.; Scuseria, GE.; Robb, MA.; Cheeseman, JR.; Montgomery, JA.; Vreven, T.; Kudin, KN.; Burant, JC.; Millam, JM.; Iyengar, SS.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, GA.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, JE.; Hratchian, HP.; Cross, JB.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, RE.; Yazyev, O.; Austin, AJ.; Cammi, R.; Pomelli, C.; Ochterski, JW.; Ayala, PY.; Morokuma, K.; Voth, GA.; Salvador, P.; Dannenberg, JJ.; Zakrzewski, VG.; Dapprich, S.; Daniels, AD.; Strain, MC.; Farkas, O.; Malick, DK.; Rabuck, AD.; Raghavachari, K.; Foresman, JB.; Ortiz, JV.; Cui, Q.; Baboul, AG.; Clifford, S.; Cioslowski, J.; Stefanov, BB.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, RL.; Fox, DJ.; Keith, T.; Al-Laham, MA.; Peng, CY.; Nanayakkara, A.; Challacombe, M.; Gill, PMW.; Johnson, B.; Chen, W.; Wong, MW.; Gonzalez, C.; Pople, JA. Gaussian 03, Revision B.05. Gaussian, Inc.; Pittsburgh PA: 2003.
16. Morgunova E, Tuuttila A, Bergmann U, Isupov M, Lindqvist Y, Schneider G, Tryggvason K. Structure of human pro-matrix metalloproteinase-2: Activation mechanism revealed. *Science* 1999;284:1667–1670. [PubMed: 10356396]
17. Ozhogina OA, Grishaev A, Bominaar EL, Patthy L, Trexler M, Llinas M. NMR Solution Structure of the Neurotrypsin Kringle Domain. *Biochemistry* 2008;47:12290–12298. [PubMed: 18956887]
18. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983;22:2577–2637. [PubMed: 6667333]
19. Petersen MTN, Jonson PH, Petersen SB. Amino acid neighbours and detailed conformational analysis of cysteines in proteins. *Protein Engineering* 1999;12:535–548. [PubMed: 10436079]
20. Gorbitz CH. Conformational properties of disulfide bridges .2. Rotational potentials of diethyl disulfide. *Journal of Physical Organic Chemistry* 1994;7:259–267.
21. Schmidt B, Ho L, Hogg PJ. Allosteric disulfide bonds. *Biochemistry* 2006;45:7429–7433. [PubMed: 16768438]
22. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner P. A New Force-Field for Molecular Mechanical Simulation of Nucleic-Acids and Proteins. *Journal of the American Chemical Society* 1984;106:765–784.
23. Stathakis P, Lay AJ, Fitzgerald M, Schlieker C, Matthias LJ, Hogg PJ. Angiostatin formation involves disulfide bond reduction and proteolysis in kringle 5 of plasmin. *Journal of Biological Chemistry* 1999;274:8910–8916. [PubMed: 10085135]
24. Stephens PJ, Jollie DR, Warshel A. Protein control of redox potentials of iron-sulfur proteins. *Chemical Reviews* 1996;96:2491–2513. [PubMed: 11848834]

Abbreviations

apoA/K4	apolipoprotein A kringle 4
DFT	density functional theory
FA12	factor XIIa
FN2	fibronectin type 2 domain
HABP2	hyaluronan-binding protein 2
HGF/K1	hepatocyte growth factor kringle 1
HGFA	hepatocyte growth factor activator
K	kringle domain
M6PR	mannose 6-phosphate receptor
MMP2/2	second domain of matrix metalloproteinase 2

NT/K	neurotrypsin kringle
PGN/K4	plasminogen kringle 4
PGN/K5	plasminogen kringle 5
PDB	Protein Data Bank
PPII	left-handed polyproline II helix
SCOP	structural classification of proteins
SDF	small disulfide-rich fold
THRB/K1	prothrombin kringle 1
tPA/K2	tissue plasminogen activator kringle 2
RMSD	root-mean-squares deviation
ROR1	tyrosine-protein kinase transmembrane receptor 1
uPA/K	urokinase-type plasminogen activator kringle

Fibronectin type 2

```

.....67.50. 9 00.8.65. 7 ..09..96.....00677.796.8556680
FN2/1 VTQTYGGNSNGEPV.L...PFTYNGRT.....F....YSCTTEGRQDGHLCSTTSNYEQDQKYSFC
FN2/2 LVQTRGGNSNGALCH.F...PFLYNNHN.....Y....TDCTSEGRDMMKWCSTTQNYDADQKFGFC
MMP2/1 VVRVKYGNADGEYCK.F...PFLNGKE.....Y....NSCTDTGRSDGFLWCSTTYNFEKDGKYGFC
MMP2/2 ALFTMGGAEGQPK.F...PFRFQTS.....Y....DSCTTEGRTDGYRWCSTTEDYDRDKKYGFC
MMP2/3 AMSTVGNGSEGAPV.F...PFTFLGNK.....Y....ESCTSAGRSDGKMWCAATANYDDDRKWGFC
PDC109/a TQDGPALPEDEECV.F...PFVYRNK.....H....FDCTVHGSFL..PWCSLDADY.VGR.WKYC
PDC109/b AQRDYAKCV.F...PFIYGGKK.....Y....ETCTKIGSMWM.SWCSLSPNYDKDRAWKYC
M6PR PETDDGVPOV.F...PFIENGKS.....Y....EECIIESRAK..LWCSTTADYDRDHEWGFC
HGFA EDGRPCR.F...PFRYGGRM.....L....HACTSEGSAGR.KWCATTHNYDRDRAWGYC
FA12 VTGEPCH.F...PFQYHRQL.....Y....HKCTHKGRPGQPWCATTPNFDQDQRWGYC
5..8...0..6....75.....6....5.066565....600.65.....56560

```

Kringle

```

NT/K C..GAGEPW.C..NATNLGVPCLHWDEVPPFLERSPPASWAE...LRGQPHNFCRSPDGAGR.PWCIFYRNAQ.GKVDWGYCD...C
PGN/K1 CKTGNGKNYRGTMSTKNGITCQKWSSTSPHRPRFSPATHPSEG..LE...ENYCRNPDNDPQGPWCYTTD...PEKRYDYCDILEC
PGN/K2 CMHCSGENYDGKISKTMSTGLCQAWDSQSPAHGYIPSKFPNKN..LK...KNYCRNPDREL.PWCFTTD...PNKRWELCDIPRC
PGN/K3 CLKGTGENYRGNAVAVTVSGHTCQHWSAQTPHHTNRTPENFPCKN..LD...ENYCRNPDGKRA.PWCFTTN...SQVRWEYCKIPSC
PGN/K4 CYHGDGQSYRGTSSTTTTGKKQSWSSMTPHRHQKTPENYPNAG..LT...MNYCRNPDADKG.PWCFTTD...PSVRWEYCNLKKC
PGN/K5 CMFGNGKGYRGKRATTVTGTPQDWAQEPHRSIFTPETNPRAG.LE...KNYCRNPDGDVGGPWCYTTN...PRKLYDYCDVPQC
apoA/K4 CYHGDGQSYRGFSSTTVTGRTQSWSSMTPHWHQRTTEYYPNGG..LT...RNYCRNPDAEIR.PWCYTM...PSVRWEYCNLTQC
THRB/K1 CAEGLCTNYRGHVNIITRSGIEQLWRSRYPHKPEINSTTHPGAD..LQ...ENFCRNPDSTTGWCYTTD...PTVRQECISIPVC
THRB/K2 CVPDRGQQYQGR LAVTTHGLPCLAWASQAQALSKHQDFNSAVQ..LV...ENFCRNPDGDEEGVWCYVAG...KPGDFGYCDLNYC
tPA/K1 CYEDQGISYRGTWSTAESGAETNWNSSALAQKPYSGRRPDAIRLGLGN..HNYCRNPDRLSK.PWCYVFKA..GKYSSEFCSTPAC
tPA/K2 CYFGNGSAYRGTHSLTESGASCLPWNSMILIGKYYTAQNPSAQALGLGK..HNYCRNPDGDAK.PWCHVLKN..RRLTWEYCDVPSC
uPA/K CYEGNGHFYRGKASTDTMGRPLPWNSATVLQQTYHAHRSDALQLGLGK..HNYCRNPDNRRR.PWCYVQVG..LKPLVQECMVHDC
HGF/K1 CIIGKRSYKGTVSITKSGIKQFPWSSMIPHEHSFLPSSYRGKD..LQ...ENYCRNPDGEEGGPWCFTSN...PEVRYEVCIDIPQC
HGFA/K CFLNGTGYRGVASTSASGLSLAWNSDLLYQELHVDVSGAALLGLGP..HAYCRNPDNDER.PWCYVVKD..SALSWEYCRLEAC
FA12/K CYDGRGLSYRGRLARTLSGAPQPWASEATYRNVTAEQARNWG...LGG..HAFCRNPDNDIR.PWCYVFLNR..DRLSWEYCDLAQC
HABP2/K CYVGDGYSYRGKMNRVTNOHACLYWNHLLQLQENYNMFMEDAETHGIGE..HNFCRNPDADK.PWCFTKVTN.DKVWWEYCDVSAC
ROR1/K CYNSTGVDRGTVSVTKSGRQCPWNSQYPHTHTFTALRPE...LNGG.HSYCRNPDGNQKEAPWCFTLD...ENFKSDLCIDIPAC
0..8.0..970..6.8.59..06.058..65.5.....9....5880090955...90076.5.....576066...0

```

2 10 18 30 40 47 58 72 74
(1) (22) (51) (63) (75) (80)

cystine b cystine c
cystine a

Figure 1.

Multiple sequence alignments of the kringle superfamily. The conservation scores are calculated for each column of the multiple alignments between and within the FN2 and K families and given in the middle, top and bottom rows, respectively. Positions with the conservation scores less than 0.5 are indicated with a dot (5 corresponds to a rounded score of 0.5, 6 to a rounded score of 0.6, etc., and 0 to the highest score of 1.0). The columns are color-coded according to the conservancy between and within the families. The positions conserved for both families are shaded in green; the shade increases in the order of conservancy as 0.5 (light) < ... < 1.0 (dark). The columns colored in blue correspond to positions that are specifically conserved within the families. Sequential numbering corresponds to NT/K; alternative numbering based on the PGN/K5 sequence, which starts at the first cysteine residue, is given in parentheses.

Figure 2.

Comparison of NT/K with representative structures from the kringle and FN2 families: PGN/K4 (code 1pk4), NT/K (code 2k51), and MMP2/2 (code 1ck7). The structures were superimposed using cystines *b* and *c*. β -Strands are shown in orange. The disulfides (yellow) are labeled *a*, *b* and *c*. Loop 1 of NT/K corresponds to residues 20-46, loop 2 to residues 47-56, and loop 3 to residues 57-72.

Figure 3.

(A) Schematic representation of the β -secondary structure in NT/K. Hydrogen bonds observed in the structure are marked by dotted lines. Hydrogen bond arrows point from donor to acceptor. Disulfide bonds, labeled SS-a, SS-b and SS-c, are indicated by bold lines. Antiparallel bridge (i, j) ($i < j$) at 'narrow-pair' is defined as $\{\text{Hbond}(j \rightarrow i), \text{Hbond}(i \rightarrow j)\}$. Antiparallel bridge (i, j) ($i < j$) at 'wide-pair' is defined as $\{\text{Hbond}(i+1 \rightarrow (j-1)), \text{Hbond}(j+1 \rightarrow (i-1))\}$ [10]. A ladder is a set of one or more consecutive bridges of identical type. A, B, C and D are ladders, D1, D3 and D5 are wide pair bridges, D2, D4 and D6 are narrow pair bridges, and T1, T2, and T3 are turns. N and C denote the N- and C-termini of the backbone. (B) Scheme showing the disposition of the disulfides (bold lines) relative to the hydrogen bond based arrays (dashed lines) in the kringle fold. The central array is labeled i . The N- and C-terminal half-cystine residues of disulfide j ($j = 1, 2$) are labeled N_j and C_j , respectively. (C) The inner disulfides in NT/K showing the crossing of the two S-S bonds.

Figure 4.

(**A**) Dihedrals of diethyl disulfide viewed along their rotation axes. The atom of the view axis in the forefront is indicated by a large circle and the one in the background by a black circular dot. Bonds connected to the atom in the forefront are solid and bonds connected to the atom in the background are dashed. Alternative values for the dihedrals are shown in gray. The number of orientations for each dihedral is indicated in the bottom row. (**B**) Structures of the six conformational classes of diethyl disulfide obtained by geometry optimizations, using DFT. Only the right-handed conformation of each class is shown. All structures are viewed from the perspective of the (quasi) C_2 axis of the central CSSC fragment.

Figure 5.

Energies of rotamers of diethyl disulfide versus $\text{C}^\alpha\text{-C}^\alpha$ distance obtained by relaxed scans of DFT total energies. The dots indicate disulfides b (at 5.9 $\text{\AA}$) and c (at 6.5 $\text{\AA}$).

Figure 6. Dihedrals for conformational classes ggg (dots) and tgg (solid) of diethyl disulfide versus C^α - C^α distance obtained in scans of Figure 5. $\chi_2' \approx \chi_2$ for class ggg.

Table 1Angles^a between flanking backbones between disulfide bonds of inner cystines in the kringle superfamily

Protein	Code	Orientation angles, κ (°) ^a	
		cystine <i>b</i>	cystine <i>c</i>
<i>K family</i>			
NT/K	2k4r	57	94
PGN/K1	1ki0	53	106
PGN/K2	1ki0	57	103
PGN/K3	1ki0	62	99
PGN/K4	1pk4	59	122
PGN/K5	5hpg	48	96
tPA/K2	1tpk	50	113
uPA/K	2i9b	55	121
apoA/K4	1i71	60	117
HGF/K1	1bht	52	103
THRB/K1	1nl1	54	102
Average (SD) ^b		55(4)	107(10)
<i>FN2 family</i>			
MMP2/1	1ck7	54	110
MMP2/2	1ck7	47	97
MMP2/3	1ck7	51	117
PDC109/a	1h8p	62	120
PDC109/b	1h8p	56	112
M6PR	2v5o	68	111
Average (SD) ^b		56(6)	111(9)

^a κ is the angle between vectors $N_1 \rightarrow C_1$ and $N_2 \rightarrow C_2$ where N_i and C_i are the amide and carbonyl atoms of half-cystine *i* [5]. The κ values classify the backbone orientations in cystines *b* and *c* as parallel (*P*) [0° to 60°] and orthogonal (*O*) [60° to 120°].

^b The numbers in parentheses are the standard deviations in the least significant digits.

Table 2

Geometry for cystine-pair cluster^a in the kringle superfamily

Protein	Code	Cluster dihedrals		Distance ^d , C ^α -C ^{α'} (Å)			
		C ^α _{bc} (°) _b	ν ^{S^γ} _{bc} (°) _b	θ ^{S^γ} _{bc} (°) _c	18 _b -47 _c	18 _b -72 _c	58 _b -72 _c
<i>K family</i>							
NT/K	2k4r	-96 (4)	-146 (6)	35	7.5 (2)	9.5 (4)	7.1 (1)
PGN/K1	1ki0	-91	-142	35	7.5	9.3	7.3
PGN/K2	1ki0	-93	-147	31	7.4	9.4	7.4
PGN/K3	1ki0	-98	-151	26	7.3	9.2	7.3
PGN/K4	1pk4	-90	-158	18	7.4	8.5	6.8
PGN/K5	5hpg	-88	-145	30	7.3	8.8	7.4
tPA/K2	1tpk	-82	-143	31	7.5	8.6	6.9
uPA/K	2i9b	-85	-140	35	7.6	8.8	7.1
apoA/K4	li71	-88	-149	26	7.3	8.5	6.9
HGF/K1	1bht	-93	-146	32	7.4	9.4	7.4
THRB/K1	1nl1	-89	-140	38	7.5	9.2	7.4
Average (SD) ^e		-90 (5)	-146(5)	31 (6)	7.4 (1)	9.0 (4)	7.2 (2)
<i>FN2 family</i>							
MMP2/1	1ck7	-77	-127	43	7.7	8.0	6.8
MMP2/2	1ck7	-71	-88	76	7.7	8.5	7.5
MMP2/3	1ck7	-73	-117	51	8.4	8.1	6.8
PDC109/b	1h8p	-76	-112	55	8.2	8.1	6.6
M6PR	2v5o	-86	-114	52	8.6	8.2	7.0
Average (SD) ^e		-76 (4)	-112 (14)	55 (11)	8.1 (4)	8.2 (2)	6.9 (3)

^aCystine *b*: Cys18_b—Cys58_b and cystine *c*: Cys47_c—Cys72_c (sequential numbering corresponds to NT/K in Figure 1).

^bCluster dihedrals are defined as $C_b^{\alpha} - C_b^{\alpha'} - C_c^{\alpha} - C_c^{\alpha'}$ where C_b^{α} is the N-terminal half-cystine residue forming the first cystine *b*, $C_b^{\alpha'}$ is the C-terminal residue, C_c^{α} is the N-terminal half-cystine residue forming the second cystine *c*, $C_c^{\alpha'}$ is the C-terminal residue. $\nu_{bc}^{S^{\gamma}}$ is calculated from vectors joining the S^{γ} atoms $S_b^{\gamma} - S_b^{\gamma'} - S_c^{\gamma} - S_c^{\gamma'}$ [11].

^c θ_{bc}^{γ} is the dihedral angle between the inner disulfide bonds, $S_b^{\gamma'} - S_b^{\gamma'} - S_c^{\gamma} - S_c^{\gamma}$.

^d The shortest distance among $C_b^{\alpha} - C_c^{\alpha}$, $C_b^{\alpha} - C_c^{\alpha'}$, $C_b^{\alpha'} - C_c^{\alpha}$, and $C_b^{\alpha'} - C_c^{\alpha'}$ is for 58b-47c in all kringles.

^e The numbers in parentheses are the standard deviations in the least significant digits.

Table 3

Geometrical parameters of disulfides in the kringle superfamily

Cystine ^d	Code	Distance(Å) C ^α -C ^{α'}	Distance(Å) C ^β -C ^{β'}	χ ₁	χ ₂	χ ₃	χ _{2'}	χ _{1'}	φ	ψ	φ'	ψ'
Side chain dihedrals <i>b</i> (°)												
Backbone dihedrals (°)												
Cystine <i>b</i>												
<i>K</i> family												
NT/K ^c	2k4r	5.9 (1)	3.8	-65 (6)	-69 (5)	-80 (3)	-172 (4)	71 (2)	-88 (8)	158 (7)	-134 (3)	196 (2)
PGN/K1	1ki0	5.7	3.6	-67	-71	-80	-176	72	-68	150	-155	155
PGN/K2	1ki0	5.7	3.6	-71	-65	-81	-165	65	-73	141	-162	160
PGN/K3	1ki0	5.6	3.8	-75	-54	-85	-172	72	-71	141	-169	156
PGN/K4	1pk4	5.5	3.5	-70	-64	-73	-167	64	-74	143	-159	156
PGN/K5	5hpg	5.6	3.6	-70	-65	-77	-171	60	-75	152	-164	170
tPA/K2	1tpk	5.6	3.5	-57	-77	-71	-168	57	-60	145	-158	166
uPA/K	2fd6	5.2	3.8	-69	-75	-85	-182	79	-70	160	-181	162
uPA/K	2i9b	5.9	3.7	-71	-78	-85	-172	70	-82	151	-166	157
apoA/K4	1i71	5.7	3.6	-70	-65	-80	-172	72	-71	146	-169	158
HGF/K1	1bht	5.8	3.6	-66	-78	-78	-172	69	-70	155	-157	163
THRB/K1	1nl1	5.7	3.6	-71	-67	-79	-170	67	-66	147	-157	159
<i>FN2</i> family												
MMP2/1	1ck7	5.4	3.1	-41	-82	-76	-163	60	-65	161	-154	165
MMP2/2	1ck7	5.6	3.7	-67	-67	-82	-169	69	-57	166	-166	153
MMP2/3	1ck7	5.9	3.6	-45	-93	-69	-179	59	-78	170	-150	161
PDC109/b	1h8p	5.6	3.7	-70	-59	-85	-168	68	-57	147	-146	157
M6PR	2v5o	6.0	3.9	-62	-72	-93	-173	88	-71	152	-136	162
Cystine <i>c</i>												
<i>K</i> family												
NT/K ^c	2k4r	6.5 (1)	3.9 (1)	-58 (5)	-97 (6)	-96 (8)	-157 (7)	-57 (6)	-55 (5)	161 (8)	-102 (13)	34 (47)
PGN/K1	1ki0	6.5	3.7	-50	-120	-86	-170	-52	-70	134	-108	163
PGN/K2	1ki0	6.5	3.6	-59	-117	-82	-163	-59	-61	136	-110	158
PGN/K3	1ki0	6.5	3.7	-57	-113	-84	-162	-61	-59	140	-101	159
PGN/K4	1pk4	6.1	3.5	-67	-94	-82	-136	-86	-60	132	-85	157

Cystine ^a	Code	Distance(Å)		Distance (Å)		Side chain dihedrals <i>b</i> (°)			Backbone dihedrals (°)			
		C ^α -C ^{α'}	C ^β -C ^{β'}	χ ₁	χ ₂	χ ₃	χ ₂ '	χ ₁ '	φ	ψ	φ'	ψ'
PGN/K4 ^d	1pk4	6.1		-66	-175	86	103	-172				
PGN/K5	5hpg	6.2	3.5	-62	-105	-77	-165	-58	-61	128	-110	158
tPA/K2	1pk	6.3	3.6	-57	-102	-80	-153	-83	-61	127	-92	158
uPA/K	2fd6	6.3	3.7	-71	-180	88	102	-168	-55	131	-82	156
uPA/K	29fb	6.6	4.0	-59	-109	-110	-162	-56	-58	136	-81	186
apoA/K4	1i71	6.3	3.7	-64	-98	-89	-152	-75	-58	132	-84	158
HGF/K1	1bht	6.4	3.6	-66	-115	-79	-160	-61	-68	124	-102	161
THRB/K1	1nl1	6.4	3.6	-61	-122	-81	-170	-50	-62	140	-115	170
<i>FN2 family</i>												
MMP2/1	1ck7	6.6	3.7	-47	-139	-105	-166	-37	-80	153	-88	137
MMP2/2	1ck7	6.8	3.8	-45	-138	-77	-175	-56	-75	157	-124	72
MMP2/3	1ck7	6.9	4.2	-40	-135	-132	-170	-28	-66	136	-89	154
PDC109/b	1h8p	6.6	3.7	-64	-118	-88	-157	-60	-68	159	-103	-40
M6PR	2v5o	6.7	4.0	-57	-112	-98	-165	-71	-64	137	-70	135

^aCystine *b*: Cys18*b*—Cys58*b* and Cystine *c*: Cys47*c*—Cys72*c* (sequential numbering corresponds to NT/K in Figure 1).

^bTorsion angles are defined as $C^{\alpha} \overset{\chi_1}{=} C^{\beta} \overset{\chi_2}{=} S^{\gamma} \overset{\chi_3}{=} S^{\gamma'} \overset{\chi'_2}{=} C^{\beta'} \overset{\chi'_1}{=} C^{\alpha'}$. The numbers in parentheses are the standard deviations in the least significant digits.

^cExperimental χ₁ dihedral angles refer to values deduced directly from NMR parameters (i.e. NOE intensities of cross-peaks involving the H^N, H^α, and H^β protons intrasidue NOE intensities), and not influenced by any computational algorithms but left-handed conformers (χ₂³ = -90°) have been enforced in the final refinement [17].

^dRef. [6].

Table 4

List of equilibrium conformations of diethylene disulfide in vacuum calculated with density functional theory ^{a, b, c}

Class	Descriptor	Symm.	Frequency ^d	χ_2 (°)	χ_3 (°)	χ_2' (°)	C ^{α} -C ^{α} (Å)	C ^{β} -C ^{β} (Å)	ΔE^e cm ⁻¹
1	ggg <i>spiral</i>	C ₂	2	69	89	69	5.84	3.86	0
2	tgg <i>trans-spiral</i>	C ₁	4	69	89	178	5.84	3.85	108
3	ggg' <i>corner</i>	C ₁	4	68	97	-70	4.86	4.01	158
4	tgt <i>trans</i>	C ₂	2	177	89	177	6.60	3.85	206
5	tgg' <i>hook</i>	C ₁	4	175	97	-69	5.10	4.00	255
6	g'gg' <i>staple</i>	C ₂	2	-75	109	-75	3.73	4.19	559

^aGeometry optimizations using B3LYP/6-311G*.

^bColumns: Numbering of classes based on relative energies of conformations, labeling of classes based on dihedrals (see text), descriptive names, molecular symmetry group, number of rotamers in a class for a protein-bound disulfide, dihedral angles C ^{α} C ^{β} SS', C ^{β} SS'C ^{β} , SS'C ^{β} C ^{α} , C ^{β} -C ^{β} , and relative total energies. Energies and geometrical parameter values are obtained for optimized structures.

^cOther distances: S-S = 2.09-2.10 Å, C-S = 1.85 Å, C-C = 1.52 Å.

^dThe numbers of rotamers in this column add up to 18, that is the number of combinations for (χ_2 , χ_3 , χ_2') (cf. Figure 4A).

^eThe additivity in the dihedral distortion energies is also evident from the similarity of the sections of the potential energy surface along the trajectories ggg → tgg and tgg → tgt (Figure S1).

Distribution of disulfide conformations in proteins over the six-class taxonomy and averages for dihedral angles and distances

Table 5

Class	Descriptor	total ^a	% ^b	χ_2 (°)	χ_3 (°)	χ'_2 (°)	C ^u -C ^u (Å)
1	ggg spiral	190	42	72	84	72	5.8
2	tgg trans-spiral	78	17	180	85	75	5.8
3	ggg' corner	90	20	78	97	-72	5.0
4	tgt trans	20	4	169	89	169	6.5
5	tgg' hook	45	10	-176	98	-77	5.0
6	g'gg' staple	30	7	-90	106	-90	4.1

^aTotal number of disulfide conformations in the data set (see Methods).

^bPercentage of disulfides conformations in the data set.

Table 6
Correlation between disulfide classifications according to three-class and six-class taxonomies ^a

		six-class taxonomy					
		<i>spiral</i>	<i>tran-spiral</i>	<i>corner</i>	<i>trans</i>	<i>hook</i>	<i>staple</i>
		R- <i>ggg</i> L- <i>ggg</i>	R- <i>ttg</i> R- <i>ggt</i> L- <i>ttg</i> L- <i>ggt</i>	R- <i>ggg</i> R- <i>g'</i> <i>gg</i> L- <i>ggg</i> L- <i>g'gg</i>	R- <i>tgt</i> L- <i>tgt</i>	R- <i>tgg</i> R- <i>g'</i> <i>'gt</i> L- <i>tgg</i> L- <i>g'gt</i>	R- <i>g'gg</i> L- <i>g'gg</i>
three-class taxonomy	<i>spiral</i>	X	X	—	X	—	—
	<i>hook</i>	—	X	X	X	X	—
	<i>short hook</i>	—	—	—	X	X	X
		(+++) (-- -) ^b (++ -) (- + +) (+ - -) (- - +) (- + -) (+ - +)					

^aX indicates that the intersection of the three-class (row) and the six-class taxonomy class (column) is not empty; — indicates that the intersection is empty.

^bR- and L-spiral are listed as different classes in PROMOTIF [13].

Table 7Definition of descriptors used for classifying disulfide rotamers in proteins ^a

Descriptor	This work ggg	H&S ^b	H&T ^c	S&H&H ^d
<i>spiral</i>	ggg	+++	+++	+++
<i>trans-spiral</i>	tgg	—	—	—
<i>corner</i>	ggg'	++-	—	—
<i>trans</i>	tgt ^e	—	—	—
<i>hook</i>	tgg'	—	++-	++-
<i>staple</i>	g'gg'	-+-	—	-+-
<i>short hook</i>	—	—	-+-	—

^aThe classifications of H&S [11], H&T [13], and S&H&H [21] are based on the signs of dihedrals χ_2 , χ_3 , and χ_2' (+ or -); in our classification, χ_2 and χ_2' are either g, g', or t. Only one representative rotamer of each class is given; — indicates that the corresponding descriptor is not used by the authors listed in the column header.

^bRef. [11].

^cRef. [13].

^dRef. [21].

^eTrans-gauche-trans rotamer was identified by Richardson in immunoglobulins [12].